Le Frido

Les quelques premières années de mathématiques

Contributeurs :
Laurent Claessens
Lilian Besson
Bertrand Desmons

14 septembre 2020
d269c527683bea856400b963ac10d172c47c9d8c

Copyright 2011-2019
Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the chapter entitled "GNU Free Documentation License".

Thèmes

Ceci est une sorte d'index thématique.

1 : Cardinalité
2 : tribu, algèbre de parties, λ-systèmes et co.
3 : théorie de la mesure
4 : intégration
5 : suites et séries
6 : polynôme de Taylor
7 : normes
8 : caractérisations séquentielles
9 : topologie produit
10 : espaces métriques, normés
11 : gaussienne
12 : compacts
13 : densité
14 : espaces de fonctions
15 : fonctions Lipschitz
16 : formule des accroissements finis
17 : limite et continuité
18 : différentiabilité
19 : points fixes
20 : théorèmes de Stokes, Green et compagnie
21 : permuter des limites
22 : applications continues et bornées
23 : inégalités
24 : connexité
25 : suite de Cauchy, espace complet
26 : application réciproque
27 : déduire la nullité d'une fonction depuis son intégrale
28 : équations différentielles
29 : injections
30 : logarithme
31 : inversion locale, fonction implicite
32 : convexité
33 : fonction puissance
34 : dualité
35 : opérations sur les distributions
36 : convolution
37 : séries de Fourier
38 : transformée de Fourier
39 : méthode de Newton
40 : méthodes de calcul
41 : espaces vectoriels
42 : définie positive
43 : norme matricielle, norme opérateur et rayon spectral
44 : série de matrices
45 : rang
46 : extension de corps et polynômes
47 : décomposition de matrices
48 : systèmes d'équations linéaires
49 : formes bilinéaires et quadratiques
50 : arithmétique modulo, théorème de Bézout
51 : polynômes
52 : zoologie de l'algèbre
53 : invariants de similitude
54 : réduction, diagonalisation
55 : endomorphismes cycliques
56 : déterminant
57 : polynôme d'endomorphismes
58 : exponentielle
59 : types d'anneaux
60 : sous-groupes
61 : groupe symétrique
62 : action de groupe
63 : classification de groupes
64 : produit semi-direct de groupes
65 : théorie des représentations
66 : isométries
67 : caractérisation de distributions en probabilités
68 : théorème central limite
69 : lemme de transfert
70 : probabilités et espérances conditionnelles
71 : dénombrements
72 : enveloppes
73 : équations diophantiennes
74 : techniques de calcul

Thème 1 : Cardinalité Le Frido ne définit pas la notion de nombre cardinal ; ça nous mènerait trop loin. Au lieu de cela, nous allons nous contenter des notions d'équipotence, surpotence et subpotence, et démonter un certain nombre de résultats en utilisant sans retenue le lemme de Zorn.

(1) Définition d'équipotence, surpotence et subpotence, notations $A \succ B$ et $A \approx B$, définition 1.27.

(2) Toute partie d'une ensemble fini est finie, lemme 1.35.

(3) Si A est un ensemble fini ou dénombrable, alors il existe une surjection $\mathbb{N} \to A$, lemme 1.36.

(4) Si A est un ensemble infini et si $f\colon A \to B$ est une application injective, alors $f(A)$ est infini, lemme 1.37

(5) Toute partie infinie de $\mathbb{N}$ est dénombrable, proposition 1.38

(6) Une bijection $\mathbb{N} \to \mathbb{N} \times \mathbb{N}$, proposition 1.41.

(7) Une décomposition de $\mathbb{N}$ en une infinité de parties équipotentes à $\mathbb{N}$, corollaire 1.42.

(8) Si il existe une surjection $\mathbb{N} \to A$, alors A est fini ou dénombrable, lemme 1.43.

(9) Une union dénombrable d'ensembles finis ou dénombrables est finie ou dénombrable, proposition 1.44

(10) Tout ensemble infini contient une partie en bijection avec $\mathbb{N}$, proposition 1.47.

(11) Toute partie d'un ensemble fini est finie, et toute partie d'un ensemble dénombrable est finie ou dénombrable, proposition 1.48.

(12) Si $A \geq B$ et $B \geq A$, alors $A \approx B$, théorème de Cantor-Schröder-Bernstein 1.51

(13) Le théorème de Cantor 1.54 dit qu'il n'existe pas de surjection d'un ensemble vers son ensemble des parties. On en déduit qu'il n'existe pas d'ensemble contenant tous les ensembles (corolaire 1.56).

(14) Si A est infini et si $A \geq B$, alors $A \approx A \cup B$ par le lemme 1.58.

(15) Si S est un ensemble infini alors il existe une bijection $\varphi\colon \{0,1\} \times S \to S$, proposition 1.59.

(16) Si A est infini, alors $A \times \mathbb{N} \approx A$, proposition 1.61.

(17) Si A est infini et si $B \prec A$, alors $A \backslash B \approx A$, lemme 1.63.

(18) Si A est infini, alors $A \approx A \times A$, théorème 1.67.

Il y a aussi des résultats de cardinalité autour des extensions de corps.

(1) Si $\mathbb{K}$ est un corps infini, alors $\mathbb{K}[X] \approx \mathbb{K}$.

(2) Le théorème de Steinitz 6.125 affirme que tout corps admet une unique clôture algébrique. La preuve utilise pas mal de cardinalité ainsi que le lemme de Zorn.

Thème 2 : tribu, algèbre de parties, λ-systèmes et co. Il existe des centaines de notions de mesures et de classes de parties.

(1) Le plus souvent lorsque nous parlons de mesure est que nous parlons de mesure positive, définition 13.20 sur un espace mesuré avec une tribu, définition 13.1.

(2) Une mesure extérieure est la définition 13.13

(3) Une algèbre de partie : définition 13.16. Une mesure sur une algèbre de parties : définition 13.14. L'intérêt est que si on connait une mesure sur une algèbre de parties, elle se prolonge en une mesure sur la tribu engendrée par le théorème de prolongement de Hahn 13.68.

(4) Un λ-système : définition 13.30.

(5) Une mesure complexe : définition 13.201.

En théorie de l'intégration, si X est une partie de $\mathbb{R}^n$, la convention est de considérer des fonctions

$$f\colon \big(X, \mathcal{L}\mathrm{eb}(X)\big) \to \big(\mathbb{R}, \mathcal{B}\mathrm{or}(\mathbb{R})\big).$$

Voir les points 13.106 et 13.150 pour les conventions à ce propos.

À propos d'applications mesurables :

(1) Définition d'une application mesurable, définition 13.41.

(2) Une fonction continue est borélienne, théorème 13.49.

À propos de tribu induite :

(1) Définition 13.6.

(2) Les boréliens induits sont bien les boréliens de la topologie induite : $\mathcal{B}\mathrm{or}(Y) = \mathcal{B}\mathrm{or}(X)_Y$, théorème 13.50.

Thème 3 : théorie de la mesure

Mesure À propos de mesure.

(1) Mesure positive, mesure finie et σ-finie, c'est la définition 13.20.
(2) Le produit de tribus est donné par la définition 13.112,
(3) Produit d'une mesure par une fonction, définition 13.189.
(4) le produit d'espaces mesurés est donné par la définition 13.211.
(5) Mesure de Lebesgue sur $\mathbb{R}$, définition 13.127.
(6) Une partie de $\mathbb{R}$ non mesurable au sens de Lebesgue : l'exemple 13.141.
(7) Mesure de Lebesgue sur $\mathbb{R}^N$, définition 13.213.
(8) Mesure à densité, définition 13.189.

Théorèmes d'approximation Il est important de pouvoir approcher des fonctions continues ou L^p par des fonctions étagées, sinon on ne parvient pas à faire tourner la machine de l'intégration de Lebesgue.

(1) Si $(S, \mathcal{A}, \mu)$ est un espace mesuré et si $f\colon S \to [0, +\infty]$ est une fonction mesurable, le théorème fondamental d'approximation 13.105 dit qu'il existe une suite croissante de fonctions étagées qui converge vers f.
(2) Les fonctions simples sont denses dans L^p, proposition 26.44.
(3) Encadrement d'un borélien A par un fermé F et un ouvert V par le lemme 13.204 : $F \subset A \subset V$ avec $\mu(V \backslash F) < \epsilon$.
(4) Approximation L^p de la fonction caractéristique d'un borélien par une fonction continue par le théorème 13.206.

Thème 4 : intégration À propos d'intégration.

L'ordre dans lequel les choses sont faites
— Nous commençons par considérer des fonctions $f\colon \Omega \to [0, +\infty]$ dans la définition 13.151.
— Nous donnerons ensuite quelques propriétés restreintes au fonctions à valeurs positives, par exemple
  (1) La convergence monotone 13.160,
  (2) Lemme de Fatou 13.164.
  (3) (presque) linéarité pour les fonctions positives, théorème 13.165.
— La définition pour les fonctions à valeurs dans $\mathbb{R}$ puis $\mathbb{C}$ est 13.168.
— Pour les fonctions à valeurs dans un espace vectoriel, c'est la définition 13.177.

Quelque résultats (1) Intégrale associée à une mesure, définition 13.151
(2) L'existence d'une primitive pour toute fonction continue est le théorème 11.322.
(3) La définition d'une primitive est la définition 11.136.
(4) Primitive et intégrale, proposition 13.235.
(5) Intégrale impropre, définition 13.248.

Intégrale et mesure (1) L'intégrale de la fonction 1 donne la mesure : $\int_B 1 d\mu = \mu(B)$, c'est le lemme 13.157.
(2) Le théorème de Radon-Nikodym 13.198 donne une densité pour certaines mesures.
(3) Le produit d'une mesure par une fonction donné par la définition 13.189 introduit aussi une densité : $(w \cdot \mu)(A) = \int_A w d\mu$.

Autre résultats (1) Si $A, B \subset \Omega$ sont des parties disjointes, alors $\int_{A \cup B} f = \int_A f + \int_B f$, proposition 13.174.
(2) La σ-additivité dénombrable, $\int_{\bigcup_i A_i} f d\mu = \sum_{i=0}^{\infty} \int_{A_i} f d\mu$ est dans les propositions 13.187 et 13.188.

Thème 5 : suites et séries

Suites Les suites réelles sont en général dans la proposition 7.129 et ce qui s'ensuit. Cette proposition est souvent prise comme définition lorsque seules les suites réelles ne sont considérées.

(1) Les suites adjacentes, c'est la définition 7.142.

(2) Les séries alternées, théorème 10.72. Il s'agit de dire que $\sum_{k=0}^{\infty}(-1)^k a_k$ converge quand a_k est décroissante et tend vers zéro.

(3) Le concept de suite adjacente sert à étudier la série de Taylor de $\ln(x+1)$, voir le lemme 14.73 et ce qui l'entoure.

(4) La définition de la convergence absolue est la définition 10.51.

(5) Une suite réelle croissante et majorée converge, proposition 7.137.

(6) Toute suite dans un compact admet une sous-suite convergente, théorème 7.138.

(7) Pour tout réel, il existe une suite croissante de rationnels qui y converge, proposition 7.122.

Calcul de suites (1) Somme : $x_n + y_n \to x + y$ est la proposition 7.133.

Série Les séries sont en général dans la section 10.3.

(1) Quelques séries usuelles en 10.4.3 : série harmonique, géométrique, de Riemann, et la mythique arithmético-géométrique.

(1a) La série est associative : $\sum_k (a_k + b_k) = \sum_k a_k + \sum_k b_k$. C'est la proposition 10.60.

(1b) La série harmonique diverge : $\sum_k \frac{1}{k} = \infty$, exemple 10.66.

(1c) La série géométrique : $\sum_{k=0}^{N} q^k = \frac{1-q^{N+1}}{1-q}$, exemple 10.67.

(1d) Une autre cool série : $\sum_{k=1}^{N} \frac{1}{k(k+1)} = \frac{N}{N+1}$, lemme 10.71.

(2) Critère des séries alternées, théorème 10.72.

(3) Convergence d'une série implique convergence vers zéro du terme général, proposition 10.58.

Sommes infinies En ce qui concerne les sommes finies, la notation $\sum_{i=1}^{N}$ est définie en 1.89. Pour permuter les termes d'une somme avec un élément du groupe symétrique, nous avons la proposition 1.91.

Voici quelques résultats à propos de sommes infinies.

(1) Une somme indexée par un ensemble quelconque est la définition 10.91.

(2) La définition de la somme d'une infinité de termes est donnée par la définition 10.48.

(3) Une somme de termes positifs indexée par un ensemble indénombrable est toujours infinie par le lemme 10.96.

(4) si la série converge, on peut regrouper ses termes sans modifier la convergence ni la somme (associativité) ; Pour les sommes infinies l'associativité et la commutativité dans une série sont perdues. Néanmoins, il subsiste que

(4a) si la série converge absolument, on peut modifier l'ordre des termes sans modifier la convergence ni la somme (commutativité, proposition 10.89).

(5) Permuter une somme infinie avec une application linéaire : $f(\sum_{i\in I} v_i) = \sum_{i\in I} f(v_i)$, c'est la proposition 10.102.

Fonction analytique La fonction $\sum_{k=0}^{\infty} a_n z^n$ est holomorphe dans son disque de convergence par la proposition 14.23.

Thème 6 : polynôme de Taylor

Énoncés Il existe de nombreux énoncés du théorème de Taylor, et en particulier beaucoup de formules pour le reste.

(1) Énoncé : théorème 11.375.

(2) Une majoration du reste est dans le théorème 14.34

(3) De classe C^2 sur $\mathbb{R}^n$, proposition 11.382.

(4) Avec un reste donné par un point dans $]x, a[$, proposition 11.386.

(5) Avec reste intégral, proposition 19.147 et théorème 19.144 pour le cas simple $\mathbb{R} \to \mathbb{R}$.

(6) Le polynôme de Taylor généralise à l'utilisation de toutes les dérivées disponibles le résultat de développement limité donné par la proposition 11.113.

(7) Pour les fonctions holomorphes, il y a le théorème 25.26 qui donne une série de Taylor sur un disque de convergence.

Utilisation Des polynômes de Taylor sont utilisés pour démontrer des théorèmes par-ci par-là.

(1) Il est utilisé pour justifier la méthode de Newton autour de l'équation (33.95).

(2) On utilise pas mal de Taylor dans les résultats liant extrémum et différentielle/hessienne. Par exemple la proposition 16.70.

Quelques développements Voici quelques développements limités à savoir. Ils sont calculables en utilisant la formule de Taylor-Young (proposition 11.391).

$$e^x = \sum_{k=0}^{n} \frac{x^k}{k!} + x^n \alpha(x) \qquad \text{ordre } n, \text{proposition } 14.68$$

$$\cos(x) = \sum_{k=0}^{p} \frac{(-1)^k x^{2k}}{(2k)!} + x^{2p+1} \alpha(x) \qquad \text{ordre } 2p+1, \text{proposition } 17.57$$

$$\sin(x) = \sum_{k=0}^{p} \frac{(-1)^k x^{2k+1}}{(2k+1)!} + x^{2p+2} \alpha(x) \qquad \text{ordre } 2p+1, \text{proposition } 17.57$$

$$\ln(1+x) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} x^k + \alpha(x) x^n \qquad \text{ordre } n, \text{proposition } 14.72$$

$$\ln(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k \qquad \text{exact proposition } 14.74$$

$$\ln(2) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \qquad \text{exact proposition } 14.74$$

$$(1+x)^l = \sum_{k=0}^{l} \binom{l}{k} x^k \qquad \text{exact si } l \text{ est entier.}$$

$$(1+x)^\alpha = 1 + \sum_{k=1}^{n} \frac{\alpha(\alpha-1)\dots(\alpha-k+1)}{k!} x^k + x^n \alpha(x) \qquad \text{ordre } n.$$

Dans toutes ces formules, la fonction $\alpha \colon \mathbb{R} \to \mathbb{R}$ vérifie $\lim_{t \to 0} \alpha(t) = 0$.

Le développement limité en 0 d'une fonction paire ne contient que les puissances de x d'exposant paire. Voir comme exemple le développement de la fonction cosinus.

Thème 7 : normes

Définition Espace vectotiel normé : définition 7.113.

Équivalence de norme (1) Définition de l'équivalence de norme 7.221.

(2) La proposition 7.223 sur l'équivalence des normes $\|.\|_2$, $\|.\|_1$ et $\|.\|_\infty$ dans $\mathbb{R}^n$.

(3) En général pour les normes $\|.\|_p$, il y a des inégalités dans 11.364 et 16.96 ; voir aussi le thème 7.

(4) La proposition 16.106 donne l'inégalité $\|x\|_q \leqslant n^{\frac{1}{q}-\frac{1}{p}}\|x\|_p$ dès que $0 < q < p$.

(5) Toutes les normes sur un espace vectoriel de dimension finie sont équivalentes par le théorème 7.224.

(6) Montrer que le problème $a - b$ est stable dans l'exemple 33.26.

(7) La proposition 10.15 donnant $\rho(A) \leqslant \|A\|$ utilise l'équivalence de toutes les normes sur un espace vectoriel de dimension finie (théorème 7.224.).

Norme opérateur et d'algèbre voir le thème 43.

Thème 8 : caractérisations séquentielles

(1) Fonction séquentiellement continue, définition 7.77.

(2) La continuité implique la continuité séquentielle, proposition 7.79 et corollaire 7.80.

(3) Pour des espaces métriques, la continuité séquentielle d'une fonction est équivalente à la continuité, proposition 7.268. Une version spéciale pour $\mathbb{R}^m$ est donnée par le théorème 11.156.

Thème 9 : topologie produit

(1) La définition de la topologie produit est 7.12.

(2) Pour les espaces vectoriels normés, le produit est donné par la définition 7.230.

(3) L'équivalence entre la topologie de la norme produit et la topologie produit est le lemme 7.230.

(4) Quand V et W sont des espaces métriques, la topologie considérée sur $V \times W$ est celle de la définition 7.230. C'est à la fois la topologie de la norme produit et la topologie produit.

(5) La convergence dans un espace vectoriel est si et seulement si il y a convergence composante par composante, proposition 7.38.

Thème 10 : espaces métriques, normés

(1) Un espace métrique est un espace muni d'une distance, définition 7.93.

(2) La distance entre un point et un ensemble est la définition 7.107.

(3) Le théorème-définition 7.94 donne la topologie sur un espace métrique en disant que les boules ouvertes sont une base de la topologie (définition 7.59).

(4) La définition de la convergence d'une suite est la définition 7.10.

(5) Dans un espace vectoriel normé, une application est continue si et seulement si elle est bornée, proposition 10.18.

(6) Un espace vectoriel topologique qui possède en tout point une base dénombrable de topologie accepte une distance, théorème 7.288.

Thème 11 : gaussienne

(1) Le calcul de l'intégrale

$$\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}$$

est fait de deux façons dans l'exemple 13.265. Dans les deux cas, le théorème de Fubini 13.262 est utilisé.

(2) Le lemme 28.22 calcule la transformée de Fourier de $g_\epsilon(x) = e^{-\epsilon\|x\|^2}$ qui donne $\hat{g}_\epsilon(\xi) = \left(\frac{\pi}{\epsilon}\right)^{d/2} e^{-\|\xi\|^2/4\epsilon}$.

(3) Le lemme 28.23 donne une suite régularisante à base de gaussienne.

(4) Elle est utilisée pour régulariser une intégrale dans la preuve de la formule d'inversion de Fourier 28.26

Thème 12 : compacts

Propriétés générales Quelques propriétés de compacts.

(1) La définition d'un ensemble compact est la définition 7.48.

(2) Les compacts sont les fermés bornés par le théorème 7.127.

(3) Le théorème de Borel-Lebesgue 7.124 dit qu'un intervalle[1] de $\mathbb{R}$ est compact si et seulement si il est de la forme $[a, b]$.

(4) Théorème des fermés emboîtés dans le cas compact, corolaire 7.58. À ne pas confondre avec celui dans le cas des espaces métrique, théorème 7.304.

(5) L'image d'un compact par une fonction continue est un compact, théorème 7.92.

(6) Suites dans un compact

(6a) Toute suite dans un compact admet une sous-suite convergente, théorème 7.138.

(6b) Dans $\mathbb{R}^n$, toute suite dans un compact admet une sous-suite convergente, théorème 7.155. La démonstration de ce théorèma est non seulement plus compliquée que le cas général, mais utilise en plus le cas dans $\mathbb{R}$; lequel cas n'est pas démontré de façon directe dans le Frido.

(6c) Un espace métrique est compact si et seulement si toute suite contient une sous-suite convergente. C'est le théorème de Bolzano-Weierstrass 7.103. La démonstration de ce théorème est indépendante.

(7) Une fonction continue sur un compact est bornée et atteint ses bornes, théorème 7.105.

(8) Une fonction continue sur un compact y est uniformément continue, théorème de Heine 11.90.

Produits de compacts À propos de produits de compacts. C'est un compact dans tous les cas métriques[2].

(1) Les produits d'espaces métriques compacts sont compacts. Il s'agit du théorème de Tykhonov que nous verrons ce résultat dans les cas suivants.

— $\mathbb{R}$, lemme 7.126.

— Produit fini d'espaces métriques compacts, théorème 7.318.

— Produit dénombrable d'espaces métriques compacts, théorème 7.320.

Thème 13 : densité

(1) Densité de $\mathbb{Q}$ dans $\mathbb{R}$, proposition 7.121.

(2) Densité des polynômes dans $\left(C^0([0,1]), \|.\|_\infty\right)$, théorème de Bernstein 35.141.

(3) Densité des polynômes dans $\left(C^0(I), \|.\|_\infty\right)$ lorsque $I = [a, b]$, corolaire 35.142.

(4) Densité de $\mathscr{D}(\mathbb{R}^d)$ dans $L^p(\mathbb{R}^d)$ pour $1 \leqslant p < \infty$, théorème 26.47.

(5) Densité de $\mathscr{S}(\mathbb{R}^d)$ dans l'espace de Sobolev $H^s(\mathbb{R}^d)$, proposition 30.15.

(6) Densité de $\mathscr{D}(\mathbb{R}^d)$ dans l'espace de Sobolev $H^s(\mathbb{R}^d)$, proposition 30.17.
Cela est utilisé pour le théorème de trace 30.19.

(7) Les applications étagées dans les applications mesurables (qui plus est avec limite croissante), théorème fondamental d'approximation 26.50.

(8) Les fonctions continues à support compact dans $L^2(I)$, théorème 26.51.

(9) Les polynômes trigonométriques sont denses dans $L^p(S^1)$ pour $1 \leqslant p < \infty$. Deux démonstrations indépendantes par le théorème 27.6 et le théorème 26.70.

1. Définition 1.13.
2. Si vous connaissez des exemples non métriques de produits de compacts qui ne sont pas compacts, écrivez-moi.

Les densités sont bien entendu utilisées pour prouver des formules sur un espace en sachant qu'elles sont vraies sur une partie dense. Mais également pour étendre une application définie seulement sur une partie dense. C'est par exemple ce qui est fait pour définir la trace γ_0 sur les espaces de Sobolev $H^s(\mathbb{R}^d)$ en utilisant le théorème d'extension 16.121.

Comme presque tous les théorèmes importants, le théorème de Stone-Weierstrass possède de nombreuses formulations à divers degrés de généralité.

— Le lemme 11.313 le donne pour la racine carré.

— Le théorème 11.319 donne la densité des polynômes dans les fonctions continues sur un compact.

— Le théorème 11.316 est une généralisation qui donne la densité uniforme d'une sous-algèbre de $C(X, \mathbb{R})$ dès que X sépare les points.

— Le théorème 11.317 donne le même résultat pour la densité dans $C(X, \mathbb{C})$.

— Le lemme 27.1 est une version pour les polynômes trigonométriques.

— Le lemme 11.313 est un cas particulier du théorème 11.319, mais nous en donnons une démonstration indépendante afin d'isoler la preuve de la généralisation 11.317. Une version pour les polynômes trigonométriques sera donnée dans le lemme 27.1.

Le théorème de Stone-Weierstrass est utilisé, entre autres nombreuses choses, pour prouver la densité des polynômes trigonométriques dans les fonctions continues sur S^1, voir la proposition 26.87.

Thème 14 : espaces de fonctions En ce qui concerne les densités, voir le thème 13.

Topologie Les espaces de fonctions sont souvent munis de topologies définies par des semi-normes.

(1) La topologie des semi-normes est la définition 7.326.

(2) La définition 29.10 donne les topologies sur $C^\infty(\Omega)$, $\mathscr{D}(K)$ et $\mathscr{D}(\Omega)$.

(3) La topologie $*$-faible sur $\mathscr{D}'(\Omega)$ est donnée par la définition 29.17.

L'espace $L^2\big([0, 2\pi]\big)$ C'est un espace très important, entre autres parce qu'il est de Hilbert et est bien adapté à la transformée de Fourier.

(1) Un rappel de la construction en 26.72.

(2) Le produit scalaire $\langle f, g\rangle$ est donné en (26.331) et la base trigonométrique est (26.332).

(3) La densité des polynômes trigonométriques dans $L^p(S^2)$ est le théorème 26.70 ou le théorème 27.6, au choix.

(4) Une conséquence de cette densité est que le système trigonométrique est une base hilbertienne de L^2 par le lemme 26.107.

L'espace L^2 est discuté en analyse fonctionnelle, dans la section 26.4 et les suivantes parce que l'étude de L^2 utilise entre autres l'inégalité de Hölder 26.33.

Le fait que L^2 soit une espace de Hilbert est utilisé dans la preuve du théorème de représentation de Riesz 26.128.

Thème 15 : fonctions Lipschitz

(1) Définition : 11.259.

(2) La notion de Lipschitz est utilisée pour définir la stabilité d'un problème, définition 33.25.

Thème 16 : formule des accroissements finis Il en existe plusieurs formes :

(1) Une version adaptée aux espaces de dimension finie est le théorème 11.257.

(2) L'existence de $c \in]a, b[$ tel que

$$f'(c) = \frac{f(b) - f(a)}{b - a} \tag{-2.2}$$

est le théorème des accroissements finis proprement dit. C'est le théorème 11.130.

(3) Au premier ordre, proposition 11.113.

(4) Pour les fonctions $\mathbb{R} \to \mathbb{R}$ en le théorème 11.130.

(5) Une généralisation pour les intervalles non bornés : théorème 11.131.

(6) Espaces vectoriels normés, théorème 10.145

Thème 17 : limite et continuité

(1) Limite d'une fonction en un point : définition 7.67. Il n'y a pas unicité en général comme le montre l'exemple 7.33 dans un espace non séparé.

(2) La proposition 7.335 donne l'unicité de la limite dans le cas des espaces duaux pour la topologie $*$-faible. La proposition 7.70 nous dira qu'il y a unicité dès que l'espace d'arrivée est séparé.

(3) Définition de la continuité d'une fonction en un point et sur une partie de l'espace de départ : définiton 7.71.

(4) Continuité sur une partie si et seulement si continue en chaque point, c'est le théorème 7.74.

(5) Voir l'exemple 11.64 traité en détail.

(6) La fonction $f(x, y) = x + y$ est continue, lemme 7.270.

Thème 18 : différentiabilité

(1) La différentielle est définie en général pour des espaces vectoriels normés par la proposition 10.127

(2) Nous parlons de différentielle en dimension finie et donnons une interprétation géométrique en 11.20.1.

(3) La recherche d'extrémums d'une fonction sur $\mathbb{R}^n$ passe par la seconde différentielle, proposition 16.70.

(4) Lien entre différentielle seconde (hessienne) et convexité en la proposition 16.94 et le corolaire 16.96.

(5) Une fonction est de classe C^1 si et seulement si ses dérivées partielles sont continues, théorème 11.243.

(6) Une fonction est C^n si et seulement si ses dérivées partielles sont C^{n-1}, c'est le théorème 11.273.

(7) La différentielle est liée aux dérivées partielles par les formules données au lemme 11.196

$$df_a(u) = \frac{\partial f}{\partial u}(a) = \frac{d}{dt}\Big[f(a + tu)\Big]_{t=0} = \sum_{i=1}^{m} u_i \frac{\partial f}{\partial x_i}(a) = \nabla f(a) \cdot u.$$

Thème 19 : points fixes

(1) Il y a plusieurs théorèmes de points fixes.

Théorème de Picard 16.34 donne un point fixe comme limite d'itérés d'une fonction Lipschitz. Il aura pour conséquence le théorème de Cauchy-Lipschitz 16.40, l'équation de Fredholm, théorème 16.39 et le théorème d'inversion locale dans le cas des espaces de Banach 16.48.

Théorème de Brouwer qui donne un point fixe pour une application d'une boule vers elle-même. Nous allons donner plusieurs versions et preuves.

(1a) Dans $\mathbb{R}^n$ en version C^∞ via le théorème de Stokes, proposition 19.30.

(1b) Dans $\mathbb{R}^n$ en version continue, en s'appuyant sur le cas C^∞ et en faisant un passage à la limite, théorème 19.31.

(1c) Dans $\mathbb{R}^2$ via l'homotopie, théorème 25.18. Oui, c'est très loin. Et c'est normal parce que ça va utiliser la formule de l'indice qui est de l'analyse complexe [3].

Théorème de Markov-Kakutani 19.36 qui donne un point fixe à une application continue d'un convexe fermé borné dans lui-même. Ce théorème donnera la mesure de Haar 19.37 sur les groupes compacts.

Théorème de Schauder C'est une version valable en dimension infinie du théorème de Brouwer. Théorème 19.32

(2) Pour les équations différentielles

(2a) Le théorème de Schauder a pour conséquence le théorème de Cauchy-Arzela 19.33 pour les équations différentielles.

(2b) Le théorème de Schauder 19.32 permet de démontrer une version du théorème de Cauchy-Lipschitz (théorème 16.40) sans la condition Lipschitz, mais alors sans unicité de la solution. Notons que de ce point de vue nous sommes dans la même situation que la différence entre le théorème de Brouwer et celui de Picard : hors hypothèse de type « contraction », point d'unicité.

(3) En calcul numérique

— La convergence d'une méthode de point fixe est donnée par la proposition 33.47.

— La convergence quadratique de la méthode de Newton est donnée par le théorème 33.53.

— En calcul numérique, section 33.5

— Méthode de Newton comme méthode de point fixe, sous-section 33.6.2.

(4) D'autres utilisations de points fixes.

— Processus de Galton-Watson, théorème 37.53.

— Dans le théorème de Max-Milgram 24.59, le théorème de Picard est utilisé.

Thème 20 : théorèmes de Stokes, Green et compagnie

(1) Forme générale, théorème 19.71.

(2) Rotationnel et circulation, théorème 23.8.

Le théorème de Stokes peut être utilisé pour montrer le théorème de Brower, proposition 19.30.

Thème 21 : permuter des limites

Fonctions définies par une intégrale Les théorèmes sur les fonctions définies par une intégrale, section 16.4. Nous avons entre autres

(1) $\partial_i \int_B f = \int_B \partial_i f$, avec B compact, proposition 16.25.

(2) Si f est majorée par une fonction ne dépendant pas de x, nous avons le théorème 16.15 pour la continuité de $x \mapsto \int_\Omega f(x,\omega)d\mu(\omega)$.

(3) Pour la fonction $F(x) = \int_\Omega f(x,\omega)d\mu(\omega)$, nous avons la dérivation sous l'intégrale par la formule de Leibnitz

$$F'(a) = \int_\Omega \frac{\partial f}{\partial x}(a,\omega)d\mu(\omega) \tag{-2.3}$$

démontrée en le théorème 16.18.

3. On aime bien parce que ça ne demande pas Stokes, mais quand même hein, c'est pas gratos non plus.

(4) Si l'intégrale est uniformément convergente, nous avons le théorème 16.16 qui donne la continuité de $F(x) = \int_\Omega f(x,\omega)d\mu(\omega)$.

(5) Pour dériver $\int_B g(t,z)dt$ avec B compact dans $\mathbb{R}$ et $g\colon \mathbb{R}\times\mathbb{C}\to\mathbb{C}$, il faut aller voir la proposition 25.16.

(6) En ce qui concerne le x dans la borne, le théorème 13.235 lie primitive et intégrale en montrant que $F(x) = \int_a^x f(t)dt$ est une primitive de f (sous certaines conditions). Le théorème fondamental de l'analyse 13.236 en est une conséquence.

Conegence monotone Théorème de la convergence monotone, théorème 13.160.

Fubini Le théorème de Fubini permet non seulement de permuter des intégrales, mais également des sommes parce que ces dernières peuvent être vues comme des intégrales sur $\mathbb{N}$ muni de la tribu des parties et de la mesure de comptage [4]. Nous utilisons cette technique pour permute une somme et une intégrale dans l'équation (25.124).

L'utilisation de Fubini pour permuter des intégrales (sur deux variables différentes) ou deux sommes est expliquée dans 13.263.

C'est par exemple utilisé pour permuter deux sommes dans le cadre des chaines de Markov en 37.8.

— le théorème de Fubini-Tonelli 13.259 demande que la fonction soit mesurable et positive ;

— le théorème de Fubini 13.262 demande que la fonction soit intégrable (mais pas spécialement positive) ;

— le corolaire 13.261 demande l'intégrabilité de la valeur absolue des intégrales partielles pour déduire que la fonction elle-même est intégrable.

Limite et dérivées, différentielle (1) Permuter limite et dérivée, théorème 11.311.

(2) Permuter limite et dérivées partielles, théorème 11.312.

(3) Permuter limite et différentielle, théorème 14.6.

Quelques remarques sur les techniques de démonstration.

(1) Le résultat fondamental 11.311 est démontré sans recourir à des intégrales. Une preuve alternative, plus courte, avec des intégrales est donnée en 13.243.

(2) Les résultats un peu plus élaborés 11.312 et 14.6 sont prouvés avec des intégrales, mais devraient pouvoir être adaptés.

Somme et dérivée Permuter somme et différentielle, théorème 14.6.

Limite et mesure Une mesure n'est pas toujours une limite, mais la définition d'une mesure positive sur un espace mesurable parle de permuter limite et mesure : définition 13.20(2).

Thème 22 : applications continues et bornées

(1) Une application linéaire non continue : exemple 10.28 de $e_k \mapsto ke_k$. Les dérivées partielles sont calculées en (24.135).

(2) La dérivation sur les polynômes (exemple 10.29) donne un autre exemple d'application linéaire non continue.

(3) Une application linéaire est bornée si et seulement si elle est continue, proposition 10.18.

Thème 23 : inégalités

Inégalité de Jensen (1) Une version discrète pour $f\big(\sum_i \lambda_i x_i\big)$, la proposition 16.100.

(2) Une version intégrale pour $f\big(\int \alpha d\mu\big)$, la proposition 26.31.

(3) Une version pour l'espérance conditionnelle, la proposition 35.60.

Inégalité pour les normes ℓ^p (1) Hölder pour L^p : $\|fg\|_1 \leqslant \|f\|_p\|g\|_q$, proposition 26.33.

4. Mesure de comptage, définition 13.230.

(2) Hölder pour ℓ^p : $\|x\|_q \leqslant n^{\frac{1}{q}-\frac{1}{p}}\|x\|_p$, proposition 16.106.

Inégalité de Minkowsky (1) Pour une forme quadratique[5] q sur $\mathbb{R}^n$ nous avons $\sqrt{q(x+y)} \leqslant \sqrt{q(x)} + \sqrt{q(y)}$. Proposition 9.261.

(2) Si $1 \leqslant p < \infty$ et si $f, g \in L^p(\Omega, \mathcal{A}, \mu)$ alors $\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$. Proposition 26.37.

(3) L'inégalité de Minkowsky sous forme intégrale s'écrit sous forme déballée

$$\left[\int_X \left(\int_Y |f(x,y)| d\nu(y)\right)^p d\mu(x)\right]^{1/p} \leqslant \int_Y \left(\int_X |f(x,y)|^p d\mu(x)\right)^{1/p} d\nu(y).$$

ou sous forme compacte

$$\left\| x \mapsto \int_Y f(x,y) d\nu(y) \right\|_p \leqslant \int_Y \|f_y\|_p d\nu(y)$$

Transformée de Fourier Pour tout $f \in L^1(\mathbb{R}^n)$ nous avons $\|\hat{f}\|_\infty \leqslant \|f\|_1$, lemme 28.12.

Inégalité des normes Inégalité de normes : si $f \in L^p$ et $g \in L^1$, alors $\|f * g\|_p \leqslant \|f\|_p \|g\|_1$, proposition 26.56.

Thème 24 : connexité

(1) Définition 7.42

(2) L'image d'un connexe par une fonction continue est connexe, proposition 7.90.

(3) Connexité par arcs, définition 7.161.

(4) Une partie de $\mathbb{R}^2$ qui est connexe, mais pas connexe par arcs, proposition 20.56.

(5) Une partie de $\mathbb{R}$ est connexe si et seulement si elle est un intervalle, proposition 7.153.

(6) Le groupe $\mathrm{SL}(n, \mathbb{K})$ est connexe par arcs : proposition 12.16.

(7) Le groupe $\mathrm{GL}(n, \mathbb{C})$ est connexe par arcs : proposition 12.17.

(8) Le groupe $\mathrm{GL}(n, \mathbb{C})$ est connexe par arcs, proposition 12.17.

(9) Le groupe $\mathrm{GL}(n, \mathbb{R})$ a exactement deux composantes connexes par arcs, proposition 12.18.

(10) Le groupe $\mathrm{O}(n, \mathbb{R})$ n'est pas connexe, lemme 12.13.

(11) Les groupes $\mathrm{U}(n)$ et $\mathrm{SU}(n)$ sont connexes par arcs, lemme 12.14.

(12) Pour tout $n \geqslant 2$, le groupe $\mathrm{SO}(n)$ est connexe, le groupe $\mathrm{O}(n)$ a deux composantes connexes, proposition 12.3.

(13) Connexité des formes quadratiques de signature donnée, proposition 16.112.

Thème 25 : suite de Cauchy, espace complet Nous parlons d'espaces topologiques complets. À ne pas confondre avec un espace mesuré complet, définition 13.56.

(1) Corps complet : définition 1.109(6), espace métrique complet : définition 7.277.

(2) La définition 7.276 donne la notion de suite de Cauchy dans un espace métrique.

(3) La définition 7.274 donne la notion de suite de τ-Cauchy dans un espace vectoriel topologique.

(4) Deux espaces métriques (avec une distance) peuvent être isomorphes en tant qu'espaces topologiques, mais ne pas avoir les mêmes suites de Cauchy, exemple 7.280.

(5) La proposition 7.281 donne l'équivalence entre les suites de Cauchy et les suites τ-Cauchy dans le cas des espaces vectoriels topologiques *normés*.

(6) L'exemple 7.280 est un exemple pire que simplement une suite de Cauchy qui ne converge pas. Le problème de convergence de cette suite n'est pas simplement que la limite n'est pas dans l'espace ; c'est que la suite de Cauchy donnée ne convergerait même pas dans $\mathbb{R}$.

5. Définition 9.160.

(7) Le théorème 16.125 est un théorème de complétion d'un espace métrique.

(8) Dans $\mathbb{R}$, une suite est convergente si et seulement si elle est de Cauchy, théorème 7.294(2).

(9) Toute suite convergente dans un espace métrique est de Cauchy, proposition 7.282.

Quelques espaces qui sont complets sont listés ci-dessous. Attention : la complétude est bien une propriété de la métrique ; le même ensemble peut être complet pour une distance et pas pour une autre. Souvent, cependant la distance à considérer est donnée par le contexte.

(1) Les réels $\mathbb{R}$, théorème 7.294.

(2) Un espace vectoriel normé sur un corps complet est complet, proposition 7.298.

(3) La proposition 11.296 donne quelques espaces complets. Soit X un espace topologique métrique (Y, d) un espace espace métrique complet. Alors les espaces
   (3a) $\big(C_b^0(X, Y), \|.\|_\infty\big)$
   (3b) $\big(C_0^0(X, Y), \|.\|_\infty\big)$
   (3c) $\big(C_0^k(X, Y), \|.\|_\infty\big)$
   sont complets.

(4) Le lemme 11.297 dit que $\big(C^0(A, B), \|.\|_\infty\big)$ est complet dès que A est compact et B est complet.

(5) L'espace $\mathscr{D}(K)$ est complet tant pour la topologie des semi-normes que pour la topologie métrique (qui sont les mêmes). C'est la proposition 29.14.

(6) L'espace $\mathscr{S}(\Omega)$ est complet et métrisable par la proposition 29.56.

(7) L'espace $L^p(\Omega, \mathcal{A}, \mu)$ par le théorème 26.40.

La limite uniforme d'une suite de fonctions dérivables n'est pas spécialement dérivable. Même si les fonctions sont de classe C^∞, la limite n'est pas spécialement mieux que continue. En effet, le théorème de Stone-Weierstrass 11.319 nous dit que les polynômes (qui sont C^∞) sont denses dans les fonctions continues sur un compact pour la norme uniforme. Vous ne pouvez donc pas espérer que $\big(C^p(X, Y), \|.\|_\infty\big)$ soit complet en général.

Thème 26 : application réciproque

(1) Définition 7.82.

(2) Dans le cas des réels, des exemples sont donnés en 11.7.

(3) Continuité, proposition 7.84.

(4) Théorème de la bijection 11.61 (qui contient aussi de la continuité).

(5) Dérivabilité, proposition 11.120.

Thème 27 : déduire la nullité d'une fonction depuis son intégrale Des résultats qui disent que si $\int f = 0$ c'est que $f = 0$ dans un sens ou dans un autre.

(1) Il y a le lemme 13.179 qui dit ça.

(2) Un lemme du genre dans L^2 existe aussi pour $\int f\varphi = 0$ pour tout φ. C'est le lemme 26.59.

(3) Et encore un pour L^p dans la proposition 26.132.

(4) Si $\int f\chi = 0$ pour tout χ à support compact alors $f = 0$ presque partout, proposition 29.1.

(5) La proposition 26.22 donne $f = 0$ dans L^p lorsque $\int fg = 0$ pour tout $g \in L^q$.

(6) Une fonction $h \in C_c^\infty(I)$ admet une primitive dans $C_c^\infty(I)$ si et seulement si $\int_I h = 0$. Théorème 16.2.

Thème 28 : équations différentielles L'utilisation des théorèmes de point fixe pour l'existence de solutions à des équations différentielles est fait dans le chapitre sur les points fixes.

(1) Le théorème de Schauder a pour conséquence le théorème de Cauchy-Arzela 19.33 pour les équations différentielles.

(2) Le théorème de Schauder 19.32 permet de démontrer une version du théorème de Cauchy-Lipschitz (théorème 16.40) sans la condition Lipschitz

(3) Le théorème de Cauchy-Lipschitz 16.40 est utilisé à plusieurs endroits :
   — Pour calculer la transformée de Fourier de $e^{-x^2/2}$ dans le lemme 28.22.
(4) Théorème de stabilité de Lyapunov 31.34.
(5) Le système proie-prédateur de Lotka-Volterra 31.35
(6) Équation de Schrödinger, théorème 31.41.
(7) L'équation $(x - x_0)^\alpha u = 0$ pour $u \in \mathscr{D}'(\mathbb{R})$, théorème 29.64.
(8) La proposition 31.37 donne un résultat sur $y'' + qy = 0$ à partir d'une hypothèse de croissance.
(9) Équation de Hill $y'' + qy = 0$, proposition 31.39.

Thème 29 : injections

(1) L'espace de Sobolev $H^1(I)$ s'injecte de façon compacte dans $C^0(\bar{I})$, proposition 30.6.
(2) L'espace de Sobolev $H^1(I)$ s'injecte de façon continue dans $L^2(I)$, proposition 30.6.
(3) L'espace $L^2(\Omega)$ s'injecte continument dans $\mathscr{D}'(\Omega)$ (les distributions), proposition 29.20.

Thème 30 : logarithme

(1) Le logarithme pour les réels strictement positifs $\ln\colon]0, \infty[\to \mathbb{R}$ est donné en la définition 14.58.
(2) Les principales propriétés sont dans la proposition 14.60 : $\ln(xy) = \ln(x) + \ln(y)$ etc.
(3) La proposition 14.74 donne la série

$$\ln(1 + x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k. \tag{-2.4}$$

(4) L'exemple 19.146 donne l'encadrement $0.644 \leqslant \ln(2) \leqslant 0.846$.
(5) La proposition 14.96 dit que toute matrice complexe admet un logarithme. En particulier une série explicite est donnée pour le logarithme d'un bloc de Jordan.
(6) Sur les complexes, le logarithme $\ln\colon \mathbb{C}^* \to \mathbb{C}$ est la définition 25.50. Attention : ce n'est pas la seule définition possible.
(7) La série harmonique diverge à vitesse logarithmique, et la série des inverses des nombres premiers, c'est encore plus lent : théorème 14.85.

Thème 31 : inversion locale, fonction implicite

Des énoncés Il existe plusieurs énoncés à différent niveaux de généralité.

(1) Inversion locale dans $\mathbb{R}^n$: théorème 16.47. Pour un Banach c'est le théorème 16.48.
(2) Fonction implicite dans un Banach : théorème 16.49.

Des utilisations (1) Utilisé pour montrer que le flot d'une équation différentielle est un C^p-difféomorphisme local, voir 31.30.
(2) Pour le théorème de Von Neumann 16.62.

Thème 32 : convexité L'essentiel des résultats sur les fonctions convexes sont dans la section 16.11. On a surtout :

(1) Définition des fonctions convexes : 16.76 et 16.92 en dimension supérieure.
(2) En termes de différentielles, 16.93 pour la différentielle première et 16.96 pour la hessienne.
(3) Une courbe paramétrée convexe est la définition 20.90.
(4) L'enveloppe convexe d'une courbe fermée simple et convexe : 20.92.
(5) Courbure et convexité d'une courbe paramétrée : section 20.13.4.

(6) Une courbe paramétrée convexe est localement le graphe d'une fonction convexe par le lemme 20.91.

(7) La convexité est utilisée dans la méthode du gradient à pas optimal de la proposition 16.109.

En termes de parties convexes, on :

(1) Définition 8.25 d'une partie convexe d'un espace vectoriel.

(2) Une boule est convexe, proposition 8.26.

Thème 33 : fonction puissance Il y a beaucoup de choses à dire...

Définition Nous considérons, pour $a > 0$, la fonction $g_a \colon \mathbb{R} \to \mathbb{R}$ donnée par $g_a(x) = a^x$. La définition de cette fonction se fait en de nombreuses étapes.

(1) a^n pour $n \in \mathbb{N}$ en la définition 1.95.

(2) a^n pour $n \in \mathbb{Z}$ en la définition 11.323.

(3) $a^{1/n}$ pour $n \in \mathbb{Z}$ en la définition 11.325.

(4) a^q pour $q \in \mathbb{Q}$ en la définition 11.325.

(5) $\sqrt[n]{x}$ en la définition 11.327.

(6) La fonction g_a est Cauchy-continue sur $\mathbb{Q}$, c'est la proposition 11.338.

(7) a^x pour $a > 0$ et $x \in \mathbb{R}$ en la définition 11.340.

(8) a^z pour $a > 0$ et $z \in \mathbb{C}$ en la définition 17.4.

Quelques propriétés (1) Pour tout $q \in \mathbb{Q}$, il y a un $\sqrt{q}$ dans $\mathbb{R}$, proposition 1.166.

(2) Pour $a > 0$ et $x, y \in \mathbb{R}$ nous avons $a^x a^y = a^{x+y}$, proposition 11.341.

(3) Si $a > 0$ et $x, y \in \mathbb{R}$ nous avons $(a^x)^y = (a^y)^x = a^{xy}$ par la proposition 11.351.

(4) La fonction puissance est strictement croissante (en ses deux arguments), proposition 11.346.

(5) La formule $a^{-x} = 1/a^x$ est la proposition 11.341(3).

Dérivation Comme toutes les choses sur la fonction puissance, les preuves sont assez différentes selon que l'on parle de a^x ou de x^a.

(1) La fonction puissance est strictement croissante, proposition 11.346

(2) La fonction a^x est dérivable et sa dérivée vérifie $g_a'(x) = g_a(x)g_a'(0)$, proposition 11.354.

(3) La formule de dérivation pour $x \mapsto x^q$ avec $q \in \mathbb{Q}$ est la proposition 11.363.

(4) La dérivation de $x \mapsto x^\alpha$ avec $\alpha \in \mathbb{R}$ est la proposition 13.245. Si elle est tellement loin, c'est parce qu'elle nécessite de permuter une limite de fonctions avec une dérivée.

(5) Pour la formule générale de dérivation de $x \mapsto a^x$ demande de savoir les logarithmes (proposition 14.66).

L'équation fonctionnelle L'exponentielle et plus généralement la fonction puissance $g_a(x) = a^x$ peut être introduite au moyen d'une équation fonctionnelle au lieu de l'équation différentielle usuelle. Cette fameuse équation fonctionnelle est

$$f(x + y) = f(x)f(y) \tag{-2.5}$$

en la définition 11.356.

(1) L'équivalence entre l'équation fonctionnelle et l'équation différentielle est donnée par la proposition 11.361.

(2) La fonction $g_a(x) = a^x$ vérifie l'équation fonctionnelle $g_a(x + y) = g_a(x)g_a(y)$ et les conséquences. C'est la définition 11.356 et les choses qui suivent.

(3) L'équation fonctionnelle pour une fonction continue $f \colon \mathbb{R} \to S^1$ est traitée dans la proposition 11.367.

Une définition alternative de la fonction puissance serait de poser directement

$$a^x = e^{x\ln(a)}.$$

De là les propriétés se déduisent facilement. Dans cette approche, les choses se mettent dans l'ordre suivant :

— Définir $\exp(x)$ par sa série pour tout x.

— Démontrer que $\exp(q) = \exp(1)^q$ pour tout rationnel q (première partie de la proposition 14.57).

— Définir $e = \exp(1)$.

— Définir, pour x irrationnel, $a^x = \exp(x\ln(a))$.

— Prouver que $e^x = \exp(x)$ pour tout x.

Thème 34 : dualité Ne pas confondre dual algébrique et dual topologique d'un espace vectoriel.

(1) Définition de la base duale 4.130.

(2) Base préduale (existencec, unicité) : proposition 4.133.

Dual topologique et algébrique Ils sont définis par 4.129. Le dual algébrique est l'ensemble des formes linéaires, et le dual topologique ne considère que les formes linéaires continues (en dimension infinie, les applications linéaires ne sont pas toutes continues).

Topologie Une topologie possible sur le dual d'un espace vectoriel topologique est celle $*$-faible de la définition 7.333.

Nous comparons les topologies faibles et de la norme en la section 10.9.

Théorèmes de dualité Quelques théorèmes établissent des dualités entre des espaces courants.

(1) Le théorème de représentation de Riesz 24.17 pour les espaces de Hilbert.

(2) La proposition 26.128 pour les espaces $L^p\big([0,1]\big)$ avec $1 < p < 2$.

(3) Le théorème de représentation de Riesz 26.130 pour les espaces L^p en général.

Tous ces théorèmes donnent la dualité par l'application $\Phi_x = \langle x, . \rangle$.

Thème 35 : opérations sur les distributions

(1) Convolution d'une distribution par une fonction, définition par l'équation (29.96).

(2) Dérivation d'une distribution, proposition-définition 29.22.

(3) Produit d'une distribution par une fonction, définition 29.21.

Thème 36 : convolution

(1) Définition 26.52, et principales propriétés sur $L^1(\mathbb{R})$ dans le théorème 26.53.

(2) Inégalité de normes : si $f \in L^p$ et $g \in L^1$, alors $\|f * g\|_p \leqslant \|f\|_p \|g\|_1$, proposition 26.56.

(3) $\varphi \in L^1(\mathbb{R})$ et $\psi \in \mathscr{S}(\mathbb{R})$, alors $\varphi * \psi \in \mathscr{S}(\mathbb{R})$, proposition 26.155.

(4) Les suites régularisantes : $\lim_{n\to\infty} \rho_n * f = f$ dans la proposition 28.18.

(5) Convolution d'une distribution par une fonction, définition par l'équation (29.96).

Thème 37 : séries de Fourier

— Formule sommatoire de Poisson, proposition 28.15.

— Inégalité isopérimétrique, théorème 27.19.

— Fonction continue et périodique dont la série de Fourier ne converge pas, proposition 27.18.

— Nous allons montrer la convergence de $\sum_{k\in\mathbb{Z}} c_k(f)e^{inx}$ vers $f(x)$ dans divers cas :

(1) Si f est continue et périodique, convergence au sens de Cesaro, théorème de Fejèr 27.5.

(2) Convergence au sens $L^2\Big([0,2\pi]\Big)$ dans le théorème 26.108.

(3) Si f est continue, périodique et si sa série de Fourier converge uniformément, théorème 27.13.

(4) Si f est périodique et la série des coefficients converge absolument pour tout x, proposition 27.14.

(5) Si f est périodique et de classe C^1, théorème 27.15.

Il est cependant faux de croire que la continuité et la périodicité suffisent à obtenir une convergence, comme le montre dans la proposition 27.18.

Thème 38 : transformée de Fourier

(1) Définition sur L^1, définition 28.1.

(2) La transformée de Fourier d'une fonction $L^1(\mathbb{R}^d)$ est continue, proposition 28.11.

(3) L'espace de Schwartz est stable par transformée de Fourier. L'application $\mathcal{F}\colon \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ est continue. Proposition 28.20

(4) L'application $\mathcal{F}\colon \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ est une bijection. Formule d'inversion, proposition 28.26.

Thème 39 : méthode de Newton

(1) Nous parlons un petit peu de méthode de Newton en dimension 1 dans 33.6.

(2) La méthode de Newton fonctionne bien avec les fonctions convexes par la proposition 33.55.

(3) La méthode de Newton en dimension n est le théorème 33.61.

(4) Un intervalle de convergence autour de α s'obtient par majoration de $|g'|$, proposition 33.47.

(5) Un intervalle de convergence quadratique s'obtient par majoration de $|g''|$, théorème 33.53.

(6) En calcul numérique, section 33.6.

Thème 40 : méthodes de calcul

(1) Théorème de Rothstein-Trager 19.93.

(2) Algorithme des facteurs invariants 4.113.

(3) Méthode de Newton, théorème 33.61

(4) Calcul d'intégrale par suite équirépartie 27.8.

Thème 41 : espaces vectoriels

(1) Existence d'une base. Pour un espace vectoriel quelconque, proposition 4.20.

(2) Théorème de la base incomplète. Pour un espace vectoriel quelconque, théorème 4.21.

Thème 42 : définie positive

(1) Une application bilinéaire est définie positive lorsque $g(u,u) \geqslant 0$ et $g(u,u) = 0$ si et seulement si $u = 0$ est la définition 7.185.

(2) Un opérateur ou une matrice est défini positif si toutes ses valeurs propres sont positives, c'est la définition 9.152.

(3) Pour une matrice symétrique, définie positive si et seulement si $\langle Ax, x\rangle > 0$ pour tout x. C'est le lemme 9.157.

(4) Une application linéaire est définie positive si et seulement si sa matrice associée l'est. C'est la proposition 9.156.

Remarque : nous ne définissons pas la notion de matrice définie positive dans le cas d'une matrice non symétrique.

Thème 43 : norme matricielle, norme opérateur et rayon spectral Quelques définitions

(1) Définition de la norme opérateur : définition 10.3.

(2) Définition du rayon spectral 10.8.

La norme matricielle n'est rien d'autre que la norme opérateur de l'application linéaire donnée par la matrice.

(1) Lien entre norme matricielle et rayon spectral, le théorème 10.21 assure que $\|A\|_2 = \sqrt{\rho(A^tA)}$.

(2) Lien entre valeurs propres et norme opérateur : le lemme 10.22 pour les matrices symétriques strictement définies positives donne $\|A\|_2 = \lambda_{max}$.

(3) Pour toute norme algébrique nous avons $\rho(A) \leqslant \|A\|$, proposition 10.15.

(4) Dans le cadre du conditionnement de matrice. Voir en particulier la proposition 33.108 qui utilise le théorème 10.21.

(5) Rayon spectral et convergence de méthode itérative, proposition 33.144.

Pour la norme opérateur nous avons les résultats suivants.

(1) La majoration $\|Au\| \leqslant \|A\|\|u\|$ est le lemme 10.10.

(2) Définition d'une algèbre : 3.83 et pour une norme d'algèbre : 10.7.

(3) La norme opérateur est une norme d'algèbre, lemme 10.13.

(4) Pour des espaces vectoriels normés, être borné est équivalent à être continu : proposition 10.18.

(5) Le lemme à propos d'exponentielle de matrice 14.119 donne :

$$\|e^{tA}\| \leqslant P\big(|t|\big) \sum_{i=1}^{r} e^{t\,\mathrm{Re}(\lambda_i)}.$$

La norme opérateur est utilisée pour donner une norme sur les produits tensoriels, définition 10.123.

Une norme matricielle donne une topologie. Il y a donc également des liens entre rayon spectral et convergence de série. Dans cette optique, pour les séries de matrices, voir le thème 44.

Thème 44 : série de matrices

(1) Rayon spectral et norme opérateur : thème 43.

(2) Exponentielle de matrices : thème 58.

(3) Série entière de matrices : section 14.12.

(4) Pour la série $\sum_k A^k = (1-A)^{-1}$.
   - Pour un espace de Banach : proposition 10.152.
   - Pour les matrices nilpotentes : proposition 9.123.
   - En lien avec le rayon spectral (si et seulement si $\rho(A) < 1$) dans la proposition 14.115.
   - Le lemme 14.28 parle de la série entière $\sum_{n\in\mathbb{N}} z^{nk} = (1-z^k)^{-1}$.

   Cette série est utilisée entre autres dans la proposition 33.164 pour prouver qu'une M-matrice irréductible vérifie $A^{-1} > 0$.

Thème 45 : rang

(1) Définition pour une application linéaire : 4.38, pour une matrice : 4.103. L'équivalence est la proposition 4.106.

(2) Le théorème du rang, théorème 4.39

(3) Pour une applicaton liénaire entre deux espaces vectoriels de même dimension finie, il est équivalent d'être injectif, surjectif ou bijectif, c'est le corollaire 4.41.

(4) Pour prouver que des matrices sont équivalentes et pour les mettre sous des formes canoniques, nous avons le lemme 4.111 et son corolaire 4.112.

(5) Tout hyperplan de $\mathbb{M}(n, \mathbb{K})$ coupe $\mathrm{GL}(n, \mathbb{K})$, corolaire 4.112. Cela utilise la forme canonique sus-mentionnée.

(6) Le lien entre application duale et orthogonal de la proposition 4.142 utilise la notion de rang.

(7) Le lemme 9.232 parle de commutant et utilise la notion de rang. Ce lemme sert à prouver diverses conditions équivalentes à être un endomorphisme cyclique dans le théorème 9.233.

Thème 46 : extension de corps et polynômes

(1) Définition d'une extension de corps 6.52.

(2) Pour l'extension du corps de base d'un espace vectoriel et les propriétés d'extension des applications linéaires, voir la section 9.12.

(3) Extension de corps de base et similitude d'application linéaire (ou de matrices, c'est la même chose), théorème 9.235.

(4) Extension de corps de base et cyclicité des applications linéaires, corolaire 9.234.

(5) À propos d'extensions de $\mathbb{Q}$, le lemme 6.167.

(6) Corps de rupture : définition 6.104 existence par la proposition 6.110. Il n'y a pas unicité.

(7) Corps de décomposition : définition 6.131. Attention : le plus souvent nous parlons de corps de décomposition d'un seul polynôme. Cette définition est un peu surfaite. Existence par la proposition 6.132 qui le donne même comme extension par toutes les racines, et unicité à isomorphisme près par le théorème 6.134, énoncé de façon plus simple (mais pas indépendante !) en la proposition 6.135.

(8) Si $\mathbb{K}$ est algébrique clos et si $\alpha\colon \mathbb{K} \to \mathbb{L}$ est une extension algébrique, alors $\alpha(\mathbb{K}) = \mathbb{L}$ par lemme 6.70.

Un trio de résultats d'enfer est :

(1) Dans un anneau principal qui n'est pas un corps, un idéal est maximal si et seulement si il est engendré par un irréductible (proposition 3.124).

(2) Dans un anneau, un idéal I est maximal si et seulement si A/I est un corps (proposition 3.56)

(3) Si $\mathbb{K}$ est un corps, $\mathbb{K}[X]$ est principal (lemme 3.178).

Thème 47 : décomposition de matrices

(1) Décomposition de Bruhat, théorème 12.36.

(2) Décomposition de Dunford, théorème 9.196.

(3) Décomposition polaire 12.30 et la proposition 16.58 pour la régularité.

Thème 48 : systèmes d'équations linéaires

— Algorithme des facteurs invariants 4.113.

— La méthode du gradient à pas optimal permet de résoudre par itérations $Ax = b$ lorsque A est symétrique strictement définie positive. Il s'agit de minimiser une fonction bien choisit. Propositions 16.108 pour l'existence et 16.109 pour la méthode.

Thème 49 : formes bilinéaires et quadratiques

(1) Les formes bilinéaires sont définies en 7.182.

(2) Forme quadratique, définition 9.160. Sa matrice, définition 9.164.

(3) Équivalence de forme quadratiques, définition 9.187. Deux formes quadratiques sont équivalentes si et seulement si elles ont même signature, proposition 9.188.

(4) Une isométrie d'une forme bilinéaire est affine ou linéaire, théorème 9.249.

(5) Forme bilinéaire dégénérée, définition 9.246.

(6) Une forme bilinéaire est non-dégénérée si et seulement si sa matrice associée est inversible, c'est la proposition 9.248.

(7) Une isométrie d'une forme bilinéaire est linéaire ou affine par le théorème 9.249.

(8) Toute forme quadratique admet des bases orthogonales, théorème 9.174 pour le cas général ; proposition 9.175 pour le cas de $\mathbb{R}^n$, en se basant sur le théorème spectral.

(9) Théorème de Sylvester à propos de signature (définition 9.180) de forme quadratique réelle : 9.186.

(10) Base q-orthogonale pour une forme quadratique, théorème 9.174.

Thème 50 : arithmétique modulo, théorème de Bézout

(1) Pour $\mathbb{Z}^*$ c'est le théorème 3.13.

(2) Théorème de Bézout dans un anneau principal : corolaire 3.133.

(3) Théorème de Bézout dans un anneau de polynômes : théorème 6.41.

(4) En parlant des racines de l'unité et des générateurs du groupe unitaire dans le lemme 18.4. Au passage nous y parlerons de solfège.

(5) La proposition 3.15 qui donne tout entier assez grand comme combinaisons de a et b à coefficients positifs est utilisée en chaines de Markov, voir la définition 37.47 et ce qui suit.

(6) PGCD et PPCM sont dans la définition 1.81.

(7) Calcul effectif du PGCD puis des coefficients de Bézout : sous-sections 3.3.4.1 et 3.3.4.2.

Thème 51 : polynômes

Définitions Soient un anneau A, un corps $\mathbb{K}$, une extension $\mathbb{L}$ de $\mathbb{K}$ et un élément $\alpha \in \mathbb{L}$.

(1) La définition la plus formelle est en tant que module produit $A^{(\mathbb{N})}$, définition 3.86. Le produit et l'évaluation sont définis en 3.87 et la formule $(PQ)(x) = P(x)Q(x)$ dans 3.88.

(2) $A[X]$, définition 3.86 ; l'anneau $\mathbb{K}[X]$ a même définition parce que c'est un cas particulier. L'évaluation d'un polynôme en un élément de l'anneau, $P(\alpha)$ est définie en 3.87.

(3) Liens entre $\mathbb{K}[\alpha]$, $\mathbb{K}[X]$, $\mathbb{K}(\alpha)$ et $\mathbb{K}(X)$ lorsque α est transcendant, proposition 6.84. Et la proposition 6.87 pour le cas où α est algébrique[6].

(4) Si A est un anneau et si α est un élément d'une extension de A (comme anneau), nous écrivons $A[\alpha]$ pour le plus petit sous-anneau de B contenant A et α. C'est la définition 3.170.

(5) $\mathbb{K}(X)$, le corps des fractions de $\mathbb{K}[X]$, définition 6.74. Si $R = P/Q$ dans $\mathbb{K}(X)$, l'évaluation est $R(\alpha) = P(\alpha)Q(\alpha)^{-1}$, définition 6.75.

(6) $\mathbb{K}(\alpha)_{\mathbb{L}}$ est le plus petit corps de $\mathbb{L}$ contenant $\mathbb{K}$ et α, définition 6.76.

(7) À propos de polynômes à plusieurs variables.

— Anneau de polynômes : $A[X_1, \dots, X_n]$ est la définition 3.91.

— Corps engendré : $\mathbb{K}(\alpha_1, \dots, \alpha_n)$ est la définition 6.128.

— Corps des fractions rationnelles : $\mathbb{K}(X_1, \dots, X_n)$ est la définition 6.129.

Coefficients dans un anneau commutatif (1) Les polynômes à coefficients dans un anneau commutatif sont à la section 3.14.

Coefficients dans un corps Les polynômes à coefficients dans un corps sont à la section 6.3.

(1) Nous parlons de l'idéal des polynômes annulateurs dans le théorème 6.37.

(2) Le théorème 6.37 dit que $\mathbb{K}[X]$ est un anneau principal et que tous ses idéaux sont engendrés par un unique polynôme unitaire.

6. Définition 6.62.

(3) Le polynôme minimal est irréductible, proposition 6.59.

(4) Quelques formules sur le pgcd, lemme 6.50.

Polynôme primitif (1) Un polynôme est irréductible sur A si et seulement si irréductible et primitif sur le corps des fractions, corolaire 3.196.

Polynôme d'endomorphisme C'est la section 9.6.

Racines et factorisation (1) Si $\mathbb{A}$ est un anneau, la proposition 3.188 factorise une racine.

(2) Si $\mathbb{A}$ est un anneau, la proposition 3.192 factorise une racine avec sa multiplicité.

(3) Si $\mathbb{A}$ est un anneau, le théorème 3.194 factorise plusieurs racines avec leurs multiplicités.

(4) Le théorème 3.194 nous indique que lorsqu'on a autant de racines (multiplicité comprise) que le degré, alors nous avons toutes les racines.

(5) Si $\mathbb{K}$ est un corps et α une racine dans une extension, le polynôme minimal de α divise tout polynôme annulateur par la proposition 6.85.

(6) Le théorème 6.101 annule un polynôme de degré n ayant $n+1$ racines distinctes.

(7) La proposition 6.170 nous annule un polynôme à plusieurs variables lorsqu'il a trop de racines.

(8) En analyse complexe, le principe des zéros isolés 16.126 annule en gros toute série entière possédant un zéro non isolé.

(9) Polynômes irréductibles sur $\mathbb{F}_q$.

Thème 52 : zoologie de l'algèbre Nous listons ici un peu tous les termes qui arrivent en algèbre.

Éléments Pour les éléments, nous avons :

(0a) Élément irréductible en 3.110.

(0b) Élément premier en 3.135.

Anneaux Pour les anneaux, nous avons :

(0a) Anneau factoriel en 3.114.

(0b) Anneau principal en 3.118.

(0c) Anneau intègre en 1.88.

(0d) Anneau noetherien en 3.141.

Idéaux Pour les idéaux, nous avons :

(0a) Idéal principal en 3.117.

(0b) Idéal premier en 3.119.

Thème 53 : invariants de similitude

(1) Théorème 9.221.

(2) Pour prouver que la similitude d'applications linéaires résiste à l'extension du corps de base, théorème 9.235.

(3) Pour prouver que la dimension du commutant d'un endomorphisme de E est de dimension au moins $\dim(E)$, lemme 9.232.

(4) Nous verrons dans la remarque 9.222 à propos des invariants de similitude que toute matrice est semblable à la matrice bloc-diagonale constituées des matrices compagnon (définition 9.216) de la suite des polynômes minimals.

Thème 54 : réduction, diagonalisation Des résultats qui parlent diagonalisation

(1) Définition d'un endomorphisme diagonalisable : 9.125.

(2) Conditions équivalentes au fait d'être diagonalisable en termes de polynôme minimal, y compris la décomposition en espaces propres : théorème 9.128.

(3) Diagonalisation simultanée 9.131, pseudo-diagonalisation simultanée 9.254.

(4) Diagonalisation d'exponentielle 14.97 utilisant Dunford.

(5) Décomposition polaire théorème 12.30. $M = SQ$, S est symétrique, réelle, définie positive, Q est orthogonale.

(6) Décomposition de Dunford 9.196. $u = s + n$ où s est diagonalisable et n est nilpotent, $[s, n] = 0$.

(7) Réduction de Jordan (bloc-diagonale) 9.224.

(8) L'algorithme des facteurs invariants 4.113 donne $U = PDQ$ avec P et Q inversibles, D diagonale, sans hypothèse sur U. De plus les éléments de D forment une chaine d'éléments qui se divisent l'un l'autre.

Le théorème spectral et ses variantes :

(1) Théorème spectral, matrice symétrique, théorème 9.150. Via le lemme de Schur complexe 9.137.

(2) Théorème spectral autoadjoint (c'est le même, mais vu sans matrices), théorème 9.267

(3) Théorème spectral hermitien, lemme 9.144.

(4) Théorème spectral, matrice normales, théorème 9.147.

Pour les résultats de décomposition dont une partie est diagonale, voir le thème 47 sur les décompositions. Réduction de quadriques :

(1) Réduction de Gauss, théorème 9.171.

Thème 55 : endomorphismes cycliques

(1) Définition 9.99.

(2) Son lien avec le commutant donné dans la proposition 9.230 et le théorème 9.233.

(3) Utilisation dans le théorème de Frobenius (invariants de similitude), théorème 9.221.

Thème 56 : déterminant

(1) Déterminant d'une matrice : définition 4.74.

(2) Déterminant d'un endomorphisme 9.8.

(3) Principales propriétés algébriques du déterminant : la proposition 9.10.

(4) Déterminant et manipulations de lignes et colonnes, section 4.3.10 et les propositions qui précèdent à partir du lemme 4.76 qui dit que $\det(A) = \det(A^t)$.

(5) Les n-formes alternées forment un espace de dimension 1, proposition 9.3.

(6) Déterminant d'une famille de vecteurs 9.4.

(7) Calcul d'un déterminant de taille 2×2 : équation (4.102).

(8) Interprétations géométriques

  (8a) À propos d'orthogonalité, le déterminant est très lié au produit vectoriel en dimension 3. Et il le généralise en dimension supérieure.

    i. Liaison au produit vectoriel (orthogonalité) dans la proposition 7.214.

    ii. En particulier le lemme 7.215 nous dit comment un déterminant donne un vecteur orthogonal à une famille donnée de vecteurs.

  (8b) Déterminant et aires, volumes

i. Déterminant et mesure de Lebesgue : théorème 13.254.

ii. Aire du parallélogramme : proposition 17.44.

iii. Volume du parallélépipède avec le produit mixte et le déterminant 3×3, 17.45.

Tant que nous en sommes dans les interprétations géométriques, il faut lier déterminant, produit vectoriel, orthogonalité et mesure en notant que l'élément de volume lors de l'intégration en dimension 3 est donné par (19.186) : $dS = \|T_u \times T_v\|$ qui est la norme du produit vectoriel des vecteurs tangents au paramétrage.

Nous voyons dans l'équation (19.183) que l'élément de volume pour une partie de dimension n dans $\mathbb{R}^m$ (à l'occasion d'y intégrer une fonction) est donné par un déterminant mettant en jeu les vecteurs tangents du paramétrage.

(9) Le déterminant de Vandermonde est à la proposition 9.12. Il est utilisé à divers endroits :

(9a) Pour prouver que u est nilpotente si et seulement si $\mathrm{Tr}(u^p) = 0$ pour tout p (lemme 9.121)

(9b) Pour prouver qu'un endomorphisme possédant $\dim(E)$ valeurs propres distinctes est cyclique (proposition 9.230).

Thème 57 : polynôme d'endomorphismes

(1) Endomorphismes cycliques et commutant dans le cas diagonalisable, proposition 9.230.

(2) Racine carrée d'une matrice hermitienne positive, proposition 12.25.

(3) Théorème de Burnside sur les sous-groupes d'exposant fini de $\mathrm{GL}(n, \mathbb{C})$, théorème 9.242.

(4) Décomposition de Dunford, théorème 9.196.

(5) Algorithme des facteurs invariants 4.113.

Thème 58 : exponentielle Toutes les exponentielles sont définies par la série

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!},$$

tant que la somme a un sens.

Réels Voici le plan que nous suivons dans le Frido :

— L'exponentielle est définie par sa série en 14.37.

— Nous démontrons qu'elle vérifie l'équation différentielle $y' = y$, $y(0) = 1$ (théorème 14.54).

— Nous démontrons l'unicité de la solution à cette équation différentielle.

— Nous démontrons qu'elle est égale à $x \mapsto y(1)^x$. Cela donne la définition du nombre e comme valant $y(1)$.

— Nous définissons le logarithme comme l'application réciproque de l'exponentielle (définition 14.58).

— Les fonctions trigonométriques (sinus et cosinus) sont définies par leurs séries. Il est alors montré que $e^{ix} = \cos(x) + i\sin(x)$.

Propriétés — La formule $a^{-x} = 1/a^x$ est la proposition 11.341(3).

Complexes (1) Le fait que $e^{i\theta}$ donne tous les nombres complexes de norme 1 est la proposition 17.48.

(2) Le groupe des racines de l'unité est donné par l'équation (18.1).

Algèbre normée commutative Pour la définition c'est la proposition 14.37 et pour la régularité C^∞ c'est la proposition 14.41.

Idem non commutatif Il y a une tentative de théorème 14.42, mais c'est principalement pour les matrices qu'il y a des résultats.

Matrices De nombreux résultats sont disponibles pour les exponentielles de matrices.

(1) Les sections 10.8 et 14.4.4 parlent d'exponentielle de matrices.

(2) L'exponentielle donne lieu à une fonction de classe C^∞, proposition 14.116.

(3) Le lemme à propos d'exponentielle de matrice 14.119 donne :

$$\|e^{tA}\| \leqslant P\big(|t|\big) \sum_{i=1}^{r} e^{t\,\mathrm{Re}(\lambda_i)}.$$

(4) La proposition 14.97 : si $A \in \mathbb{M}(n, \mathbb{R})$ a un polynôme caractéristique scindé, alors A est diagonalisable si et seulement si e^A est diagonalisable.

(5) La section 14.12.3 parle des fonctions exponentielle et logarithme pour les matrices. Entre autres la dérivation et les séries.

(6) Pour résoudre des équations différentielles linéaires : sous-section 31.6.1.

(7) La proposition 14.96 dit que l'exponentielle est surjective sur $\mathrm{GL}(n, \mathbb{C})$.

(8) La proposition 10.155 : si u est un endomorphisme, alors $\exp(u)$ est un polynôme en u.

(9) Calcul effectif : sous-section 14.12.4.

(10) Proposition 12.21 : si $A \in \mathbb{M}(n, \mathbb{C})$ alors $e^{\mathrm{Tr}(A)} = \det(e^A)$.

(11) Les séries entières de matrices sont traitées autour de la proposition 14.113.

Thème 59 : types d'anneaux

(1) Définition d'anneau : définition 1.73.

(2) La définition d'anneau principal est 3.118 ; pour un idéal principal, c'est 3.117.

(3) $\mathbb{Z}$ est intègre, exemple 3.97, principal et euclidien (proposition 3.154).

(4) $\mathbb{Z}[X]$ n'est pas principal (voir (5)).

(5) Si A est un anneau intègre qui n'est pas un corps, alors $A[X]$ n'est pas principal, lemme 3.155.

(6) L'anneau des fonctions holomorphes sur un compact donné est principal, proposition 25.20.

(7) L'anneau $\mathbb{Z}[i\sqrt{3}]$ n'est pas factoriel, exemple 3.116.

(8) L'anneau $\mathbb{Z}[i\sqrt{5}]$ n'est ni factoriel ni principal, exemple 3.145.

(9) Tous les idéaux de $\mathbb{Z}/6\mathbb{Z}$ sont principaux, mais $\mathbb{Z}/6\mathbb{Z}$ n'est pas principal. Exemple 3.128.

Thème 60 : sous-groupes

(1) Théorème de Burnside sur les sous-groupes d'exposant fini de $\mathrm{GL}(n, \mathbb{C})$, théorème 9.242.

(2) Sous-groupes compacts de $\mathrm{GL}(n, \mathbb{R})$, lemme 12.37 ou proposition 12.38.

Thème 61 : groupe symétrique

(1) Définition 2.63.

(2) La signature $\epsilon\colon S_n \to \{-1, 1\}$ est l'unique homomorphisme surjectif de S_n sur $\{-1, 1\}$, proposition 2.76(1).

(3) La table des caractères du groupe symétrique S_4 est donné dans la section 15.5.

(4) Le groupe symétrique S_4 est le groupe des symétries affines du tétraèdre régulier, proposition 17.164.

(5) Le groupe alterné A_5 est l'unique groupe simple d'ordre 60, proposition 5.40.

(6) La proposition 5.30 donne la position du groupe alterné dans le groupe symétrique : A_n est un sous-groupe caractéristique de S_n et l'unique sous-groupe d'indice 2.

Thème 62 : action de groupe

(1) Définition d'une action de groupe sur un ensemble : 2.46.

(2) Action du groupe modulaire sur le demi-plan de Poincaré, théorème 22.93.

(3) La formule de Burnside (théorème 2.59) parle du nombre d'orbites pour l'action d'un groupe fini sur un ensemble fini.

(4) Des applications de la formule de Burnside : le jeu de la roulette et l'affaire du collier, 17.8.2.1 et 17.8.2.2.

(5) Le groupe symétrique S_n agit sur l'anneau $\mathbb{K}[T_1, \ldots, T_n]$, lemme 3.95.

Thème 63 : classification de groupes

(1) Structure des groupes d'ordre pq, théorème 5.25.

(2) Le groupe alterné est simple, théorème 5.36.

(3) Définition 5.6 d'un p-groupe.

(4) Théorème de Sylow 5.11.

(5) Théorème de Burnside sur les sous-groupes d'exposant fini de $\mathrm{GL}(n, \mathbb{C})$, théorème 9.242.

(6) $(\mathbb{Z}/p\mathbb{Z})^* \simeq \mathbb{Z}/(p-1)\mathbb{Z}$, corolaire 18.34.

Thème 64 : produit semi-direct de groupes

(1) Définition 2.79.

(2) Le corolaire 2.81 donne un critère pour prouver qu'un produit NH est un produit semi-direct.

(3) L'exemple 17.87 donne le groupe des isométries du carré comme un produit semi-direct.

(4) Le théorème 5.25 classifie les groupes d'ordre pq (p, q premiers distincts) à grands coups de produit semi-directs.

(5) Le théorème 17.63 donne les isométries de $\mathbb{R}^n$ par $\mathrm{Isom}(\mathbb{R}^n) = T(n) \times_\rho O(n)$ où $T(n)$ est le groupe des translations.

(6) La proposition 17.65 donne une décomposition du groupe orthogonal $\mathrm{O}(n) = \mathrm{SO}(n) \times_\rho C_2$ où $C_2 = \{\mathrm{Id}, R\}$ où R est de déterminant -1.

(7) La proposition 8.58 donne $\mathrm{Aff}(\mathbb{R}^n) = T(n) \times_\rho \mathrm{GL}(n, \mathbb{R})$ où $\mathrm{Aff}(\mathbb{R}^n)$ est le groupe des applications affines bijectives de $\mathbb{R}^n$.

Thème 65 : théorie des représentations

(1) Définition 4.149.

(2) Table des caractères du groupe diédral, section 17.16.

(3) Table des caractères du groupe symétrique S_4, section 15.5.

Thème 66 : isométries Il y a $(\mathbb{R}^n, \|.\|)$ et $\mathbb{R}^n, d$.

Les isométries de $\|.\|$ sont linéaires, tandis que les isométries de la distance contiennent aussi les translations et les rotations de centre différent de l'origine.

Ne pas confondre une isométrie d'un espace affine avec une isométrie d'un espace euclidien. Les isométries d'un espace euclidien préservent le produit scalaire et fixent donc l'origine (lemme 7.197). Les isométries des espaces affines par contre conservent les distances (définition 8.59) et peuvent donc déplacer l'origine de l'espace vectoriel sur lequel il est modelé ; typiquement les translations sont des isométries de l'espace affine mais pas de l'espace euclidien.

Parfois, lorsqu'on coupe les cheveux en quatre, il faut faire attention en parlant de $\mathbb{R}^n$: soit on en parle comme d'un espace métrique (muni de la distance), soit on en parle comme d'un espace normé (muni de la norme ou du produit scalaire).

Général Quelques résultats généraux et en vrac à propos d'isométries.

(1) Définition d'une isométrie pour une forme bilinéaire, 9.178. Pour une forme quadratique : définition 9.177.

(2) Définition du groupe orthogonal 9.40, et le spécial orthogonal $\mathrm{SO}(n)$ en la définition 9.43. Le groupe $\mathrm{SO}(2)$ est le groupe des rotations, par corolaire 17.133.

(3) La rotation $R_A(\theta)$ d'un angle θ autour du point $A \in \mathbb{R}^2$ est donnée par la définition 17.122.

(4) La proposition 17.134 donne à toute rotation $R_0(\theta)$ une matrice de la forme connue. C'est autour de cela que nous définissons les angles, définition 17.152.

(5) Le groupe orthogonal est le groupe des isométries de $\mathbb{R}^n$, proposition 9.42.

(6) Les isométries de l'espace euclidien sont affines, 9.249.

(7) Les isométries de l'espace euclidien comme produit semi-direct : $\mathrm{Isom}(\mathbb{R}^n) \simeq T(n) \times_\rho \mathrm{O}(n)$, théorème 17.63.

(8) Isométries du cube, section 5.7.

(9) Nous parlons des isométries affines du tétraèdre régulier dans la proposition 17.164.

Groupe diédral Le groupe diédral est un peu central dans la théorie des isométries de $(\mathbb{R}^2, d)$ parce que beaucoup de sous-groups finis des isométries de $(\mathbb{R}^2, d)$ sont en fait isomorphes au groupe diédral.

(1) Générateurs du groupe diédral, proposition 17.85.

(2) Un sous-groupe fini des isométries de $(\mathbb{R}^2, d)$ contenant au moins une réflexion est isomorphe au groupe diédral par le théorème 17.169.

(3) Le théorème 17.171 dit que le groupe des isométries propres d'une partie quelconque de $(\mathbb{R}^2, d)$ est soit cyclique soit isomorphe au groupe diédral.

Isométries et réflexions Dans un espace euclidien, toute isométrie peut être décomposée en réflexions autour d'hyperplans. Voici quelques énoncés à ce propos.

(1) Définition d'une réflexion dans $\mathbb{R}^2$ 17.116.

(2) La caractérisation en termes de projection orthogonale est le lemme 17.103 ; en termes de médiatrice c'est le lemme 17.108.

(3) Définition d'un hyperplan 9.236.

(4) En dimension 2, une rotation est définie comme composée de deux réflexions en la définition 17.113.

(5) En dimension 2, les réflexions ont un déterminant -1 par le lemme 17.126.

(6) Les isométries du plan $(\mathbb{R}^2, d)$ sont données dans le théorème 17.159, et sont au plus 3 réflexions par le théorème 17.157.

(7) Décomposition des isométries de $\mathbb{R}^n$ en réflexions par le lemme 17.70.

(8) En particulier, les éléments de $\mathrm{SO}(3)$ sont des compositions de deux réflexions par le corolaire 17.72.

(9) Une isométrie de $\mathbb{R}^n$ préserve l'orientation si et seulement si est elle composition d'un nombre pair de réflexions. C'est le théorème 17.74.

Sous-groupe fini (1) Les sous-groupes finis des isométries de $(\mathbb{R}^2, d)$ sont cycliques, théorème 17.169.

(2) Les sous-groupes finis de $\mathrm{SO}(3)$ sont listés dans 17.193.

(3) Les sous-groupes finis de $\mathrm{SO}(2)$ sont cycliques, lemme 17.137.

Thème 67 : caractérisation de distributions en probabilités

(1) La probabilité conjointe est la définition 35.19.

(2) La fonction de répartition est la définition 35.61.

(3) La fonction caractéristique est la définition 35.63.

Thème 68 : théorème central limite

(1) Pour les processus de Poisson, théorème 39.5.

Thème 69 : lemme de transfert Il y a deux résultats qui portent ce nom. Le premier est dans la théorie de Fourier, le résultat $\hat{f}' = i\xi\hat{f}$.

(1) Lemme 28.19 sur $\mathscr{S}(\mathbb{R}^d)$

(2) Lemme 30.10 pour L^2.

L'autre lemme de transfert est en théorie des tribus, le résultat $\sigma\big(f^{-1}(\mathcal{C})\big) = f^{-1}\big(\sigma(\mathcal{C})\big)$ du lemme 13.47. Celui-ci est d'ailleurs plutôt nommé « lemme de transport ».

Thème 70 : probabilités et espérances conditionnelles Les deux définitions de base, sur lesquelles se basent toutes les choses conditionnelles sont :

— L'espérance conditionnelle d'une variable aléatoire sachant une tribu : $E(X|\mathcal{F})$ de la définition 35.37.

Les autres sont listées ci-dessous.

Probabilité conditionnelle .

Plusieurs probabilités conditionnelles.

— D'un événement en sachant un autre : la définition 35.33 donne

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Cela est la définition de base. L'autre est une définition dérivée.

— D'un événement vis-à-vis d'une variable aléatoire discrète. C'est par la définition 35.55 qui définit la variable aléatoire

$$P(A|X)(\omega) = P(A|X = X(\omega)).$$

Dans le cas continu, c'est la définition 35.56 :

$$P(A|X) = P(A|\sigma(X)) = E(\mathbb{1}_A|\sigma(X)).$$

— D'un événement par rapport à une tribu. C'est la variable aléatoire

$$P(A|\mathcal{F}) = E(\mathbb{1}_A|\mathcal{F}).$$

Espérances conditionnelles Plusieurs espérances conditionnelles.

— D'une variable aléatoire par rapport à un évélement, définition 35.49 :

$$E(X|A) = \frac{E(X\mathbb{1}_A)}{P(A)}. \tag{-2.6}$$

— d'une variable aléatoire par rapport à une tribu. La variable aléatoire $E(X|\mathcal{F})$ est la variable aléatoire $\mathcal{F}$-mesurable telle que

$$\int_B E(X|\mathcal{F}) = \int_B X$$

pour tout $X \in \mathcal{F}$. Si $X \in L^2(\Omega, \mathcal{A}, P)$ alors $E(X|\mathcal{F}) = \mathtt{proj}_K(X)$ où K est le sous-ensemble de $L^2(\Omega, \mathcal{A}, P)$ des fonctions $\mathcal{F}$-mesurables (théorème 35.37). Cela au sens des projections orthogonales.

— d'une variable aléatoire par rapport à une autre. La définition 35.39 est une variation sur le même thème :

$$E(X|Y) = E(X|\sigma(Y)),$$

Notons que partout, si X est une variable aléatoire, la notation « sachant X » est un raccourci pour dire « sachant la tribu engendrée par X ».

Quelque formules.

(1) Pour l'espérance conditionnelle d'une variable aléatoire prenant seulement une quantité dénombrable de valeurs : $E(X|A) = \sum_{k=0}^{\infty} y_k P(X = y_k|A)$ par le lemme 35.50.

(2) La probabilité conditionnelle se factorise par rapport à l'union disjointe par le lemme 35.36 : $P\big(\bigcup_{i=0}^{\infty} A_i|B\big) = \sum_{i=0}^{\infty} P(A_i|B)$.

Thème 71 : dénombrements

— Coloriage de roulette (17.8.2.1) et composition de colliers (17.8.2.2).

— Nombres de Bell, théorème 14.126.

— Le dénombrement des solutions de l'équation $\alpha_1 n_1 + \dots \alpha_p n_p = n$ utilise des séries entières et des décompositions de fractions en éléments simples, théorème 18.29.

Thème 72 : enveloppes

(1) L'ellipse de John-Loewner donne un ellipsoïde de volume minimum autour d'un compact dans $\mathbb{R}^n$, théorème 16.117.

(2) Le cercle circonscrit à une courbe donne un cercle de rayon minimal contenant une courbe fermée simple, proposition 20.89.

(3) Enveloppe convexe du groupe orthogonal 12.35.

(4) Enveloppe convexe d'une courbe fermée plane comme intersection des demi-plans tangents, proposition 20.95.

Thème 73 : équations diophantiennes

(1) Équation $ax + by = c$ dans $\mathbb{N}$, équation (3.62).

(2) Dans 3.3.10, nous résolvons $ax + by = c$ en utilisant Bézout (théorème 3.13).

(3) L'exemple 3.159 donne une application de la pure notion de modulo pour $x^2 = 3y^2 + 8$. Pas de solutions.

(4) L'exemple 3.160 résout l'équation $x^2 + 2 = y^3$ en parlant de l'extension $\mathbb{Z}[i\sqrt{2}]$ et de stathme.

(5) Les propositions 3.164 et 3.166 parlent de triplets pythagoriciens.

(6) Le dénombrement des solutions de l'équation $\alpha_1 n_1 + \dots \alpha_p n_p = n$ utilise des séries entières et des décompositions de fractions en éléments simples, théorème 18.29.

(7) La proposition 1.41 donne une bijection $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$ en résolvant dans $\mathbb{N}$ (entre autres) l'équation $k = y^2 + x$ pour k fixé.

Thème 74 : techniques de calcul Il y en a pour tous les gouts.

Primitives et intégrales Toute la section 16.12 donne des trucs et astuces pour trouver des primitives et des intégrales.

Limite à deux variables Les exemples de limites à plusieurs variables font souvent intervenir des coordonnées polaires (du théorème 17.195) ou autres fonctions trigonométriques. Ils sont donc placés beaucoup plus bas que la théorie.

— Méthode du développement asymptotique, sous-section 17.13.2.

— Méthode des coordonnées polaires, sous-section 17.13.1.

— Utilisation du théorème de la fonction implicite, dans l'exemple 17.197.

-2.1 Conventions sur les matrices et changement de bases

Le lien entre matrice et application linéaire est donné par la définition 4.65. L'application d'une matrice à un vecteur est (4.84). Le lien le plus simple entre l'application linéaire et les éléments de matrice est donné par la proposition 4.68. Voici les relations :

$$T_{\alpha i} = T(e_i)_\alpha \tag{-2.7a}$$

$$T(e_i) = \sum_\alpha T_{\alpha i} f_\alpha \tag{-2.7b}$$

$$T(x) = \sum_{i\alpha} T_{\alpha i} x_i f_\alpha \tag{-2.7c}$$

$$T(x)_\alpha = \sum_i T_{\alpha i} x_i. \tag{-2.7d}$$

Lorsque nous avons une base orthonormée nous avons aussi les propositions 4.123 et 4.123 qui donnent des formules avec produit scalaire :

(1) $T_{\alpha i} = e_\alpha \cdot T(e_i)$

(2) $x \cdot Ay = \sum_{kl} A_{kl} x_k y_l$.

où le point est le produit scalaire usuel de $\mathbb{R}^n$.

-2.1.1 Le changement de base

Soit un espace vectoriel V muni de deux bases $(e_i)_{i=1,\dots,n}$ et $(f_\alpha)_{\alpha=1,\dots,n}$. Le lemme 4.114 donne le lien entre les vecteurs de base :

(1) $f_\alpha = \sum_i Q_{i\alpha} e_i$

(2) $e_i = \sum_\alpha Q^{-1}_{\alpha i} f_\alpha$

La proposition 4.115 donne un certain nombre de formules pour les coordonnées des vecteur :

(1) $y_\alpha = \sum_i Q^{-1}_{\alpha i} x_i$

(2) $x_i = \sum_\alpha Q_{i\alpha} y_\alpha$.

(3) $x_i = (Qy)_i$

(4) $x = Qy$

La transformation de la matrice d'une application linéaire lors d'un changement de base est la proposition 4.118. Soit une application linéaire $T\colon V \to V$ de matrices A et B dans les bases $\{e_i\}$ et $\{f_\alpha\}$. Si les bases sont liées par $f_\alpha = \sum_i Q_{i\alpha} e_i$, alors les matrices A et B sont liées par

$$B = Q^{-1} A Q. \tag{-2.8}$$

-2.1.2 Changement de base : matrice d'une forme bilinéaire

La proposition 9.169 fait le changement de matrice d'une forme bilinéaire lors d'un changement de base. Si la matrice de q dans la base $\{e_i\}$ est A et celle dans la base $\{f_\alpha\}$ est B, alors

$$B = Q^t A Q. \tag{-2.9}$$

Pour comparaison avec la loi de transformation des matrices des applications liénaires, voir la remarque 9.170.

Table des matières

Thématique **2**
-2.1 Conventions sur les matrices et changement de bases . . . 30
-2.1.1 Le changement de base . . . 30
-2.1.2 Changement de base : matrice d'une forme bilinéaire . . . 30

Table des matières **31**

Index **61**

Liste des notations **84**

0 Introduction (Vol 1) **89**
0.1 Auteurs, contributeurs, sources et remerciements . . . 89
0.1.1 Ceux qui ont travaillé sur le Frido . . . 89
0.1.2 Aide directe, mais pas volontairement sur le Frido . . . 90
0.1.3 Des gens qui ont fait un travail qui m'a bien servi . . . 91
0.2 Originalité . . . 91
0.3 Les choses qui doivent vous faire tiquer . . . 92
0.4 Quelques choix qui peuvent provoquer des quiproquos . . . 92
0.5 Autres choix pas spécialement standards . . . 92
0.6 Sage est là pour vous aider . . . 93
0.6.1 Lancez-vous dans Sage . . . 93
0.6.2 Exemples de ce que Sage peut faire pour vous . . . 93
0.7 Comment contribuer et aider ? . . . 94
0.7.1 Des preuves qui manquent . . . 95
0.7.2 Du texte qui manque . . . 95
0.7.3 Mes questions de géométrie . . . 95
0.7.4 Mes questions d'algèbre . . . 96
0.7.5 Mes questions d'analyse . . . 97
0.7.6 Mes questions d'analyse numérique . . . 100
0.7.7 Mes questions de probabilité et statistiques. . . . 100
0.7.8 Mes questions de LaTeX et programmation . . . 100
0.7.9 Numérique . . . 101
0.8 Taper du code pour le Frido . . . 101
0.8.1 Pour compiler le document vous même . . . 101
0.8.2 Nommage des fichiers `tex` . . . 101
0.8.3 Inclure des exemples de code . . . 101
0.8.4 Pour les exercices . . . 101
0.9 Les politiques éditoriales . . . 102
0.9.1 Licence libre . . . 102
0.9.2 pdflatex . . . 102
0.9.3 utf8 . . . 102
0.9.4 Notations . . . 102
0.9.5 De la bibliographie . . . 102
0.9.6 Faire des références à tout . . . 102

0.9.7 Des listes de liens internes . . . 102
0.9.8 Pas de références vers le futur . . . 102
0.9.9 Écriture inclusive . . . 103
0.10 Vérifier si vous n'avez pas fait de bêtises . . . 103
0.11 Acceptation des contributions . . . 103
0.11.1 Attention aux expressions rationnelles . . . 103
0.11.2 Pas de modifications massives, automatiques pour des raisons cosmétiques 103

1 Construction des ensembles de nombres (Vol 1) 105
1.1 Quelques éléments sur les ensembles . . . 105
1.1.1 Petit mot d'introduction . . . 105
1.1.2 Ensemble ordonné . . . 106
1.1.3 Lemme de Zorn . . . 107
1.1.4 Complémentaire . . . 107
1.1.5 Relations d'équivalence . . . 108
1.2 Les naturels . . . 109
1.3 Quelques résultats de cardinalité . . . 109
1.3.1 Équipotence, surpotence, subpotence . . . 109
1.3.2 Un peu d'infinité . . . 110
1.3.3 Dénombrabilité et ensemble des naturels . . . 111
1.3.4 Théorème de Cantor-Schröder-Bernstein . . . 115
1.3.5 Comparabilité cardinale . . . 117
1.3.6 Théorème de Cantor, ensemble des ensembles . . . 117
1.3.7 Ajouter ou soustraire des cardinalités . . . 118
1.4 Groupes . . . 124
1.4.1 Définition, unicité du neutre . . . 124
1.5 Anneaux . . . 125
1.6 Les entiers . . . 127
1.6.1 Quelques autres résultats . . . 127
1.6.2 Anneau intègre . . . 128
1.7 Somme à valeurs dans un groupe abélien . . . 128
1.7.1 Somme à valeurs dans un anneau . . . 128
1.7.2 Somme à valeurs dans un groupe abélien . . . 129
1.7.3 Fonction puissance . . . 130
1.8 Corps . . . 131
1.8.1 Définitions, morphismes . . . 131
1.8.2 Corps des fractions . . . 132
1.8.3 Suites de Cauchy dans un corps totalement ordonné . . . 133
1.9 Les rationnels . . . 136
1.9.1 Suites de Cauchy dans les rationnels . . . 136
1.9.2 Insuffisance des rationnels . . . 138
1.10 Les réels . . . 141
1.10.1 L'ensemble . . . 141
1.10.2 Relation d'ordre . . . 144
1.10.3 Complétude . . . 147
1.10.4 Intervalles . . . 149
1.10.5 Maximum, supremum et compagnie . . . 150
1.10.6 Racines . . . 153
1.11 Les complexes . . . 154
1.11.1 Définitions . . . 155

2 Théorie des groupes (Vol 1) 157
2.1 Groupes . . . 157
2.2 Sous-groupe normal . . . 159

2.2.1 Classes de conjugaison . . . 160
2.3 Groupe dérivé . . . 161
2.4 Théorèmes d'isomorphismes . . . 161
2.5 Indice d'un sous-groupe et ordre des éléments . . . 164
2.6 Suite de composition . . . 165
2.7 Groupes résolubles . . . 168
2.8 Action de groupes . . . 169
2.9 Permutations, groupe symétrique . . . 174
2.9.1 Décomposition en cycles . . . 174
2.10 Produit semi-direct de groupes . . . 177
2.11 Groupe de torsion . . . 179
2.12 Famille presque nulle . . . 179

3 Anneaux (Vol 1) 181
3.1 Inversibles et nilpotents . . . 181
3.2 PGCD, PPCM et éléments inversibles . . . 182
3.3 Le groupe et anneau des entiers . . . 182
3.3.1 Division euclidienne . . . 182
3.3.2 Sous-groupes de $(\mathbb{Z}, +)$ . . . 183
3.3.3 PGCD, PPCM et Bézout . . . 183
3.3.4 Calcul effectif du PGCD et de Bézout . . . 186
3.3.5 Générateurs . . . 188
3.3.6 Générateurs pour le groupe multiplicatif . . . 188
3.3.7 Décomposition en facteurs premiers . . . 189
3.3.8 Ordre d'un élément dans un groupe fini . . . 191
3.3.9 Écriture des fractions . . . 193
3.3.10 Équation diophantienne linéaire à deux inconnues . . . 194
3.3.11 Quotients . . . 195
3.4 Binôme de Newton et morphisme de Frobenius . . . 196
3.5 Idéal dans un anneau . . . 196
3.5.1 Résultats supplémentaires sur l'anneau des entiers . . . 198
3.6 Caractéristique . . . 198
3.7 Module sur un anneau . . . 200
3.7.1 Module produit . . . 201
3.7.2 Sous-module . . . 202
3.8 Polynômes . . . 204
3.8.1 Polynômes d'une variable . . . 204
3.8.2 La notation $A[X]$ . . . 205
3.8.3 Polynômes de plusieurs variables . . . 206
3.8.4 Action du groupe symétrique . . . 207
3.9 Anneau intègre . . . 207
3.9.1 Caractéristique d'un anneau intègre . . . 208
3.9.2 Divisibilité et classes d'association . . . 208
3.9.3 PGCD et PPCM . . . 209
3.9.4 Anneaux intègres et corps . . . 210
3.9.5 Élément irréductible . . . 210
3.10 Anneau factoriel . . . 210
3.11 Anneau principal et idéal premier . . . 211
3.11.1 Bézout . . . 215
3.11.2 Élément premier . . . 216
3.11.3 Anneau noethérien . . . 217
3.12 Anneau $\mathbb{Z}/6\mathbb{Z}$ . . . 218
3.13 Anneau euclidien . . . 220
3.13.1 Équations diophantiennes . . . 223

3.13.2 Triplets pythagoriciens et équation de Fermat pour $n = 4$ . . . 224
3.14 Polynômes à coefficients dans un anneau commutatif . . . 227
3.14.1 Monômes . . . 227
3.14.2 Évaluation . . . 227
3.14.3 Polynômes sur un anneau intègre . . . 228
3.14.4 Division euclidienne . . . 228
3.14.5 Polynôme primitif . . . 229
3.14.6 Racines des polynômes . . . 230
3.14.7 Quelques identités . . . 232

4 Espaces vectoriels (début) (Vol 1) 233
4.1 Parties libres, génératrices, bases et dimension . . . 233
4.1.1 Et en dimension infinie . . . 239
4.1.2 Espace librement engendré . . . 240
4.2 Applications linéaires . . . 241
4.2.1 Définition . . . 241
4.2.2 Linéarité et bases . . . 244
4.2.3 Rang . . . 245
4.2.4 Injection, surjection . . . 248
4.3 Matrices . . . 250
4.3.1 Définitions . . . 250
4.3.2 Application linéaire associée . . . 252
4.3.3 Déterminant . . . 256
4.3.4 Déterminant en petite dimension . . . 256
4.3.5 Manipulations de lignes et de colonnes . . . 256
4.3.6 Réduction de Gauss . . . 262
4.3.7 Matrices inversibles . . . 264
4.3.8 Inversibilité et déterminant . . . 265
4.3.9 Quelques ensembles de matrices particuliers . . . 266
4.3.10 Déterminant et combinaisons de lignes et colonnes . . . 266
4.3.11 Transvections . . . 268
4.3.12 Mineur, rang . . . 269
4.3.13 Matrices équivalentes et semblables . . . 270
4.3.14 Algorithme des facteurs invariants . . . 271
4.4 Changement de base . . . 273
4.4.1 Changement de base : vecteurs de base . . . 273
4.4.2 Changement de base : coordonnées . . . 273
4.4.3 Changement de base : matrice d'une application linéaire . . . 274
4.5 Espaces de polynômes . . . 275
4.5.1 Projection et orthogonalité . . . 276
4.6 Dualité . . . 278
4.6.1 Orthogonal . . . 280
4.6.2 Transposée : pas d'approche naïve . . . 280
4.6.3 Transposée : la bonne approche . . . 282
4.6.4 Polynômes de Lagrange . . . 284
4.6.5 Dual de $\mathbb{M}(n, \mathbb{K})$ . . . 284
4.7 Représentation de groupe . . . 286

5 Classification de certains groupes (Vol 1) 287
5.1 Théorèmes de Sylow . . . 287
5.2 Groupe monogène . . . 291
5.3 Automorphismes du groupe $\mathbb{Z}/n\mathbb{Z}$ . . . 292
5.4 Groupes abéliens finis . . . 293
5.5 Groupes d'ordre pq . . . 296

5.6 Groupe symétrique, groupe alterné . . . 299
5.6.1 Le groupe alterné . . . 299
5.6.2 Sous-groupes normaux . . . 305
5.6.3 Indice . . . 306
5.7 Isométriques du cube . . . 308

6 Corps (Vol 1) 311
6.1 Généralités . . . 311
6.1.1 Corps ordonnés . . . 311
6.1.2 Automorphismes de $\mathbb{R}$ et $\mathbb{C}$ . . . 311
6.1.3 Corps premier . . . 313
6.1.4 Petit théorème de Fermat . . . 313
6.2 Théorème des deux carrés . . . 314
6.2.1 Un peu de structure dans $\mathbb{Z}[i]$ . . . 314
6.2.2 Résultats chinois . . . 318
6.3 Polynômes à coefficients dans un corps . . . 320
6.3.1 Irréductibilité . . . 320
6.3.2 Idéaux . . . 321
6.3.3 Bézout . . . 324
6.3.4 Lemme et théorème de Gauss . . . 324
6.3.5 Polynômes sur un corps et pgcd . . . 325
6.4 Extension de corps . . . 327
6.4.1 Extension et polynôme minimal . . . 328
6.4.2 Extensions algébriques et éléments transcendants . . . 330
6.4.3 Extension algébrique et polynôme minimal . . . 332
6.4.4 Extensions et polynômes . . . 333
6.4.5 Racines de polynômes . . . 343
6.4.6 Corps de rupture . . . 344
6.4.7 Pile d'extensions . . . 346
6.4.8 Clôture algébrique . . . 347
6.4.9 Polynômes à plusieurs variables . . . 353
6.4.10 Corps de décomposition . . . 353
6.4.11 Non irréductible ou pas corps ? . . . 357
6.4.12 Clôture algébrique . . . 358
6.4.13 Extensions séparables . . . 359
6.5 Idéal maximum . . . 363
6.5.1 Idéal maximum . . . 363
6.6 Polynômes symétriques et alternés . . . 365
6.6.1 Polynômes symétriques, alternés ou semi-symétriques . . . 365
6.6.2 Polynôme symétrique élémentaire . . . 365
6.6.3 Relations coefficients racines . . . 367
6.7 Minuscule morceau sur la théorie de Galois . . . 368

7 Topologie générale (Vol 1) 371
7.1 Éléments généraux de topologie . . . 371
7.1.1 Définitions et propriétés de base . . . 371
7.1.2 Quelques exemples . . . 372
7.1.3 Adhérence, fermeture, intérieur, point d'accumulation et isolé . . . 375
7.2 Suites et convergence . . . 377
7.2.1 Convergence dans un fermé . . . 377
7.2.2 Pour des limites uniques : séparabilité . . . 378
7.2.3 Fonctions équivalentes . . . 379
7.3 Connexité . . . 380
7.4 Compacité . . . 381

7.4.1 Définition et notions connexes . . . 381
7.4.2 Base de topologie . . . 383
7.4.3 Quelques propriétés . . . 384
7.4.4 Compactifié d'Alexandrov . . . 385
7.5 Limites et continuité de fonctions . . . 385
7.5.1 Limites . . . 385
7.5.2 Continuité . . . 386
7.5.3 Continuité et topologie induite . . . 389
7.5.4 Continuité et connexité . . . 390
7.5.5 Continuité et compacité . . . 391
7.6 Topologie, distances et normes . . . 391
7.6.1 Distance et topologie métrique . . . 392
7.6.2 Distance à un ensemble . . . 396
7.6.3 Norme . . . 396
7.7 Topologie sur l'ensemble des réels . . . 398
7.7.1 Compacité pour les réels . . . 399
7.7.2 Conséquence : les fermés bornés sont compacts . . . 401
7.7.3 Suites et limites dans les réels . . . 402
7.7.4 Opérations sur les limites . . . 403
7.7.5 Exemples . . . 403
7.7.6 Suites croissantes et bornées . . . 404
7.7.7 Suites adjacentes . . . 405
7.7.8 Limite supérieure et inférieure . . . 406
7.7.9 Ouverts, voisinage, topologie . . . 407
7.7.10 Intervalles et connexité . . . 408
7.7.11 Recouvrement par des intervalles ouverts . . . 411
7.7.12 Connexité par arcs . . . 412
7.7.13 Un peu de connexité par arcs . . . 413
7.7.14 Des exemples . . . 414
7.7.15 Topologie de la droite réelle complétée . . . 416
7.7.16 Quelques mots à propos de la droite réelle complétée . . . 417
7.7.17 Limite pointée ou épointée ? . . . 417
7.8 Formes bilinéaires et quadratiques . . . 420
7.9 Produit scalaire, produit hermitien . . . 420
7.9.1 Norme, produit scalaire et Cauchy-Schwarz (cas réel) . . . 421
7.9.2 Cauchy-Schwarz etc. cas complexe . . . 425
7.9.3 Projection et orthogonalité . . . 426
7.9.4 Théorème de Pythagore . . . 427
7.9.5 Produit vectoriel . . . 428
7.9.6 Produit mixte . . . 431
7.9.7 Procédé de Gram-Schmidt . . . 433
7.9.8 Approximation . . . 433
7.9.9 Quelques exemples de normes sur $\mathbb{R}^n$ . . . 434
7.10 Équivalence des normes . . . 435
7.10.1 En dimension finie . . . 436
7.10.2 Contre-exemple en dimension infinie . . . 437
7.11 Produit fini d'espaces vectoriels normés . . . 438
7.11.1 Distance et norme produit . . . 438
7.12 Topologie réelle en dimension n . . . 440
7.12.1 Ouverts et fermés . . . 440
7.12.2 Intérieur, adhérence et frontière . . . 441
7.12.3 Point d'accumulation, point isolé . . . 442
7.12.4 Limite de suite . . . 443

7.13 Topologie et distance . . . 444
7.13.1 Suites et espaces métriques . . . 445
7.13.2 Espace métrisable . . . 448
7.14 Suites de Cauchy, métrique et espaces complets . . . 448
7.14.1 Généralités . . . 448
7.14.2 Espace topologique métrique . . . 450
7.15 Topologie et espace vectoriel . . . 451
7.15.1 Espace vectoriel topologique . . . 451
7.15.2 Équivalence entre Cauchy et $\tau-$Cauchy . . . 454
7.16 Norme ; espace vectoriel normé . . . 456
7.17 Espaces métriques . . . 458
7.17.1 Espaces métrisables . . . 458
7.17.2 Fonctions continues . . . 458
7.17.3 Ensembles enchaînés . . . 464
7.17.4 Produit fini d'espaces métriques . . . 464
7.17.5 Équicontinuité . . . 465
7.17.6 Continuité uniforme . . . 466
7.18 Ensembles nulle part denses . . . 466
7.19 Topologie des semi-normes . . . 467
7.19.1 Espace dual . . . 469
7.19.2 Espace $C^k(\mathbb{R}, E')$ . . . 470
7.20 Espaces de Baire . . . 471

8 Espaces affines (Vol 1) 473
8.1 Vecteurs agissant sur un espace . . . 473
8.2 Repères cartésiens affines . . . 474
8.3 Classification affine des conique . . . 474
8.4 Applications affines . . . 476
8.4.1 Autres propriétés . . . 477
8.5 Isomorphismes . . . 478
8.6 Sous espaces affines . . . 479
8.7 Barycentre . . . 481
8.7.1 Sous-espaces affines . . . 481
8.7.2 Enveloppe convexe . . . 483
8.7.3 Applications affines et barycentre . . . 486
8.8 Repères, coordonnées cartésiennes et barycentriques . . . 487
8.8.1 Équation de droite . . . 490
8.8.2 Associativité, coordonnées barycentriques dans un triangle . . . 491
8.9 Applications affines sur $\mathbb{R}^n$ . . . 492
8.9.1 Structure de groupe pour les applications affines . . . 493
8.10 Isométries . . . 494

9 Espaces vectoriels (encore) (Vol 1) 495
9.1 Déterminants . . . 495
9.1.1 Formes multilinéaires alternées . . . 495
9.1.2 Déterminant d'une famille de vecteurs . . . 496
9.1.3 Déterminant d'un endomorphisme . . . 499
9.1.4 Déterminant de Vandermonde . . . 501
9.1.5 Déterminant de Gram . . . 504
9.1.6 Déterminant de Cauchy . . . 504
9.1.7 Matrice de Sylvester . . . 504
9.1.8 Théorème de Kronecker . . . 507
9.2 Orientation . . . 509
9.2.1 Cas vectoriel . . . 509

9.2.2 Cas affine . . . 510
9.3 Hermitien, orthogonal, adjoint . . . 511
9.3.1 Opérateur orthogonal, matrice orthogonale . . . 513
9.4 Topologie . . . 514
9.4.1 Boules et sphères . . . 514
9.4.2 Ouverts, fermés, intérieur et adhérence . . . 516
9.4.3 Point isolé, point d'accumulation . . . 522
9.4.4 Des exemples . . . 523
9.5 Valeur propre et vecteur propre . . . 524
9.5.1 Généralités . . . 524
9.5.2 Dans le vif du sujet . . . 526
9.6 Polynômes d'endomorphismes . . . 526
9.6.1 Polynômes d'endomorphismes . . . 526
9.6.2 Polynôme minimal et minimal ponctuel . . . 528
9.6.3 Polynôme caractéristique . . . 534
9.7 Diagonalisation et trigonalisation . . . 538
9.7.1 Matrices semblables . . . 538
9.7.2 Endomorphismes nilpotents . . . 538
9.7.3 Endomorphismes diagonalisables . . . 542
9.7.4 Diagonalisation : cas complexe, pas toujours . . . 545
9.7.5 Trigonalisation : généralités . . . 545
9.7.6 Trigonalisation : cas complexe . . . 546
9.7.7 Diagonalisation : cas complexe, ce qu'on a . . . 548
9.7.8 Diagonalisation : cas réel . . . 550
9.8 Formes bilinéaires et quadratiques . . . 553
9.8.1 Généralités . . . 554
9.8.2 Matrice associée à une forme bilinéaire . . . 554
9.8.3 Changement de base : matrice d'une forme bilinéaire . . . 555
9.8.4 Réduction de Gauss . . . 556
9.8.5 Orthogonalité . . . 559
9.8.6 Diagonalisation . . . 560
9.8.7 Isométrie, forme quadratique et bilinéaire . . . 561
9.9 Signature, théorème de Sylvester . . . 562
9.9.1 Invariance de la trace . . . 565
9.10 Fonctions . . . 565
9.11 Sous espaces caractéristiques . . . 566
9.11.1 Théorèmes de décomposition . . . 567
9.11.2 Diverses conséquences . . . 569
9.11.3 Valeurs singulières . . . 570
9.12 Extension du corps de base . . . 570
9.12.1 Extension des applications linéaires . . . 570
9.12.2 Projections . . . 572
9.12.3 Rang, polynôme minimal, polynôme caractéristique . . . 575
9.13 Frobenius et Jordan . . . 576
9.13.1 Matrice compagnon . . . 576
9.13.2 Réduction de Frobenius . . . 577
9.13.3 Forme normale de Jordan . . . 580
9.14 Commutant et endomorphismes cycliques . . . 581
9.14.1 Endomorphisme cyclique . . . 581
9.14.2 Commutant : cas diagonalisable . . . 582
9.14.3 Commutant : cas général . . . 584
9.15 Hyperplans et formes linéaires . . . 587
9.15.1 Trouver la matrice d'une symétrie donnée . . . 588

9.16 Théorème de Burnside . . . 590
9.16.1 Théorème de Lie-Kolchin . . . 592
9.17 Retour sur les formes bilinéaires et quadratiques . . . 595
9.17.1 Dégénérescence d'une forme bilinéaire . . . 595
9.17.2 Isométries . . . 595
9.17.3 Pseudo-réduction simultanée . . . 597
9.17.4 Topologie . . . 598
9.17.5 Diagonalisation . . . 599
9.17.6 Isotropie . . . 600
9.17.7 Inégalité de Minkowski . . . 600
9.17.8 Ellipsoïde . . . 601
9.18 Théorème spectral autoadjoint . . . 602
9.19 Système d'équations linéaires : méthode de Gauss . . . 605

10 Espaces vectoriels normés (Vol 2) 607
10.1 Norme opérateur . . . 607
10.1.1 Norme d'algèbre . . . 609
10.1.2 Matrices, spectre et norme . . . 610
10.1.3 Rayon spectral . . . 611
10.1.4 Normes de matrices et d'applications linéaires . . . 615
10.1.5 Application linéaire continue et bornée . . . 617
10.1.6 Suites . . . 620
10.1.7 Continuité du produit de matrices . . . 621
10.2 Applications multilinéaires . . . 622
10.3 Séries . . . 624
10.3.1 Les trois types de convergence . . . 625
10.4 Série réelle . . . 627
10.4.1 Critères de convergence absolue . . . 628
10.4.2 Critères de convergence simple . . . 629
10.4.3 Quelques séries usuelles . . . 630
10.4.4 Séries alternées . . . 632
10.4.5 Moyenne de Cesaro . . . 633
10.4.6 Écriture décimale d'un nombre . . . 633
10.4.7 Théorème de Banach-Steinhaus . . . 637
10.4.8 Convergence forte . . . 639
10.5 Sommes de familles infinies . . . 640
10.5.1 Convergence commutative . . . 640
10.5.2 Somme non dénombrables . . . 642
10.5.3 Sommes dénombrables . . . 643
10.6 Produit tensoriel d'espaces vectoriels . . . 646
10.6.1 Somme directe d'espaces vectoriels . . . 647
10.6.2 Les produits tensoriels . . . 648
10.6.3 Le produit tensoriel . . . 649
10.6.4 Bases . . . 653
10.6.5 Norme . . . 654
10.6.6 Applications bilinéaires, matrices et produit tensoriel . . . 655
10.6.7 Application d'opérateurs . . . 656
10.7 Calcul différentiel dans un espace vectoriel normé . . . 656
10.7.1 Définition de la différentielle . . . 656
10.7.2 Accroissements finis . . . 657
10.7.3 (non ?) Différentiabilité des applications linéaires . . . 657
10.7.4 Dérivation en chaine et formule de Leibnitz . . . 658
10.7.5 Différentiation de produit . . . 663
10.7.6 Formule des accroissements finis . . . 666

10.7.7 Applications multilinéaires . . . 668
10.7.8 Différentielle partielle . . . 670
10.7.9 L'inverse, sa différentielle . . . 671
10.8 Exponentielle de matrice . . . 673
10.9 Espace dual . . . 674
10.9.1 Topologies . . . 674
10.9.2 Réflexivité . . . 676
10.9.3 Module de continuité . . . 677
10.10 Mini introduction aux nombres p-adiques . . . 678
10.10.1 La flèche d'Achille . . . 678
10.10.2 La tortue et Achille . . . 678
10.10.3 Dans les nombres p-adiques, c'est vrai . . . 679

11 Analyse réelle (Vol 2) 681
11.1 Intervalles . . . 681
11.2 Application réciproque . . . 682
11.2.1 Définitions . . . 682
11.2.2 Graphe de la fonction réciproque . . . 683
11.3 Limite de fonctions . . . 683
11.3.1 Définition . . . 683
11.3.2 Quelques règles de calcul . . . 685
11.3.3 Limite en l'infini . . . 688
11.3.4 Limite en des nombres . . . 689
11.3.5 Limites quand tout va bien . . . 690
11.3.6 Limites de fonctions . . . 691
11.3.7 Limite à gauche et à droite . . . 692
11.4 Limite en compactifié d'Alexandrov . . . 692
11.5 Continuité . . . 693
11.5.1 Opération sur la continuité . . . 694
11.5.2 La fonction la moins continue du monde . . . 695
11.5.3 Approche topologique . . . 695
11.5.4 Continuité de la racine carrée, invitation à la topologie induite . . . 697
11.5.5 Prolongement par continuité . . . 699
11.5.6 Prolongement par continuité . . . 700
11.5.7 Théorème de la bijection . . . 701
11.6 Limite et continuité . . . 702
11.6.1 Règles simples de calcul . . . 704
11.6.2 Prolongement des rationnels vers les réels . . . 705
11.7 Espace des fonctions continues . . . 709
11.8 Uniforme continuité . . . 713
11.9 Fonctions sur un compact . . . 715
11.10 Polynômes . . . 716
11.10.1 Polynômes sur les réels . . . 716
11.10.2 Polynômes sur les complexes . . . 716
11.11 Dérivée : exemples introductifs . . . 722
11.11.1 La vitesse . . . 722
11.11.2 La tangente à une courbe . . . 722
11.11.3 L'aire en dessous d'une courbe . . . 723
11.12 Dérivation de fonctions réelles . . . 724
11.12.1 Exemples . . . 726
11.12.2 Interprétation géométrique : tangente . . . 727
11.12.3 Interprétation géométrique : approximation affine . . . 727
11.12.4 Développement limité au premier ordre . . . 728
11.13 Règles de calcul . . . 729

11.13.1 Dérivée de la réciproque . . . 732
11.14 Dérivation et croissance . . . 734
11.14.1 Théorèmes de Rolle et des accroissements finis . . . 736
11.14.2 Règle de l'Hospital . . . 738
11.14.3 Dérivée et primitive . . . 740
11.15 Fonctions de plusieurs variables . . . 741
11.15.1 Graphes de fonctions à plusieurs variables . . . 744
11.15.2 Courbes de niveau . . . 745
11.16 Limites à plusieurs variables . . . 749
11.16.1 Caractérisation de la limite par les suites . . . 751
11.16.2 Règle de l'étau . . . 752
11.16.3 Méthode des chemins . . . 753
11.17 Dérivée directionnelle . . . 755
11.17.1 Dérivée partielle et directionnelles . . . 755
11.17.2 Gradient : direction de plus grande pente . . . 759
11.17.3 Gradient : orthogonal au plan tangent . . . 760
11.18 Dérivée directionnelle de fonctions composées . . . 760
11.19 Formes différentielles . . . 762
11.19.1 Décomposition dans la base duale . . . 762
11.19.2 L'isomorphisme musical . . . 763
11.20 Différentielle . . . 764
11.20.1 Exemples introductifs . . . 764
11.20.2 Différentielle . . . 765
11.20.3 Matrice de la différentielle . . . 765
11.20.4 Quelques propriétés . . . 766
11.20.5 Différentielle, dual et forme différentielle . . . 767
11.20.6 Ce n'est pas la différentielle extérieure . . . 768
11.20.7 Fonctions composées . . . 768
11.20.8 Continuité, dérivabilité et différentiabilité . . . 769
11.20.9 Calcul de valeurs approchées . . . 772
11.20.10 Différentielle et tangente . . . 774
11.20.11 Prouver qu'une fonction n'est pas différentiable . . . 775
11.20.12 Gradient . . . 780
11.20.13 Linéarité . . . 780
11.21 Produit . . . 780
11.21.1 Difficulté d'ordre supérieur . . . 782
11.21.2 Solution : produit tensoriel . . . 783
11.21.3 Formes bilinéaires . . . 783
11.22 Différentielle de fonction composée . . . 784
11.23 Autres trucs sur la différentielle . . . 786
11.23.1 Différentielle et dérivées partielles . . . 786
11.23.2 Plan tangent . . . 786
11.23.3 Calcul de différentielles . . . 786
11.23.4 Notes idéologiques quant au concept de plan tangent . . . 787
11.23.5 Gradient et recherche du plan tangent . . . 788
11.23.6 Projection orthogonale . . . 789
11.24 Jacobienne . . . 790
11.24.1 Rappels et définitions . . . 790
11.25 Fonctions de classe C^1 . . . 791
11.26 Différentielle et dérivée complexe . . . 796
11.26.1 Quelques règles de calcul . . . 800
11.27 Théorèmes des accroissements finis . . . 800
11.28 Fonctions Lipschitziennes . . . 801

11.29 Différentielles d'ordre supérieur . . . 802
11.29.1 Différentielle et dérivées partielles . . . 802
11.29.2 Espaces d'applications multilinéaires et identifications . . . 806
11.29.3 Identification des espaces d'applications multilinéaires (vieux) . . . 810
11.29.4 Fonctions différentiables plusieurs fois . . . 811
11.29.5 Différentielle seconde, fonction de classe C^2 . . . 811
11.29.6 Ordre supérieur . . . 814
11.30 Suites et séries : généralités . . . 818
11.30.1 Convergence uniforme . . . 818
11.30.2 Série de fonctions . . . 822
11.31 Permuter limite et dérivée . . . 824
11.32 Densité des polynômes . . . 826
11.32.1 Théorème de Stone-Weierstrass . . . 826
11.33 Primitive de fonction continue . . . 830
11.34 La fonction puissance . . . 831
11.34.1 Sur les naturels . . . 831
11.34.2 Sur les rationnels, racines . . . 833
11.34.3 Dérivation de la fonction puissance (première) . . . 845
11.34.4 Équation fonctionnelle . . . 847
11.34.5 Dérivation de la fonction puissance (seconde) . . . 849
11.34.6 Hölder . . . 850
11.34.7 Vers les complexes . . . 851
11.35 Polynômes de Taylor . . . 852
11.35.1 Fonctions « petit o » . . . 856
11.35.2 Autres formulations . . . 858
11.35.3 Formule et reste . . . 858
11.35.4 Reste intégral . . . 859
11.36 Développement limité autour de zéro . . . 859
11.36.1 Généralités . . . 859
11.36.2 Formule de Taylor-Young . . . 860
11.36.3 Règles de calcul . . . 862
11.37 Développement ailleurs qu'à l'origine . . . 864
11.38 Développement au voisinage de l'infini . . . 864
11.38.1 La fonction puissance : remarques pour la suite . . . 864
11.39 Fonctions réelles de deux variables réelles . . . 864
11.39.1 Limites de fonctions à deux variables . . . 865
11.39.2 Dérivées partielles . . . 866
11.39.3 Différentielle et accroissement . . . 867
11.40 Les fonctions à valeurs vectorielles . . . 867
11.41 Fonctions vectorielles de plusieurs variables . . . 868
11.42 Limites à plusieurs variables . . . 868
11.43 Champs de vecteurs . . . 872
11.43.1 Matrice jacobienne . . . 872
11.44 Divergence, rotationnel et l'opérateur nabla . . . 873
11.45 Interprétation de la divergence . . . 875
11.46 Quelques formules de Leibnitz . . . 877

12 Analyse sur des groupes (Vol 2) 879
12.1 Action de groupe et connexité . . . 879
12.2 Espaces de matrices . . . 881
12.2.1 Dilatations et transvections . . . 881
12.2.2 Connexité de certains groupes . . . 887
12.2.3 Densité . . . 889
12.2.4 Racine carrée d'une matrice hermitienne positive . . . 891

12.2.5 Racine carrée d'une matrice symétrique positive . . . 892
12.2.6 Décomposition polaires : cas réel . . . 893
12.2.7 Enveloppe convexe . . . 895
12.2.8 Décomposition de Bruhat . . . 897
12.3 Sous-groupes du groupe linéaire . . . 899

13 Tribus, théorie de la mesure, intégration (Vol 2) 903
13.1 Tribus . . . 903
13.1.1 Généralités . . . 903
13.1.2 Tribu induite . . . 904
13.1.3 Tribu de Baire . . . 905
13.2 Théorie de la mesure . . . 907
13.2.1 Mesure sur un ensemble de parties . . . 907
13.2.2 Mesure sur une algèbre de parties . . . 907
13.2.3 Mesure sur une tribu, espace mesuré . . . 909
13.2.4 Mesure extérieure . . . 915
13.3 Applications mesurables . . . 917
13.3.1 Propriétés . . . 917
13.3.2 D'une tribu à l'autre . . . 918
13.4 Espace mesuré complet . . . 921
13.4.1 Partie négligeable . . . 921
13.4.2 Prolongement . . . 930
13.5 Tribu borélienne . . . 932
13.5.1 Mesure image . . . 934
13.5.2 Régularité d'une mesure . . . 934
13.5.3 Théorème de récurrence . . . 938
13.6 Mesurabilité des fonctions à valeurs réelles . . . 939
13.6.1 Fonctions à valeurs réelles sur un espace mesurable . . . 940
13.6.2 Fonction étagée . . . 945
13.6.3 Fonctions réelles à variables réelles . . . 948
13.7 Tribu produit, mesure produit . . . 949
13.7.1 Produit d'espaces mesurables . . . 949
13.7.2 Le cas des boréliens . . . 950
13.8 Mesure de Lebesgue sur $\mathbb{R}$ . . . 951
13.8.1 Mesure et tribu de Lebesgue . . . 956
13.8.2 Propriétés de la mesure de Lebesgue . . . 957
13.8.3 Fonctions mesurables . . . 961
13.8.4 Ensemble de Vitali (non mesurable) . . . 962
13.8.5 Ensemble de Cantor . . . 962
13.8.6 Mesure positive sans intervalle . . . 965
13.9 Intégrale par rapport à une mesure . . . 965
13.9.1 Définition pour les fonctions à valeurs positives . . . 966
13.9.2 Premières propriétés . . . 967
13.9.3 Propriétés plus avancées . . . 970
13.9.4 Fonctions à valeurs réelles . . . 974
13.9.5 Additivité de l'intégrale . . . 975
13.9.6 Fonctions à valeurs vectorielles (dimension finie) . . . 976
13.9.7 Quelques propriétés . . . 978
13.9.8 Permuter limite et intégrale . . . 979
13.9.9 Additivité de l'intégrale de Lebesgque . . . 981
13.9.10 Produit d'une mesure par une fonction (mesure à densité) . . . 983
13.9.11 Mesure et topologie . . . 985
13.10 Propriétés des intégrales . . . 986
13.11 Mesure à densité . . . 988

13.11.1 Théorème de Radon-Nikodym . . . 988
13.11.2 Mesure complexe . . . 989
13.11.3 Théorème d'approximation . . . 989
13.12 Produit de mesures . . . 993
13.13 Tribu et mesure de Lebesgue sur $\mathbb{R}^d$ . . . 998
13.13.1 Ensembles négligeables . . . 999
13.13.2 Parties et fonctions mesurables . . . 1000
13.13.3 Propriétés d'unicité . . . 1000
13.13.4 Régularité . . . 1002
13.14 Propriétés de l'intégrale de Lebesgue . . . 1003
13.14.1 Quelques limites dans les bornes . . . 1003
13.14.2 Mesure de comptage et série . . . 1005
13.14.3 Théorème de la moyenne . . . 1005
13.14.4 Primitives et intégrales . . . 1006
13.14.5 Exemples et applications . . . 1008
13.14.6 Permuter limite et dérivée . . . 1009
13.14.7 Intégrales impropres . . . 1011
13.15 Changement de variables dans une intégrale multiple . . . 1013
13.15.1 Des lemmes . . . 1013
13.15.2 Déterminant et mesure de Lebesgue . . . 1014
13.15.3 Le théorème et sa démonstration . . . 1015
13.15.4 Exemples . . . 1020
13.16 Changement d'espace mesuré . . . 1020
13.17 Théorème de Fubini-Tonelli et de Fubini . . . 1022

14 Suites et séries de fonctions (Vol 2) 1031
14.1 Séries de fonctions . . . 1031
14.1.1 Intégration de séries de fonctions . . . 1031
14.1.2 Différentiabilité . . . 1032
14.2 Séries entières . . . 1035
14.2.1 Disque de convergence . . . 1035
14.2.2 Propriétés de la somme . . . 1038
14.2.3 Dérivation . . . 1042
14.2.4 Intégration . . . 1045
14.3 Séries de Taylor . . . 1045
14.3.1 Polynôme de Taylor d'une série entière . . . 1045
14.3.2 Une majoration pour le reste . . . 1046
14.3.3 Fonctions analytiques . . . 1047
14.4 Exponentielle sur une algèbre normée . . . 1048
14.4.1 Définition . . . 1048
14.4.2 Différentielles . . . 1050
14.4.3 Séries dans une algèbre normée . . . 1053
14.4.4 Exponentielle de matrice . . . 1054
14.5 Exponentielle et logarithme dans les réels . . . 1057
14.5.1 L'équation différentielle . . . 1057
14.5.2 Existence . . . 1058
14.5.3 Le nombre de Neper e . . . 1059
14.5.4 Application réciproque : logarithme . . . 1059
14.5.5 Approximations numériques de e . . . 1061
14.5.6 Résumé des propriétés de l'exponentielle . . . 1062
14.5.7 Dérivée de la fonction puissance . . . 1063
14.5.8 Dérivée du logarithme . . . 1064
14.5.9 Taylor pour l'exponentielle . . . 1064
14.5.10 Analycité . . . 1065

14.5.11 Autres propriétés et petits calculs . . . 1065
14.5.12 Taylor pour le logarithme . . . 1066
14.5.13 Développements et calcul de limites . . . 1069
14.6 Vitesses des puissances, de l'exponentielle et du logarithme . . . 1070
14.6.1 Un peu de théorie . . . 1070
14.6.2 Nombres premiers . . . 1073
14.6.3 Quelques limites . . . 1074
14.7 Trigonométrie hyperbolique . . . 1076
14.8 Séries entières de matrices . . . 1077
14.8.1 Différentiabilité . . . 1077
14.9 Exponentielle de matrices . . . 1080
14.9.1 Diagonalisabilité d'exponentielle . . . 1081
14.10 Étude d'asymptote . . . 1082
14.11 Développement en série . . . 1083
14.11.1 Série génératrice d'une suite . . . 1083
14.11.2 Développement en série et Taylor . . . 1083
14.11.3 Resommer une série . . . 1085
14.12 Séries entières de matrices . . . 1090
14.12.1 Rayon de convergence . . . 1090
14.12.2 Convergence et rayon spectral . . . 1091
14.12.3 Exponentielle et logarithme de matrice . . . 1093
14.12.4 Calcul effectif de l'exponentielle d'une matrice . . . 1096
14.13 Lemme de Borel . . . 1096
14.13.1 Fonctions plateaux . . . 1096
14.13.2 Le lemme de Borel . . . 1098
14.14 Nombres de Bell . . . 1100

15 Représentations et caractères (Vol 2) 1103
15.1 Représentations et caractères . . . 1103
15.1.1 Crochet de dualité et transformée de Fourier . . . 1105
15.1.2 Groupes non abéliens . . . 1106
15.1.3 Représentations linéaires des groupes finis . . . 1106
15.1.4 Module . . . 1107
15.1.5 Structure hermitienne . . . 1109
15.1.6 Caractères . . . 1110
15.2 Équivalence de représentations et caractères . . . 1110
15.2.1 Représentation régulière . . . 1113
15.2.2 Caractères et représentations : suite et fin . . . 1114
15.3 Représentation produit tensoriel . . . 1116
15.4 Exemple sur le groupe symétrique . . . 1117
15.5 Table des caractères du groupe symétrique S_4 . . . 1117
15.5.1 Calculs à partir de rien ou presque . . . 1117
15.5.2 Représentation de S_4 via les isométries du tétraèdre . . . 1119
15.5.3 À propos de la représentation ρ_u . . . 1119

16 Encore de l'analyse (et c'est pas fini) (Vol 2) 1123
16.1 Densité des polynômes . . . 1123
16.2 Primitive et intégrale . . . 1123
16.2.1 Théorème taubérien de Hardy-Littlewood . . . 1124
16.2.2 Théorème de Müntz . . . 1127
16.3 Intégrales convergeant uniformément . . . 1130
16.3.1 Définition et propriété . . . 1130
16.3.2 Critères de convergence uniforme . . . 1131
16.4 Fonctions définies par une intégrale . . . 1131

16.4.1 Continuité sous l'intégrale . . . 1132
16.4.2 Le coup du compact . . . 1133
16.4.3 Dérivabilité sous l'intégrale . . . 1134
16.4.4 Absolue continuité . . . 1135
16.4.5 Différentiabilité sous l'intégrale . . . 1137
16.5 Deux théorèmes de point fixe . . . 1140
16.5.1 Points fixes attractifs et répulsifs . . . 1140
16.5.2 Picard . . . 1141
16.6 Théorèmes de point fixes et équations différentielles . . . 1144
16.6.1 Théorème de Cauchy-Lipschitz . . . 1144
16.7 Théorèmes d'inversion locale et de la fonction implicite . . . 1151
16.7.1 Mise en situation . . . 1151
16.7.2 Théorème d'inversion locale . . . 1152
16.7.3 Théorème de la fonction implicite . . . 1155
16.7.4 Exemple . . . 1157
16.8 Décomposition polaire (régularité) . . . 1159
16.9 Théorème de Von Neumann . . . 1161
16.10 Recherche d'extrémums . . . 1163
16.10.1 Extrema à une variable . . . 1163
16.10.2 Extrema libre . . . 1165
16.10.3 Extrema et Hessienne . . . 1165
16.10.4 Un peu de recettes de cuisine . . . 1167
16.10.5 Extrema liés . . . 1168
16.11 Fonctions convexes . . . 1170
16.11.1 Inégalité des pentes . . . 1170
16.11.2 Convexité et régularité . . . 1172
16.11.3 Dérivées d'une fonction convexe . . . 1172
16.11.4 Graphe d'une fonction convexe . . . 1174
16.11.5 Convexité et hessienne . . . 1179
16.11.6 Quelques inégalités . . . 1185
16.12 Trucs et astuces de calcul d'intégrales . . . 1191
16.12.1 Quelques intégrales « usuelles » . . . 1191
16.12.2 Reformer un carré au dénominateur . . . 1193
16.12.3 Décomposition en fractions simples . . . 1194
16.13 Algorithme du gradient à pas optimal . . . 1194
16.14 Formes quadratiques, signature, et lemme de Morse . . . 1198
16.15 Ellipsoïde de John-Loewner . . . 1202
16.16 Prolongement de fonctions . . . 1208
16.16.1 Encore du prolongement . . . 1210
16.17 Complétion d'un espace métrique . . . 1212
16.17.1 Principe des zéros isolés . . . 1214
16.18 Un petit extra . . . 1215

17 Trigonométrie, isométries (Vol 3) 1217
17.1 Trigonométrie . . . 1217
17.1.1 Définitions, périodicité et quelques valeurs remarquables . . . 1217
17.1.2 Fonction puissance (pour les complexes) . . . 1218
17.1.3 Formules de trigonométrie . . . 1219
17.2 Très modeste approximation de π . . . 1228
17.3 Cercle trigonométriques . . . 1229
17.3.1 Les fonctions tangente et arc tangente . . . 1229
17.3.2 La fonction arc sinus . . . 1232
17.3.3 La fonction arc cosinus . . . 1233
17.3.4 Une meilleure approximation de π . . . 1235

17.3.5 Forme trigonométrique des nombres complexes . . . 1236
17.3.6 Angle entre deux vecteurs . . . 1236
17.3.7 Aire du parallélogramme . . . 1237
17.4 Paramétrisation du cercle . . . 1239
17.4.1 Bijection continue . . . 1239
17.4.2 Inverse . . . 1241
17.4.3 Cercle trigonométrique . . . 1242
17.4.4 Du point de vue de la tribu, mesure et co. . . . 1242
17.5 Exemples trigonométriques . . . 1245
17.5.1 Quelques équations trigonométriques . . . 1245
17.5.2 Développements en série . . . 1246
17.5.3 Intégration . . . 1247
17.5.4 Changement de variables dans une intégrale . . . 1248
17.6 Isométries de l'espace euclidien . . . 1248
17.6.1 Structure du groupe $\mathrm{Isom}(\mathbb{R}^n)$ . . . 1249
17.7 Isométries dans $\mathbb{R}^n$ . . . 1251
17.8 Groupes finis d'isométries . . . 1255
17.8.1 Groupe diédral . . . 1257
17.8.2 Applications : du dénombrement . . . 1264
17.8.3 Points fixés par une affinité . . . 1265
17.9 Classification des isométries dans $\mathbb{R}^2$ . . . 1266
17.9.1 Réflexions . . . 1266
17.9.2 Segment, plan médiateur et équidistance . . . 1267
17.9.3 Translations et réflexions . . . 1268
17.9.4 Rotations . . . 1269
17.9.5 Rotation d'un angle donné . . . 1272
17.9.6 Rotations vectorielles . . . 1274
17.9.7 Matrice des transformations orthogonales . . . 1276
17.9.8 Rotations, SO(2) et matrice de rotation . . . 1278
17.9.9 Rotation et application affine . . . 1279
17.9.10 Angle entre deux droites . . . 1280
17.9.11 Angle orienté . . . 1280
17.9.12 Angles et nombres complexes . . . 1284
17.9.13 Classification . . . 1286
17.9.14 Classification des isométries de $\mathbb{R}$ . . . 1289
17.9.15 Isométries du tétraèdre régulier . . . 1290
17.9.16 Représentation de S_4 via les isométries du tétraèdre . . . 1291
17.10 Transformations de Lorentz . . . 1292
17.10.1 Sous-groupe fini d'isométries du plan . . . 1295
17.10.2 Relations trigonométriques dans un triangle rectangle . . . 1299
17.10.3 Pavages du plan . . . 1300
17.11 Un peu de structure de $\mathrm{O}(n)$ . . . 1316
17.11.1 Valeurs propres dans $\mathrm{O}(n)$ . . . 1316
17.11.2 Sous-groupes finis de SO(3) . . . 1319
17.12 Systèmes de coordonnées . . . 1328
17.12.1 Coordonnées polaires . . . 1328
17.12.2 Coordonnées cylindriques . . . 1335
17.12.3 Coordonnées sphériques . . . 1336
17.13 Calcul de limites . . . 1337
17.13.1 Méthode des coordonnées polaires . . . 1338
17.13.2 Méthode du développement asymptotique . . . 1341
17.14 Quelques intégrales avec de la trigonométrie . . . 1342
17.14.1 Changement de variables . . . 1342

17.14.2 Coordonnées polaires . . . 1343
17.14.3 Coordonnées cylindriques . . . 1345
17.14.4 Coordonnées sphériques . . . 1346
17.14.5 Un autre système utile . . . 1347
17.15 Aire d'une surface de révolution . . . 1348
17.16 Table de caractères du groupe diédral . . . 1350
17.16.1 Représentations de dimension un . . . 1350
17.16.2 Représentations de dimension deux . . . 1351
17.16.3 Le compte pour n pair . . . 1353
17.16.4 Le compte pour n impair . . . 1353

18 Corps finis, racines de l'unité (Vol 3) 1355
18.1 Le groupe des racines de l'unité dans les nombres complexes . . . 1355
18.1.1 Le groupe . . . 1355
18.1.2 Fonction indicatrice d'Euler (première partie) . . . 1357
18.1.3 Introduction par les racines de l'unité . . . 1357
18.1.4 Générateurs . . . 1359
18.1.5 Fonction indicatrice d'Euler (propriétés) . . . 1359
18.1.6 Décomposition en fractions simples . . . 1359
18.2 Chiffrement RSA . . . 1360
18.2.1 Mise en place par Bob . . . 1360
18.2.2 Chiffrement . . . 1360
18.2.3 Déchiffrement . . . 1360
18.2.4 Une imprudence à ne pas commettre . . . 1361
18.2.5 Problèmes calculatoires . . . 1361
18.2.6 La solidité de RSA . . . 1362
18.2.7 Note non mathématique pour doucher l'enthousiasme . . . 1362
18.3 Polynômes cyclotomiques . . . 1362
18.3.1 Définitions et propriétés . . . 1362
18.3.2 Nombres premiers . . . 1366
18.4 Dénombrement des solutions d'une équation diophantienne . . . 1368
18.5 Corps finis . . . 1370
18.5.1 Théorème de Wedderburn . . . 1370
18.5.2 Existence, unicité . . . 1372
18.5.3 Symboles de Legendre et carrés . . . 1374
18.5.4 Théorème de Chevalley-Warning . . . 1380
18.5.5 Contenu d'un polynôme . . . 1382
18.5.6 Théorème de l'élément primitif . . . 1383
18.5.7 Construction de $\mathbb{F}_q$ . . . 1387
18.5.8 Exemple : étude de $\mathbb{F}_{16}$ . . . 1390
18.5.9 Polynômes irréductibles sur $\mathbb{F}_q$ . . . 1392
18.5.10 Matrices . . . 1394
18.6 Constructions à la règle et au compas . . . 1396
18.6.1 Quelques constructions . . . 1396
18.6.2 Nombres constructibles . . . 1398
18.6.3 Polygones constructibles . . . 1399

19 Intégration sur des variétés (Vol 3) 1405
19.1 Variétés . . . 1405
19.1.1 Introduction . . . 1405
19.1.2 Définition, carte . . . 1405
19.1.3 Ancienne définition . . . 1406
19.1.4 Espace tangent . . . 1408
19.2 Intégration . . . 1408

19.2.1 Résultats préliminaires . . . 1408
19.2.2 Quelques expressions pour l'élément de volume . . . 1410
19.3 Intégrale sur une variété . . . 1412
19.3.1 Mesure sur une carte . . . 1412
19.3.2 Intégrale sur une carte . . . 1413
19.3.3 Exemples . . . 1414
19.3.4 Orientation . . . 1414
19.3.5 Formes différentielles . . . 1416
19.3.6 Intégrale d'une fonction sur une sous-variété . . . 1417
19.4 Longueur, aire, volumes etc. . . . 1418
19.4.1 Quelques aires faciles . . . 1418
19.5 Autres théorèmes de points fixes . . . 1419
19.5.1 Brouwer . . . 1419
19.5.2 Théorème de Schauder . . . 1421
19.5.3 Théorème de Cauchy-Arzella . . . 1422
19.5.4 Théorème de Markov-Kakutani et mesure de Haar . . . 1423
19.6 Intégrales curvilignes . . . 1424
19.6.1 Chemins de classe C^1 . . . 1424
19.6.2 Intégrer une fonction . . . 1424
19.6.3 Intégrer un champ de vecteurs . . . 1425
19.6.4 Intégrer une forme différentielle sur un chemin . . . 1426
19.6.5 Intégration d'une forme différentielle sur un chemin . . . 1426
19.6.6 Interprétation physique : travail . . . 1428
19.6.7 Intégrer un champ de vecteurs sur un bord en $2D$ . . . 1429
19.6.8 Intégrer une forme différentielle sur un bord en $2D$ . . . 1429
19.6.9 Intégrer une forme différentielle sur un bord en $3D$ . . . 1429
19.6.10 Intégrer un champ de vecteurs sur un bord en $3D$ . . . 1429
19.6.11 Dérivées croisées et forme différentielle exacte . . . 1430
19.7 Surfaces paramétrées . . . 1431
19.7.1 Graphe d'une fonction . . . 1432
19.7.2 Intégrale sur une partie de $\mathbb{R}^m$ . . . 1433
19.8 Intégrales de surface . . . 1434
19.8.1 Intégrale d'un champ de vecteurs . . . 1434
19.9 Intégrales de surface . . . 1434
19.9.1 Aire d'une surface paramétrée . . . 1434
19.9.2 Intégrale d'une fonction sur une surface . . . 1436
19.9.3 Intégrale d'une 2-forme . . . 1436
19.10 Flux d'un champ de vecteurs à travers une surface . . . 1437
19.11 Divergence, Green, Stokes . . . 1439
19.11.1 Théorème de la divergence . . . 1440
19.11.2 Formule de Green . . . 1441
19.11.3 Formule de Stokes . . . 1442
19.12 Résumé des intégrales vues . . . 1443
19.12.1 L'intégrale d'une fonction sur les réels . . . 1443
19.12.2 Intégrale d'une fonction sur un chemin . . . 1444
19.12.3 Intégrale d'une fonction sur une surface . . . 1444
19.12.4 Intégrale d'une fonction sur un volume . . . 1444
19.12.5 Conclusion pour les fonctions . . . 1445
19.12.6 Circulation d'un champ de vecteurs . . . 1445
19.12.7 Flux d'un champ de vecteurs . . . 1445
19.12.8 Conclusion pour les champs de vecteurs . . . 1446
19.12.9 Attention pour les surfaces fermées ! . . . 1446
19.13 Formes différentielles exactes et fermées . . . 1448

19.14 Théorème d'Abel angulaire . . . 1450
19.14.1 Passage à la limite sous le signe intégral . . . 1453
19.14.2 Intégrale en dimension un . . . 1453
19.14.3 Intégrales convergentes . . . 1454
19.14.4 La méthode de Rothstein-Trager . . . 1454
19.15 Rappel sur les intégrales usuelles . . . 1460
19.16 Intégrales le long de chemins . . . 1460
19.16.1 Circulation d'un champ de vecteur . . . 1460
19.17 Circulation d'un champ conservatif . . . 1462
19.18 Intégration de fonction à deux variables . . . 1464
19.18.1 Intégration sur un domaine rectangulaire . . . 1464
19.18.2 Intégration sur un domaine non rectangulaire . . . 1465
19.18.3 Changement de variables . . . 1467
19.19 Les intégrales triples . . . 1467
19.19.1 Volume . . . 1469
19.20 Un petit peu plus formel . . . 1470
19.20.1 Intégration sur un domaine non rectangulaire . . . 1470
19.20.2 Changement de variables . . . 1471
19.21 Aire et primitive . . . 1472
19.21.1 Longueur d'arc de courbe . . . 1473
19.21.2 Aire de révolution . . . 1473
19.22 L'aire en dessous d'une courbe . . . 1474
19.23 Propriétés des intégrales . . . 1474
19.24 Techniques d'intégration . . . 1475
19.24.1 Intégration par parties . . . 1476
19.24.2 Changement de variables – pour trouver des primitives . . . 1478
19.24.3 Changement de variables – pour calculer des intégrales . . . 1479
19.24.4 Intégrations des fractions rationnelles réduites . . . 1481
19.24.5 Quelques formules à connaitre . . . 1482
19.24.6 Approximation de $\ln(2)$ . . . 1482
19.25 Constructions plus naïves de l'intégrale dans le cas réel . . . 1485
19.25.1 Mesure de Lebesgue, version rapide . . . 1485
19.25.2 Pavés et subdivisions . . . 1486
19.25.3 Intégrale d'une fonction en escalier . . . 1489
19.25.4 Intégrales partielles . . . 1489
19.25.5 Réduction d'une intégrale multiple . . . 1490
19.25.6 Propriétés de l'intégrale . . . 1491
19.25.7 Intégrales multiples, cas général . . . 1492
19.25.8 Réduction d'une intégrale multiple . . . 1493
19.25.9 Intégrales sur des parties de $\mathbb{R}^2$ . . . 1494
19.25.10 Intégrales sur des parties de $\mathbb{R}^3$ . . . 1497
19.25.11 Fonctions et ensembles non bornés . . . 1499
19.25.12 Lemme de Morse . . . 1500
19.26 Autres intégrales sympathiques . . . 1502
19.26.1 Intégrale de Wallis . . . 1502
19.26.2 Formule de Stirling . . . 1506
19.26.3 La fonction sinus cardinal, intégrale de Dirichlet . . . 1509

20 Arcs paramétrés (Vol 3) 1519
20.1 Définitions . . . 1519
20.2 Longueur d'arc . . . 1519
20.3 Abscisse curviligne . . . 1522
20.3.1 Formule intégrale de la longueur . . . 1523
20.4 Suite du chapitre . . . 1528

20.5 Autres exemples . . . 1529
20.6 Élément de longueur . . . 1530
20.6.1 Élément de longueur : cartésiennes . . . 1530
20.6.2 Élément de longueur : polaires (1) . . . 1531
20.6.3 Élément de longueur : polaires (2) . . . 1531
20.6.4 Approximation de la longueur par des cordes . . . 1533
20.7 Arc géométrique . . . 1535
20.7.1 Abscisse curviligne et paramétrage normal . . . 1536
20.7.2 Tangente à une courbe paramétrée . . . 1541
20.8 Un peu de topologie . . . 1542
20.9 Repère de Frenet . . . 1545
20.9.1 Torsion . . . 1547
20.10 Hors des coordonnées normales . . . 1548
20.11 Tracer des courbes paramétriques dans $\mathbb{R}^2$ . . . 1551
20.12 Courbes planes . . . 1552
20.12.1 Angle . . . 1552
20.12.2 Courbure signée . . . 1553
20.12.3 Degré, indice et homotopie . . . 1556
20.13 Courbes fermées planes . . . 1562
20.13.1 Cercle circonscrit . . . 1562
20.13.2 Description locale . . . 1564
20.13.3 Enveloppe convexe . . . 1565
20.13.4 Courbure et convexité . . . 1568
20.13.5 Théorème des quatre sommets . . . 1569
20.13.6 Le théorème de Jordan . . . 1571

21 Géométrie hyperbolique (Vol 3) 1573
21.1 Inversion . . . 1573
21.1.1 Cercles perpendiculaires . . . 1573
21.1.2 Inversion . . . 1574

22 Espaces projectifs (Vol 3) 1579
22.1 Sous espaces projectifs . . . 1579
22.2 Espace projectifs comme « complétés » d'espaces affines . . . 1581
22.3 Théorème de Pappus . . . 1584
22.4 Homographies . . . 1585
22.4.1 Homographies . . . 1585
22.4.2 Le groupe projectif . . . 1586
22.4.3 Repères projectifs . . . 1587
22.4.4 Identifications $P(\mathbb{K}^2)$ vers $\mathbb{K} \cup \{\infty\}$ . . . 1591
22.4.5 Birapport . . . 1592
22.5 Coordonnées homogènes . . . 1598
22.5.1 Curiosité : matrice de translation . . . 1598
22.5.2 Dualité . . . 1599
22.5.3 Polynômes . . . 1600
22.6 La sphère de Riemann $P_1(\mathbb{C})$ . . . 1602
22.6.1 Éléments de géométrie dans $P_1(\mathbb{C})$ . . . 1603
22.6.2 Homographies . . . 1606
22.6.3 Birapport . . . 1612
22.6.4 Division harmonique . . . 1615
22.6.5 Groupe circulaire . . . 1619
22.6.6 Action du groupe modulaire . . . 1621

23 Analyse vectorielle (Vol 3) 1627

23.1 Le théorème de Green . . . 1627
23.2 Théorème de la divergence dans le plan . . . 1631
23.2.1 La convention de sens de parcours . . . 1631
23.2.2 Théorème de la divergence . . . 1632
23.3 Théorème de Stokes . . . 1632
23.4 Théorème de Gauss . . . 1634
23.5 Coordonnées curvilignes . . . 1636
23.5.1 Base locale . . . 1636
23.5.2 Importance de l'orthogonalité . . . 1636
23.5.3 Coordonnées polaires . . . 1638
23.5.4 Coordonnées cylindriques . . . 1638
23.5.5 Coordonnées sphériques . . . 1639
23.5.6 Gradient en coordonnées curvilignes . . . 1639
23.5.7 Divergence en coordonnées curvilignes . . . 1640
23.5.8 Laplacien en coordonnées curvilignes orthogonales . . . 1643
23.5.9 Rotationnel en coordonnées curvilignes orthogonales . . . 1643
23.6 Les formules . . . 1644
23.6.1 Coordonnées polaires . . . 1644
23.6.2 Coordonnées cylindriques . . . 1644
23.6.3 Coordonnées sphériques . . . 1645

24 Espaces de Hilbert (Vol 3) 1647
24.1 Espaces de Hilbert . . . 1647
24.1.1 Sous-espace vectoriel fermé ? ? ? . . . 1648
24.2 Théorème de la projection . . . 1649
24.3 Systèmes orthogonaux et bases . . . 1652
24.3.1 Orthogonal d'une partie . . . 1652
24.3.2 Dual, théorème de représentation de Riesz . . . 1653
24.3.3 Séparabilité . . . 1655
24.3.4 Base hilbertienne . . . 1657
24.3.5 Décomposition dans une base hilbertienne . . . 1661
24.3.6 Digression sur les normes opérateurs . . . 1666
24.3.7 Applications linéaires et continuité . . . 1667
24.4 Théorème de Kochen-Specker . . . 1669
24.5 Théorème de Lax-Milgram . . . 1670

25 Analyse complexe (Vol 3) 1675
25.1 Fonctions holomorphes . . . 1675
25.1.1 Équations de Cauchy-Riemann . . . 1675
25.1.2 Intégrales sur des chemins fermés . . . 1677
25.1.3 Lacets, indice et homotopie . . . 1679
25.1.4 Théorème de Cauchy et analycité . . . 1680
25.1.5 Théorème de Brouwer en dimension 2 . . . 1683
25.1.6 Principe des zéros isolés . . . 1684
25.1.7 Prolongement de fonctions holomorphes . . . 1686
25.1.8 Théorème de Runge . . . 1687
25.2 Intégrales de fonctions holomorphes . . . 1689
25.2.1 Mesure de Radon . . . 1693
25.3 Conditions équivalentes à l'holomorphie . . . 1694
25.4 Singularités, pôles et méromorphe . . . 1695
25.5 Fonctions d'Euler . . . 1696
25.5.1 Euler et factorielle . . . 1699
25.6 Partition d'un entier en parts fixées . . . 1700
25.7 Exponentielle et logarithme complexe . . . 1702

25.7.1 Propriétés de l'exponentielle . . . 1702
25.7.2 Intégrale de Fresnel . . . 1703
25.7.3 Logarithme complexe . . . 1705
25.8 Théorème de Weierstrass . . . 1715

26 Analyse fonctionnelle (Vol 4) 1717
26.1 Théorème d'Ascoli . . . 1717
26.2 Espaces de Lebesgue L^p . . . 1719
26.2.1 Généralités . . . 1719
26.2.2 Un peu de convergence de suites . . . 1722
26.2.3 L'espace L^∞ . . . 1724
26.2.4 Quelques identifications . . . 1725
26.2.5 Inégalité de Jensen, Hölder et de Minkowski . . . 1726
26.2.6 Ni inclusions ni inégalités . . . 1728
26.2.7 Complétude . . . 1730
26.2.8 Théorèmes d'approximation . . . 1735
26.2.9 Densité des fonctions infiniment dérivables à support compact . . . 1736
26.2.10 Approximation . . . 1739
26.3 Convolution . . . 1740
26.3.1 Approximation de l'unité . . . 1742
26.3.2 Densité des polynômes trigonométriques . . . 1744
26.4 Espaces L^2, généralités . . . 1746
26.5 L'espace $L^2(\mathbb{R}^d)$ . . . 1747
26.6 L'espace $L^2(S^1)$ . . . 1748
26.6.1 Espace mesuré . . . 1748
26.6.2 Topologie . . . 1749
26.6.3 Système trigonométrique . . . 1755
26.6.4 Convolution . . . 1756
26.6.5 Approximation de l'unité . . . 1758
26.6.6 Base hilbertienne (suite des polynômes trigonométriques) . . . 1762
26.6.7 Convolution, bis . . . 1763
26.7 L'espace $L^2\big([a,b]\big)$ . . . 1764
26.8 Sur $[0,2\pi[$ . . . 1766
26.9 Sur $[-T,T[$ . . . 1766
26.9.1 Le cas dans $[0,2\pi]$ . . . 1766
26.10 Théorème de la projection normale . . . 1769
26.10.1 Espace uniformément convexe . . . 1769
26.10.2 Des inégalités . . . 1771
26.10.3 Inégalités de Clarkson . . . 1777
26.10.4 Uniforme convexité des espaces de Lebesgue . . . 1779
26.10.5 Théorème de la projection normale . . . 1780
26.11 Dualité et théorème de représentation de Riesz . . . 1782
26.12 Théorèmes de Hahn-Banach . . . 1791
26.13 Théorème de Tietze . . . 1794
26.14 Espace de Schwartz . . . 1797
26.14.1 Topologie . . . 1798
26.14.2 Produit de convolution . . . 1801
26.15 Théorème de Montel . . . 1801
26.16 Espaces de Bergman . . . 1802

27 Séries de Fourier (Vol 4) 1807
27.1 Densité des polynômes trigonométriques . . . 1807
27.1.1 Convergence pour les fonctions continues (via Weierstrass) . . . 1807
27.1.2 Convergence pour les fonctions continues (via Fejér) . . . 1807

27.1.3 Densité dans L^p . . . 1810
27.1.4 Suite équirépartie, critère de Weyl . . . 1811
27.2 Fonctions de Dirichlet . . . 1813
27.3 Coefficients et série de Fourier . . . 1814
27.3.1 Le contre-exemple que nous attendions tous . . . 1817
27.3.2 Inégalité isopérimétrique . . . 1819
27.3.3 À propos des coefficients . . . 1820

28 Transformation de Fourier (Vol 4) 1821
28.1 Transformée de Fourier sur $L^1(\mathbb{R}^d)$ . . . 1822
28.1.1 Formule sommatoire de Poisson . . . 1825
28.2 Suite régularisante . . . 1828
28.3 Transformée de Fourier dans l'espace de Schwartz . . . 1829
28.3.1 Quelques transformées de Fourier . . . 1832
28.3.2 Formule d'inversion . . . 1834
28.4 Transformée de Fourier sur $L^2(\mathbb{R}^d)$ . . . 1838
28.4.1 Le problème . . . 1838
28.4.2 Extension de $L^1 \cap L^2$ vers L^2 . . . 1839

29 Distributions (Vol 4) 1843
29.1 Dérivée faible . . . 1844
29.1.1 Dérivée partielle au sens faible . . . 1844
29.1.2 Dérivée faible partielle . . . 1846
29.2 Topologie et convergence sur des espaces de fonctions . . . 1846
29.3 Distributions . . . 1849
29.3.1 Multiplication d'une distribution par une fonction . . . 1851
29.3.2 Dérivée de distribution . . . 1851
29.3.3 Ordre et support d'une distribution . . . 1852
29.4 Distributions tempérées . . . 1855
29.4.1 Topologie . . . 1857
29.4.2 Distributions associées à des fonctions . . . 1857
29.4.3 Composition avec une fonction . . . 1857
29.4.4 Transformée de Fourier d'une distribution tempérée . . . 1858
29.4.5 Convolution d'une distribution par une fonction . . . 1858
29.4.6 Approximation de la distribution de Dirac . . . 1859
29.4.7 Peigne de Dirac . . . 1862
29.5 L'espace $C^\infty(\mathbb{R}, \mathscr{D}'(\mathbb{R}^d))$ . . . 1863
29.5.1 Dérivation . . . 1865
29.6 L'espace $C^\infty(\mathbb{R}, \mathscr{S}'(\mathbb{R}^d))$ . . . 1865
29.6.1 Propriétés générales . . . 1865
29.6.2 Dérivation . . . 1868
29.7 Une équation de distribution . . . 1870

30 Espaces de Sobolev, équations elliptiques (Vol 4) 1873
30.1 Espaces de Sobolev . . . 1873
30.1.1 Sur un intervalle de $\mathbb{R}$ . . . 1873
30.1.2 Sur un ouvert de $\mathbb{R}^n$ . . . 1878
30.1.3 Espace de Sobolev fractionnaire . . . 1880
30.2 Trace . . . 1882
30.3 Théorème de plongement . . . 1885

31 Équations différentielles ordinaires (Vol 4) 1889
31.1 Équation homogène, solution particulière . . . 1890
31.2 Que faire avec $f(z)dz = g(t)dt$? . . . 1890

31.3 Équations linéaires du premier ordre . . . 1891
31.3.1 Pourquoi la variation des constantes fonctionne toujours ? . . . 1893
31.4 Équations à variables séparées . . . 1894
31.4.1 La méthode rapide . . . 1894
31.4.2 La méthode plus propre . . . 1895
31.4.3 Les théorèmes . . . 1895
31.5 Équations linéaires d'ordre supérieur . . . 1897
31.5.1 Équations et systèmes linéaires à coefficients constants . . . 1897
31.5.2 Si les coefficients ne sont pas constants ? . . . 1897
31.6 Système d'équations linéaires . . . 1898
31.6.1 La magie de l'exponentielle... . . . 1898
31.6.2 ...mais la difficulté . . . 1898
31.6.3 La recette . . . 1898
31.6.4 Système d'équations linéaires avec matrice constante . . . 1899
31.6.5 Système d'équations linéaires avec matrice non constante . . . 1900
31.7 Réduction de l'ordre . . . 1900
31.8 Autour de Cauchy-Lipschitz . . . 1902
31.8.1 Fuite des compacts et explosion en temps fini . . . 1902
31.8.2 Écart entre deux conditions initiales . . . 1904
31.8.3 Flot d'un champ de vecteurs . . . 1905
31.8.4 Stabilité de Lyapunov . . . 1918
31.8.5 Système proies-prédateurs de Lotka-Volterra . . . 1922
31.9 Équation du second ordre . . . 1925
31.9.1 Wronskien . . . 1925
31.9.2 Avec second membre . . . 1926
31.9.3 Équation $y'' + q(t)y = 0$ . . . 1926
31.9.4 Équation de Hill . . . 1928
31.10 Différents types d'équations différentielles . . . 1931
31.10.1 Équation homogène . . . 1931
31.10.2 Équation de Bernoulli . . . 1931
31.10.3 Équation de Riccati . . . 1932
31.10.4 Équation différentielle exacte . . . 1932
31.11 Distributions pour les équations différentielles . . . 1933
31.11.1 Équation de Schrödinger . . . 1933
31.12 Équations différentielles du premier ordre . . . 1937
31.13 Premier ordre, variables séparables . . . 1939
31.14 Équations différentielles linéaires du premier ordre . . . 1941
31.14.1 Méthode de variation de la constante . . . 1943
31.15 Équations différentielles linéaires du second ordre . . . 1944
31.15.1 Équations différentielles linéaires du second ordre homogènes à coefficients constants . . . 1945
31.15.2 Linéaires du second ordre à coefficients constants, non homogènes . . . 1946
31.16 Fonction de Green . . . 1948

32 Équations aux dérivées partielles (Vol 4) 1951
32.1 Symbole principal, équation des caractéristiques . . . 1951
32.2 Méthode des caractéristiques pour l'ordre 1 . . . 1951
32.2.1 Un exemple complet un peu minimal . . . 1952
32.2.2 Un théorème d'existence et d'unicité . . . 1954
32.3 Méthode des caractéristiques pour l'ordre 2 . . . 1957
32.3.1 Principe général . . . 1957
32.3.2 Exemple : l'équation d'onde . . . 1958
32.4 Classification des équations du second ordre . . . 1960
32.4.1 Problème au limite . . . 1961

32.5 Principe du maximum . . . 1962
32.6 Quelques exemples . . . 1965
32.6.1 Un changement de variables . . . 1965

33 Numérique (Vol 4) 1969
33.1 Introduction . . . 1969
33.2 Représentations numériques . . . 1969
33.2.1 Entier relatif en complément à deux (binaire) . . . 1969
33.2.2 Représentation en virgule flottante . . . 1971
33.2.3 Simple précision, IEEE-754 . . . 1971
33.3 Problèmes pour écrire des nombres . . . 1974
33.3.1 Troncature : la base . . . 1974
33.3.2 Troncature : le drift . . . 1975
33.3.3 Quelques bonnes règles . . . 1976
33.3.4 Erreur de “cancellation” . . . 1976
33.3.5 Calcul d’une dérivée . . . 1978
33.3.6 Erreur d’absorption . . . 1978
33.4 Conditionnement et stabilité . . . 1979
33.4.1 Comment choisir et penser le K ? . . . 1982
33.5 Un peu de points fixes . . . 1983
33.5.1 Choix de la fonction à point fixe . . . 1983
33.5.2 Convergence quadratique . . . 1984
33.5.3 Convergence . . . 1986
33.6 Méthode de Newton . . . 1987
33.6.1 « Justification » par la formule par Taylor . . . 1987
33.6.2 « Justification » par points fixes . . . 1988
33.6.3 Convergence de la méthode de Newton . . . 1988
33.6.4 Formalisation de l’algorithme . . . 1990
33.6.5 Caractéristiques . . . 1991
33.6.6 Exemple de la racine carrée . . . 1992
33.6.7 Si multiplicité . . . 1992
33.6.8 Et la dérivée ? . . . 1993
33.6.9 Méthode de Newton : le cas général . . . 1993
33.7 Estimation de l’ordre de convergence . . . 1995
33.8 Autres méthodes . . . 1995
33.8.1 Méthode de Schröder . . . 1995
33.8.2 Halley . . . 1996
33.9 Méthode des sécantes variables . . . 1996
33.9.1 Aitken . . . 1997
33.10 Équations algébrique . . . 1997
33.10.1 Résoudre un système linéaire . . . 1997
33.10.2 Caractéristiques . . . 1998
33.10.3 Définitions . . . 1998
33.11 Équations non linéaires . . . 1999
33.11.1 Méthode de bisection . . . 2000
33.12 Efficacité . . . 2002
33.13 Exemples sous forme d’exercices . . . 2003
33.14 Approximations de fonctions . . . 2006
33.14.1 Critère d’interpolation . . . 2006
33.14.2 Base de Newton . . . 2007
33.14.3 Méthode des minimums quadratiques . . . 2009
33.14.4 Notre espace de Hilbert . . . 2010
33.14.5 Droite de régression . . . 2011
33.15 Conditionnement d’une matrice . . . 2012

33.15.1 Perturbation du vecteur . . . 2014
33.15.2 Perturbation de la matrice . . . 2015
33.16 Système linéaires (généralités) . . . 2016
33.16.1 Les méthodes directes . . . 2017
33.16.2 Méthodes itératives . . . 2017
33.17 Système linéaires (méthodes directes) . . . 2018
33.17.1 Inversion de matrice triangulaire . . . 2018
33.17.2 Transformation gaussienne . . . 2018
33.17.3 Méthode de Gauss pour résoudre des systèmes d'équations linéaires . . . 2019
33.17.4 Méthode de Gauss sans pivot (décomposition LU) . . . 2020
33.17.5 Matrice de permutation élémentaire . . . 2024
33.18 Méthode de Gauss avec pivot partiel (décomposition PLU) . . . 2025
33.18.1 L'idée . . . 2025
33.18.2 Le théorème . . . 2026
33.18.3 D'un point de vue algorithmique . . . 2030
33.18.4 Exemples . . . 2031
33.19 Résolution de systèmes linéaires (suite) . . . 2034
33.19.1 Déterminant . . . 2034
33.19.2 Plusieurs termes indépendants . . . 2034
33.19.3 Cholesky . . . 2035
33.20 Système linéaire (méthodes itératives) . . . 2038
33.20.1 La méthode générale . . . 2039
33.20.2 Jacobi . . . 2039
33.20.3 Gauss-Seidel . . . 2040
33.20.4 Autres . . . 2040
33.21 Indices connectés, matrice irréductible . . . 2040
33.22 Localisation des valeurs propres . . . 2041
33.22.1 Matrices à diagonale dominante . . . 2044
33.22.2 M-matrice . . . 2047

34 Méthode des différences finies (Vol 4) 2051
34.1 Problèmes de dimension un . . . 2051
34.1.1 Un schéma à cinq points . . . 2052
34.1.2 Exemple . . . 2056
34.2 Problèmes de dimension deux . . . 2057
34.2.1 Discrétisation en croix . . . 2057
34.2.2 Discrétisation en carré . . . 2058
34.2.3 Résolution de la discrétisation en croix . . . 2059
34.3 Consistance, convergence . . . 2061
34.3.1 Définitions, mise en place . . . 2061
34.3.2 Exemple . . . 2063
34.3.3 Consistance, stabilité et convergence . . . 2064
34.3.4 Exemple : schéma à cinq points, laplacien en croix . . . 2065
34.4 Autres laplaciens . . . 2066
34.4.1 Travail avec le laplacien à 9 points . . . 2069

35 Variables aléatoires et théorie des probabilités (Vol 4) 2071
35.1 Espace de probabilité . . . 2071
35.2 Variables aléatoires . . . 2072
35.2.1 Indépendance . . . 2072
35.2.2 Lois conjointes et indépendance . . . 2075
35.2.3 Somme et produit de variables aléatoires indépendantes . . . 2076
35.2.4 Espérance . . . 2078
35.2.5 Variance . . . 2079

35.2.6 Covariance . . . 2080
35.2.7 Probabilité conditionnelle : événements . . . 2081
35.2.8 Espérance conditionnelle . . . 2083
35.2.9 Probabilité conditionnelle : tribu . . . 2090
35.2.10 Variables de Rademacher indépendantes . . . 2093
35.2.11 Un petit paradoxe . . . 2094
35.2.12 Inégalité de Jensen . . . 2101
35.2.13 Fonction de répartition . . . 2101
35.2.14 Fonction caractéristique . . . 2101
35.2.15 Fonction génératrice des moments, transformée de Laplace . . . 2103
35.2.16 Loi d'une variable aléatoire . . . 2104
35.2.17 Changement de variables . . . 2105
35.3 Convergence . . . 2105
35.4 Loi des grands nombres, théorème central limite . . . 2110
35.4.1 Loi des grands nombres . . . 2110
35.4.2 Théorème central limite . . . 2112
35.4.3 Marche aléatoire . . . 2116
35.5 Les lois usuelles . . . 2116
35.5.1 Loi de Bernoulli . . . 2116
35.5.2 Loi binomiale . . . 2117
35.5.3 Loi multinomiale . . . 2118
35.5.4 Loi géométrique . . . 2118
35.5.5 Loi de Poisson . . . 2119
35.5.6 Loi exponentielle . . . 2119
35.5.7 Approximation de la binomiale par une Poisson . . . 2122
35.5.8 Loi de Poisson et loi exponentielle . . . 2123
35.5.9 Loi normale . . . 2125
35.5.10 Vecteurs gaussiens . . . 2126
35.5.11 Variable aléatoire de Rademacher . . . 2130
35.5.12 Loi de Student . . . 2132
35.5.13 Indépendance, covariance et variance de somme . . . 2133
35.6 Estimation des grands écarts . . . 2133
35.7 Simulations de réalisations de variables aléatoires . . . 2137
35.7.1 Générateur uniforme . . . 2137
35.7.2 Simulation par inversion . . . 2137
35.7.3 Algorithme de Box-Muller . . . 2138
35.7.4 Méthode du rejet . . . 2139
35.7.5 Simuler une loi géométrique à l'ordinateur . . . 2141
35.7.6 Simuler une loi exponentielle à l'ordinateur . . . 2141
35.7.7 Simuler une loi de Poisson à l'ordinateur . . . 2141
35.8 Sage . . . 2142
35.8.1 Loi exponentielle . . . 2142
35.8.2 Inverser des lois . . . 2142
35.9 Monte-Carlo . . . 2143
35.9.1 Intervalle de confiance . . . 2144
35.10 Résultats qui se démontrent avec des variables aléatoires . . . 2146
35.10.1 Nombres normaux . . . 2146
35.10.2 Théorème de Bernstein . . . 2148

36 Statistiques (Vol 4) 2153
36.1 Notations et hypothèses . . . 2153
36.2 Modèle statistique . . . 2153
36.3 Modèles d'échantillonnages . . . 2156
36.4 Estimation ponctuelle . . . 2159

36.5 Statistiques et estimateurs . . . 2161
36.5.1 Qualité des estimateurs . . . 2161
36.5.2 Méthode des moments . . . 2162
36.5.3 Méthode de substitution . . . 2164
36.5.4 Méthode du maximum de vraisemblance . . . 2164
36.5.5 Exemples sous forme d'exercices . . . 2164
36.5.6 Estimation d'une fonction de répartition . . . 2166
36.5.7 Exemples sous forme d'exercices . . . 2167
36.5.8 Espérance et variance d'un estimateur . . . 2169
36.6 Estimation par intervalle de confiance . . . 2169
36.6.1 Région de confiance . . . 2172
36.6.2 Fonction pivotale . . . 2172
36.6.3 Sondage de proportion . . . 2176
36.7 Estimer une densité lorsqu'on ne sait rien . . . 2177
36.7.1 Distance entre des mesures . . . 2177
36.7.2 Estimateur par fenêtres glissantes . . . 2178
36.8 Test d'hypothèses, prise de décision . . . 2180
36.8.1 Exemple : qualité des pièces d'usine . . . 2180
36.8.2 Exemple : la résistance d'un fil . . . 2180
36.8.3 Vocabulaire et théorie . . . 2181
36.8.4 Risque de première et seconde espèce . . . 2182
36.8.5 Modèle paramétrique de loi gaussienne . . . 2183
36.9 Tests paramétriques . . . 2184
36.10 Tests d'adéquation . . . 2185

37 Chaînes de Markov à temps discret (Vol 4) 2191
37.1 Généralités . . . 2191
37.2 Chaînes de Markov sur un ensemble fini . . . 2194
37.3 Marche aléatoire sur $\mathbb{Z}$ . . . 2196
37.3.1 Chaînes de Markov homogènes . . . 2198
37.3.2 Graphe de transition . . . 2199
37.3.3 Chaîne de Markov définie par récurrence . . . 2200
37.4 Classification des états . . . 2202
37.4.1 Chaînes irréductibles . . . 2204
37.4.2 Nombre de visites . . . 2206
37.5 Mesure invariante . . . 2210
37.6 Convergence vers l'équilibre . . . 2212
37.7 Processus de Galton-Watson . . . 2215

38 Martingales (Vol 4) 2219
38.1 Convergence de martingales . . . 2219
38.2 Temps d'arrêt et martingale terminée . . . 2222
38.3 Décomposition de martingales . . . 2224
38.4 Problème de la ruine du joueur . . . 2226
38.4.1 Le cas où la pièce est truquée . . . 2227
38.4.2 Le cas où la pièce est non truquée . . . 2230
38.4.3 Un petit complément . . . 2232

39 Processus de Poisson (Vol 4) 2235
39.1 Processus de Poisson . . . 2235
39.2 Quelques trucs sur la simulation . . . 2238
39.2.1 Le théorème central limite pour Markov . . . 2239
39.2.2 Feuille 5 . . . 2239
39.2.3 Feuille 6 . . . 2239

39.2.4 Feuille 7 . . . 2240
39.2.5 Simuler des lois conditionnelles . . . 2240

40 Langages (Vol 4) 2241
40.1 Langages . . . 2241
40.1.1 Alphabets et mots . . . 2241
40.1.2 Langage . . . 2242

41 Utilisation dans les autres sciences (Vol 4) 2245
41.1 Démystification du MRUA . . . 2245
41.1.1 Preuve de la formule . . . 2245
41.1.2 Interprétation graphique . . . 2246
41.2 Relativité en mécanique newtonienne . . . 2246
41.2.1 Relativité du mouvement . . . 2246
41.2.2 Bob et Alice . . . 2246
41.3 Invariance de la vitesse de la lumière . . . 2247
41.3.1 Champ de gravitation et électrique . . . 2247
41.3.2 Support du champ : pas d'éther . . . 2248
41.3.3 Le problème . . . 2248
41.4 Conséquences . . . 2248
41.4.1 Ligne d'univers . . . 2248
41.4.2 Transformations de Lorentz . . . 2249
41.4.3 Conditions d'existence . . . 2252
41.4.4 La notion d'intervalle . . . 2253
41.4.5 Le cône de lumière d'un point . . . 2253
41.4.6 Contraction des longueurs . . . 2254
41.4.7 Dilatation des intervalles de temps . . . 2254
41.4.8 Invariance de l'intervalle . . . 2255
41.4.9 Vitesse limite . . . 2260
41.5 Applications . . . 2260
41.5.1 Le GPS . . . 2260
41.5.2 Les ondes électromagnétiques . . . 2260
41.6 Mécanique relativiste . . . 2260
41.6.1 Des problèmes, toujours des problèmes . . . 2261
41.6.2 Loi d'addition des vitesses . . . 2261
41.6.3 L'action d'une force . . . 2262
41.6.4 Équivalence entre la masse et l'énergie . . . 2263
41.7 Principe de correspondance . . . 2263

42 Exemples avec Sage (Vol 4) 2265
42.1 Graphiques . . . 2265
42.1.1 Autres . . . 2265

43 Épilogue : la constante de Weiner (Vol 4) 2285

44 GNU Free Documentation License (Vol 4) 2287

Bibliographie 2295

Index

λ-système, 912
p-Sylow, 287
p-groupe, 287
q-orthogonal, 554

Abel
    angulaire, 1450
    convergence radiale, 1040
abscisse
    curviligne, 1536
absolument continue, 1135
absorbant, 2202
abélianisé, 161
abélien
    groupe, 124
accroissement, 867
accroissements finie
    dérivée partielle, 758
accumulation
    dans espace vectoriel normé, 522
accélération d'un chemin, 1519
action, 169
    adjointe, 170
    de groupe
        Wedderburn, 1370
    domaine fondamental, 172
    fidèle, 170
    libre, 173
    transitive, 173
action de groupe, 502
    sur des matrices, 1500
adhérence, 519
adjoint, 511
affine
    application, 476
    espace, 473
    sous-espace, 479
aire, 1496
aire dans $\mathbb{R}^2$, 1418
algorithme, 1998
    consistant, 1998
    convergeant, 1998
    facteurs invariants, 271
    fortement consistant, 1998
    stable, 1998
algèbre, 203
    de parties, 907
    engendrée, 203
algébrique, 330
algébriquement
    indépendant, 369
alignement
    dans un espace projectif, 1580
alphabet, 2241
alterné
    groupe, 299
    polynôme, 365
alternée
    forme linéaire, 495
analytique
    au sens complexe, 852
angle
    d'une courbe, 1555
    entre deux droites, 1284
    orienté de vecteurs, 1280
angle entre deux vecteurs, 1236
Anneau
    $\mathbb{Z}/n\mathbb{Z}$
        polynôme cyclotomique, 1364
anneau, 125
    $\mathbb{Z}/n\mathbb{Z}$, 297, 1360, 1368, 1373
    de séries formelles, 1100
        utilisation, 1700
    euclidien
        facteurs invariants, 271
    factoriel, 210
    noethérien, 217
    principal, 215, 315, 532
        utilisation, 317
    quotient par un idéal, 126
    à division, 311
anneau commutatif, 125
anneau intègre, 128
application
    de classe C^k, 803
    différentiable, 656, 770, 800, 803, 1154, 1500
        extrémums lié, 1168
    définie positive, 420

en escalier, 1488
linéaire
théorème de Banach-Steinhaus, 638
mesurable, 911
multilinéaire, 622
ouverte, 619
semi-définie positive, 421
tangente, 774
application affine, 248
application bilinéaire, 420
application réciproque, 389
approximation
de fonctions
par des polynômes, 2148
de l'unité, 1742
par polynômes, 1125
polynomiale, 1687
apériodique
chaine de Markov, 2214
état d'une chaine de Markov, 2213
arc
géométriques, 1535
paramétré, 1519
arc cosinus, 1234
arc sinus, 1232
arc tangente, 1230
archimédien, 134
associée
subdivision, 1488
associés
éléments d'un anneau, 209
asymptotiquement pivotale, 2172
attractif
point fixe, 1140
automorphisme
d'espace vectoriel, 242
axiome
du choix, 106

Baire
espace, 471
théorème, 466, 471
tribu, 905
Banach
espace, 449
barycentre
cas affine, 481
cas vectoriel, 1255
enveloppe convexe, 485
base, 234
d'un module, 202
de Newton, 2007
de topologie, 383
dénombrable, 393
espace métrique, 393
duale, 278
espace préhilbertien, 1657
hilbertienne, 1657
utilisation, 1819
locale, 1636
base associée à un repère cartésien, 474
base canonique de $\mathbb{R}^m$, 238
base préduale, 279
Bergman (espace), 1802
Bernoulli, 2117
somme, 2134
Berry-Esséen (borne), 2115
Bessel
inégalité, 1656
biais
d'estimateur, 2162
bien
conditionné, 1980
enchaîné, 464
bijection, 106, 389
bilinéaire, 622
binormale, 1546
birapport, 1592
birapport dans $\mathbb{C} \cup \{\infty\}$, 1612
birégulier
point sur une courbe, 1533
Bolzano-Weierstrass
espaces métriques, 394
bon
ordre, 106
bord, 441
borné, 393
partie de V, 514
temps d'arrêt, 2222
bornée
différentielle, 791
partie de $\mathbb{R}^m$, 713
suite, 404
borélienne
fonction, 917
tribu, 932
boréliens, 932
boule
avec semi-normes, 467
fermée, 440, 514
ouverte, 392, 440, 514
boule dans un corps, 134
Bruhat (décomposition), 897
Burnisde
formule, 173
Bézout
anneau principal, 215

calcul effectif, 187
nombres entiers, 184
polynômes, 324

canonique
base, 238
décomposition, 109
espace affine, 474
Cantor
ensemble, 962
caractère, 1110
abélien, 1103
de S_4, 1117
groupe diédral, 1350
irréductible, 1110
caractéristique
d'un anneau, 199
d'une équation différentielle, 1951
polynôme, 534
sous-groupe, 157
cardinal, 110
cardioïde, 1533
carré
dans un corps fini, 1374
carrée
matrice, 250
carte, 1406
catégorie
ensemble de première, 466
Cauchy
critère
uniforme, 818
déterminant, 504
formule, 1680
produit, 1038
suite, 454
théorème, 287
Cauchy-continue, 706
Cauchy-Riemann, 1675
Cauchy-Schwarz, 422, 1647
Cayley
théorème, 287
cellule d'un pavage, 1488
centrale (application), 1110
centralisateur, 126, 157
centre
d'un anneau, 126
d'un groupe, 157
d'une rotation, 1270
cercle
circonscrit à une courbe, 1562
dans la sphère de Riemann, 1603
cercle-droite, 1604
cercles
perpendiculaires, 1573
Cesaro
moyenne, 633
chaine, 464
de Markov, 2191
apériodique, 2214
convergence, 2214
finie, 2194
homogène, 2191
irréductible, 2199
récurrente positive, 2209
régulière, 2194
champ
conservatif, 1462
de vecteurs, 1425
champ dérivant d'un potentiel, 1462
changement de variable, 1535
Chasles, 473
Chemin
classe C^2, 1424
chemin, 1424, 1519
dans $\mathbb{R}^p$, 1519
circulation, 1460
classe
d'association, 209
de conjugaison, 160
classe C^1, 657
classe C^0, 814
classe C^p, 814
classe d'association, 209
classe de conjugaison
dans S_4 , 176
clôture algébrique, 331
codimension, 238
coefficients
de Fourier, 1744
coefficients binomiaux, 196
coefficients de Fourier, 1756
coercion, 1670
coercive, 1194
colinéarité, 1579
combinaison
convexe, 481
combinatoire, 1264
commutant, 581
commutateur
dans un groupe, 161
commutatif
groupe, 124
compacité, 394, 461, 464, 1691, 1715
sous-groupes du groupe linéaire, 900
théorème de Dini, 820
utilisation, 1204

théorème de Montel, 1801
compact, 381, 517
arc paramétré, 1519
boule unité, 402
et fonction continue, 394, 409
fermé et borné, 401
implique fermé, 384
intervalle $[a, b]$, 400
le coup du, 1133
localement, 382
opérateur, 1717
produit dénombrable, 465
produit fini, 464
quasi, 381
relatif, 1717
relativement, 382
suite exhaustive, 463
séquentiellement, 382
complet
$\mathbb{R}$
corps, 149
espace métrique, 455
corps, 134
espace mesuré, 921
espace topologique, 449
métrique, 449
complète
famille de projecteurs, 203
complément
à deux, 1970
complémentaire, 107
complétion
projective, 1582
complétude, 1210, 1212, 1730
espaces L^p, 1731
complété
d'un espace métrique, 1212
composante, 867
composante connexe, 380
composition
suite de, 165
concaténation
de langages, 2242
de mots, 2241
concave, 1170
log-concave, 1177
condition initiale, 1938
conditionnement
absolu, 1979
d'une matrice inversible, 2012
relatif asymptotique, 1998
conjugués
éléments d'une extension, 330
connectés
indices d'une matrice, 2040
connexe par arc, 412
connexité, 464
définition, 380
et intervalles, 408
fonction holomorphe, 1215
indice d'une courbe, 1679
le groupe $\mathrm{GL}^+(n, \mathbb{R})$, 888
par arc
fonction différentiable, 800
points d'accumulation, 461
prolongement analytique, 1215
signature d'une forme quadratique, 1199
théorème de Runge, 1687
théorème des valeurs intermédiaires, 696
utilisation
Brouwer, 1683
Conservative, 1428
consistance
estimateur, 2161
ordre, 2062
constructible
angle, 1399
point, 1396
réel, 1396
construction
des réels, 142
contenu, 229
continue
fonction entre espaces métriques, 459
fonction entre espaces topologiques, 386
forme différentielle, 763
sur espace métrique, 698
uniformément, 714
continue sur $\mathbb{R}$, 693
continuité
fonction définie par une intégrale, 1132
séquentielle, 388
contraction, 1141
convergeant
schéma discret, 2062
convergence
commutative, 640
dans un espace vectoriel normé, 402
de martingales, 2223
de suite, 373
en loi, 2106
en probabilité, 2106
ordre, 1995
presque surement, 2105
quadratique, 1984
rapidité, 1073, 1826, 1827, 1994, 2134

suite
dans un corps, 134
suite dans $\mathbb{R}^m$, 443
suite numérique, 402, 1125, 1811
Abel angulaire, 1450
uniforme, 818
intégrale, 1130
théorème de Dini, 820
convergence absolue, 625
convergence forte, 639
convergence normale, 625
convergence uniforme
série de fonctions, 625
convergent
estimateur, 2161
convexe
courbe plane, 1564
fonction sur $\mathbb{R}^n$, 1179
convexité
barycentre, 483
enveloppe de $O(n)$, 896
fonction, 1170
inégalité de Jensen, 1185
locale, 1793
méthode de Newton, 1990
utilisation, 1204
convolution, 2077, 2217
convolution sur S^1, 1756
coordonnées
cartésiennes
dans un espace affine, 488
curvilignes, 1636
cylindrique, 1335
dans un espace affine, 474
homogène, 1598
sphériques, 1336
coordonnées barycentrique, 489
coordonnées polaires, 1330
corps, 131
archimédien, 134
complet, 134
de décomposition, 353
de rupture, 344
polynôme cyclotomique, 1364
des fractions, 132, 133
des fractions rationnelles
utilisation, 1700
extension, 360, 366
fini, 1373, 1377, 1381
Wedderburn, 1370
formellement réel, 311
ordonné, 133
premier, 313
corps algébriquement clos, 331
cosinus, 1217
hyperbolique, 1076
courbe, 1519
de Jordan, 1571
efficacité, 2182
fermée, 1552
simple, 1552
étude métrique, 1819
courbe de niveau, 742, 745
courbure, 1546
signée, 1553
totale, 1554
covariance, 2080
critique
Galton-Watson, 2217
point, 1165
point d'un arc, 1533
région, 2181
valeur, 2182
critère
Abel, 1036
Abel pour intégrales, 1131
Cauchy
uniforme, 818
de Cauchy, 149, 455
Weierstrass, 1131
série de fonctions, 823
critère du quotient, 629
cyclique
endomorphisme, 530
matrice, 530
cycloïde
coordonnées normales, 1538
longueur, 1532
Cône de lumière, 2253

d, 756
degré
application $S^1 \to S^1$, 1556
d'un polynôme, 205
d'une représentation, 286
extension de corps, 328
dense, 377
nulle part, 466
densité, 1212
conjointe, 2076
d'une variable aléatoire, 2072
dans un espace de fonction
critère de Weyl, 1811
de $\mathbb{Q}$ dans $\mathbb{R}$, 398
utilisation, 1176
de $\mathrm{GL}(n, \mathbb{R})$ dans $\mathbb{M}(n, \mathbb{R})$, 889
de $\mathscr{D}(\mathbb{R}^n)$ dans $L^1(\mathbb{R}^n)$, 1799

de $C_c^\infty(\mathbb{R}^d)$ dans $L^p(\mathbb{R}^d)$, 1737
de $L^2\big([0,1]\big)$ dans $L^p\big([0,1]\big)$, 1738
de $S^+(n,\mathbb{R})$ dans $S^{++}(n,\mathbb{R})$, 893
des fonctions étagées dans L^p, 1737
des polynômes
dans $C_c^0[0,1]$, 2148
matrices diagonalisables dans $\mathbb{M}(n,\mathbb{C})$, 889
mesure, 983
points extrémaux dans $\mathcal{L}$, 895
prolongement, 1210
densité d'une mesure, 983
diagonale
dominante, 2044
fortement dominante, 2044
strictement dominante, 2044
diagonalisable, 542
et polynôme minimum scindé, 542
exponentielle, 1081
diagonalisation
cas complexe, 548
cas réel, 551
endomorphisme autoadjoint, 602
simultanée, 544
diamètre, 713
difféomorphisme, 607, 1342, 1471
de classe C^k, 811
différence
centrée, 2051
divisée, 2008
progressive, 2051
régressive, 2051
différentiabilité, 800
différentielle, 656
de $u \mapsto u^{-1}$, 671
partielle, 670
totale, 867
dilatation, 882
dilatation (matrice), 268
dimension, 238
n-formes multilinéaires alternées, 495
définition, 238
sous espace affine, 479
utilisation, 485
direction, 756, 1541
sous-espace affine, 479
Dirichlet
noyau, 1807
théorème, 1808
théorème (sur les nombres premiers), 1368
disque de convergence, 1036, 1037
distance, 392
associée à une norme, 397
compatible, 452
entre deux mesures de probabilités, 2177
invariante, 452
point et ensemble, 445
distance discrète, 445
distance produit, 438
distingué
sous-groupe, 157
distribution, 1849
de Dirac, 1856
produit par une fonction, 1851
tempérée, 1855
équation de Schrödinger, 1933
divergence, 873
diviseur
dans un anneau, 128
de zéro, 128
de zéro à droite, 128
polynôme, 229
diviseur de zéro, 128
division
euclidienne, 183, 228
harmonique, 1615
diédral, 1265
domaine, 681
fondamental d'une action, 172
dominé
modèle statistique, 2160
dominée
convergence (Lebesgue), 980
mesure, 988
droite, 248
dans la sphère de Riemann, 1603
projective, 1579
droite réelle complétée, 417
droite vectorielle, 248
dual
d'un espace de Hilbert, 1653
de L^p, 1784
de $L^p(\Omega)$, 1789
de $\mathbb{M}(n,\mathbb{K})$, 284
dual algébrique, 278
dual topologique, 618, 674
Dunford
décomposition, 568
dyadique, 999
développement, 147
décalage, 1971
décomposition
Bruhat, 897
canonique, 109
corps, 353
Dunford, 568
application, 1081

exponentielle de matrice, 569
Jordan
et exponentielle de matrice, 1080
polaire, 894
primaire, 567
sous-espaces caractéristiques, 567
spectrale, 567
décomposition décimale, 636
dénombrable, 110
à l'infini, 382
dénombrement, 1264
partitions de $\{1, \ldots, n\}$, 1100
déplacement, 1300
dérivabilité
fonction définie par une intégrale, 1134
lemme de Borel, 1098
dérivable, 724
au sens complexe, 796
fonction, 1889
dérivation
au sens des distribution
Sobolev, 1876
dérivé
groupe, 161
dérivée
dans Sobolev $H^1(I)$, 1873
directionnelle, 755, 759
distributionnelle, 1851
faible, 1846
fonction à valeurs dans E', 471
partielle, 755, 757, 866
seconde, 725
dérivée faible, 1844
déterminant, 495
Cauchy, 1128
d'un endomorphisme, 499
d'une famille de vecteurs, 496
de Cauchy, 504
et inversibilité, 500
forme linéaire alternée, 495
Gram, 504, 1128
interprétation géométrique, 1014
matrice, 256
résultant, 506, 1455
utilisation, 1204
Vandermonde, 501
détermination
logarithme, 1712
principale, 1712
développable
en série entière, 1083
développement
asymptotique, 864
limité
en zéro, 859
fonction holomorphe, 1675
premier ordre, 728
Taylor, 1500

écart-type, 2080
échantillon, 2154, 2156
effectif
empirique, 2186
efficacité
courbe, 2182
d'une méthode itérative, 2002
élément
de surface, 1436
de torsion, 179
élément de surface, 1412
ellipsoïde, 601
elliptique
équation aux dérivées partielles, 1960
endomorphisme, 242
cyclique, 530
diagonalisable, 591, 891, 1929
Dunford, 568
diagonalisation, 551
décomposition
polaire, 894
nilpotent
Dunford, 568
préservant une forme quadratique, 902
sous-espace stable, 568, 1929
endomorphisme direct, 510
endomorphisme préserve l'orientation, 510
engendré, 196
λ-système, 912
corps, extension, 341
idéal dans un anneau, 196
sous-espace affine, 479
sous-groupe, 157
tribu, 904
par une variable aléatoire, 2073
ensemble
de Cantor, 962
différence symétrique, 107
infini, 110
ensemble connexe, 380
ensemble des mots, 2241
ensemble quotient, 108
ensembles
disjoints, 105
entrelacement, 1110
enveloppe
convexe, 483
équation

différentielle
linéaire, 1891
ordinaire d'ordre 1, 1889
variables séparées, 1894
générale de degré n, 369
équi-intégrable, 2223
équivalence
arcs paramétrés, 1535
chemin, 1527
classe de fonctions, 1719
de représentations, 1110
de suites, 404
homotopie, 1679
norme, 436
suite de composition, 167
erreur, 2009, 2038
assignation, 1974
de consistance, 2062
discrétisation, 2062
quadratique, 2012
troncature, 1974
erreur relative, 1974
escalier, 945
Espace
de Sobolev, 1878
espace
L^2
Sobolev, 1876
L^p, 1722
Banach, 449
de Baire, 471
de Bergman, 1802
de fonctions
L^p, 1731
Sobolev H^1, 1876
de Hilbert
espace de Sobolev H^1, 1876
de probabilité, 2071
de Schwartz, 1797, 1855
de Sobolev, 1873, 1880
euclidien, 433
mesurable, 903
mesuré, 909
complété, 923
métrique, 392
base de topologie, 393
projectif, 1579
propre, 526
séparé, 378
tangent, 1408
topologique
métrisable, 458
vectoriel, 200
dimension, 495
vectoriel topologique
métrisable, 448
espace topologique, 371
espace vectoriel
topologique, 451
espace vectoriel normé, 396
espérance, 2078
conditionnelle, 2083, 2085
variable aléatoire, 2089
événement, 2085
estimateur, 2161
biais, 2162
consistant, 2161
convergent, 2161
de fonction de répartition, 2167
maximum de vraisemblance, 2164
estimation
des grands écarts, 2134
étrangers
polynômes, 229
Euclide
algorithme étendu, 186
lemme, 189
euclidien
anneau, 220
espace, 421
Euler
indicatrice, 1357
événement, 2071
exact
intervalle de confiance, 2170
excès
intervalle de confiance, 2170
exhaustive (suite de compacts), 463
exponentielle, 1062
convergence, 631
de matrice, 673, 1080, 1081, 1093
utilisation, 1162
existence, 1058
rapide, 1362
unicité, 846
exposant, 293, 591
d'un groupe, 160
extension
corps de base, 570
de corps, 327, 366
algébrique, 339
finie, 1383
simple, 341
utilisation, 1400
isométrie, 1211
séparable, 361

extension algébrique, 331
extension algébriquement clos, 331
extrapolation, 2007
extrémal
point dans un convexe, 895
extrémité
d'un intervalle, 408
extrémum, 1500
lié, 1168
local
relatif, 1168
extrémums, 1166
volume d'un ellipsoïde, 1204

facteur
intégrant, 1933
factoriel
anneau, 210
factorisation
de polynôme, 231, 343
faisceau de droites, 1600
famille
sommable, 642
famille trigonométrique
sur S^1, 1755
famille équicontinue, 465
Fatou, 971
Fejér
noyau, 1807
fermeture, 519
fermeture séquentielle, 446
fermé, 371, 441, 517
dans un compact, 384
fidèle (action), 170
filtration, 2219
fine
subdivision, 1520
fixateur, 169
flot, 1905, 1951
flux
d'un champ de vecteur, 1440
flux d'un champ de vecteurs, 1438
fonction, 681
Γ d'Euler, 1696
borélienne, 917
caractéristique, 1490
d'une variable aléatoire, 2101
continue
par morceaux, 949
égales, 447
convexe, 1170, 1176
croissante, 681
de Dirichlet, 1813
de Möbius, 1392
de répartition, 2101
décroissante, 681
définie par une intégrale, 1097, 1131, 1136, 1137, 1715
Γ d'Euler, 1696
utilisation, 2121
en escalier intégrable, 1489
génératrice, 2103
holomorphe, 796, 1715
théorème de Montel, 1801
image, 681
monotone, 681
par morceaux, 949
méromorphe
Γ d'Euler, 1696
valeurs vectorielles, 867
à décroissance rapide, 1797
étagée, 1736
fonction continue en un point, 386
fonction de classe C^1, 791
fonction dérivée, 724
fonctionnelle
énergie, 1673
fonctions équivalentes, 379
fondamental
domaine d'une action, 172
forme
bilinéaire, 420
non dégénérée, 595
différentielle, 762
exacte, 1448
fermée, 1448
linéaire
différentielle, 1168
quadratique, 554, 565, 1199, 1201, 1500
groupe orthogonal, 902
forme bilinéaire symétrique, 420
forme canonique
fonction simple, 946
forme linéaire, 278
alternée, 495
formule
Bayes, 2082
Burnside, 173
d'expulsion (produit vectoriel), 431
de Cauchy, 1680
Hadamard, 1037
inversion Möbius, 1392
probabilité totales, 2082
sommatoire de Poisson, 1826
Taylor
reste intégral, 1485
utilisation, 1994

formule de Stirling, 1507
Fourier, 1826
série
utilisation, 1819
transformée
groupe abélien fini, 1105
fraction
rationnelle
intégration, 1455
fraction rationnelle, 1481
fractions
rationnelles, 132
fractions (corps), 133
Fredholm
équation, 1143
Frenet
formules, 1547
Fresnel
intégrale, 1703
Frobenius
morphisme, 200
Frobénius
réduction, 578
frontière, 377, 441, 521
fréquence
empirique, 2186, 2206
Fubini
théorème
dans $\mathbb{R}^n$, 1029

Galton-Watson
sous-critique, 2217
sur-critique, 2217
Gauss
lemme
polynômes, 324
somme de, 1375
Gershgorin
disque, 2042
gradient, 771, 780, 790
Gram (déterminant), 504
Gram-Schmidt, 433
graphe, 681, 727
de transition (chaine de Markov), 2199
fonction, 741
fonction de deux variables, 744
groupe, 124
p-groupe, 287
$\mathrm{GL}(n, \mathbb{R})$, 1201
action, 1621
utilisation, 902
agissant sur un ensemble
diédral, 1257
alterné, 299
circulaire, 1619
de Galois, 368
de permutation, 1350
de permutations, 1264
caractères de S_4, 1117
de torsion, 179
des isométries
espace métrique, 444
des symétries, 1297
diédral, 1257, 1265
générateurs (preuve), 1258
générateurs (utilisation), 1350
dérivé, 161
de $\mathrm{GL}(n, \mathbb{K})$, 887
de $\mathrm{SL}(n, \mathbb{K})$, 887
du groupe alterné, 302
du groupe symétrique, 300
en géométrie, 1257
et géométrie, 495, 1264, 1621
isométries du cube, 308
fini, 289, 297, 1264, 1360, 1370, 1373
alterné, 301
diédral, 1257
Wedderburn, 1370
linéaire, 897
décomposition polaire, 894
enveloppe convexe de $\Omega(n)$, 896
hyperplan, 285
sous-groupes compacts, 900
modulaire, 1621
orthogonal, 513
d'une forme quadratique, 902
partie génératrice, 301, 1360, 1621
permutation, 495, 502, 897, 1360
diédral, 1257
projectif, 1586
quotient, 165
résoluble, 168
spécial orthogonal, 514
symétrique, 174
action sur un triangle, 1107
groupe cyclique, 159
groupe de pavage, 1301
groupe dérivé
de $\mathrm{GL}(n, \mathbb{C})$, 881
Grönwall (lemme), 1889, 1890
générateur, 159
génératrice
partie d'un module, 202
géométrie
avec des groupes, 1258, 1621
avec nombres complexes, 1258, 1621
géométrique

avec des nombres complexes, 1819

Hadamard
conditions, 1962
formule, 1037
Hardy-Littlewood (théorème), 1125
Hausdorff, 378
Heine (théorème), 714
hermitienne, 421
hessienne, 812
Hilbert
espace, 1647
holomorphe, 796
sur un compact, 796
homographie, 1585, 1621
sur $\mathbb{C} \cup \{\infty\}$, 1607
homogène
chaine de Markov, 2191
homotopie, 1558
homéomorphisme, 389
hyperbolique
équation aux dérivées partielles, 1960
hyperplan, 587, 1251
de $\mathbb{M}(n, \mathbb{K})$, 286
sépare
au sens strict, 1793
séparer
au sens large, 1793
hypothèse
alternative, 2181
composite, 2181
multiple, 2181
nulle, 2181
simple, 2181

identifiable, 2160
identité de polarisation, 554
idéal
bilatère, 126
dans un anneau, 126
maximal, 198
principal
à droite, 211
à gauche, 211
image, 681
incompressible
champ de vecteur, 875
indicatrice d'Euler, 1357
indice, 164
d'une courbe dans $\mathbb{C}$, 1679
de rotation, 1557
indice d'inertie, 562
inductif, 107
induite
topologie, 518
tribu, 904
indécomposable
module, 203
indépendance, 2075
affine, 487
algébrique, 369
projective, 1587
sous tribus, 2072
variables aléatoires, 2073
événements, 2072
utilisation, 2146, 2226
infimum, 150
inférence statistique, 2153
injection, 106, 388
interpolation, 2007
Intervalle, 2253
intervalle, 107, 412, 681
fermé, 681
longueur, 952
ouvert, 681
intervalle de confiance
asymptotique, 2174
intégrable, 974
fonction non en escalier, 1495
fonction positive, 1499
fonction à valeurs vectorielles, 977
intégrale
calcul, 1811
convergente, 1011, 1454
d'une fonction sur une carte, 1413
d'une fonction sur une variété, 1418
d'une forme différentielle, 1427
fonction en escalier, 1489
Fresnel, 1703
impropre, 1011
sur un chemin, 1424
intégrale d'une fonction, 966
intégrale d'une forme différentielle, 1426
intégrale de Dirichlet, 1517
intégrale sur une carte, 1409
intégration
fraction rationnelle, 1455
intérieur, 376, 441
d'un ensemble, 516
point, 516
invariance cyclique
trace, 252
invariant
de similitude, 578
invariante
mesure
pour une chaine de Markov, 2210

inverse
dans un groupe, 124
inverse généralisé, 2138
inversion, 1574
dans le groupe symétrique, 174
inversion dans $\mathbb{C} \cup \{\infty\}$, 1606
involution, 544
inégalité
arithmético-géométrique, 1186
Bessel, 1656
Cauchy-Schwarz, 422, 1647
de Khintchine, 2131
de la moyenne, 800
des pentes, 1171
Hölder, 1727
utilisation, 2079, 2149
isopérimétrique, 1819
Jensen, 1185
espérance conditionnelle, 2101
pour une somme, 1185
version intégrale, 1726
Kantorovitch, 1186
Markov, 2110
Minkowski, 1728
triangulaire, 392, 396
irrationalité
$\sqrt{2}$, 138
irréductible
chaine de Markov, 2199
dans un anneau, 210
module, 203
représentation, 1108
isobarycentre, 481
isolé
point dans un espace vectoriel normé, 522
isomorphisme, 648
$(\mathbb{Z}/p\mathbb{Z})^* \simeq \mathbb{Z}/(p-1)\mathbb{Z}$, 1373
d'anneaux, 126
d'espaces topologiques, 389
de corps, 131
espace affine, 478
espace vectoriel normé, 614
espaces vectoriels, 242
isométrie
de forme quadratique, 561
de l'espace euclidien $\mathbb{R}^2$, 1258
espace euclidien
isométries du cube, 308
groupe, 444
isométrie (forme bilinéaire), 561
isométrie d'espaces métriques, 444
isotrope
cône, 600
totalement, 600
isotrope (vecteur), 600

jacobien, 790, 872
Jordan
chemin, 1441
courbe, 1571
réduction, 580
Jordan-Hölder, 165

Kronecker, 238

lacet, 1679
Lagrange
multiplicateur, 1168
polynôme, 284
théorème, 164
lagrangien, 1168
langage, 2242
itéré, 2243
itéré strict, 2243
unité, 2242
vide, 2242
Laplace
transformée, 2103
laplacien, 1643
Legendre
symbole, 1374
Leibnitz, 729
applications entre espaces vectoriels normés, 669
lemme
Borel, 1098
d'Euclide, 189
de Borel-Cantelli, 2109
de Gauss
contenu de polynôme, 1382
pour des entiers, 189
de Morse, 1500
de Schreider, 167
de Slutsky, 2107
de transport, 918
de Zorn, 107
des noyaux, 527
Fatou, 971
Gauss
dans un anneau principal, 216
polynômes, 324
Grönwall, 1889, 1890
Hadamard, 1139
regroupement, 2075
Schur complexe, 546
Schur réel, 550
lettres, 2241

Levi-Civita, 1637
libre
action, 173
partie, 233
partie d'un module, 202
librement engendré, 240
Ligne d'univers, 2249
limite, 566
d'ensembles, 911
d'une fonction, 385
de fonctions holomorphes, 1715
de suite
espace topologique, 373
fonction de plusieurs variables, 691
inférieure, 406
inversion, 1100, 1125, 1715, 1722
permutation
utilisation, 2103
suite dans $\mathbb{R}^m$, 443
suite numérique, 402
supérieure, 406
unicité, 386
limite pointée, 417
limite à droite, 692
linéaire
application, 241
linéairement indépendant
module, 200
Lipschitz, 1141
localement, 802
Lipschitzienne, 801
localement
intégrable, 1011
log-concave, 1177
logarithme
complexe, 1708
dans $\mathbb{C}$, 1705
de matrice, 1093
sur les réels positifs, 1059
loi
χ^2, 2132
binomiale
comportement asymptotique, 2134
conjointe, 2075
d'une variable aléatoire, 2104
de Poisson, 2119
des grands nombres
forte, 2111
pour les chaines de Markov, 2212
processus de Poisson, 2236
utilisation, 2134, 2146
marginale, 2075
normale
vecteur gaussien, 2126
parente, 2153, 2154
parente d'un échantillon, 2156
réciprocité quadratique, 1377
sans mémoire, 2121
Student, 2132
longueur
arc géométrique, 1536
d'un arc paramétré compact, 1520
d'un intervalle, 952
d'une arrête, 1486
élément de, 1530
longueur d'arc, 1522
Lotka-Volterra, 1922

M-matrice, 2047
maigre, 905
maigre (ensemble), 466
majorant, 106, 150
essentiel, 1724
Markov
inégalité, 2110, 2159
martingale, 2219
bornée dans $L^2(\Omega)$, 2220
matrice, 250, 897, 1621
compagnon, 576
creuse, 2017
cyclique, 530
de dilatation, 268
de permutation, 268
de similitude, 797
de Sylvester, 504
de transvection, 268
dense, 2017
hermitienne
racine carrée, 891
jacobienne, 780, 790
normale, 548
orthogonale, 266, 513
permutation
élémentaire, 2024
racine carrée, 891
réductible, 2040
semblable, 891
semblables, 565, 1201
stochastique, 2193
symétrique, 1201
réelle, 1199
trigonalisable, 545
équivalence, 270
dans le groupe linéaire, 902
matrice d'une application linéaire, 253
matrice d'une forme bilinéaire, 554
matrice de transition, 2192

matrice-colonne, 251
matrice-ligne, 250
matrices
  similitude, 270
maximale
  partie orthonormale, 1658
maximum, 106
  global, 1163
  local, 1163
mesurable
  application, 911
  au sens de m^*, 916
  ensemble, 909
  fonction, 917
  Lebesgue, 1485
mesure
  σ-finie, 909
  absolument continue, 988
  angle entre vecteurs, 1281
  complexe, 989
  dans une carte, 1412
  de Borel, 937
  de comptage, 1005, 1028
  de Haar, 1423
  de Lebesgue, 956
  de Radon, 937, 1693
  externe, 1485
  extérieure, 907
  finie, 909
  image, 934
  positive, 909
  probabilité, 2071
  produit, 996
  régulière, 937
    extérieure, 937
    intérieure, 937
  sur un ensemble de parties, 907
mesure σ-finie, 907
mesure finie, 907
mesure à densité, 983
minimum
  ensemble ordonné, 106
minorant, 106, 150
modulaire (groupe), 1621
module
  de continuité, 677
  indécomposable, 203
  irréductible, 203
  simple, 203
  à gauche, 200
module produit, 201
modèle
  paramétrique, 2154
  statistique, 2153
  échantillonnage, 2154, 2156
moment, 2078
  fonction génératrice, 2103
monogène, 159
  extension de corps, 341
monotonie, 1927
monoïde, 315
monôme, 227
morceau
  fonction continue ou monotone, 949
morphisme
  d'algèbres, 203
  d'anneaux, 125
  de corps, 131
  espace vectoriel normé, 614
  Frobenius, 200
morphisme de produits tensoriels, 648
mot, 2241
  longueur d'un mot, 2241
  mot vide, 2241
  nombre d'occurrences, 2241
mot non vide, 2241
moyenne
  de Cesaro, 633
  empirique, 2080
  empirique d'un échantillon, 2156
  quadratique, 2080
multiplicateur
  de Lagrange, 1168
multiplicité
  racine d'un polynôme, 230
  racine de $f(x) = 0$, 1987
  valeur propre
    algébrique, 721
    géométrique, 721
méromorphe, 1695
Méthode
  de Newton, 1988
méthode
  des chemins, 753
  Newton, 1994
    cas convexe, 1990
métrisable
  espace vectoriel topologique, 448

nabla, 788
neutre
  dans un groupe, 124
Newton
  méthode, 1994
nilpotent, 128
niveau de confiance, 2170
nombre

complexe
norme 1, 1370
de Fermat, 1400
dénormalisé, 1972
normal, 2146
normalisée, 1972
premier, 184, 297, 315, 1073, 1360, 1368, 1373
dans leur ensemble, 184
deux nombres entre eux, 184
théorème des deux carrés, 317
tours d'une courbe plane, 1556
nombre premier
polynôme cyclotomique, 1364
non dénombrable, 110
normal
arc paramétré, 1536
endomorphisme, 548
nombre, 2146
sous-groupe, 157
normal extérieur
vecteur, 1440
normale
loi réduite, 2125
principale, 1546
normalisateur, 157
norme, 396
d'algèbre, 609
d'application linéaire, 607
d'une application linéaire, 607
euclidienne, 434
dans $\mathbb{R}^m$, 604
supremum, 434
sur $\mathbb{Z}[i\sqrt{5}]$, 218
vecteur, 277
équivalence, 436
normé
espace vectoriel, 397
noyau
application vers un groupe, 161
d'une forme bilinéaire, 600
Dirichlet, 1807
Fejér, 1807
vers un espace vectoriel, 242
nulle part dense, 466
négatif, 145
négligeable
partie d'un espace mesuré, 921

observation, 2071
opposés
chemins, 1527
opérateur
autoadjoint, 513
définit positif, 551
hermitien, 513
opérateurs
compatibles, 1669
orbite
d'un point sous une action, 169
ordre, 106
bon ordre, 106
d'un groupe, 159
d'un polynôme, 229
d'une matrice carrée, 250
dans un corps, 133
de convergence d'un schéma, 2062
distribution, 1852
partiel, 106
sur un anneau factoriel, 211
total, 106
élément, 159
orientable
variété, 1415
orientation, 509, 1414, 1434
orientation affine, 510
origine
abscisse curviligne, 1536
repère affine, 488
orthogonal, 604, 1652
coordonnées curviligne, 1636
famille de projecteurs, 203
matrice, 513
opérateur, 513
sous-espace, 280
vecteur, 277
orthonormé, 605
système, 1655
oscillation
d'une fonction, 709
d'une fonction en un point, 709
osculateur (cercle), 1551
ouvert, 371, 441, 517

parabolique
équation aux dérivées partielles, 1960
parallèle
sous-espaces affines, 479
paramétrage, 1527
normale, 1536
paramétrages
admissible, 1535
Parseval, 1662
partie
génératrice, 233
régulière, 859
totale, 1655
partie convexe, 480

partie génératrice, 159
partie libre
  module, 200
partie linéaire, 476
partition
  d'un entier en parts fixées, 1700
  de l'unité, 1417
  dénombrable mesurable, 941
pavable, 1486
pavage du plan, 1300
pavé, 1485, 1486
Pearson
  théorème, 2186
peigne de Dirac, 1862
permutation, 174
  matrice, 268
permutation paire, 176
permuter
  différentielle et intégrale
    $\mathbb{R}^n$, 1138
  dérivée et intégrale
    $\mathbb{R}^n$, 1137
    dans $\mathbb{R}$, 1134
    dans $\mathbb{R}$ avec les bornes, 1136
  dérivée et limite, 826
  intégrale
    et série, 1028
  limite et intégrale
    convergence dominée, 980
    convergence monotone, 970
    espace mesuré, 1132, 1133
  somme et intégrale, 971, 1031
  série entière et dérivation, 1043
  série entière et intégration, 1045
permuter limite et intégrale, 1131
petit théorème de Fermat, 313
pgcd
  calcul effectif, 187
  dans un anneau intègre, 127
  polynômes, 326
pivotale, 2172
plan
  projectif, 1579
  tangent, 789, 867
plan vectoriel, 248
Plancherel, 1662
plongement, 1210
Poincaré (demi-plan), 1621
point
  d'équilibre
    stable, 1918
  pondéré, 481
point adhérent, 376
point critique
  définition, 1502
point d'accumulation, 377
point fixe, 2217
  attractif, 1140
  Brouwer, 1419
  Picard, 1141
  Schauder, 1421
point intérieur, 375
point isolé, 377
Poisson
  formule sommatoire, 1826
  processus, 2235
polynôme
  alterné, 365
  annulateur, 526, 529
  caractéristique, 534, 547
  contenu, 229
  cyclotomique, 1362
    irréductibilité, 1364
    propriétés, 1363
  d'endomorphisme, 891
    décomposition de Dunford, 568
  de Bernstein, 2148
  irréductible
    sur $\mathbb{F}_q$, 1393
    séparable, 359
  Lagrange, 284
  minimal, 332, 528
    d'un élément d'une extension, 328
    ponctuel, 529
  racines, 366
  scindé, 321
  semi-symétrique, 365
  symétrique, 365, 366, 502, 507, 1381
    élémentaire, 365, 367
  séparable, 359
  trigonométrique, 1744
  à plusieurs indéterminées, 507, 1381
polynôme de plusieurs variables, 206
polynôme de Taylor, 852
polynôme scindé, 321
polynôme trigonométrique, 1755
polynômes, 204
portée
  mesure, 988
positif, 145
potentiel, 1431, 1462
ppcm
  dans un anneau intègre, 127
premier
  corps, 313
  deux polynômes entre eux, 229

deux éléments d'un anneau principal, 215
idéal, 212
sous corps, 313
premier temps d'atteinte, 2202
premier type
région solide, 1497
presque
nulle, 179
partout, 911
surjective, 1794
primitif
polynôme, 229
au sens du pgcd, 230
racine, 1386
triplet pythagoricien, 224
élément d'un corps, 1374
élément d'une extension de corps, 341
primitive, 724, 1460
de fonction continue, 830
et intégrale, 1012
fonction, 740
primitive et intégrale, 1006
principal
anneau, 211
idéal, 211
principe
prolongement analytique, 1215
zéros isolés, 1214
Principe de correspondance, 2264
probabilité
conditionnelle, 2081
problème
aux limites d'évolution, 1961
aux limites stationnaires, 1961
bien posé, 1961
limite de Dirichlet, 1961
limite de Von Neumann, 1961
problème de Cauchy, 1938
processus
adapté à une filtration, 2219
arrêté, 2224
croissant prévisible, 2224
Galton-Watson, 2215
Poisson, 2235
sans mémoire, 2121
produit
d'espaces vectoriels normés, 439
d'une mesure par une fonction, 983
de Cauchy, 1038
de convolution, 1740
et Fourier, 1823
de langages, 2242
de mots, 2241
distribution et fonction, 1851
espaces mesurés, 998
espaces topologiques, 374
mixte, 429, 431
scalaire
en général, 421
sur $\mathbb{M}(n, \mathbb{R})$, 616
sur $\mathbb{R}^n$, 603
semi-direct, 177
tensoriel
de représentations, 1116
vectoriel, 428
produit hermitien, 421
produit pseudo-scalaire, 421
produit remarquable, 1688
produit scalaire sur $\mathbb{R}^n$, 277
produit tensoriel, 648, 649
projecteur
dans un module, 203
projectif
complétion, 1582
droite, 1579, 1580
espace, 1579
groupe, 1586
hyperplan, 1580
plan, 1579
repère, 1587
sous-espace, 1579
projection
orthogonale, 1650
projection normale, 1769
prolongement
analytique, 1215
utilisation, 1802
de fonctions, 1210
lemme de Borel, 1098
méromorphe de la fonction Γ, 1696
par continuité, 700
dans $H^1(I)$, 1876
par densité, 1210
théorème de Hahn, 930
propriété d'intersection non vide, 382
propriété universelle, 648
prébase, 373
précision
simple, 1971, 1972
préhilbertien, 1647
préserve l'orientation, 510
puissance
d'un langage, 2243
d'un mot, 2242
d'un point, 1573
d'un test, 2182

d'une inversion, 1606
pôle, 1320

quasi-compact, 381
quaternion, 363
quotient, 183, 229
dans une suite de composition, 165
de groupe, 165
de groupes, 296
quotient d'un espace vectoriel, 646

racine
carré
de matrice hermitienne, 891
carré de matrice
hermitienne positive, 891
d'un polynôme, 230
de l'unité, 591, 1258, 1355, 1364, 1368, 1621
primitive, 1357
utilisation, 1363
de polynôme, 343
multiple, 1987
primitive, 1386
simple, 1987
racine carrée, 697
raffinement, 1520
subdivision d'un pavé, 1487
rang, 245, 495, 564
classe d'équivalence, 270
diagonalisation, 551
différentielle, 1168
utilisation, 1654
rang d'une forme quadratique, 562
rang d'une matrice, 269
rare, 905
rationnels, 136
rayon
de convergence, 1036
de courbure, 1546
de torsion, 1547
spectral, 569, 609
recouvrement, 381
rectangle
produit de tribus, 949
rectifiable, 1520
arc géométrique, 1536
rejet
région dans une prise de décision, 2181
relation d'équivalence, 108
relations
coefficient-racines, 367
de Chasles, 473, 1475
relativement
compact, 382, 1717
relèvement, 1556
Représentation
virgule flottante normalisée, 1971
représentation, 286
de groupe fini
caractères de S_4, 1117
fidèle, 286
groupe diédral, 1350
irréductible, 1108
produit tensoriel, 1116
régulière gauche, 1113
virgule fixe, 1971
repère
affine, 488
cartésien
espace affine, 474
de Frenet, 1546
projectif, 1587
reste, 183, 229
d'un développement limité, 859
risque
première espèce, 2182
quadratique, 2161
seconde espèce, 2182
rotation
en dimension 2, 1269
rotation d'angle θ, 1273
rotation-homothétie, 1605
rupture
corps, 344
règle du produit nul, 128
réciproque
continuité, 389
dérivabilité, 733
récurrent
nul, 2203
point d'un système dynamique, 938
positif, 2203
état, 2203
réduction
d'endomorphisme, 568
Frobénius, 578
Jordan, 580
réel, 142
réflexif, 676
réflexion, 1289
dans $\mathbb{R}^2$, 1266
glissée, 1287
par rapport à un hyperplan, 1251
région
critique, 2181
de confiance exact, 2172
de rejet, 2181

régularité
d'une mesure, 937
extérieure de la mesure de Lebesgue, 959
intérieure de la mesure de Lebesgue, 961
régulier
arc, 1533
chemin, 1441
point d'un arc, 1533
polyèdre, 1290
régulier à droite, 128
régulière
surface, 1432
répulsif
point fixe, 1140
résidu
méthode itérative, 2038
résolvante, 1899
résultant, 505
utilisation, 507, 1455

Schrödinger, 1933
Schur (théorème), 1110
schéma
consistant, 2062
schéma numérique, 2062
section, 757
de graphe, 742
propriété des, 993
segment
dans $\mathbb{R}^p$, 408
dans un espace affine, 481
semblables
matrices, 538
semi-définie positive, 551
semi-norme, 467
semi-simple
endomorphisme, 531
semi-symétrique
polynôme, 365
sesquilinéaire, 421
signature
d'une permutation, 174
forme quadratique, 562
similitude, 797
simple
extension de corps, 341
fonction, 945
groupe, 157
module, 203
singularité, 1695
effaçable, 1695
pôle, 1695
sinus, 1217
hyperbolique, 1076
sinus cardinal, 1193, 1509
solfège, 1356
solution
générale, 1937
particulière, 1937
somme
inférieure, 1492
partielle, 624
supérieure, 1492
somme directe, 647
somme directe (de représentations), 1107
somme partielles
Abel angulaire, 1450
sommet, 1570
sous anneau, 196
sous arc, 1519
sous-additivité
sur algèbre de parties, 908
sous-espace
affine engendré par une partie, 479
caractéristique, 566
sous-groupe
caractéristique, 157
distingué, 297
dans le groupe alterné, 301
engendré, 157
normal, 165, 296
sous-martingale, 2219
sous-module, 202
sous-module engendré, 201
sous-suite, 713
spectre
matrice hermitienne, 548
matrice symétrique réelle, 551
spectre d'un endomorphisme, 526
sphère, 514
de Riemann, 1602
stabilisateur, 169
stabilité
d'un point d'équilibre, 1918
Lyapunov, 1918
stable, 1979
schéma numérique, 2065
stathme
sur $\mathbb{Z}[i]$, 314
stathme euclidien, 220
stationnaire
chaine de Markov, 2211
statistique, 2161
statistiques
descriptives, 2153
strictement
convexe

sur $\mathbb{R}^n$, 1179
strictement convexe, 1170
structure
complexe, 1553
structure d'anneau canonique, 126
Student, 2132, 2174
subdivision, 1487
associée à une fonction, 1488
d'un intervalle, 1520
subordonnée
norme, 607
subpotent, 109
suite, 179
arithmético-géométrique, 631
de Cauchy, 448, 449
dans un corps, 134
de fonctions, 1722
théorème de Montel, 1801
de fonctions intégrables, 1131, 1715
de Jordan-Hölder, 165
définie par itération, 1994
exacte, 177
régularisante, 1828
équirépartie, 1811
critère de Weyl, 1811
suite de composition, 165
suite de variables aléatoires de Bernoulli, 2215
suites adjacentes, 405
support, 1489
d'une permutation, 174
distribution, 1852
famille d'éléments, 179
supremum, 150
d'une suite d'ensembles, 903
sur-martingale, 2219
surface paramétrée, 1431
surjection, 106, 388
surpotent, 109
Sylow
p-Sylow, 287
Sylvester (matrice), 504
symbole
de Legendre, 1374
symbole principal, 1951
symétrique
polynôme, 365
système
fondamental, 1897
orthonormé, 1655
trigonométrique, 1655, 1744
séparable, 1655
espace topologique, 378
extension de corps, 361
polynôme irréductible, 359
polynôme non constant, 359
élément d'une extension, 361
sépare
les points, 827
séquentiellement fermé, 446
série
dans un espace vectoriel normé, 624
de Fourier, 1814, 1826
utilisation, 1819
de Laurent, 1696
de puissance, 1035
donnant $(1-A)^{-1}$, 671
entière, 1035, 1100, 1826
Abel angulaire, 1450
fonctions holomorphes, 1681
processus de Markov, 2217
utilisation, 1700, 2103
fonctions, 1125, 1826
génératrice d'une suite, 1083
utilisation, 1700
géométrique, 630
harmonique, 630
nombres, 1125
numérique, 1073, 1100
Riemann, 630
Taylor, 1084
série convergente, 624
série de Taylor, 852
série divergente, 624

tangent
vecteur unitaire, 1546
tangente, 1229, 1541
Tangente hyperbolique, 2258
tangente hyperbolique, 1077
tangente à un chemin, 1408
taubérien, 1124
taux d'accroissement, 1170
Taylor
série entière, 1084
temps d'arrêt, 2222
temps de retour, 2202
terminée
martingale, 2223
test, 2181
bilatéral, 2182
unilatéral, 2182
théorème
accroissements finis, 667
dans $\mathbb{R}$, 737
forme générale, 800
Ascoli, 1717
Baire, 466

Banach-Steinhaus, 638
  avec semi-normes, 1718
base incomplète, 237
Beppo-Levi, 970
Bolzano-Weierstrass, 394
Borel-Cantelli, 2071
Borel-Lebesgue, 401
Brouwer, 1420
  dimension 2, 1683
Bézout
  polynômes, 324
  utilisation, 505
Carathéodory, 485
Cauchy
  groupe, 191
Cauchy-Arzela, 1422
Cauchy-Lipschitz, 1144
Cayley-Hamilton, 537, 890
central limite, 2113
  processus de Poisson, 2237
Chevalley-Warning, 1381
chinois, 319
  anneau des polynômes, 320
  anneau principal, 215
Cochran, 2158
Cochrane, 2159
convergence
  dominée de Lebesgue, 980
  monotone, 970
de Baire, 471
de Jordan, 1571
de représentation de Riesz, 1654
des deux carrés, 317
  version faible, 315
Dini, 820
Dirichlet, 1808
  forme faible, 1368
Doob, 2223
du rang, 245
décomposition des noyaux
  et exponentielle de matrice, 1096
extension d'isométrie, 1211
Fejér, 1809
fonction implicite dans $\mathbb{R}^n$, 1157
fonction implicite dans Banach, 1156
Fubini
  dans $\mathbb{R}^n$, 1029
  espace mesuré, 1024
  version compacte dans $\mathbb{R}^2$, 1464
Fubini-Tonelli, 1022
fuite des compacts, 1902
Gauss
  polynômes, 324
Gauss-Wantzel, 1400
Glivenko-Cantelli, 2167
Hahn-Banach, 1792
Hardy-Littlewood, 1125
Heine, 714
incidence, 1580
inversion locale, 1154
  utilisation, 1168, 1500
isomorphisme
  second, 162
  troisième, 163
isomorphisme de Banach, 1717
Kronecker, 507
Lagrange, 164
Lie-Kolchin, 593
Lotka-Volterra, 1923
Markov-Takutani, 1423
Montel, 1801
Pappus
  affine, 1584
  projectif, 1584
Pearson, 2186
petit de Fermat, 313
Picard, 1141
point fixe
  Brouwer, 1683
projection
  cas vectoriel, 1650
  partie fermée convexe, 1649
prolongement de Hahn, 930
prolongement de Riemann, 1695
Radon-Nikodym, 988
  complexe, 989
Rolle, 736
Rothstein-Trager, 1455
Runge, 1687
Schauder, 1421
Schur, 1110
spectral, 567
  autoadjoint, 602
  matrice symétrique, 551
  matrices normales, 548
stabilité de Lyapunov, 1918
Stone-Weierstrass, 827, 830
Sylvester, 564
taubérien, 1124
taubérien faible, 1452
transfert, 2104
Tykhonov, 464
  dénombrable, 465
  fini, 464
valeurs intermédiaires, 696
Von Neumann, 1162

Wedderburn, 1370
Weierstrass, 394
élément primitif, 362, 1383, 1384
théorème des extrémums liés, 1168
théorème fondamental du calcul intégral, 1007
topologie, 371, 518
$*$-faible, 469, 1850
p-adique, 679
discrète, 372
engendrée par une famille, 373
et semi-normes, 467
faible, 609
forte, 608
grossière, 372
induite, 374
métrique, 392
produit, 374
sur $\mathscr{D}(\Omega)$, 1846
sur $\mathscr{D}(K)$, 1846
sur $C^\infty(\Omega)$, 1846
sur dual topologique, 469
usuelle sur $\mathbb{R}^n$, 518
topologique
somme directe, 1652
torsion, 1547
d'un groupe, 179
totale, 1655
trace, 1882
dual de $\mathbb{M}(n, \mathbb{K})$, 284
endomorphisme, 539
matrice, 538
produit scalaire sur $\mathbb{M}(n, \mathbb{R})$, 616
unicité pour la propriété de trace, 285
transcendant, 330
transformation
Fourier, 1826
gaussienne, 2018
transformée
de Cauchy, 1693
de Fourier, 1821, 2102
continuité, 1824
groupe abélien fini, 1105
Fourier
distribution tempérée, 1858
Laplace, 2103
transient
état, 2203
transition
probabilité, 2191
transitive, 173
transitoire
état, 2203
transposition, 174
transposé, 512
transposée, 282
transvection, 882
transvection (matrice), 268
transversale, 172
tribu, 903
borélienne, 932
de Baire, 905
de Lebesgue, 956
engendrée, 904
par un événement, 2073
par une application, 918
par une variable aléatoire, 2073
induite, 904
produit, 949
tribu de Lebesgue sur S^1, 1244
trigonalisation
et polynôme caractéristique, 545
simultanée, 592, 593
triplet
pythagoricien, 224
type
fini
en algèbre, 363
espace vectoriel, 236
tétraèdre, 1290

unicité
des mesures, 914
uniformément continue, 466
uniformément convexe, 1769
unipotent, 128
unitaire
normale principale, 1546

valeur
principale (distribution), 1856
propre, 526
singulière, 570
valeur absolue
p-adique, 679
valeur principale, 1706
valeur propre
d'une forme quadratique, 599
valuation
p-adique, 679
d'un polynôme, 228
Vandermonde (déterminant), 501
variable
de décision, 2182
variable aléatoire, 2072
absolument continue, 2072
Bernoulli
marche aléatoire, 2196

utilisation, 2148
binomiale
utilisation, 2226
centrée, 2079
de Bernoulli
utilisation, 2226
de Rademacher, 2130
variance, 2079
empirique, 2080, 2157
empirique corrigée, 2157
vecteur gaussien, 2126
variation des constantes, 1893, 1897
variété, 1168, 1405
variété
orientée, 1415
vecteur
cyclique, 530
gaussien, 2126
propre, 526
unitaire normal, 1546
unitaire tangent, 1541
Vitali (ensemble), 962
vitesse d'un chemin, 1519
voisinage, 371, 518
volume
d'une région solide, 1498
région bornée dans $\mathbb{R}^3$, 1496
volume dans $\mathbb{R}^3$, 1418
vraisemblance, 2164

Weiner
constante, 2285
Wronskien, 1925

Éther, 2248
Événement, 2253
élément
inversible
dans un anneau, 128
élément premier, 216
élémentaire
polynôme symétrique, 365
équation
aux variations, 1911
de Riccati, 1932
des classes, 172
des orbites, 172
différentielle
Hill, 1928
homogène, 1931
système, 1929
étude qualitative, 1929
diophantienne, 194, 223, 226
Fredholm, 1143
orbite-stabilisateur, 171
équation différentielle
linéaire du premier ordre, 1942
linéaire du premier ordre, homogène, 1942
linéaire du second ordre, 1944
linéaire du second ordre, homogène, 1945
premier ordre, 1937
second ordre, 1937
variables séparables, 1939
équation exponentielle, 846
équation fonctionnelle, 846
équation homogène associée, 1942
équilibre
point point une équation différentielle, 1918
équipotent, 109
équivalence
de norme, 436
équivalence de forme quadratiques, 565
étagée
fonction, 945
état
apériodique, 2213
récurrent, 2203
récurrent positif, 2203
transitoire, 2203
étoile de Kleene, 2243
étranger
dans leur ensemble, 229
évaluation
polynôme plusieurs variables, 206

Liste des notations

$N \lhd G$ Le sous-groupe N est normal dans G, page 157

Algèbre

$[\mathbb{L} : \mathbb{K}]$ degré d'une extension de corps, page 328

$\mathcal{L}(E, F)$ Ensemble des applications linéaires de E dans F, page 241

$\mathbb{K}(A)$ corps contenant $\mathbb{K}$ et A, page 341

$\mathbb{K}[A]$ anneau contenant $\mathbb{K}$ et A, page 341

$\mathbb{N}_0$ les naturels non nuls : $\mathbb{N}_0 = \mathbb{N}\backslash\{0\}$, page 110

$\mathcal{M}_{n\times m}$ l'ensemble des matrices $n \times m$, page 242

∇f gradient de la fonction f, page 780

$\mathtt{proj}_V$ projection de $V \times W$ sur V, page 619

$\mathrm{Span}(A)$ l'ensemble des combinaisons linéaires finies d'éléments de A, page 233

$df_a(u)$ Application de la différentielle de f sur le vecteur u, page 765

$f^{(n)}$ La n-ième dérivée de la fonction f, page 854

$o(x)$ fonction tendant rapidement vers zéro, page 856

(p) idéal engendré par p, page 196

$\mathbb{F}_p$ lorsque p est premier, page 313

$\mathbb{F}_{p^n}$ corps fini à p^n éléments, page 1372

$\mathrm{Frac}(\mathbb{A})$ Le corps des fractions de l'anneau $\mathbb{A}$, page 133

$\mathrm{Fun}(X, Y)$ les applications de X vers Y, page 126

$\mathrm{S}(E)$ Les opérateurs autoadjoints de E, page 513

U_n Le groupe des racines n^{e}de l'unité., page 1355

$\mathrm{res}(P, Q)$ résultat des polynômes P et Q, page 505

$\sqrt{A}$ racine d'une matrice hermitienne positive, page 891

$\theta_\alpha(P)$ la multiplicité de α par rapport à P, page 230

$A[X]$ tous les polynômes de degré fini à coefficients dans A, page 205

$A_n[X]$ les polynômes à coefficients dans A et de degré inférieur à n, page 227

$C(P)$ matrice compagnon, page 577

$D \mid P$ D divise P, page 229

$E_\lambda(u)$ Espace propre de u, page 526

$mat_{\mathcal{B}}(q)$ matrice de q dans la base $\mathcal{B}$, page 1202

$U(A)$ ensemble des inversibles, page 128

Ensembles de matrices

$S^+(n,\mathbb{R})$ matrices symétriques définies positives, page 552

$S^{++}(n,\mathbb{R})$ matrices symétriques strictement définies positives, page 552

$\mathrm{Aut}(E)$ automorphisme de l'espace vectoriel E, page 242

$\mathrm{End}(E)$ les endomorphismes de E, page 242

$L(E,F)$ applications linéaires bornées (continues), page 614

$\mathrm{O}(n,\mathbb{R})$ le groupe des matrices orthogonales, page 513

$\Omega(E)$ formes quadratiques non dégénérées, page 1198

$Q(E)$ formes quadratiques réelles sur E, page 554

$Q^+(E)$ formes quadratiques positives, page 1198

$Q^{++}(E)$ formes quadratiques strictement définies positives, page 1198

$S_n^{p,q}(\mathbb{R})$ matrices symétriques réelles de signature (p,q), page 1199

$\beta(s)$ Vecteur unitaire de la binormale, page 1546

$\gamma \sim g$ Équivalence d'arcs paramétrés, page 1535

$\nu(s)$ Vecteur unitaire de la normale principale, page 1546

$c(s)$ rayon de courbure, page 1546

$t(s)$ Torsion, page 1547

Géométrie

$P(E)$ l'espace projectif de E, page 1579

$(x_0 : \ldots : x_n)$ coordonnées homogènes dans un espace projectif, page 1598

$\mathcal{C}\mathrm{onv}(A)$ enveloppe convexe, page 483

$\mathbb{C}[G]$ combinaisons d'éléments de G à coefficients dans $\mathbb{C}$, page 1107

$\mathrm{PGL}(E)$ groupe projectif, page 1586

B^o orthogonal dans le dual, page 280

$P_1(\mathbb{C})$ sphère de Riemann, page 1602

Chaînes de Markov

$\pi(x)$ lié au temps de retour, page 2205

Probabilités et statistique

$\sigma(X)$ La tribu engendrée par la variable aléatoire X, page 2073

$a \wedge b$ $\min(a,b)$, page 2223

K_X matrice de covariance d'un vecteur gaussien, page 2126

$m(\mathcal{A})$ Ensemble des fonctions $\mathcal{A}$-mesurables, page 917

Théorie des groupes

$(G/H)_g$ classes à gauche, page 170

$[a]_p$ ensemble des $a + kp$, page 195

$[G,G]$ groupe dérivé, page 161

$[g, h]$ commutateur dans un groupe, page 161

$\mathrm{Aff}(\mathbb{R}^n)$ Le groupe des applications affines bijectives de $\mathbb{R}^n$., page 493

gr groupe engendré, page 157

$\hat{G}$ groupe des caractères de G, page 1103

σ_x réflexion par rapport à x, page 1289

A_n groupe alterné, page 299

$D(G)$ groupe dérivé, page 161

D_n groupe diédral, page 1257

G^{ab} groupe abélianisé de G, page 161

$N \times_\phi H$ produit semi-direct, page 177

S_n le groupe symétrique, page 174

Topologie et théorie des ensembles

$\mathrm{Adh}(A)$ adhérence de A, page 376

$\complement A$ Le complémentaire de l'ensemble A, page 107

$\mathrm{Diam}(A)$ Diamètre de la partie A, page 713

$\mathrm{Int}(A)$ intérieur de A, page 375

∂A La frontière de l'ensemble A, page 521

$A \Delta B$ différence symétrique, page 107

A^c complémentaire de A, page 107

Analyse

$\mathrm{Isom}(X)$ Le groupe des isométries de X, page 444

$\mu \ll \nu$ La mesure μ est absolument continue par rapport à la mesure ν., page 988

$C^\infty\big(\mathbb{R}, \mathscr{S}'(\mathbb{R}^d)\big)$ Fonctions à valeurs dans les distributions., page 1933

$C^\infty(I, \mathscr{D}'(\mathbb{R}^d))$ fonctions à valeurs dans les distributions, page 1863

$(S, \hat{\mathcal{F}}, \hat{\mu})$ complété de l'espace mesuré $(S, \hat{\mathcal{F}}, \hat{\mu})$, page 923

$\mathcal{L}(E, F)$ Les applications linéaires de E vers F, page 657

$\mathcal{L}^{(n)}(V, W)$ L'espace des applications n-linéaires $V^n \to W$, page 811

$\arg(z)$ La valeur principale de l'argument de $z \in \mathbb{C}$, page 1706

L Les applications linéaires continues de E vers F, page 657

$\mathbb{D}_b$ l'ensemble de écritures décimales en base b, page 633

$\mathbb{R}$ l'ensemble des réels, page 142

$\mathbb{R}^+$ les réels positifs ou nuls, page 144

exp exponentielle, page 1062

$\mathscr{H}'$ dual, page 1653

$\liminf a_n$ limite inférieure, page 406

$\limsup a_n$ limite supérieure, page 406

$\mathcal{L}^p$ espace de Lebesgue, sans les classes, page 1719

$\mu \perp \nu$ mesures perpendiculaires, page 988

μ^* La mesure extérieure associée à la mesure μ, page 915

∂_z,$\partial_{\bar{z}}$ dérivées partielles d'une fonction complexe, page 1675

$\mathtt{proj}_K(x)$ projection orthogonale de x sur y, page 1650

$\sigma(\mathcal{A})$ tribu engendrée par $\mathcal{D}$, page 904

$\mathscr{D}(\Omega)$ Les fonctions C^∞ à support compact sur Ω, page 1849

$\mathscr{S}'(\mathbb{R}^d)$ espace des distributions tempérées, page 1855

$A^2(\Omega)$ espace de Bergman, page 1802

$A^\perp$ orthogonal d'une partie., page 1652

$C_c(I)$ fonctions continues à support compact dans I, page 1739

$f \sim g$ fonctions ayant des limites équivalentes, page 753

$H^1(\Omega)$ espace de Sobolev sur Ω, page 1878

$H^1(I)$ espace de Sobolev, page 1873

$H^m(M)$ espace de Sobolev, page 1880

$L^1_{loc}(I)$ fonctions intégrables sur les compacts de I, page 1873

L^p espace de Lebesgue avec les classes, page 1721

$M_i\varphi$ La fonction $x \mapsto x_i\varphi(x)$, page 1797

$S_n f$ somme partielle de série de Fourier, page 1767

Chapitre 0

Introduction

0.1 Auteurs, contributeurs, sources et remerciements

Les remerciements, dans chaque catégorie, sont mis dans l'ordre chronologique approximatif. Les noms en couvertures sont ceux qui ont fourni du code LaTeX (typiquement : un patch via github), par ordre chronologique approximatif d'entrée dans le projet.

0.1.1 Ceux qui ont travaillé sur le Frido

(1) Carlotta Donadello pour l'ensemble du cours de CTU de géométrie analytique 2010-2011. Une grosse partie de « analyse réelle » vient de là.

(2) Les étudiants de géométrie analytique en CTU 2010-2011 ont détecté d'innombrables coquilles. Les étudiants du cours présentiel de géométrie analytique 2011-2012 ont signalé un certain nombre d'incorrections dans les exercices et les corrigés. Les agrégatifs de Besançon 2011-2012 pour leurs plans et leurs développements.

(3) Lilian Besson pour m'avoir signalé un paquet de fautes, et quelques points pas clairs en statistiques.

(4) Plouf qui m'a signalé une coquille dans le fil la-selection-scientifique-de-la-semaine-numero-106.

(5) Benjamin de Block pour des coquilles et une mise au point sur les conventions à propos de $\mathbb{R}^+$ et $(\mathbb{R}^+)^*$.

(6) Olivier Garet pour avoir répondu à plein de questions de probabilités.

(7) François Gannaz pour de la relecture et une version plus claire de la preuve (et de l'énoncé) de la proposition 18.6.

(8) Danarmk pour des réponses à des questions dans les commentaires (allongement pour éviter un Overfull hbox) `http://linuxfr.org/nodes/110155/comments/1675589`. Et aussi pour une discussion à propos de la topologie sur $\mathscr{D}(\Omega)$.

(9) Cédric Boutilier pour des réponses à des questions de probabilité statistique. `https://github.com/LaurentClaessens/mazhe/issues/16`

(10) Remsirems pour des réponses à des questions d'analyse [1]

(11) Bertrand Desmons pour plusieurs patchs rendant plus clairs de nombreux passages sur les suites de Cauchy dans $\mathbb{Q}$.

(12) Anthony Ollivier pour m'avoir fait remarquer qu'il n'est pas vrai que $A[X]$ est euclidien lorsque A est intègre (contre-exemple : $A = \mathbb{Z}$). Ça fait une faute de moins dans le Frido.

(13) ybailly pour avoir détecté un bon nombre de coquilles dans la partie sur les ensembles de nombres.

(14) Éric Guirbal pour le remplacement de `frenchb` par `french`.

1. `http://linuxfr.org/nodes/110155/comments/1675813`

(15) cdrcprds pour une réponse à une question d'algèbre, démonstration à l'appui à propos de pgcd. Également pour sa relecture sans pitié de la partie sur la cardinalité (en particulier $A \approx A \backslash B$) et pour avoir pointé l'utilisé du théorème de comparabilité cardinale.

(16) Antoine Bensalah pour avoir répondu à une question sur Lax-Milgram tout en même temps que pointé une erreur dans la démonstration et fourni l'exemple 24.60 sur l'optimalité de l'inégalité.

(17) Guillaume Deschamps pour ses remarques à propos du fait que le chapitre « constructions des ensembles » est très ardu.

(18) Guillaume Barriere pour sa relecture attentive jusqu'aux corps.

(19) Samy Clementz pour avoir découvert une faute dans la définition de mesure positive sur un espace mesurable.

(20) Sylvain Rousseau pour avoir clarifié une construction dans le théorème de Bower version C^∞.

(21) Maxmax pour des typos dans l'index thématique.

(22) Laurent Choulette pour une typo dans les propriétés du neutre d'un groupe.

(23) Pierre Lairez pour la démonstration du théorème d'inversion de limite et de dérivée 11.311 sans passer par les intégrales (et les lemmes correspondants à propos du module de continuité).

(24) Gregory Berhuy pour des réponses d'algèbres dans les catégories facile, moyen et difficile.

(25) Benoît Tran pour avoir signalé un paquet de typos dans la démonstration de l'ellipsoïde de John-Loewner et ses dépendances.

(26) Provatiscus pour avoir signalé un paquet de choses pas claires, et surtout pour avoir trouvé une faute dans la démonstration du fait qu'une fonction continue sur $\mathbb{Q}$ se prolonge en une fonction continue sur $\mathbb{R}$. Et pour cause : cet énoncé est faux. `https://github.com/LaurentClaessens/mazhe/issues/124`

(27) William pour l'environnement `example` qui gère correctement le triangle.

(28) Colin Pitrat pour de nombreuses remarques, typos et relecture de théorèmes.

(29) Bruno Turgeon pour une très belle moisson de fautes de frappe (euphémisme pour dire « mon ignorance crasse de l'orthographe »).

(30) Sacha Dhénin pour une belle quantité de fautes de frappes et pour avoir soulevé quelque points pas clairs (par exemple la définition du sous-groupe engendré dans le cas de la partie vide).

0.1.2 Aide directe, mais pas volontairement sur le Frido

(1) Plein de monde pour diverses contributions à des énoncés d'exercices. Pierre Bieliavsky pour les énoncés d'analyse numérique (MAT1151 à Louvain la Neuve 2009-2010). Jonathan Di Cosmo pour certaines corrections de MAT1151. François Lemeux, exercices sur les normes de matrices et correction de coquilles. Martin Meyer et Mustapha Mokhtar-Kharroubi pour certains exercices du cours *Outils mathématiques* (surtout ceux des DS et examens).

(2) Nicolas Richard et Ivik Swan pour les parties des exercices et rappels de calcul différentiel et intégral (Université libre de Bruxelles, 2003-2004) qui leurs reviennent.

(3) Carlotta Donadello pour la partie géométrie analytique : topologie dans $\mathbb{R}^n$, courbes, intégrales, limites. (Université de Franche-Comté 2010-2012)

(4) Le forum usenet de math, en particulier pour la construction des corps fini dans la fil « Vérifier qu'un polynôme est primitif » initié le 20 décembre 2011.

(5) Mihai Bostan nous a donné ses notes manuscrites de son cours présentiel de géométrie analytique 2009-2010. (Presque) Toute la structure du cours de géométrie analytique lui est due (qui est maintenant fondue un peu partout dans les chapitres d'analyse).

0.1.3 Des gens qui ont fait un travail qui m'a bien servi

(1) Arnaud Girand pour avoir mis ses développements bien faits en ligne. Une bonne vingtaine de résultats un peu partout dans ces notes viennent de lui.

(2) Le site `http://www.les-mathematiques.net` m'a donné les preuves de nombreux résultats.

(3) Pierre Monmarché pour son document en ligne tout plein de développements, et des réponses à des questions.

(4) Tous les contributeurs du Wikipédia francophone (et aussi un peu l'anglophone) doivent être remerciés. J'en ai pompé des quantités astronomiques ; des articles utilisés sont cités à divers endroits du texte, mais ce n'est absolument pas exhaustif. Il n'y a à peu près pas un résultat important de ces notes dont je n'aie pas lu la page Wikipédia, et souvent plusieurs pages connexes.

(5) Les intervenants du fil « Antisymétrisation, alterné, déterminant et caractéristique » sur `les-mathematiques.net` m'ont bien aidé pour la section sur les déterminants 9.1 (bien qu'ils ne le savent pas).

(6) Xavier Mauquoy pour l'énoncé et la preuve du théorème 3.43.

(7) David Revoy pour les dessins de Pepper&Carrot de la couverture.

J'ai souvent donné entre parenthèses à côté des mots « théorème », « lemme » ou « proposition » une ou plusieurs références vers les sources de la preuve que je donne. Ce sont parfois des liens vers la bibliographie ; parfois aussi des liens hypertextes vers des sites, des blogs, etc. Tous ces gens ont fait du bon boulot. Sans toute cette « communauté », l'internet serait mort [2].

0.2 Originalité

Ces notes ne sont pas originales par leur contenu : ce sont toutes des choses qu'on trouve facilement sur internet ; je crois que la bibliographie est éloquente à ce sujet. Ce cours se distingue des autres sur les points suivants.

La longueur J'ai décidé de ne pas me soucier de la taille du fichier. Il fera cinq mille pages s'il le faut, mais il restera en un bloc. Étant donné qu'il n'existe qu'une seule mathématique, il ne m'a pas semblé intéressant de produire une division artificielle entre l'analyse, la géométrie ou l'algèbre. Tous les résultats d'une branche peuvent (et sont) être utilisés dans toutes les autres branches.

Dans cette optique, je me suis évertué à ne créer que des références « vers le haut ». À moins d'oubli de ma part [3], il n'y a aucun endroit du texte qui dépend d'un lemme démontré plus bas. Le fait qu'un théorème B soit plus bas qu'un théorème A signifie qu'on peut démontrer A sans savoir B.

La licence Ce document est publié sous une licence libre. Elle vous donne explicitement le droit de copier, modifier et redistribuer.

Les mises à jour Ce document est régulièrement mis à jour. Des fautes d'orthographe sont corrigées (presque) chaque jour. Si vous me signalez une faute de mathématique, elle sera corrigée.

Transparence Je ne fais pas semblant que ces notes soient parfaites. Les points sur lesquels je ne suis pas sûr, les preuves que j'ai inventées moi-même sont clairement indiqués pour inciter le lecteur à redoubler de prudence. Une liste de questions à résoudre est inclue en la section 0.7. Voir 0.3 pour plus de détails.

2. Cette dernière phrase doit être comprise comme un appel à ne pas utiliser Moodle et autres iCampus pour diffuser vos cours de math, ou en tout cas pas comme moyen exclusif.

3. Par exemple pour les théorèmes pour lesquels je n'ai pas lu ni a fortiori écrit de preuves.

0.3 Les choses qui doivent vous faire tiquer

Un cours de math doit toujours être lu attentivement, surtout si vous avez l'intention de resservir à un jury le fruit de vos lectures. Dans ce livre, trois éléments doivent vous faire redoubler de prudence.

La référence [1] D'abord les références à [1] indiquent qu'une bonne partie de ce qui suit est de l'invention personnelle de l'auteur. Cela ne veut évidemment pas dire que c'est moi qui ait découvert le résultat. Ça veut dire que je n'ai pas trouvé le résultat ou certaines parties de la preuve.

Les notes en bas de page Certaines notes en bas de page sont écrites dans une fonte spéciale [4]. Elles indiquent des points sur lesquels je doute ou des étapes de calculs que je ne parviens pas à reproduire en suivant mes sources. Lorsque vous voyez une telle note, redoublez de prudence, allez voir la source, et écrivez-moi si vous pouvez résoudre le problème.

Les environnements dédiés Et enfin certains problèmes sont indiqués plus longuement dans un environnement dédié en petits caractères comme ceci :

Problèmes et choses à faire

Les choses écrites comme ceci sont des questions ou des éléments sur lesquels j'ai un doute. Lisez-les attentivement. Ces notes mentionnent des points que personnellement je n'oserais pas affirmer plein d'aplomb à un jury d'agrégation.

0.4 Quelques choix qui peuvent provoquer des quiproquos

Comme tout cours de mathématique, ce cours fait des choix qui sont parfois discutables. Voici quelques points sur lesquels les choix faits ici ne sont peut-être pas ceux fait par tout le monde. Ce sont donc des points sur lesquels vous devez faire attention pour éviter les quiproquos lors par exemple d'un oral dans un concours.

(1) Nous utilisons la définition usuelle de limite d'une fonction en un point. Elle diffère de celle donnée par le ministère de l'enseignement en France. Si votre but est de passer un concours d'enseignement en France, vous devriez lire 7.7.17 ; dans tous les autres cas, la définition prise ici est celle qu'il vous faut.

(2) Un compact est une partie d'un espace topologique pour lequel tout recouvrement par des ouverts admet un sous-recouvrement fini. Le fait d'être séparable n'est pas inclus dans la définition de compact. De nombreux textes français incluent la séparabilité dans la compacité.

(3) Le logarithme sur $\mathbb{C}$ est une application $\ln\colon \mathbb{C}^* \to \mathbb{C}$ définie partout sauf en zéro. Elle n'est donc pas continue sur la fameuse demi-droite. À ne pas confondre avec une *détermination* du logarithme qui est par définition continue et donc non définie sur la demi-droite.

Cela est un choix très discutable. La raison de donner à la notation « ln » cette signification est simplement de suivre l'usage de Sage. Pour Sage, $\ln(-1)$ existe et vaut $i\pi$.

Voir les remarques 25.62.

(4) Le mot « corps » n'implique pas la commutativité, et nous n'utilisons pas la terminologie « anneau à division ». Voir la remarque 1.98 et la discussion 6.1.

0.5 Autres choix pas spécialement standards

Nous listons ici quelques choix qui n'induisent pas de différences ou d'incompatibilité avec les autres cours, mais qui doivent être compris et justifiés.

(1) Nous n'utilisons pas les notations $o(x)$ ou autres $O(N^2)$. D'abord parce que je n'ai jamais très bien compris comment elles fonctionnent, et ensuite (surtout) parce que ces notations induisent en erreur. Ce sont des notations qui cachent sous des notations a peu près intuitive l'utilisation de théorèmes pas simples.

4. Les notes comme celle-ci signifient qu'il y a certaines chose dont je ne suis pas sûr.

Écrire

$$f(x) = P(x) + o(x^2), \tag{0.1}$$

c'est un peu comme quand on écrit (horreur !)

$$F(x) = \int f(x)dx + C. \tag{0.2}$$

Où est le x à droite ? Quel est le statut de C ?

Même chose pour la notation $f(x) = P(x) + o(x^2)$. Le x de $o(x^2)$ est-il le x qu'on a à gauche ? Si $g(x) = Q(x) + o(x^2)$, est-ce le même o que celui de f ?

(2) Nous allons être plus calme avec la notation $A[X]$ pour les polynômes sur l'anneau A, et encore moins $A[X_1, \ldots, X_n]$ pour les polynômes de n variables. Au lieu de cela nous utilisons $\mathcal{P}(A)$ et $\mathcal{P}_n(A)$.

Est-ce que vous diriez que $A[X] = A[Y]$? Quelle est exactement la nature de X dans $P = X^2 + 1$ ou dans $P(X) = X^2 + 1$? Si $P \in A[X]$ vaut $P(X) = X^2$ et si $Q \in A[Y]$ vaut $Q(Y) = Y^2$, est-ce que vous oseriez écrire $P = Q$?

0.6 Sage est là pour vous aider

Il existe de nombreux logiciels de mathématique. Notre préféré est Sage pour une raison très précise : en tant que langage de programmation, Sage est python qui est un langage généraliste. La syntaxe et la structure de Sage ne sont pas *ad hoc* pour faire de math, et ce qu'on apprend en Sage peut être recyclé pour faire n'importe quoi : navigateur web, script de manipulation de texte, traitement d'image, réseau neuronaux, ...

Sage est un logiciel disponible pour l'épreuve de modélisation de l'agrégation de mathématique ; il y a donc de bonnes chances que vous en ayez l'usage.

0.6.1 Lancez-vous dans Sage

(1) Aller sur `http://www.sagemath.org`,

(2) créer un compte,

(3) créer des feuilles de calcul et s'amuser !!

Il y a beaucoup de documentation sur le site officiel [5], et nous vous conseillons particulièrement le livre [2].

Si vous comptez utiliser régulièrement ce logiciel, je vous recommande *chaudement* de l'installer sur votre ordinateur.

0.6.2 Exemples de ce que Sage peut faire pour vous

Ce livre est émaillé de petits bouts de code en Sage montrant ses différentes fonctionnalités là où nous en avons besoin [6]. Voici une liste (non exhaustive) de ce que Sage peut faire pour vous.

(1) Calculer des limites de fonctions, exemples 42.1 et 42.2.

(2) Tracer des graphes de fonctions, exemple 42.2.

(3) Tracer des courbes en trois dimensions, voir exemple 11.145. Notez que pour cela vous devez installer aussi le logiciel Jmol. Pour Ubuntu, c'est dans le paquet `icedtea6-plugin`.

(4) Calculer des dérivées partielles de fonctions à plusieurs variables, voir exemple 42.3.

(5) Résoudre des systèmes d'équations linéaires. Voir les exemples 42.4 et 42.5. Lire aussi la documentation.

(6) Tout savoir d'une forme quadratique, voir exemple 42.6.

5. http://www.sagemath.org
6. Soit un vrai besoin comme tracer un graphique en 3D, soit de la paresse comme calculer une grosse dérivée.

(7) Calculer la matrice hessienne de fonctions de deux variables, déterminer les points critiques, déterminer le genre de la matrice hessienne aux points critiques et écrire les extrémums de la fonctions (sous réserve d'être capable de résoudre certaines équations), voir les exemples 42.7 et 42.8.

(8) Indiquer une infinité de solutions à une équation en utilisant des paramètres. L'exemple 42.9 montre ça avec une équation algébrique. Un exemple concernant des fonctions trigonométriques :

```
sage: solve(sin(x)/cos(x)==1,x,to_poly_solve=True)
[x == 1/4*pi + pi*z1]
sage: solve(sin(x)**2==cos(x)**2,x,to_poly_solve=True)
[sin(x) == cos(x), x == -1/4*pi + 2*pi*z86, x == 3/4*pi + 2*pi*z84]
```

Notez l'option `to_poly_solve=true` dans `solve`.

(9) Calculer des dérivées symboliquement, voir exemple 42.10.

(10) Calculer des approximations numériques comme dans l'exemple 42.11.

(11) Calculer dans un corps de polynômes modulo comme $\mathbb{F}_p[X]/P$ où P est un polynôme à coefficients dans $\mathbb{F}_p$. Voir l'exemple 18.65.

Sage peut en général faire tout ce que vous êtes capable de faire à l'entrée en master et probablement bien plus, à la notable exception des limites à deux variables.

Remarque 0.1.
Sage peut toutefois vous induire en erreur si vous n'y prenez pas garde parce qu'il sait des choses en mathématique que vous ne savez pas. Par conséquent il peut parfois vous donner des réponses (mathématiquement exactes) auxquelles vous ne vous attendez pas. Voir par exemple 14.111 pour le logarithme de nombres négatifs. Et aussi ceci :

```

 SageMath version 7.3, Release Date: 2016-08-04
 Type "notebook()" for the browser-based notebook interface.
 Type "help()" for help.

sage: limit(1/x,x=0)
Infinity
sage: limit(1/x**2,x=0)
+Infinity
```

tex/sage/sageSnip017.sage

Sage fait une différence entre `Infinity` et `+Infinity` et donne

$$\lim_{x\to 0}\frac{1}{x}=\infty \tag{0.3}$$

ainsi que

$$\lim_{x\to 0}\frac{1}{x^2}=+\infty. \tag{0.4}$$

Voir aussi la compactification en un point d'Alexandroff 7.66.

0.7 Comment contribuer et aider ?

Ces notes ne sont pas relues de façon systématique. Aucune garantie. Merci de me signaler toute faute ou remarque : le relecteur c'est toi. Voici une petite liste de questions que je me pose

ou de choses écrites dont je ne suis pas certain. Si vous avez un avis ou une réponse à un des points, merci de vous faire connaitre.

Les questions ouvertes sont divisées en trois niveaux de difficulté :

(1) Niveau facile : un étudiant de licence devrait pouvoir le faire.
(2) Niveau moyen : un candidat à l'agrégation de mathématique devrait pouvoir le faire.
(3) Demande probablement de connaissances avancées en mathématique ; au moins être tout à fait à l'aise avec le niveau d'agrégation.

Quel que soit votre niveau, vous pouvez faire ceci :

(1) M'écrire pour me signaler toutes les fautes que vous voyez, même si vous n'êtes pas sûr.
(2) Si vous n'êtes pas expert, me signaler tous les endroits qui vous semblent obscurs. Vu que ces notes sont destinées à *apprendre*, les avis des non-experts sont très importants.
(3) Mettre une copie de (ou un lien vers) ces notes sur votre site.

0.7.1 Des preuves qui manquent

Vous trouverez un peu partout des énoncés sans preuves. Certaines sont surement très faciles, et d'autres probablement assez compliquées. N'hésitez pas à rédiger une preuve et me l'envoyer.

Vous pouvez m'envoyer vos preuves sous forme de « c'est bien fait dans tel cours », avec une URL.

Ne me dites juste pas « c'est bien fait dans tel *livre* ». Je ne travaille pas à l'université, et je n'ai pas accès à une bibliothèque universitaire ; je n'ai donc pas réellement accès à ces fameux « livres » dont tout le monde parle.

0.7.2 Du texte qui manque

Vous remarquerez que de nombreuses pages du Frido sont des enchainements de théorèmes et démonstrations sans articulations. Autrement dit, il manque ce qu'à l'agrégation on irait à l'oral quand on présente le plan. Si vous avez des idées de choses à ajouter ici où là, faites le moi savoir.

0.7.3 Mes questions de géométrie

0.7.3.1 Facile

(1) Donner une interprétation en termes de plans, droites ou je ne sais quoi de géométrique au 4-cycle parmi les permutations des sommets du tétraèdre. D'où sort géométriquement la matrice (17.439) ?

0.7.3.2 Moyen

(1) Démontrer la proposition 8.4 qui donne les relations de Chasles pour un espace affine.
(2) En géométrie projective, dans la sphère de Riemann $\hat{\mathbb{C}} = P_1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$ est-ce qu'il existe une notion de cercle dont le centre est ∞ ? Voir le point 22.65.
(3) Géométrie projective. Tout le monde semble définir le birapport en identifiant $P(\mathbb{K}^2)$ à $\hat{\mathbb{K}} = \mathbb{K} \cup \{\infty\}$. Bien entendu, personne ne semble s'être attribué la mission d'expliciter la dépendance du birapport en le choix de l'identification. Je le fais à la définition 22.40.

   Mais cette définition dépend du choix d'identification $\varphi\colon P(\mathbb{K}^2) \to \hat{\mathbb{K}}$, comme le montre l'exemple 22.42. J'ai donc défini des classes d'identifications possibles $A(\varphi)$ en 22.38. Et je démontre la proposition 22.43 que si $\varphi_a \in A(\varphi)$ alors les birapports construits à partir de φ et φ_a sont identiques.

   Question : pourquoi personne ne semble faire ce travail ? En quoi l'identification φ_0 que tout le monde utilise est plus canonique qu'une autre ? Est-ce que l'on peut décrire simplement les classes $A(\varphi)$? Le groupe qui conserve le birapport associé à φ est-il isomorphe au groupe qui conserve le birapport associé à φ' ? Quels que soient φ et φ' ?

   Suis-je la seule personne au monde à m'être demandé si le birapport était un objet canonique ?

(4) En géométrie projective, dans $P_1(\mathbb{C}) = \hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, si ℓ est une droite dans $\mathbb{C}$, est-ce que la droite correspondante dans $\hat{\mathbb{C}}$ contient le point ∞ ? Moi j'ai envie de dire que ∞ est sur toutes les droites. Voir le problème 23 et la remarque 22.63.

(5) Géométrie affine, barycentre. Les mauvaises langues diraient que tout cela est du snobisme autour de la paresse d'écrire $\overrightarrow{xy}$ au lieu de $y - x$. Est-ce qu'il y a des cas où toute l'étude des espaces affines et des barycentres en particulier apportent *réellement* plus qu'une facilité d'écriture par rapport à travailler dans le cadre vectoriel pur ?

0.7.4 Mes questions d'algèbre

0.7.4.1 Facile

0.7.4.2 Moyen

(1) La définition 1.90 de $\sum_{i\in I} f(i)$ (I est fini et $f\colon I \to (G,+)$ est une application) demande une preuve. Comment prouver que

$$\sum_{i=1}^{n} f\big(\sigma_1(i)\big) = \sum_{i=1}^{n} f\big(\sigma_2(i)\big) \tag{0.5}$$

lorsque σ_1 et σ_2 sont des bijections ?

De façon peut-être plus simple, il faut prouver la proposition 1.91 qui exprime l'invariance d'une somme commutative finie sous le groupe des bijections des termes.

Je crains que cela demande de décomposer $\sigma_1\sigma_2^{-1}$ en cycles et de pas mal chipoter pour tout ramener à des permutations de deux termes.

0.7.4.3 Difficile

(1) Que penser de 6.80 qui dit que l'extension de corps $\mathbb{K}(\alpha)$ dépend en réalité du corps ambiant dans lequel on calcule l'extension ?

Est-ce qu'il existe des exemples moins triviaux et plus utiles que celui que je donne ?

(2) Pour quelle classe d'anneaux et de polynômes le quotient $A[X]/(P)$ est-il un corps ?

0.7.4.4 Non classées

(1) Est-ce qu'il existe une structure raisonnable d'espace vectoriel sur $\mathbb{Z}$? Est-ce qu'il existe des corps discrets infinis ?

Dans cette question, j'ai derrière la tête que dans un espace vectoriel topologique nous avons une notion de suite de Cauchy, définition 7.274. Donc dans ce cas la notion d'espace complet est une notion topologique. Or il y a l'exemple 7.280 qui donne deux distances sur $\mathbb{N}$, qui donnent la même topologique, mais l'un étant complet, l'autre non.

Si il y avait une structure vectorielle sur $\mathbb{N}$, cela créerait une contradiction. Au moins au sens où la définition 7.274 de suite de Cauchy « topologique » ne redonne pas la même que la notion « métrique » de la définition usuelle 7.276.

(2) La « décomposition en facteurs premiers » dans $\mathbb{Z}[i\sqrt{2}]$ que je donne dans l'exemple 3.157 est-elle correcte ? En particulier le lemme 3.158 ?

(3) Est-ce que la fin de la démonstration 18.25 avec cette histoire d'ensemble $\{\xi_k^q$ tel que $q \in \mathbb{N}\}$ fini est compréhensible ?

(4) Les représentations *irréductibles* sont les modules *indécomposables*. Quid des modules irréductibles ? C'est pas un peu dingue de ne pas utiliser le mot « irréductible » pour désigner les mêmes choses dans le cas des modules et celui des représentations ?

(5) Rendre rigoureuse la remarque (9.219) qui dit que les matrices dont le polynôme minimal est égal au polynôme caractéristique sont denses dans les matrices.

(6) La partie initiation de récurrence ($r = 2$) de la preuve de la proposition 8.37 à propos de convexe et de barycentre est-elle correcte ? Ce passage de l'espace affine à l'espace vectoriel sous-jacent me paraît un peu facile.

(7) Est-ce que l'énoncé et la démonstration de la proposition 6.98 sont corrects ? Si a et b sont des racines de P, alors $\mu_a\mu_b$ divise P (si $\mu_a \neq \mu_b$). Cette proposition est utilisée dans la démonstration de l'irréductibilité des polynômes cyclotomiques (proposition 18.23).

(8) À quoi sert l'hypothèse « autre que $\mathbb{F}_2$ » dans le lemme 18.75 ? Peut-être dans la notion de déterminant parce qu'en caractéristique 2, l'antisymétrie d'une forme linéaire n'implique le fait qu'elle soit alternée.

(9) L'inversibilité de la somme de Gauss (proposition 18.41) est-elle bien démontrée ?

(10) Des commentaires sur l'exemple 6.143 qui montre que $X^p - X + 1$ est irréductible sur $\mathbb{F}_p$.

(11) Les idéaux de A/I sont en bijection avec les idéaux de A contenant I. Justification de l'équation (3.87).

(12) À propos d'extensions algébriques, est-ce que la proposition 6.87 est correcte ? Est-ce qu'implicitement, il n'y a pas un sur-corps de $\mathbb{K}$ dans lequel il faut travailler ?

(13) À propos de construction à la règle et au compas. Pour l'addition d'angles, l'exemple 18.82 explique comment on construit la somme de deux angles. Le problème est que cette construction se fait par intersection de deux cercles. Une des deux intersections donne $\alpha+\beta$ et l'autre donne $\alpha - \beta$. Comment par construction peut-on choisir le bon point ?

(14) À propos de chiffrement RSA, quelle est la probabilité que le message M ne soit pas premier avec p ? Est-ce que Alice (qui est celle qui chiffre avec la clef de Bob) peut le vérifier ? Que penser des points que j'énumère à la page 1361 au dessus du problème 22 ?

(15) Isomorphisme du corps $\mathbb{R}$. Que penser de la remarque 6.5 ?

0.7.5 Mes questions d'analyse

0.7.5.1 Facile

0.7.5.2 Moyen

(1) Définition de l'intégrale. Si $f(x) = x^2$, alors $\int_{\mathbb{R}} f(x)dx$ utilise la définition de l'intégrale des fonctions partout positive en tant que supremum des fonctions étagées qui minorent f. Pas de problèmes, $\int_{\mathbb{R}} x^2dx$ existe et vaut ∞.
Par contre $g(x) = x^2 - 1$ possède une minuscule partie négative et doit donc prendre la définition des intégrales pour des fonctions de signe variable. Cette définition 13.168 demande que les intégrales des parties positives et négatives soient séparément finies. Donc $\int_{\mathbb{R}}(x^2-1)dx$ n'existe pas.
Ça me semble un peu restrictif. Il y a un problème à définir $\int_{\mathbb{R}} f$ lorsque la partie positive est infinie, mais la partie négative reste finie ?

(2) Que penser de la remarque 24.38 qui dit qu'on doit avoir un théorème de complétion de partie orthonormale en une base orthonormale pour un espace de Hilbert ? C'est vrai ?

(3) Préciser l'énoncé et donner une démonstration de la proposition 10.101 qui traite de sommes dénombrables.

(4) Explosion en temps fini. C'est le corolaire 31.18. Dans le cas où $\lim_{t\to t_{max}} \|y(t)\| = \infty$ alors la dérivée de y n'est pas non plus bornée. Correct ?

(5) Est-ce qu'il y a moyen de définir la mesure produit (de deux espaces mesurés) sans passer par l'intégrale ?
Le théorème-définition 13.210 donne le produit de mesures par la formule

$$(\mu_1 \otimes \mu_2)(A) = \int_{\Omega_1} \mu_2\big(A_2(x)\big)d\mu_1(x) = \int_{\Omega_2} \mu_1\big(A_1(y)\big)d\mu_2(y). \tag{0.6}$$

On peut faire plus abstrait ?
Sans une telle définition, l'ordre imposé est :

— mesure
— intégrale
— mesure produit.

En particulier, quand on voit l'intégrale, la mesure de Lebesgue sur $\mathbb{R}^n$ n'est pas encore définissable.

Il serait bien de pouvoir faire :

— mesure
— mesure de Lebesgue sur $\mathbb{R}$
— mesure produit
— mesure de Lebesgue sur $\mathbb{R}^n$
— intégrale.

(6) La proposition 11.354 prouve que la fonction $x \mapsto a^x$ $(a > 0)$ est dérivable. Pour ce faire, le concept de primitive est utilisé (pas jusqu'aux intégrales, cependant). Ça me semble incroyable qu'il faille ça pour prouver une dérivabilité.

Prouver que la limite

$$\lim_{\epsilon \to 0} \frac{a^\epsilon - 1}{\epsilon} \tag{0.7}$$

existe et n'est pas infinie sans recourir aux intégrales, aux primitives.

(7) Si une application $f\colon \mathbb{R} \to \mathbb{R}^2$ est continue et injective, sa réciproque est continue. Est-ce vrai ?

Plus précisément, il me faut soit une preuve soit un contre-exemple de ceci.

Proposition 0.2.
Soit une application injective et continue $f\colon [0,1] \to \mathbb{R}^n$. Nous nommons $\Gamma = f\big([0,1]\big)$ que nous munissons de la topologie induite de $\mathbb{R}^n$.
Alors $f^{-1}\colon \Gamma \to [0,1]$ est continue.

Si cela était vrai, ça me permettrait de supprimer l'hypothèse dans la proposition 20.54.

0.7.5.3 Difficile

(1) Soit $n \in \mathbb{N}$, $A \in \mathbb{R}$ et $x_0 \in \mathbb{Q}$. Nous considérons la suite[3]

$$x_{k+1} = \frac{1}{n}\left((n+1)x_k + \frac{A}{x_k^{n-1}}\right). \tag{0.8}$$

Prouver que :

— C'est une suite de Cauchy
— $x_k^n \to A$.

Il me faudrait une démonstration de cela sans passer par la méthode de Newton. Par exemple via le binôme de Newton.

Mon but serait de définir $a^{1/n}$ pour tout $n \in \mathbb{N}$ sans passer par de l'analyse (ou en tout cas pas en passant par le concept de fonction continue). Pour l'instant, c'est la définition 11.325 qui définit $a^{1/n}$. Cela se base sur un argument de fonction continue strictement croissante pour obtenir une bijection.

(2) Est-ce que la proposition 7.281 qui donne le critère $d(x_p, x_q) \leqslant \epsilon$ pour être une suite de Cauchy est valide dans un espace topologique métrique au lieu de normé ? Dans quels cas a-t-on

$$d(a,b) = d(a+u) + d(b+u) \tag{0.9}$$

lorsque d est une distance qui n'est pas spécialement induite d'une norme ?

(3) Soit V, un espace vectoriel normé. Soit $v \in V$. Est-ce qu'il existe un élément $\varphi \in V'$ (application linéaire continue) telle que $\varphi(v) = 1$ et $\|\varphi\| = 1$?

(4) Que penser de 13.154 qui tente d'expliquer pourquoi on ne définit pas l'intégrale d'une fonction non mesurable, malgré que le supremum qui la définirait existe forcément ?

(5) Changement de variables. À quel point la proposition 13.257 est-elle équivalente au théorème usuel ?

0.7.5.4 Non classées

(1) À propos de suites de Cauchy dans un espace vectoriel topologique et dans un espace métrique, est-ce que le théorème 7.291 est correct ?

Soit V un espace vectoriel topologique métrisable [7], alors il admet une métrique d compatible avec la topologie telle que une suite dans V est d-Cauchy si et seulement si elle est τ-Cauchy.

Dans cet ordre d'idée, il faut des exemples de :

— un espace vectoriel topologique métrisable et une métrique d compatible avec la topologie, mais dont les suites d-Cauchy ne sont pas celles τ-Cauchy. Et en particulier dont la complétude est différente que celle de la « bonne » métrique donnée par la théorème 7.288.

— Et aussi un exemple pour la remarque 7.292.

(2) L'exemple 16.23 parle d'inverser une intégrale et une dérivée au sens des distributions pour prouver que la dérivée de $\int_0^x g(t)dt$ par rapport à x est g. Rendre cela rigoureux.

(3) À propos du théorème de récurrence de Poincaré 13.83, l'application ϕ doit être mesurable ? Répondre à la question posée sur la page de discussion de l'article sur wikipédia.

Toujours à propos du théorème de récurrence de Poincaré, il me semble qu'il y a un énoncé qui insiste sur la compacité de l'espace des phases et une démonstration utilisant la propriété de sous-recouvrement fini. Je serais content de retrouver cela. (ce serait sans doute mettable dans la leçon sur l'utilisation de la compacité)

(4) Dans [4], on parle de la proposition 29.42 à sa page 10. Comment est-ce qu'on justifie le passage

$$\int_{\mathbb{R}^d} T\big(y \mapsto \varphi(x)\psi(x-y)\big)dx = T\Big(y \mapsto \int_{\mathbb{R}^d} \varphi(x)\psi(x-y)dx\Big). \tag{0.10}$$

Sylvie Benzoni précise que « ceci demanderai quelques justifications ». Où trouver lesdites justifications ? Il s'agit de permuter une distribution et une intégrale.

(5) Peut-on permuter une application linéaire et continue avec une somme pas spécialement dénombrable ? En supposant que $\sum_{i\in I} f(v_i)$ existe, la proposition 10.102 semble dire que oui. Est-ce correct ?

Peut-on avoir un exemple de partie sommable $\{v_i\}_{i\in I}$ et d'application linéaire continue f telle que la partie $\{f(v_i)\}$ ne soit pas sommable ?

Peut-être ceci dans un espace de Hilbert. $v_i = \frac{1}{i^2}e_i$ puis $f(e_i) = i^3 e_i$.

(6) Soit une partie orthonormale *pas spécialement dénombrable* $\{u_i\}_{i\in I}$ d'un espace de Hilbert (pas spécialement séparable). Si

$$x = \sum_{i\in I} x_i u_i, \tag{0.11}$$

puis-je prendre le produit scalaire avec u_k et le permuter avec la somme pour déduire que $x_k = \langle x, u_k \rangle$?

C'est ce que je fais dans la proposition 24.25.

(7) Théorème de point fixe et équation différentielle. Que penser de l'exemple 16.45 qui itère la contraction de Cauchy-Lipschitz pour résoudre $y'(t) = y(t)$, $y(0) = 1$? Est-ce que c'est générique comme comportement ? Est-ce que la convergence est efficace dans des cas moins triviaux ?

7. i.e. admet une base dénombrable de topologique, voir la proposition 7.289

0.7.6 Mes questions d'analyse numérique

(1) Différences finies. Il faut une analyse de consistance, stabilité et convergence du schéma à 9 points pour le laplacien donné par (34.146). Je crois qu'il est d'ordre 6, mais je n'en suis vraiment pas sûr.

0.7.7 Mes questions de probabilité et statistiques.

0.7.7.1 Facile

0.7.7.2 Moyen

(1) Soit une variable aléatoire X à valeurs réelles. Est-ce que la tribu engendrée par X est d'une façon ou d'une autre engendrée par les « courbes de niveau » de X ? C'est-à-dire par les $X^{-1}(\{t\})$ pour les $t \in \mathbb{R}$.
C'est ce qui semble ressortir de l'exemple de 35.2.10. Et intuitivement, je trouve que ça irait bien ...

0.7.7.3 Difficile

0.7.7.4 Non classées

0.7.8 Mes questions de LaTeX et programmation

0.7.8.1 Facile

(1) Comment faire en sorte que les mots commençant par « é » soient avec les « e » dans l'index, et non avant les « a » ? Il me faudrait un mécanisme plus automatique que faire `machin@truc`.
(2) Il y a des problèmes dans la table des matières. « Table des matières », « Index », et « Liste des notations » ne pointent pas vers la bonne page.

0.7.8.2 Moyen

(1) Revoir le mécaniste de l'index thématique. Il faudrait pouvoir les trier avec des titres. Mais attention : il doit arriver avant la table des matières.

0.7.8.3 Difficile

Je ne sais pas comment faire, et à mon avis il faudra innover.

(1) Si vous savez comment faire `pdf -> epub` pour créer un eBook, faites le moi savoir. Cahier des charges :
   — libre, disponible sur Ubuntu
   — en ligne de commande (en tout cas : exécutable depuis un script en python ou C++)

   Attention : le Frido étant un truc assez compliqué, avant de répondre la première chose qui vous passe par la tête, assurez-vous que votre solution fait avancer les choses sur le Frido et non sur un petit document de test.
   Nous en avons déjà un peu discuté sur https://github.com/LaurentClaessens/mazhe/issues/13. Il faudra entre autres faire un script qui remplace tous les environnements tizk des fichiers `*.pstricks` (désolé pour la convention de nommage historique) par un simple `includegraphics` du fichier `pdf` correspondant que l'on trouvera dans le répertoire `auto/pictures_tikz`.
(2) Écrire un script (en python ou autre) qui prend en argument deux numéros ou noms de chapitres et qui retourne l'ensemble des lignes du premier qui contient des `ref` ou `eqref` dont le label correspondant est dans le second.
   Attention : il faut tenir compte de `input` de façon récursive.
   Bonus : calculer le hash sha1 de chaque ligne du résultat et ne pas l'afficher si il se trouve dans la liste du fichier `commons.py`.

0.7.9 Numérique

0.7.9.1 Moyen

(1) L'erreur de cancellation provoquée par la différence $a - \tilde{a}$ lorsque a et $\tilde{a}$ sont proches n'a pas de conséquences sur l'ordre de grandeur de la réponse. Seulement des conséquences sur la valeur des chiffres significatifs. Vrai ou faux ?
Voir la remarque 33.20.

0.8 Taper du code pour le Frido

Dans cette section nous donnons quelques indications sur la façon de taper du code LaTeX pour le Frido.

Tout commence par télécharger les sources à l'adresse

https://github.com/LaurentClaessens/mazhe

0.8.1 Pour compiler le document vous même

Lisez le fichier `COMPILATION.md`.

0.8.2 Nommage des fichiers `tex`

J'ai pris l'habitude de préfixer les noms par un nombre. Par exemple `139_EspacesVecto`. Le fait est qu'il est plus simple, pour ouvrir le ficher, de taper `139<TAB>` que de se souvenir si « EspacesVecto » est écrit avec une majuscule, en français, en anglais, ...

De plus un chapitre contenant plusieurs fichiers, nous nous retrouvons rapidement avec beaucoup de fichiers nommés `EspacesVecto1`, `EspacesVecto2`, etc.

Vous trouverez le prochain numéro disponible dans `réserve.tex`.

0.8.3 Inclure des exemples de code

Pour inclure du code Sage, nous utilisons la commande `\lstinputlisting`. Ici encore, le fichier `réserve.tex` contient le prochain disponible.

0.8.4 Pour les exercices

ATTENTION : dans un futur proche, je vais supprimer tous les exercices et les mettre sous forme d'exemples. Une des raisons est de supprimer la dépendance en le paquet personnel `exocorr` qui rend compliqué la compilation du Frido par des tierces personnes.

Les exercices sont tapés dans les fichiers déjà pré-remplis `src_exocorr/exo*.tex`. Les corrections sont dans le fichier `src_exocorr/corr*.tex` correspondant. Ces fichiers ne sont pas inclus directement, mais via la macro `\Exo`.

Le fichier `réserve.tex` contient le prochain disponible.

Exemple Vous voulez créer un exercice.

— Allez voir dans `réserve.tex` la prochaine ligne `Exo` disponible.

— Mettons que ce soit `\Exo{mazhe-0018}`

— Supprimez cette ligne de `réserve.tex`, et mettez la où vous voulez voir paraitre votre exercice.

— Tapez votre exercice dans le fichier `src_exocorr/exomazhe-0018.tex` et votre correction dans le fichier `src_exocorr/corrmazhe-0018.tex`. Ces fichiers sont déjà créés et pré-remplis. Ne changez pas le code qui y est.

0.9 Les politiques éditoriales

Certaines parties de ce texte ne respectent pas les politiques éditoriales. Ce sont des erreurs de jeunesse, et j'en suis le premier triste.

0.9.1 Licence libre

Je crois que c'est clair.

0.9.2 pdflatex

Tout est compilable avec pdfLaTeX. Pas de pstricks, de psfrag ou de ps<quoiquecesoit>.

0.9.3 utf8

Je crois que c'est clair.

0.9.4 Notations

On essaie d'être cohérent dans les notations et les conventions. Pour la transformée de Fourier par exemple, je crois que la définition du produit scalaire dans L^2, des coefficients de Fourier, de la transformation et de la transformation inverse sont cohérents. Cela demande, lorsqu'on suit un livre qui ne suit pas les conventions utilisées ici, de convertir parfois massivement.

0.9.5 De la bibliographie

On évite d'écrire en haut de chapitre « les références pour ce chapitre sont ... ». Il est mieux d'écrire au niveau des théorèmes, entre parenthèses, les références.

Lorsqu'on écrit l'énoncé d'un théorème sans retranscrire la démonstration, il faut mettre une référence vers un document *en ligne* qui en contient la preuve. Il est vraiment fastidieux de chercher une preuve sur internet et de tomber sur des dizaines de documents qui donnent l'énoncé mais pas la preuve.

0.9.6 Faire des références à tout

Lorsqu'un utilise le théorème des accroissements finis, il ne faut pas écrire « d'après le théorème des accroissements finis, blablabla ». Il faut écrire un `\ref` explicite vers le résultat. Cela alourdit un peu le texte, mais lorsqu'on joue avec un texte de plus de 2000 pages, il est parfois laborieux de trouver le résultat qu'on cherche (surtout s'il existe plusieurs versions d'un résultat et que l'on veut faire référence à une version particulière).

0.9.7 Des listes de liens internes

Le début du Frido contient une espèce d'index thématique. Il serait bon de l'étoffer.

0.9.8 Pas de références vers le futur

Dans le Frido, *aucune* preuve ne peut faire une référence vers un résultat prouvé plus bas. On n'utilise pas le théorème 10 dans la démonstration du théorème 7. Cela est une contrainte forte sur le découpage en chapitres et sur l'ordre de présentation des matières.

Il est bien entendu accepté et même encouragé de mettre des notes du type « Nous verrons plus loin un théorème qui ... ». Tant que ce théorème n'est pas *utilisé*, ça va.

En faisant

```
pytex lst_frido.py --verif
```

vous aurez une liste des références vers le bas. Cette liste doit être vide ! Ce programme cherche tous les `\ref` et `\eqref` ainsi que les `\label` correspondants et vous prévient lorsque le `\label` est après le `\ref`.

Si vous pensez qu'une référence pointée doit être acceptée (par exemple parce c'est dans une des listes de liens internes), alors vous ajoutez son hash dans la liste du fichier `commons.py`. Si il s'agit vraiment d'une référence vers un résultat que vous utilisez, alors vous devez déplacer des choses. Soit votre résultat vers le bas, soit celui que vous utilisez vers le haut. Parfois cela demande de déplacer ou redécouper des chapitres entiers... Si il n'y a vraiment pas moyen, bravo, vous venez de prouver que la mathématique est logiquement inconsistante.

0.9.9 Écriture inclusive

Je suis triste de devoir le préciser, mais le Frido est écrit en français. Nous n'utiliserons donc pas de féminisation abusives, et accepterons comme correcte des tournures comme, en parlant d'une fonction, « *elle* est *un* contre-exemple », ou en parlant d'un lemme que « *il* est *une* conséquence ».

Parfois le genre d'un objet n'est pas bien défini. Par exemple $3/4$ est *la* classe d'équivalence de $(3, 4)$ dans $\mathbb{Z} \times \mathbb{Z}\backslash 0$; mais en même temps c'est *un* élément de $\mathbb{Q}$. Nous utiliserons alors, prudemment, un neutre en disant « *il* est plus petit que 1 ».

0.10 Vérifier si vous n'avez pas fait de bêtises

Lorsqu'on fait de lourdes modifications (déplacement de parties, fusion de théorèmes, etc) il est toujours possible de faire des bêtises d'au moins deux types : créer des références vers le futur et supprimer des parties (genre couper-coller en oubliant le coller). Pour s'en prémunir, le script suivant lance quelques compilations et vérifications :

```
./testing.sh
```

Aucune erreur ne devrait être signalée.

Attention : ce script fait quelques manipulations à base de `git stash` et crée une nouvelle branche (nom aléatoire assez long) pour tester votre dernière modification sans créer de commit.

0.11 Acceptation des contributions

TD ;DR : pratiquement aucun patch n'est refusé.

Le premier critère d'acceptation d'une contribution est évidemment la correction mathématique.

0.11.1 Attention aux expressions rationnelles

Si vous trouvez une faute d'orthographe, rien ne vous empêche de faire une recherche de la même faute pour la corriger d'un seul coup dans 25 fichiers. Faites toutefois attention à des remplacements automatiques sur base d'expressions rationnelles telles que

```
des [a-z]*[^s]
```

qui serait supposé détecter des erreurs de pluriel. Je vous laisse trouver au moins 5 cas sans fautes qui satisfont cette expression.

Utilisez de telles expressions pour *trouver* des fautes, pas pour les corriger.

0.11.2 Pas de modifications massives, automatiques pour des raisons cosmétiques

Il est un type de contributions que je ne vais plus accepter, ce sont les modifications massives et automatiques de *tous* les fichiers pour des raisons de « propreté » du code. Exemples :

— Supprimer automatiquement tous les espaces en bout de lignes,

— Supprimer automatiquement toutes les lignes vides en fin de fichier

— Remplacer `\ref` par `~\ref`

— Remplacer `\[` par `\begin{equation}`

— Remplacer `\og` par « (avec ou sans espaces devant ou derrière)

— ...

Ce type de substitutions automatiques créent des patch gigantesques qui prennent un temps astronomique à relire pour un bénéfice pas tellement évident.

Pire : ils ont des effets de bords pas toujours évident à détecter ou à prévoir.

N'oubliez pas que le Frido n'est pas que du LaTeX. Il est aussi divers scripts de pré et post-compilation (y compris qui hackent des fichiers intermédiaires entre deux passes de LaTeX). Ces scipts, écrits par votre très humble et très obéissant serviteur, ne sont pas parfaits et les parseurs reposent sur certaines hypothèses. Donc des choses qui ne devraient ne rien changer du point de vue de LaTeX peuvent avoir des conséquences.

Bien entendu, si vous êtes en train de taper des math et que ce genre de « malpropreté » du code vous gêne, vous pouvez corriger dans les fichiers que vous modifiez.

Chapitre 1

Construction des ensembles de nombres

1.1 Quelques éléments sur les ensembles

1.1.1 Petit mot d'introduction

1.1.
Le Frido n'est pas supposé être lu dans l'ordre de la première à la dernière page ; les matières y sont présentées dans l'ordre logique mathématique, et non dans l'ordre logique pédagogique, et encore moins par ordre de difficulté croissante.

En mathématique, si on lit une démonstration et que l'on veut vraiment tout justifier, et justifier toutes les étapes de tous les résultats utilisés, on tombe forcément un jour sur les axiomes.

Or l'axiomatique est un sujet particulièrement difficile. Nous n'allons donc pas « tout justifier » jusque là. Nous n'allons même pas préciser quel système d'axiome est utilisé. En particulier nous n'allons pas donner l'axiomatique des ensembles : nous allons supposer connus les ensembles et leurs principales propriétés.

Bref. Nous supposons avoir une théorie des ensembles qui tient la route. En particulier nous supposons connues les notions suivantes :

(1) ensemble vide,

(2) ensemble, appartenance, intersection, union,

(3) application entre deux ensembles, notation $f(x)$ pour désigner l'image de x par f,

(4) produit cartésien de plusieurs ensembles.

Ce sont toutes des choses dont la construction à partir des axiomes n'est en aucun cas évidente. En particulier, des « définitions » comme « l'intersection de deux ensembles est l'ensemble contenant exactement les éléments communs aux deux ensembles » ne sont pas correctes parce qu'elles passent à côté de l'existence et de l'unicité d'un tel ensemble.

Une petite définition cependant, que l'on utilisera :

Définition 1.2.
Deux ensembles A et B sont ***disjoints*** *si leur intersection est vide*[1] *; en d'autres termes, s'il n'existe aucun élément commun à A et B.*

1.3.
Remarquez par exemple que la première phrase de l'article de Wikipédia sur la construction de $\mathbb{N}$ est « Partant de la théorie des ensembles, on identifie 0 à l'ensemble vide, puis on construit ... ». Il est bien précisé que l'on part d'une théorie des ensembles.

1. Remarquez que les mots « intersection » et « vide » sont de ceux que nous avons décidé de ne pas définir.

1.4.
La suite de ce chapitre sera essentiellement sans exemples parce qu'avant d'avoir construit les ensembles de nombres, je ne sais pas très bien quels exemples on peut donner de quoi que ce soit.

1.1.2 Ensemble ordonné

Définition 1.5.
Soient deux ensembles E et F. Une application $f\colon E \to F$ est

(1) ***surjective*** *si pour tout $y \in F$, il existe $x \in E$ tel que $y = f(x)$;*

(2) ***injective*** *si pour tout $y \in F$, il existe au plus un $x \in E$ tel que $y = f(x)$;*

(3) ***bijective*** *si elle est à la fois injective et surjective.*

La méthode la plus courante pour démontrer qu'une application $f\colon E \to F$ est injective est de considérer $x, y \in E$ tels que $f(x) = f(y)$, et de prouver à partir de là que $x = y$. Ou alors de supposer $x \neq y$ et obtenir une contradiction.

La technique de la contradiction est évidemment la plus courante lorsque l'égalité $f(x) = g(x)$ implique une équation faisant intervenir $1/(x - y)$.

1.6.
L'**axiome du choix** que nous acceptons peut s'énoncer comme ceci[5] : Étant donné un ensemble X d'ensembles non vides, il existe une fonction définie sur X, appelée fonction de choix, qui à chacun d'entre eux associe un de ses éléments.

Définition 1.7.
Une ***relation d'ordre*** *sur un ensemble E est une relation binaire (notée $\leqslant$) sur E telle que pour tous $x, y, z \in E$,*

réflexivité *: $x \leqslant x$*

antisymétrie *: $x \leqslant y$ et $y \leqslant x$ implique $x = y$*

transitivité *: $x \leqslant y$ et $y \leqslant z$ implique $x \leqslant z$.*

Définition 1.8.
Un ensemble ordonné est ***totalement ordonné*** *si deux éléments sont toujours comparables : si $x, y \in E$ alors nous avons soit $x \leqslant y$ soit $y \leqslant x$. Si les éléments ne sont pas tous comparables, nous disons que l'ordre est* ***partiel****.*

Définition 1.9.
Soit un ensemble ordonné $(E, \leqslant)$ et une partie A de E. Nous disons que $m \in A$ est un ***minimum*** *de A si pour tout $x \in A$, l'élément m est comparable à x et $m \leqslant x$.*

Un élément $p \in E$ est un ***minorant*** *de A si pour tout $a \in A$, l'élément p et a sont comparables et $p \leqslant a$.*

Les notions de ***maximum*** *et de* ***majorant*** *sont définies de façon analogue.*

Lorsqu'une partie possède un minimum, ce dernier est nommé le « plus petit élément » de la partie. Attention : il n'en existe pas toujours. D'innombrables exemples pourront être vus lorsque nous aurons construits $\mathbb{Q}$ et $\mathbb{R}$. Typiquement les intervalles du type $]a, b[$.

Définition 1.10.
Un ensemble ordonné est ***bien ordonné*** *si toute partie non vide possède un plus petit élément : si A est une partie de E, alors $\exists x \in A, \forall y \in A, x \leqslant y$.*

1.11.
Quelques remarques.

(1) L'inégalité stricte (définie par : $x < y$ si et seulement si $x \leqslant y$ et $x \neq y$) n'est pas une relation d'ordre parce qu'elle n'est pas réflexive.

(2) Nous verrons dans la remarque 1.145 que l'intervalle $[-1, 1]$ dans $\mathbb{R}$ n'est pas bien ordonné.

(3) Un ensemble bien ordonné est forcément totalement ordonné parce que toutes les parties de la forme $\{x, y\}$ possèdent un minimum. Par conséquent x et y doivent être comparables : $x \leqslant y$ ou $y \leqslant x$.

Exemple 1.12.
Si E est un ensemble, l'inclusion est un ordre sur l'ensemble des parties de E, mais pas un ordre total parce que si X, Y sont des parties de E, alors nous n'avons pas automatiquement soit $X \subset Y$ ou $Y \subset X$. △

La notion d'ordre permet d'introduire la notion d'intervalle.

Définition 1.13.
*Soit un ensemble totalement ordonné $(E, \leqslant)$. Un **intervalle** de E est une partie I telle que tout élément compris entre deux éléments de I soit dans I. En formule, la partie I de E est un intervalle si*

$$\forall a, b \in I, (a \leqslant x \leqslant b) \Rightarrow x \in I.$$

1.1.3 Lemme de Zorn

Définition 1.14 (Ensemble inductif[6])**.**
*Un ensemble est **inductif** si tout sous-ensemble totalement ordonné admet un majorant.*

Lemme 1.15 (Lemme de Zorn[7])**.**
Tout ensemble ordonné inductif non vide admet au moins un élément maximal.

1.1.4 Complémentaire

Définition 1.16.
*Soit E, un ensemble et A, une partie de E (c'est-à-dire un sous-ensemble de E). Le **complémentaire** de l'ensemble A, dans E, noté $\complement A$ est l'ensemble des éléments de E qui ne font pas partie de A :*

$$\complement A = E \backslash A = \{x \in E \text{ tel que } x \notin A\}. \tag{1.1}$$

Nous allons aussi régulièrement noter le complémentaire de A par A^c.

Lemme 1.17.
Quelques propriétés à propos des complémentaires. Si E est un ensemble et si A et B sont des sous-ensembles de E, nous avons

(1) $\complement\complement A = A$, *en d'autres termes,* $E \backslash (E \backslash A) = A$,

(2) $\complement(A \cap B) = \complement A \cup \complement B$,

(3) $\complement(A \cup B) = \complement A \cap \complement B$,

(4) $A \backslash B = A \cap \complement B$.

Définition 1.18 (différence symétrique)**.**
*Si A et B sont des ensembles, l'ensemble $A \Delta B$ est la **différence symétrique** d'ensembles :*

$$A \Delta B = (A \cup B) \backslash (A \cap B). \tag{1.2}$$

C'est l'ensemble des éléments étant soit dans A soit dans B mais pas dans les deux.

Lemme 1.19.
Si A et B sont des ensembles nous avons

(1) $A^c \Delta B^c = A \Delta B$.

(2) $(A \Delta B) \Delta B = A$.

Démonstration. D'abord nous avons l'égalité $X^c \backslash Y^c = Y \backslash X$. Cela se prouve de façon classique en séparant deux cas selon que B soit inclus dans A ou non.

De là nous avons la première assertion :

$$A^c \Delta B^c = (A^c \cup B^c) \backslash (A^c \cap B^c) = (A \cap B)^c \backslash (A \cup B)^c = (A \cup B) \backslash (A \cap B) = A \Delta B. \tag{1.3}$$

Pour la seconde assertion, il faut remarquer que $(A\Delta B) \cup B = A \cup B$ et que $(A\Delta B) \cap B = B \backslash A$, donc

$$(A\Delta B)\Delta B = (A \cup B) \backslash (B \backslash A) = A. \tag{1.4}$$

□

1.1.5 Relations d'équivalence

Définition 1.20.
Si E est un ensemble, une ***relation d'équivalence*** *sur E est une relation $\sim$ qui est à la fois*

réflexive *$x \sim x$ pour tout $x \in E$,*

symétrique *$x \sim y$ si et seulement si $y \sim x$;*

transitive *si $x \sim y$ et $y \sim z$, alors $x \sim z$.*

Définition 1.21.
*Si E est un ensemble et si $\sim$ est une relation d'équivalence sur E, alors nous notons $E/\sim$ l'****ensemble quotient****, c'est-à-dire l'ensemble des classes d'équivalence dans E. Un élément de $E/\sim$ est de la forme*

$$[a] = \{x \in E \text{ tel que } x \sim a\}. \tag{1.5}$$

Lemme 1.22.
Soit un ensemble E et une relation d'équivalence $\sim$. Pour $a, b \in E$, nous avons $[a] = [b]$ si et seulement si $a \sim b$.

Démonstration. En deux parties.

$\Rightarrow$ Nous supposons que $[a] = [b]$. Vu que $a \sim a$, nous avons $a \in [a] = [b]$. Mais $a \in [b]$ signifie $a \sim b$, ce qu'il fallait.

$\Leftarrow$ Nous supposons que $a \sim b$, et nous démontrons que $[a] \subset [b]$ (pour l'inclusion inverse, vous devriez vous en sortir par tout seul). Si $x \in [a]$, alors $x \sim a$. Mais $a \sim b$. Donc $x \sim a \sim b$, ce qui implique $x \sim b$ par transitivité. Or dire $x \sim b$ implique $x \in [b]$.

□

Exemple 1.23.
Sur l'ensemble de tous les polygones du plan, la relation « a le même nombre de côtés » est une relation d'équivalence. Plus précisément, si P et Q sont deux polygones, nous disons que $P \sim Q$ si et seulement si P et Q ont le même nombre de côtés. Cela est une relation d'équivalence :

— un polygone P a toujours le même nombre de côtés que lui-même : $P \sim P$;

— si P a le même nombre de côtés que Q ($P \sim Q$), alors Q a le même nombre de côtés que P ($Q \sim P$) ;

— si P a le même nombre de côtés que Q ($P \sim Q$) et que Q a le même nombre de côtés que R ($Q \sim R$), alors P a le même nombre de côtés que R ($P \sim R$).

△

Exemple 1.24.
Soit f une application entre deux ensembles E et F. Nous définissons une relation d'équivalence sur E par

$$x \sim y \Leftrightarrow f(x) = f(y). \tag{1.6}$$

Nous notons par $\pi\colon E \to E/\sim$ la projection canonique. L'application

$$\begin{aligned} g\colon E/\sim &\to F \\ [x] &\mapsto f(x) \end{aligned} \tag{1.7}$$

est bien définie et injective. Elle n'est pas surjective tant que f ne l'est pas. La **décomposition canonique** de f est

$$f = g \circ \pi. \tag{1.8}$$

△

1.2 Les naturels

Toutes les constructions sont faites dans [8]. Les résultats énoncés ici sont utilisés plus bas et servent de guide à un éventuel contributeur qui voudrait écrire une partie sur la construction de $\mathbb{N}$ et $\mathbb{Z}$. Nous espérons que des preuves se trouvent dans [8]. En tout cas, le lecteur est invité à ne pas les prendre pour évidents.

Nous supposons en particulier que $\mathbb{N}$ est construit avec sa relation d'ordre. Voici quelque affirmations que nous admettons.

Lemme 1.25.
Quelques affirmations sur l'ordre dans $\mathbb{N}$.

(1) Il n'existe pas de $n \in \mathbb{N}$ tel que $n < 0$.

(2) Si $a, b \in \mathbb{N}$ vérifient $a > b$, alors il n'existe pas de x dans $\mathbb{N}$ tel que $a + x = b$.

(3) Toute partie non vide de $\mathbb{N}$ possède un unique élément minimal.

Lemme 1.26.
Toute partie non vide de $\mathbb{N}$ possède un unique minimum.

1.3 Quelques résultats de cardinalité

Je vous conseille fortement de ne pas considérer les résultats qui viennent comme évidents avant d'avoir lu quelques articles de Wikipédia sur la construction des naturels en théorie des ensembles.

1.3.1 Équipotence, surpotence, subpotence

Les notions d'équipotence, surpotence et de subpotence permettent de comparer les « tailles » des ensembles sans avoir besoin de la théorie des ordinaux. Tout ceci ne sera pas très souvent utile par la suite. Un exemple d'utilisation de ces notions est le théorème de Steinitz 6.125 qui démontre l'existence de clôture algébrique pour tout corps.

Définition 1.27 ([9, 10])**.**
Soient deux ensembles A et B.

*(1) Les ensembles A et B sont **équipotents** si il existe une bijection entre A et B. Nous notons $A \approx B$.*

*(2) L'ensemble A est **surpotent** à B si il existe une surjection de A vers B. Nous notons $A \succeq B$.*

*(3) L'ensemble A est **subpotent** à B si il existe une injection de A vers B. Nous notons $A \preceq B$.*

Nous disons également « strictement » surpotent quand il y a surpotence mais pas équipotence, et de même pour la subpotence. Les symboles $\succ$ et $\prec$ sont alors utilisés.

Proposition 1.28 ([1])**.**
L'ensemble A est subpotent à B si et seulement si B est surpotent à A.

Démonstration. En deux parties.

Premier sens Nous supposons que A est subpotent à B. Il existe une injection $\varphi\colon A \to B$. Nous définissons $f\colon B \to A$ par

$$f(x) = \begin{cases} \varphi^{-1}(x) & \text{si } x \in \varphi(A) \\ a & \text{sinon} \end{cases} \tag{1.9}$$

où a est un élément quelconque de A. Cette application est bien définie parce que φ est injective, de telle sorte que φ^{-1} est bien définie. Vu que φ est définie sur tout a, l'application f est une surjection.

L'autre sens Nous supposons que A est surpotent à B. Il existe une surjection $\varphi\colon A \to B$. Pour chaque $y \in B$, nous considérons un élément $a_y \in \varphi^{-1}(y)$. Cela existe parce que φ est surjective [2]. Nous posons alors

$$\begin{aligned} f\colon B &\to A \\ y &\mapsto a_y. \end{aligned} \tag{1.10}$$

Cela est injectif parce que $f(y_1) = f(y_2)$ implique $a_{y_1} = a_{y_2}$, c'est à dire que l'intersection $\varphi^{-1}(y_1) \cap \varphi^{-1}(y_2)$ est non vide. Vu que φ est injective, cela implique que $y_1 = y_2$.

□

Vu que l'ensemble des ensembles n'existe pas [3], nous n'allons pas énoncer le fait que ces notions donnent une relation d'ordre sur les ensembles ; il faudrait parler de classes et nous ne nous en sortirions pas. Nous allons toutefois énoncer quelque résultats qui vont dans ce sens. Pour en savoir plus, vous pouvez lire les différentes pages de Wikipédia sur les nombres cardinaux.

1.3.2 Un peu d'infinité

Définition 1.29.
*Un ensemble est **infini** s'il peut être mis en bijection avec un de ses sous-ensembles propres (c'est-à-dire différent de lui-même).*

Lemme 1.30 ([1] [4])**.**
Infinis et ses parties.

(1) Si I et J sont deux ensembles finis, alors $I \cup J$ est fini.

(2) Si I ou J est infini, alors $I \cup J$ est infini.

Proposition 1.31.
L'ensemble $\mathbb{N}$ est infini [5].

Cette proposition est à peu près prise comme définition d'un ensemble fini dans [11] qui donne également une preuve de l'équivalence avec notre définition. Rien de tout cela n'est évident [6]

Proposition-définition 1.32.
Si I est un ensemble fini, il existe un unique $N \in \mathbb{N}$ tel que I soit en bijection avec $\{0, \dots, N\}$.
*Dans ce cas, le nombre $N + 1$ est le **cardinal** de I.*

Nous ne définissons pas ce qu'est le cardinal d'un ensemble infini ; c'est très compliqué et ça ne nous servira pas.

Définition 1.33.
*Un ensemble est **dénombrable** s'il peut être mis en bijection avec $\mathbb{N}$. Il est **non dénombrable** s'il est infini et ne peut pas être mis en bijection avec $\mathbb{N}$.*

2. Mais ça demande l'axiome du choix et j'avoue être un peu étonné que ce soit indispensable ici.
3. Voir le corolaire 1.56.
4. Écrivez-moi si vous connaissez une preuve de ceci.
5. Définition 1.29.
6. Surtout que je n'ai pas défini la notation $\{0, \dots, N\}$.

Une chose vraiment amusante avec cette définition que l'on met en rapport avec la définition 1.29, c'est qu'un ensemble fini n'est ni dénombrable ni non dénombrable [7].

Lemme 1.34.
Si A est un ensemble finie et si $\sigma\colon A \to B$ est une application quelconque, alors $\sigma(A)$ est fini dans B.

Lemme 1.35.
Toute partie d'une ensemble fini est finie.

Lemme 1.36.
Si A est un ensemble fini ou dénombrable, alors il existe une surjection $\mathbb{N} \to A$.

Lemme 1.37 ([1])**.**
Si A est un ensemble infini et si $f\colon A \to B$ est une application injective, alors $f(A)$ est infini.

Démonstration. Vu que A est infini, il existe A' strictement inclus à A et une bijection $\sigma\colon A' \to A$. Nous allons prouver que la partie $f(A')$ est en bijection avec $f(A)$ tout en étant un sous-ensemble strict de $f(A)$.

Stricte inclusion Si $a \in A\backslash A'$, alors par injectivité de f, nous avons aussi $f(a) \in f(A)\backslash f(A')$. Autrement dit, $f(a)$ est un élément de $f(A)$ qui n'est pas dans $f(A')$.

La candidate bijection Vu que f est une injection nous pouvons considérer f^{-1} sur $f(A)$ est poser

$$\begin{aligned} \varphi\colon f(A') &\to f(A) \\ x &\mapsto f\Big(\sigma\big(f^{-1}(x)\big)\Big). \end{aligned} \tag{1.11}$$

Surjection Une élément de $f(A)$ est de la forme $y = f(a)$ avec $a \in A$. L'application σ est une bijection, donc nous pouvons poser $b = \sigma^{-1}(a)$. Il est alors facile de vérifier que $x = f(b)$ satisfait $\varphi(x) = f(a)$. En effet :

$$\varphi(x) = (\varphi f \sigma^{-1})(a) = (f\sigma f^{-1} f \sigma^{-1})(a) = f(a). \tag{1.12}$$

Cela prouve que φ est surjective.

Injection Soient $a, b \in A'$ tels que $\varphi\big(f(a)\big) = \varphi\big(f(b)\big)$. Nous devons prouver que $f(a) = f(b)$. Nous avons

$$(\varphi f)(a) = (f\sigma f^{-1} f)(a) = (f\sigma)(a). \tag{1.13}$$

Donc l'hypothèse dit que $(f\sigma)(a) = (f\sigma)(b)$. Vu que σ est injective, cela implique que $f(a) = f(b)$.

□

1.3.3 Dénombrabilité et ensemble des naturels

Proposition 1.38 ([1])**.**
Toute partie infinie de $\mathbb{N}$ est dénombrable.

Démonstration. Soit un ensemble infini A dans $\mathbb{N}$. Le lemme 1.26 nous indique que toute partie non vide de $\mathbb{N}$ possède un minimum. Nous en profitons pour définir $\sigma\colon \mathbb{N} \to A$ par récurrence :

$$\begin{cases} \sigma(0) = \min(A) & \text{(1.14a)} \\ \sigma(k+1) = \min\big(A\backslash\sigma\{0,\dots,k\}\big) & \text{(1.14b)} \end{cases}$$

Nous montrons un certain nombre de choses à propos de cette application.

7. Beaucoup de sources disent qu'un ensemble est dénombrable lorsqu'il est en bijection avec une partie de $\mathbb{N}$. Cela laisse la porte ouverte aux ensembles finis. Par exemple Wikipédia[12].

Elle est strictement croissante Vu que $A\backslash\alpha\{0,\dots,k\} \subset A\backslash\{0,\dots,k-1\}$, le minimum est plus grand ou égal : $\sigma(k+1) \geqslant \sigma(k)$. Mais $\sigma(k+1)$ est sélectionné dans l'ensemble $A\backslash\sigma\{0,\dots,k\}$, qui ne contient justement pas $\sigma(k)$. Donc $\sigma(k+1) \neq \sigma(k)$.

Elle est définie sur $\mathbb{N}$ Il faut montrer que pour tout k, l'ensemble $A\backslash\sigma\{0,\dots,k\}$ est non vide. Si il l'était, cela signifierait que $A \subset \sigma\{0,\dots,k\}$. Par le lemme 1.34, la partie $\sigma\{0,\dots,k\}$ est finie dans $\mathbb{N}$. Le lemme 1.35 dit alors qu'en tant que partie de $\sigma\{0,\dots,k\}$, l'ensemble A est fini. Mais comme les hypothèses disent que A est infini, nous avons une contradiction et nous concluons que σ est bien définie sur tout $\mathbb{N}$.

Elle est injective Chaque $\sigma(k)$ est sélectionné dans une partie de A qui exclu tous les $\sigma(i)$ avec $i < k$.

Elle est surjective Soit $a \in A$. Vu que σ est croissante et que $\sigma(0) \geqslant 0$, nous avons $\sigma(a) \geqslant a$. Si $\sigma(a) = a$ nous avons terminé. Supposons $\sigma(a) > a$. Alors

$$\min\big(A\backslash\sigma\{0,\dots,a\}\big) > a. \tag{1.15}$$

Si $\sigma\big(\{0,\dots,a\}\big)$ ne contenait pas a, alors $A\backslash\sigma(\{0,\dots,a\})$ le contiendrait et nous n'aurions pas l'inégalité (1.15). Donc $a \in \sigma\big(\{0,\dots,a\}\big)$ et a est bien dans l'image de σ.

□

1.39.
La proposition 1.38 pourrait être prouvée plus facilement en acceptant le théorème de Cantor-Schröder-Bernstein 1.51. Il existe une injection $A \to \mathbb{N}$ parce que A est une partie de $\mathbb{N}$. Mais vu que A est infini, il possède une partie dénombrable. Cela donne une surjection $A \to \mathbb{N}$ et donc une injection $\mathbb{N} \to A$. Le théorème de Cantor-Schröder-Bernstein conclu.

Cela dit, une telle preuve demanderait des outils plus complexes.

1.40.
La proposition suivante donne une bijection explicite entre $\mathbb{N}$ et $\mathbb{N} \times \mathbb{N}$. Elle n'a rien de transcendante, mais je ne résiste pas à la donner ici parce qu'elle est utilisée dans l'article *Un peu de programmation transfinie* de David Madore[8]. Son utilité est de pouvoir créer un langage de programmation pouvant traiter des paires d'entiers rien qu'en traitant des entiers.

Proposition 1.41 (Une bijection $\mathbb{N} \times \mathbb{N} \to \mathbb{N}$)**.**
La fonction

$$\begin{aligned} f\colon \mathbb{N} \times \mathbb{N} &\to \mathbb{N} \\ (x,y) &\mapsto \begin{cases} y^2 + x & \text{si } x < y \\ x^2 + x + y & \text{si } y \leqslant x. \end{cases} \end{aligned} \tag{1.16}$$

est une bijection.

Démonstration. Il s'agit de prouver qu'elle est injective et surjective. Dans la suite, tous les nombres sont des entiers positifs.

f est injective Pour $k \in \mathbb{N}$ donné, nous allons prouver que

(1) l'équation $f(x,y) = k$ possède au maximum une solution avec $x < y$,

(2) l'équation $f(x,y) = k$ possède au maximum une solution avec $y \leqslant x$,

(3) si $k = y^2 + x$ avec $x < y$ alors il est impossible que $k = x'^2 + x' + y'$ avec $y' \leqslant x'$.

On y va.

8. Et comme j'aime beaucoup cet article, il me fallait une excuse pour le placer ici.
http://www.madore.org/~david/weblog/d.2017-08-18.2460.html.

(1) Nous supposons $y^2 + x = t^2 + z$ avec $x < y$ et $z < t$. Pour fixer les idées, nous supposons $t > y$ et nous posons $t = y + s$ ($s \geqslant 1$). En substituant, et en isolant z,

$$\begin{aligned} z &= x - 2sy - s^2 && (1.17a)\\ &< x - 2sy && (1.17b)\\ &< x - 2sx && (1.17c)\\ &= x(1 - 2s) && (1.17d)\\ &< 0. && (1.17e) \end{aligned}$$

Impossible parce que $z \geqslant 0$.

(2) De même nous supposons $x^2 + x + y = z^2 + z + t$ avec $y \leqslant x$ et $t \leqslant z$. Nous posons $z = x + s$, et nous déballons le même genre de calculs en isolant t.

(3) Enfin nous supposons $y^2 + x = z^2 + z + t$ avec $x < y$ et $t \leqslant z$. Les plus courageux diviseront en trois cas : $y < z$, $y = z$ et $y > z$ et feront les calculs. Par exemple, pour le cas $y > z$ nous posons $y = z + s$ et nous substituons :

$$(y + s)^2 + x = z^2 + z + t \tag{1.18}$$

qui donne

$$x = z + t - 2zs - s^2 < 2z - 2zs - s^2 = 2z(1 - s) - s^2 \leqslant -s < 0 \tag{1.19}$$

parce que $s \geqslant 1$, donc $1 - s \leqslant 0$.

f est surjective Nous devons prouver que tous les éléments de $\mathbb{N}$ sont dans l'image de $\mathbb{N} \times \mathbb{N}$ par f. En premier lieu, $0 = f(0, 0)$. C'est un bon début. Soit $a \in \mathbb{N}$ non nul ; nous montrons que tous les nombres de a^2 à $(a + 1)^2$ sont des images de f. D'abord $a^2 = f(0, a)$, ensuite les nombres

$$f(1, a), f(2, a), \dots, f(a - 1, a) \tag{1.20}$$

prennent les valeurs $a^2 + 1$, ..., $a^2 + a - 1$. Enfin nous avons $f(a, 0) = a^2 + a$ et les nombres $f(a, 1), \dots, f(a, a)$ prennent les valeurs de $a^2 + a + 1$ à $a^2 + 2a = (a + 1)^2 - 1$.

□

Sachez que cette fonction s'étend aux ordinaux (mais là ce n'est plus pour rigoler).

Corolaire 1.42.
Il existe des parties $\{\mathbb{N}_i\}_{i\in\mathbb{N}}$ telles que $\bigcup_{i\in\mathbb{N}} \mathbb{N}_i = \mathbb{N}$ et que chaque $\mathbb{N}_i$ soit en bijection avec $\mathbb{N}$

Démonstration. Nous considérons la bijection $f\colon \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ donnée par (l'inverse de celle donnée) par la proposition 1.41, et nous posons

$$\mathbb{N}_i = f^{-1}(i, \mathbb{N}). \tag{1.21}$$

L'application

$$f\colon \mathbb{N}_i \to \{(i, k)\}_{k\in\mathbb{N}} \tag{1.22}$$

est une bijection. Or l'ensemble $\{(i, k)\}_{k\in\mathbb{N}}$ est évidemment en bijection avec $\mathbb{N}$. Par composition nous avons le résultat. □

Lemme 1.43 ([1]).
Si il existe une surjection $\mathbb{N} \to A$, alors A est fini ou dénombrable.

Démonstration. Pour chaque $a \in A$, l'ensemble $f^{-1}(a)$ est une partie de $\mathbb{N}$.

Une application Le lemme 1.25(3) nous permet de poser

$$\begin{aligned} \sigma\colon A &\to \mathbb{N}\\ a &\mapsto \min\big(f^{-1}(a)\big). \end{aligned} \tag{1.23}$$

σ est injective Supposons que $\sigma(a) = \sigma(b)$. Nous appelons x ce nombre :

$$x = \min\big(f^{-1}(a)\big) = \min\big(f^{-1}(b)\big). \tag{1.24}$$

Nous avons $x \in f^{-1}(a) \cap f^{-1}(b)$, ce qui implique que $f(x) = a$ et que $f(x) = b$; donc $a = b$. Donc σ est une injection.

A est infini Si A est fini, le lemme est prouvé. Donc à partir de maintenant nous supposons que A est infini. Le but est de prouver qu'il est dénombrable, c'est à dire de construire une bijection $A \to \mathbb{N}$.

$\sigma(A)$ est dénombrable Vu que $\sigma\colon A \to \mathbb{N}$ est injective et que A est infini, le lemme 1.37 dit que $\sigma(A)$ est infini dans $\mathbb{N}$. La proposition 1.38 nous dit alors que $\sigma(A)$ est dénombrable.

Soit une bijection $\varphi\colon \sigma(A) \to \mathbb{N}$.

La candidate bijection Nous posons

$$f = \varphi \circ \sigma\colon A \to \mathbb{N} \tag{1.25}$$

et nous allons prouver que c'est une bijection.

Injective Vu que φ et σ sont injectives, l'égalité $(\varphi\sigma)(a) = (\varphi\sigma)(b)$ implique immédiatement $a = b$.

Surjective Soit $k \in \mathbb{N}$. Vu que φ et σ sont des injections, nous pouvons poser $a = (\sigma^{-1}\varphi^{-1})(k)$. Il est alors immédiat que $f(a) = k$.

□

Proposition 1.44 ([1, 12])**.**
Une union dénombrable d'ensembles finis ou dénombrables est finie ou dénombrable.

Démonstration. Soient A_i des ensembles finis ou dénombrables. Nous posons $A = \bigcup_{i\in\mathbb{N}} A_i$, et nous considérons les parties $\mathbb{N}_i$ du corolaire 1.42. Vu que A_i est dénombrable ou fini et que $\mathbb{N}_i$ est dénombrable, il existe une surjection $\varphi_i\colon \mathbb{N}_i \to A_i$.

Nous définissons $s\colon \mathbb{N} \to \mathbb{N}$ par $n \in \mathbb{N}_{s(n)}$, et nous posons enfin

$$\begin{aligned} \varphi\colon \mathbb{N} &\to A \\ n &\mapsto \varphi_{s(n)}(n). \end{aligned} \tag{1.26}$$

Nous prouvons que φ est surjective.

Soit $a \in A_i$. Il existe $n \in \mathbb{N}_i$ tel que $a = \varphi_i(n)$. Mais comme $n \in \mathbb{N}_i$, nous avons $s(n) = i$. Donc

$$a = \varphi_i(n) = \varphi_{s(n)}(n) = \varphi(n). \tag{1.27}$$

Donc $\varphi\colon \mathbb{N} \to A$ est surjective.

Le lemme 1.43 conclu que A est fini ou dénombrable. □

Lemme 1.45 ([1])**.**
Si N est un ensemble dénombrable, alors il existe une bijection $g\colon \{1,2\} \times N \to N$.

Démonstration. D'abord nous définissons une bijection $\varphi\colon \{0,1\} \times \mathbb{N} \to \mathbb{N}$ par

$$\begin{aligned} \varphi\colon \{0,1\} \times \mathbb{N} &\to \mathbb{N} \\ (n,k) &\mapsto 2k + n. \end{aligned} \tag{1.28}$$

Ensuite si $f\colon \mathbb{N} \to N$ est une bijection, il suffit de poser $g(n,k) = f\big(\varphi(n,k)\big)$. □

Proposition 1.46 ([12])**.**
Si N est un ensemble dénombrable, alors pour tout $n \in \mathbb{N}$, l'ensemble N^n est dénombrable.

Les ensembles dénombrables sont les plus petits ensembles infinis possibles, comme en témoigne la proposition suivante.

Proposition 1.47.
Tout ensemble infini contient une partie en bijection avec $\mathbb{N}$.

Démonstration. Soient un ensemble infini E_0 et une partie propre E_1 en bijection avec E_0. Nous notons $\varphi\colon E_0 \to E_1$ une bijection.

Soit $x_0 \in E_0 \backslash E_1$ (axiome du choix et tout ça). Nous définissons

$$\begin{aligned} \psi\colon \mathbb{N} &\to \{\varphi^k(x_0)\} \\ n &\mapsto \varphi^n(x_0) \end{aligned} \tag{1.29}$$

et nous allons prouver que c'est une bijection. Que ce soit surjectif est immédiat. Pour l'injectivité, soit $\varphi^k(x_0) = \varphi^l(x_0)$ avec $k \neq l$. Supposons pour fixer les notations que $k > l$. Alors, vu que φ est inversible nous pouvons écrire

$$x_0 = \varphi^{k-l}(x_0) = \varphi\big(\varphi^{k-l-1}(x_0)\big) \tag{1.30}$$

où il est entendu que $\varphi^0(x_0) = x_0$. Cela signifie que x_0 est dans l'image de φ, c'est-à-dire dans E_1, ce que nous avons exclu par choix de x_0 dans $E_0 \backslash E_1$. Donc en réalité $\varphi^k(x_0) \neq \varphi^l(x_0)$ dès que $k \neq l$. □

Proposition 1.48.
Toute partie d'un ensemble fini est finie, et toute partie d'un ensemble dénombrable est finie ou dénombrable.

Lemme 1.49.
Soit un ensemble E non dénombrable ainsi qu'une application $f\colon E \to F$ où F est un ensemble quelconque. Si $f(E)$ est dénombrable (ou fini), alors il existe $y \in f(E)$ tel que $f^{-1}(y)$ est indénombrable.

Démonstration. Nous avons

$$E = \bigcup_{y \in F} f^{-1}(y). \tag{1.31}$$

Si tous les $f^{-1}(y)$ sont dénombrables, alors E est une union dénombrable (F est dénombrable) d'ensembles dénombrables. Il serait donc dénombrable (proposition 1.44), ce qui est contraire à l'hypothèse. □

1.3.4 Théorème de Cantor-Schröder-Bernstein

Lemme 1.50 ([13])**.**
Soient un ensemble A et une partie B de A. Si il existe une injection $f\colon A \to B$ alors il existe une bijection $\alpha\colon A \to B$.

Démonstration. Nous posons $Y = A \backslash B$ et nous décomposons la preuve en étapes.

Les $f^k(Y)$ sont disjoints Vu que f prend ses valeurs dans B, nous avons $f^k(Y) \subset B$ pour tout k. Et vu que $Y = A \backslash B$, nous avons

$$f^k(Y) \cap Y = \varnothing \tag{1.32}$$

pour tout k. L'application f étant injective, elle vérifie $f(C \cap D) = f(C) \cap f(D)$. Nous appliquons f^m des deux côtés de (1.32) :

$$f^{k+m}(Y) \cap f^m(Y) = \varnothing \tag{1.33}$$

pour tout $k, m \in \mathbb{N}$.

Une décomposition Vu que $f(X) \subset B$ nous avons l'égalité

$$B = f(X) \cup \big(B \backslash f(X)\big) \tag{1.34}$$

$A\backslash X = B\backslash f(X)$ Supposons $x \in A\backslash X$. Vu que $A\backslash B$ est dans X, l'élément x n'est pas dans $A\backslash B$ et donc est dans B parce qu'il est dans A. Mais x n'est pas dans X et en particulier pas dans $f(X)$ parce que $f(X) \subset X$. Donc x est dans $B\backslash f(X)$.

Dans l'autre sens, nous supposons que $x \in B\backslash f(X)$. Vu que $B \subset A$ nous avons $x \in A$. Comme x est hors de $f(X)$, il est hors des $f^k(Y)$ pour $k \geqslant 1$. Mais $x \in B$, donc x est hors de $A\backslash B = f^0(Y)$. Donc x est hors de $k^k(Y)$ pour tout $k \geqslant 0$. Donc x est hors de X.

La bijection Nous considérons l'application

$$\begin{aligned} \alpha\colon A &\to B \\ x &\mapsto \begin{cases} f(x) & \text{si } x \in X \\ x & \text{si } x \in A\backslash X. \end{cases} \end{aligned} \tag{1.35}$$

Nous démontrons dans les points suivants que α est bijective.

Injective Nous supposons $\alpha(x) = \alpha(y)$. Il y a 4 possibilités suivant que x et y soient dans X ou $A\backslash X$.

Si $x, y \in X$ alors $f(x) = f(y)$ et donc $x = y$ parce que f est injective.

Si $x \in X$ et $y \in A\backslash X$, alors $f(x) = y$. Mais $f(x) \in f(X)$ et $y \in A\backslash X = B\backslash f(X)$. Donc l'élément $f(x) = y$ est dans $f(X) \cap \big(B\backslash f(X)\big) = \emptyset$. Il n'est donc pas possible d'avoir $\alpha(x) = \alpha(y)$ avec $x \in X$ et $y \in A\backslash X$.

Si $x \in A\backslash X$ et $y \in X$, c'est la même chose.

Si $x, y \in A\backslash X$, alors $x = \alpha(x) = \alpha(y) = y$.

Injective Soit $y \in B$. Il y a deux possibilités : $y \in x$ et $y \in A\backslash X$. La première se divise en deux : $y \in Y$ et $y \in \bigcup_{k=1}^{\infty} f^k(Y)$. On y va.

$y \in Y$ Ce cas n'est pas possible parce que $y \in B$ alors que $Y = A\backslash B$.

$y \in f^k(Y)$ **avec** $k \geqslant 1$ Nous avons

$$y \in f\big(f^{k-1}(Y)\big) \subset f(X) \subset \alpha(A). \tag{1.36}$$

$y \in A\backslash X$ Alors $y = \alpha(y)$.

□

Théorème 1.51 (Cantor-Schröder-Bernstein)**.**
Soient deux ensembles A et B pour lesquels il existe des injections $f\colon A \to B$ et $g\colon B \to A$. Alors il existe une bijection entre A et B.

Démonstration. La composée $g \circ f\colon A \to A$ est injective et prend ses valeurs dans $g\big(f(A)\big) \subset g(B) \subset A$. Bref, l'application $g \circ f\colon A \to g(B)$ est injective. Le lemme 1.50 donne alors une bijection $\varphi\colon A \to g(B)$.

Nous montrons que $g^{-1} \circ \varphi\colon A \to B$ est une bijection.

Injective Nous supposons $x, y \in A$ tels que

$$g^{-1}\big(\varphi(x)\big) = g^{-1}\big(\varphi(y)\big). \tag{1.37}$$

Nous appliquons g des deux côtés : $\varphi(x) = \varphi(y)$. Vu que φ est une bijection, cela entraîne $x = y$.

Surjective Soit $b \in B$. En posant $a = \varphi^{-1}\big(g(b)\big)$ nous avons bien $(g^{-1} \circ \varphi)(a) = b$.

□

1.3.5 Comparabilité cardinale

Le théorème de comparabilité cardinale énonce que si A et B sont des ensemble, alors nous avons toujours $A \geq B$ ou $A \leq B$ (ou les deux ; dans ce cas $A \approx B$ par Cantor-Schröder-Bernstein).

Théorème 1.52 (Théorème de comparabilité cardinale[1, 14, 15])**.**
Entre deux ensembles, il existe forcément une injection de l'un dans l'autre.

Démonstration. Nous allons montrer que le graphe d'une injection de A dans B ou de B dans A est donné par un élément maximal (au sens de l'inclusion) de l'ensemble (inductif) des graphes d'injections d'une partie de A dans une partie de B.

Nous allons utiliser le lemme de Zorn 1.15 à l'ensemble [9]

$$\mathcal{A} = \left\{ (X, Y, \varphi) \text{ tel que } \begin{cases} X \subset A \\ Y \subset B \\ \varphi \colon X \to Y \text{ est injective.} \end{cases} \right\} \tag{1.38}$$

que nous ordonnons par, l'inclusion, c'est à dire par $(X_1, Y_1, \varphi_1) < (X_2, Y_2, \varphi_2)$ lorsque $X_1 \subset X_2$, $Y_1 \subset Y_2$ et $\varphi_2|_{X_1} = \varphi_1$.

Nous passons rapidement sur le fait que cet ensemble est inductif, et nous considérons tout de suite un élément maximal (X, Y, φ). Il y a deux possibilités : soit $\varphi(X) = B$, soit $\varphi(X) \neq B$.

Si $\varphi(X) = B$ Dans ce cas, $\varphi \colon X \to B$ est une surjection. L'ensemble A est donc surpotent à B. La proposition 1.28 conclu que B est subpotent à A.

Si $\varphi(X) \neq B$ Nous subdivisons en deux nouveaux cas : soi $X = A$, soit $X \neq A$.

Si $X = A$ Alors nous avons une injection $\varphi \colon A \to B$, et c'est bon.

Si $X \neq A$ Nous sommes dans le cas $X \neq A$ et $\varphi(X) \neq B$. Soient $a \in A \backslash X$ et $b \in B \backslash \varphi(X)$. Nous considérons l'application

$$\begin{aligned} \psi \colon X \cup \{a\} &\to Y \cup \{b\} \\ x &\mapsto \begin{cases} \varphi(x) & \text{si } x \in X \\ b & \text{si } x = a. \end{cases} \end{aligned} \tag{1.39}$$

Cela est une application injective. Donc le triple $(X \cup \{a\}, Y \cup \{b\}, \psi)$ majore (X, Y, φ). Nous avons une contradiction et ce cas n'est pas possible.

□

1.53.
Le théorème de comparabilité cardinale couplé au théorème de Cantor-Schröder-Bernstein nous indique que pour tout ensembles A et B, nous avons soit $A \leq B$, soit $B \leq A$. Et si $A \leq B \leq A$, alors $A \approx B$.

Nous ne sommes pas loin de dire que la relation $\leq$ donne un ordre total sur l'ensemble des ensembles. C'est très beau sauf que l'ensemble des ensembles n'existe pas [10]. Il faudrait parler de *classe* des ensembles, mais ça nous mènerait trop loin. Toujours est-il que ces deux théorèmes montrent qu'on n'est pas loin d'avoir un ordre sur les ensembles, et que cela est une des bases possibles pour développer les nombres cardinaux.

1.3.6 Théorème de Cantor, ensemble des ensembles

Théorème 1.54 (Cantor[16])**.**
Un ensemble est toujours strictement subpotent à son ensemble des parties.

9. Attention : dans le Frido, la notation $f \colon A \to B$, signifie que f est définie sur tout l'ensemble A, mais pas qu'elle est surjective sur B.
10. Corolaire 1.56.

Démonstration. Soit un ensemble E et son ensemble des parties $\mathcal{P}(E)$. Nous commençons par prouver qu'il n'existe pas de surjection $E \to \mathcal{P}(E)$. Soit en effet une application $f\colon E \to \mathcal{P}(E)$. Nous posons

$$D = \{x \in E \text{ tel que } x \notin f(x)\}. \tag{1.40}$$

Nous prouvons que D ne peut pas être dans l'image de f. Supposons que $y \in E$ soit tel que $f(y) = D$.

Si $y \in D$ Alors par définition de D, nous avons $y \notin f(y) = D$. Contradiction.

Si $y \notin D$ Alors $y \in f(y) = D$, contradiction.

Donc aucune surjection $f\colon E \to \mathcal{P}(E)$ n'existe. En particulier par de bijections.

Par ailleurs, l'application $g\colon \mathcal{P}(E) \to E$ qui fait $g(\{a\}) = a$ (et n'importe quoi d'autre sur les autres éléments de $\mathcal{P}(E)$) est une surjection $\mathcal{P}(E) \to E$.

Donc $\mathcal{P}(E)$ est toujours strictement surpotent à E. □

1.55.
Le théorème de Cantor implique en particulier qu'il existe (au moins) une infinité dénombrable d'ensemble infinis de cardinalité différentes (plus évidemment une infinité dénombrable d'ensembles finis de cardinalité différentes).

Pour tout ensemble A, il est donc possible de dire « soit E, un ensemble strictement surpotent à A ».

Corolaire 1.56.
Il n'existe pas d'ensemble contenant tous les ensembles.

Démonstration. Si E était un tel ensemble, nous aurions $\mathcal{P}(E) \subset E$ parce que les éléments de $\mathcal{P}(E)$ sont des ensembles. Or cela donnerait une surjection $E \to \mathcal{P}(E)$ alors que cela est impossible par le théorème de Cantor 1.54. □

1.3.7 Ajouter ou soustraire des cardinalités

Nous allons prouver une série de résultats que nous pourrions résumer en « ajouter ou retrancher des parties de cardinalité plus petite ne change pas la cardinalité ».

Lemme 1.57 ([1])**.**
Si A est infini et B est fini, alors $A \cup B \approx A$.

Démonstration. Nous supposons que A et B sont disjoints[11]. La proposition 1.32 nous permet de considérer une bijection $\psi\colon \{1, \dots, n\} \to B$.

Vu que A est infini, la proposition 1.47 nous permet de considérer $N \subset A$ et une bijection $\varphi\colon \mathbb{N} \to N$.

Maintenant, il s'agit seulement d'insérer B dans A en le mettant « au début » de N et en décalant les autres. La bijection est

$$\begin{aligned} f\colon A \cup B &\to A \\ x &\mapsto \begin{cases} x & \text{si } x \in A \backslash N \\ \varphi\big(\varphi^{-1}(x) + n\big) & \text{si } x \in N \\ \varphi\big(\psi^{-1}(x)\big) & \text{si } x \in B. \end{cases} \end{aligned} \tag{1.41}$$

□

Lemme 1.58.
Si A est infini et si A est surpotent à B, alors $A \approx A \cup B$.

11. Adaptez la démonstration au cas où l'intersection n'est pas vide.

Démonstration. Il existe évidemment une injection $A \to A \cup B$. Donc le théorème de Cantor-Schröder-Bernstein 1.51 nous indique que trouver une injection $A \cup B \to A$ suffira pour la peine.

Nous allons utiliser le lemme de Zorn 1.15 avec l'ensemble

$$\mathcal{A} = \left\{ (X, \varphi_X) \text{ tel que } \begin{cases} X \subset B \\ \varphi_X \colon A \cup X \to A \text{ est injective.} \end{cases} \right\} \tag{1.42}$$

muni de l'ordre de l'inclusion : $(X, \varphi_X) < (Y, \varphi_Y)$ si $X \subset Y$ et $\varphi_Y(x) = \varphi_X(x)$ pour tout $x \in A \cup X$.

$\mathcal{A}$ est inductif Soit une famille $\mathcal{F} = \{(X_i, \varphi_i)\}_{i \in I}$ complètement ordonnée indexée par l'ensemble I. En posant $X = \bigcup_{i \in I} X_i$ et $\varphi(x) = \varphi_i(x)$ dès que $x \in A \cup X_i$, l'élément (X, φ) majore $\mathcal{F}$.

Un maximum Le lemme de Zorn nous assure que $\mathcal{A}$ possède (au moins) un élément maximum. Soit un tel élément maximum (X, φ).

$X \approx B$ Ah oui, vous auriez aimé avoir $X = B$. Mais non ; il n'y a pas de garanties. Nous allons montrer que $X \approx B$, et ça suffira.

Vu que $X \subset B$, si X n'est pas équipotent à B, il est strictement inclus à B. Nous pouvons donc considérer

$$b \in B \backslash X \tag{1.43a}$$
$$a \in A \backslash \varphi(X). \tag{1.43b}$$

Nous considérons alors l'élément $(Y, \psi) \in \mathcal{A}$ défini par

$$Y = X \cup \{b\} \tag{1.44a}$$
$$\psi(x) = \begin{cases} a & \text{si } x = b \\ \varphi(x) & \text{sinon .} \end{cases} \tag{1.44b}$$

Cet élément majore (X, φ).

Donc $X \approx B$.

Résumé de la situation Nous avons $A \approx A \cup X$ ainsi que une injection $\varphi \colon A \cup X \to A$ et une bijection $\psi \colon B \to X$.

Conclusion si A est disjoint de B Si A et B sont disjoints, nous avons une bijection

$$\begin{aligned} l \colon A \cup B &\to A \\ x &\mapsto \begin{cases} \varphi(x) & \text{si } x \in A \\ \varphi\big(\psi(x)\big) & \text{si } x \in B. \end{cases} \end{aligned} \tag{1.45}$$

Conclusion si A n'est pas disjoint de B Il suffit de poser $C = B \backslash A$ et avons

$$A \cup B = [A \cup (A \cap B)] \cup C. \tag{1.46}$$

Cette union est disjointe, $A \cup (A \cap B)$ est surpotent à A et C est subpotent à B. La conclusion est donc encore valable.

□

La proposition suivante sera utilisée en théorie de la mesure, dans l'exemple 13.61. Ça utilise l'axiome du choix sous la forme du lemme de Zorn.

Proposition 1.59 ([17]).
Si S est un ensemble infini alors il existe une bijection $\varphi \colon \{0, 1\} \times S \to S$.

Démonstration. En plusieurs points.

Un gros ensemble Nous considérons l'ensemble suivant :

$$\mathcal{A} = \left\{ (X, \varphi) \text{ tel que } \begin{cases} X \subset S \text{ est infini} \\ \varphi\colon \{0,1\} \times X \to X \text{ est injective} \end{cases} \right\} \tag{1.47}$$

Nous ordonnons par l'inclusion, c'est à dire que nous disons que $(X_1, \varphi_1) < (X_2, \varphi_2)$ lorsque $X_1 \subset X_2$ et $\varphi_1(k, x) = \varphi_2(k, x)$ pour tout $(k, x) \in \{0, 1\} \times X_1$.

Il est inductif Nous prouvons que $\mathcal{A}$ est inductif[12]. Soit une partie totalement ordonnée $\mathcal{F} = \{(X_i, \varphi_i)\}_{i \in I}$ indexé par un ensemble I.

Commençons par remarquer que si $x \in X_i \cap X_j$ et si $k \in \{0, 1\}$, alors $\varphi_i(k, x) = \varphi_j(k, x)$. En effet, vu que $\mathcal{F}$ est totalement ordonné, nous pouvons supposer $(X_i, \varphi_i) < (X_j, \varphi_j)$ (sinon c'est le contraire) et nous avons alors $X_i \subset X_j$ et $\varphi_j(k, x) = \varphi_i(k, x)$.

Maintenant nous prouvons que $\mathcal{F}$ accepte un majorant dans $\mathcal{A}$. Il s'agit de construire une paire (X, φ). Nous posons d'abord

$$X = \bigcup_{i \in I} X_i. \tag{1.48}$$

D'autre part, si $x \in X$ et si $k \in \{0, 1\}$, l'ensemble

$$\{\varphi_i(k, x) \text{ tel que } i \in I, x \in X_i\} \tag{1.49}$$

contient un seul élément. Nous nommons $\varphi(k, x)$ cet élément. Cela nous définit une application $\varphi\colon \{0, 1\} \times X \to X$ qui prolonge tous les φ_i.

Donc nous avons $(X, \varphi) > (X_i, \varphi_i)$ pour tout $i \in I$.

Lemme de Zorn Vu que $\mathcal{A}$ est un ensemble inductif, nous pouvons utiliser le lemme de Zorn 1.15 et considérer un élément maximum, c'est à dire une paire (Y, φ) où $Y \subset S$ et $\varphi\colon \{0, 1\} \times Y \to Y$ est injective.

Il y a deux cas : soit Y est équipotent à S, soit il ne l'est pas.

Si Y est équipotent à S Nous considérons une bijection $\psi\colon S \to Y$ et nous montrons que l'application

$$\begin{aligned} f\colon \{0,1\} \times S &\to S \\ (n, a) &\mapsto \psi^{-1}\Big(\varphi\big((n, \psi(a))\big)\Big). \end{aligned} \tag{1.50}$$

est une injection. Il s'agit seulement d'une vérification basée sur le fait que ψ^{-1}, φ et ψ le sont.

Une injection dans l'autre sens n'est pas compliquée à construire ; par exemple $s \mapsto (1, s)$. Ayant maintenant une injection dans les deux sens, le théorème de Cantor-Bernstein 1.51 dit qu'il existe une bijection.

Si Y n'est pas équipotent à S Nous montrons que ce cas n'est pas possible. Notons que Y est certainement subpotent à S parce que $Y \subset S$.

$S \backslash Y$ est inifni Supposons le contraire. Si $S \backslash Y$ est fini, alors $Y \cup (S \backslash Y)$ est équipotent à Y par le lemme 1.57. Oui mais $Y \cup (S \backslash Y) = S$, et nous avons dit que Y n'est pas équipotent à S.

Bref, oui : $S \backslash Y$ est infini.

La contradiction La proposition 1.47 nous donne une partie dénombrable[13] $N \subset S \backslash Y$. Nous allons construire une injection $\psi\colon \{0, 1\} \times (Y \cup N) \to Y \cup N$ qui prolonge φ, contredisant ainsi la maximalité de (Y, φ).

Nous considérons une bijection $g\colon \{0, 1\} \times N \to N$ (lemme 1.45). Vu que N est dans $S \backslash Y$, les parties Y et N sont disjointes et nous pouvons écrire

$$\{0, 1\} \times (Y \cup N) = \big(\{0, 1\} \times Y\big) \cup \big(\{0, 1\} \times N\big) \tag{1.51}$$

12. Définition 1.14.
13. Et donc infinie par la proposition 1.31 que nous n'allons plus citer à chaque fois.

où l'union à droite est disjointe.

Nous pouvons donc définir ψ correctement de la façon suivante :

$$\begin{aligned}\psi\colon \{0,1\}\times(Y\cup N)&\to Y\cup N\\ (n,x)&\mapsto\begin{cases}\varphi(n,x) & \text{si } x\in Y\\ g(n,x) & \text{si } x\in N.\end{cases}\end{aligned}\tag{1.52}$$

Voila qui contredit la maximalité de (Y,φ).

□

Corolaire 1.60.
Si A est un ensemble infini, alors A possède deux sous-ensembles disjoints A_1 et A_2 qui sont tout deux en bijection avec A.

Démonstration. La proposition 1.59 donne une bijection $\varphi\colon\{1,2\}\times A\to A$. Il suffit de poser $A_1=\varphi(1,A)$ et $A_2=\varphi(2,A)$. □

Maintenant que nous pouvons mettre dans A deux copies disjointes de A, il n'est pas très étonnant que nous puissions en mettre une infinité dénombrable. C'est en substance ce que signifie la proposition suivante.

Proposition 1.61.
Si A est infini, alors $A\times\mathbb{N}\approx A$.

Démonstration. La démonstration se base sur le fait qu'à l'intérieur de A, nous pouvons construire autant de copies de A deux à deux disjointes que nous le voulons. La k^{e} « copie » sera naturellement l'image de $k\times A$.

Voyons tout cela en détail.

Ce que nous allons faire Nous allons construire, pour tout $i\geqslant 1$ des parties $A_i,B_i\subset A$ telles que

- $A_i\cap B_i=\varnothing$,
- $A_i,B_i\subset B_{i-1}$,
- $A_i\approx B_i\approx A$,
- $A_i\cap A_j=\varnothing$ si $i\neq j$

La construction Nous commençons à zéro en utilisant le corollaire 1.60 pour construire des parties disjointes A_0 et B_0 de A telles que $A_0\approx B_0\approx A$.

Ensuite, vu que $B_0\approx A$, il existe A_1 et B_1 dans B_0 tels que $A_1\cap B_1=\varnothing$ et $A_1\approx B_1\approx B_0\approx A$. Cela est notre construction pour $i=1$.

Pour la récurrence, vu que $A_i\approx B_i\approx A$, nous considérons A_{i+1} et B_{i+1} dans B_i tels que $A_{i+1}\cap B_{i+1}=\varnothing$ et $A_{i+1}\approx B_{i+1}\approx B_i\approx A$. C'est encore le corolaire 1.60 qui fait le travail.

Les propriétés Nous avons $A_i\cap A_{i+1}=\varnothing$ parce que $A_i\cap A_{i+1}\subset A_i\cap B_i=\varnothing$.

Nous devons encore montrer que $A_i\cap A_j=\varnothing$ dès que $i\neq j$. Supposons que $j>i$. Nous avons les inclusions

$$A_j\subset B_{j-1}\subset B_{j-2}\subset\ldots\subset B_i.\tag{1.53}$$

Donc $A_j\cap A_i\subset B_i\cap A_i=\varnothing$.

Une injection Nous pouvons à présent écrire une injection qui termine presque la preuve. Pour cela nous considérons pour tout i, une bijection $\psi_i\colon A\to A_i$. Ensuite nous posons

$$\begin{aligned}\varphi\colon A\times\mathbb{N}&\to A\\ (a,k)&\mapsto\psi_k(a).\end{aligned}\tag{1.54}$$

Si $\varphi(a,k)=\varphi(b,l)$, alors $\psi_k(a)=\psi_l(b)$. L'élément $\psi_k(a)$ est donc dans $A_k\cap A_l$; ce n'est possible que si $k=l$. Donc $\psi_l(a)=\psi_l(b)$. Cette dernière égalité n'est possible que si $a=b$ parce que ψ_l est une bijection.

Donc φ est une injection, et nous avons prouvé que $A\times\mathbb{N}\preceq A$.

La bijection Nous venons de prouver que $A \times \mathbb{N} \preceq A$. La surpotence $A \times \mathbb{N} \succeq A$ étant évidente, le théorème de Cantor-Schröder-Bernstein 1.51 conclu que $A \times \mathbb{N} \approx A$.

□

Lemme 1.62 ([1]).
Sois un ensemble A muni de deux sous-ensembles B et B' équipotents et disjoints. Alors $A\backslash B$ est équipotent à $A\backslash B'$.

Démonstration. Soit une bijection $\psi\colon B' \to B$. L'application

$$\begin{aligned} \varphi\colon A\backslash B &\to A\backslash B' \\ x &\mapsto \begin{cases} x & \text{si } x \notin B' \\ \psi(x) & \text{si } x \in B'. \end{cases} \end{aligned} \tag{1.55}$$

fait la bijection. □

Lemme 1.63 ([18]).
Si A est un ensemble infini et si $B \prec A$, alors $A \approx A\backslash B$.

Démonstration. Nous pouvons écrire

$$A = (A\backslash B) \cup B. \tag{1.56}$$

Le théorème de comparabilité cardinale 1.52 nous indique que soit $A\backslash B \preceq B$, soit $A\backslash B \succeq B$. Nous allons voir les deux cas.

Si $A\backslash B \succeq B$ Dans ce cas, $(A\backslash B) \cup B \approx A\backslash B$ par le lemme 1.58. Dans ce cas, notre résultat est prouvé parce que $A = (A\backslash B) \cup B \approx A\backslash B$.

Si $A\backslash B \preceq B$ Dans ce cas, le lemme 1.58 nous indique que $A = (A\backslash B) \cup B \approx B$. Mais $A \approx B$ est exclu par l'hypothèse. Ce cas est donc impossible.

□

Lemme 1.64 ([1]).
Si A est infini et si B est une partie strictement subpotente de A, alors il existe $U \subset A$ disjoint de B et équipotent à B.

Démonstration. Le lemme 1.63 nous donne une bijection $\varphi\colon A \to A\backslash B$. Il suffit alors de poser $U = \varphi(B)$. Cette partie est disjointe de B parce que φ prend ses valeurs dans $A\backslash B$. □

Lemme 1.65 ([19]).
Soit un ensemble infini A ainsi qu'un sous-ensemble $B \subset A$. Nous supposons l'existence d'une fonction surjective $f\colon B \to B \times B$.

Alors $B \preceq B \times B \preceq B \preceq A$.

Démonstration. La première est l'hypothèse sur f. La seconde est l'existence (évidente) d'une surjection $B \times B \to B$. La troisième est le fait que B soit inclus à A. □

Lemme 1.66 ([19]).
Soit un ensemble infini A ainsi qu'un sous-ensemble strictement subpotent $B \subset A$. Nous supposons l'existence d'une fonction surjective $f\colon B \to B \times B$.

Alors f peut être étendue en une injection $f\colon D \to D \times D$ où $D \subset A$ contient strictement B.

Démonstration. Par le lemme 1.64, nous considérons une partie $U \subset A$ disjointe de B et équipotent à B. Nous pouvons faire le développement

$$(B \cup U) \times (B \cup U) = (B \times B) \cup (B \times U) \cup (U \times B) \cup (U \times U). \tag{1.57}$$

Nous savons que B est surpotent à U (il est même équipotent) ; donc le lemme 1.58 nous dit que $B \cup U \approx B$. De plus il existe une bijection $B \to U$, donc

$$B \approx B \times B \approx B \times U \approx U \times B \approx U \times U. \tag{1.58}$$

Et enfin, la réunion d'ensembles équipotents est équipotent. Bref, nous avons une bijection

$$\varphi\colon U \to (U \times B) \cup (B \times U) \cup (U \times U). \tag{1.59}$$

Et enfin nous définissons

$$\begin{aligned} g\colon B \cup U &\to (B \times U) \times (B \cup U) \\ x &\mapsto \begin{cases} f(x) & \text{si } x \in B \\ \varphi(x) & \text{si } x \in U. \end{cases} \end{aligned} \tag{1.60}$$

Cette définition est bonne parce que U et B sont disjoints, et g est injective. □

Le théorème suivant est une généralisation de la proposition 1.59. Elle implique, entre autres choses, qu'il existe une bijection entre $\mathbb{R}$ et $\mathbb{R} \times \mathbb{R}$. Pour le cas de $\mathbb{N} \times \mathbb{N} \approx \mathbb{N}$, il y a la proposition 1.41 qui donne une bijection explicite et donc sans axiome du choix et sans lemme de Zorn.

Théorème 1.67.
Si A est infini, alors $A \approx A \times A$.

Démonstration. Une fois de plus, ce sera le lemme de Zorn qui va s'y coller. Soit l'ensemble

$$\mathcal{A} = \left\{ (X, \varphi) \text{ tel que } \begin{cases} X \subset A \\ \varphi\colon X \to X \times X \text{ est surjective.} \end{cases} \right\} \tag{1.61}$$

Cet ensemble est non vide parce que A est infini ; il contient donc une partie dénombrables N, et nous connaissons la surjection $\mathbb{N} \to \mathbb{N} \times \mathbb{N}$ du lemme 1.41.

Nous ordonnons $\mathcal{A}$ par l'inclusion : $(X, \varphi) \leqslant (Y, \phi)$ lorsque $X \subset Y$ et $\phi|_X = \varphi$. La tambouille usuelle montre que $\mathcal{A}$ est un ensemble inductif et le lemme de Zorn 1.15 donne l'existence d'un élément maximum que nous notons (B, φ).

Vu que B est subpotent à A (parce qu'il est inclus), soit il est strictement subpotent soit il est équipotent. Nous commençons par montrer que B ne peut pas être strictement subpotent à A.

En effet, si nous avions une surjection $B \to B \times B$, alors que B est strictement subpotent à A. Le lemme 1.66 nous dit alors que φ peut être étendue, ce qui contredirait la maximalité de (B, φ).

Donc la partie B est équipotente à A : il existe une bijection $g\colon A \to B$. Mais nous avons une surjection $B \to B \times B$ et donc aussi une injection $B \times B \to B$. Vu que nous avons par ailleurs une injection $B \to B \times B$, le théorème de Cantor-Schröder-Bernstein 1.51 nous donne une bijection $\phi\colon B \times B \to B$. Avec ça, l'application

$$\begin{aligned} f\colon A \times A &\to A \\ (a, b) &\mapsto \phi\big(g(a), g(b)\big) \end{aligned} \tag{1.62}$$

est une bijection. Donc les ensembles A et $A \times A$ sont équipotents. □

Lemme 1.68.
Si A est infini, et si pour tout $i \in \mathbb{N}$ nous avons $A_i \approx A$, alors

$$\bigcup_{i \in \mathbb{N}} A_i \approx A. \tag{1.63}$$

Démonstration. Pour chaque $i \in \mathbb{N}$ nous avons une bijection $\varphi_i\colon A_i \to A$. Nous posons

$$\begin{aligned} \varphi\colon A \times \mathbb{N} &\to \bigcup_{i=0}^{\infty} A_i \\ (a, i) &\mapsto \varphi_i(a). \end{aligned} \tag{1.64}$$

Cette application est surjective mais peut-être pas injective parce que les A_i peuvent avoir des intersections non vides. Nous avons alors le calcul

$$A \approx A \times \mathbb{N} \tag{1.65a}$$
$$\geq \bigcup_{i=0}^{\infty} A_i \tag{1.65b}$$
$$\geq A \tag{1.65c}$$

Justifications :

— Pour (1.65a), c'est la proposition 1.61.

— Pour (1.65b), c'est la surjection (1.64).

— Pour (1.65c), c'est le fait que seulement A_0 possède déjà une surjection vers A.

Donc $\bigcup_i A_i$ est à la fois surpotent et subpotent à A. Il est donc équipotent par le théorème 1.51. □

1.4 Groupes

1.4.1 Définition, unicité du neutre

Définition 1.69 (Groupe)**.**
*Un **groupe** est un ensemble G muni d'une opération interne $\cdot : G \times G \to G$ telle que*

(1) pour tous g, h, $k \in G$, $g \cdot (h \cdot k) = (g \cdot h) \cdot k$,

(2) il existe un élément $e \in G$ tel que $e \cdot g = g \cdot e = g$ pour tout $g \in G$,

(3) pour tout $g \in G$, il existe un élément $h \in G$ tel que $g \cdot h = h \cdot g = e$.

Notons que nous avons écrit $g \cdot h$ et non $\cdot(g, h)$ comme une notation purement fonctionnelle nous l'aurait suggéré. Dans les exemples concrets, selon les cas, le loi de groupe appliquée à g et h sera notée tantôt $g + h$, tantôt $g \cdot h$ ou, le plus souvent pour un groupe générique, simplement gh.

Lemme-définition 1.70 (Unicités)**.**
L'inverse et le neutre sont uniques, c'est-à-dire :

(1) il existe un unique élément $e \in G$ tel que $eg = ge = g$ pour tout $g \in G$,

(2) pour tout $g \in G$, il existe un unique élément $h \in G$ tel que $gh = hg = e$.

*Le e ainsi défini est nommé **neutre** de G. Le h tel que $gh = hg = e$ est nommé l'**inverse** de g et est noté g^{-1}.*

Démonstration. Chaque point séparément.

(1) Supposons que e_1 et e_2 vérifient la propriété. Nous avons pour tout $g \in G$: $e_1 g = g e_1 = g$. En particulier pour $g = e_2$ nous écrivons $e_1 e_2 = e_2 e_1 = e_2$. Mais en partant dans l'autre sens : $e_2 g = g e_2 = g$ avec $g = e_1$ nous avons $e_2 e_1 = e_1 e_2 = e_1$. En égalant ces deux valeurs de $e_2 e_1$ nous avons $e_1 = e_2$.

Pour la suite de la preuve nous écrivons e l'unique neutre de G.

(2) Supposons que k_1 et k_2 soient deux inverses de g. On considère alors le produit $k_1 g k_2$. Puisque $k_1 g = e$, on a $k_1 g k_2 = e k_2 = k_2$; mais, comme $g k_2 = e$, on a aussi $k_1 g k_2 = k_1 e = k_1$. Le produit est donc à la fois égal à k_1 et à k_2, et donc $k_1 = k_2$.

□

Définition 1.71.
*Un groupe G est **abélien** ou **commutatif** si pour tous g et h dans G, $gh = hg$.*

Lemme 1.72.
Si G est un groupe et si $h \in G$, alors les applications

$$\begin{aligned} L_h \colon G &\to G \\ g &\mapsto hg \end{aligned} \tag{1.66}$$

et

$$\begin{aligned} R_h \colon G &\to G \\ g &\mapsto gh \end{aligned} \tag{1.67}$$

sont des bijections.

Démonstration. D'abord si $L_h(g_1) = L_h(g_2)$, alors $hg_1 = hg_2$ et en multipliant à gauche par h^{-1} nous avons $g_1 = g_2$; donc L_h est injective. Ensuite L_h est surjective parce que si $g \in G$, alors $g = L_h(h^{-1}g)$.

Pour l'application R_h, la preuve est une simple adaptation. □

1.5 Anneaux

Définition 1.73 (Anneau[20]).
*Un **anneau**[14] est un triplet $(A, +, \cdot)$ avec les conditions*

(1) $(A, +)$ est un groupe abélien. Nous notons 0 le neutre.

(2) La multiplication est associative et nous notons 1 le neutre.

(3) La multiplication est distributive par rapport à l'addition.

*L'anneau $(A, +, \cdot)$ est **commutatif** si pour tout $a, b \in A$ nous avons $a \cdot b = b \cdot a$.*

Définition 1.74 (Morphisme d'anneaux[21]).
*Si $(A, +, \cdot)$ et $(B, +, \cdot)$ sont des anneaux, un **morphisme d'anneaux** est une application $f \colon A \to B$ telle que*

(1) $f(a + b) = f(a) + f(b)$

(2) $f(a \cdot b) = f(a) \cdot f(b)$

(3) $f(1) = 1$.

Étant bien entendu que les significations de 1, $+$ et $\cdot$ sont différentes à gauche et à droite.

Lemme 1.75.
Pour tout élément a d'un anneau nous avons $a \cdot 0 = 0$.

Démonstration. L'élément 0 est le neutre de l'addition. Il peut être écrit $1 - 1$, et en utilisant la distributivité,

$$a \cdot 0 = a \cdot (1 - 1) = a - a = 0. \tag{1.68}$$

Notons que la dernière égalité s'écrit en détail $a - a = a + (-a)$ qui donne le neutre de l'addition. □

Proposition 1.76.
Dans un anneau[15] non nul, le neutre pour l'addition est distinct du neutre pour la multiplication.

Démonstration. Supposons par contraposée que dans un anneau A, $1 = 0$. Alors, pour tout $a \in A$, on a $a = 1a = 0a = (1 - 1)a = a - a = 0$, d'où l'on déduit $-a = 0$ et par suite, $a = 0$. □

14. Nous faisons le choix qu'un anneau admet toujours un neutre pour la multiplication. Certains ouvrages parlent dans ce cas d'anneau unitaire.
15. Définition 1.73.

Soit X un ensemble et un anneau $(A, +, \times)$. Nous considérons $\mathrm{Fun}(X, A)$ l'ensemble des applications $X \to A$. Cet ensemble devient un anneau avec les définitions

$$(f+g)(x) = f(x) + g(x) \tag{1.69a}$$
$$(fg)(x) = f(x)g(x). \tag{1.69b}$$

Cela est la **structure canonique** d'anneau sur $\mathrm{Fun}(X, A)$.

Définition 1.77.
*Le **centralisateur** de $x \in A$ dans A est l'ensemble*

$$\{y \in A \text{ tel que } xy = yx\}, \tag{1.70}$$

*le **centre** de A est*

$$\{y \in A \text{ tel que } xy = yx, \forall x \in A\}. \tag{1.71}$$

Définition 1.78.
*Un **isomorphisme d'anneaux** est un morphisme bijectif.*

Définition 1.79 (Idéal dans un anneau)**.**
*Un sous-ensemble $I \subset A$ est un **idéal à gauche** à gauche si*

(1) I est un sous-groupe pour l'addition,

(2) pour tout $a \in A$, $aI \subset I$.

*Lorsqu'un ensemble est idéal à gauche et à droite, nous disons que c'est un **idéal bilatère**. Lorsque nous parlons d'idéal sans précisions, nous parlons d'idéal bilatère.*

Proposition-définition 1.80.
Soit A, un anneau, I un idéal bilatère[16] de A. Nous considérons la relation d'équivalence $x \sim y$ si et seulement si $x - y \in I$. Sur le quotient[17]

$$A/\sim = A/I, \tag{1.72}$$

nous mettons les opérations

(1) $[x] + [y] = [x + y]$

(2) $[x][y] = [xy]$.

Nous avons alors les résultats suivants :

(1) Les opérations sont bien définies,

(2) l'ensemble A/I, muni de ces opérations, est un anneau

(3) la surjection canonique $\pi\colon A \to A/I$ est un morphisme.

*Cet anneau est appelé **anneau quotient**.*

Démonstration. En plusieurs parties.

Pour (1) Nous savons que, par définition,

$$\bar{x} = \{x + i \text{ tel que } i \in I\}. \tag{1.73}$$

Calculons le produit de représentants génériques de $\bar{x}$ et de $\bar{y}$:

$$(x + i_1)(y + i_2) = xy + xi_2 + yi_1 + i_1 i_2. \tag{1.74}$$

Vu que I est un idéal nous avons $xi_2 + yi_1 + i_1 i_2 \in I$ et donc bien

$$(x + i_1)(y + i_2) \in \overline{xy}. \tag{1.75}$$

16. Définition 1.79.
17. Définition 1.21.

Pour (2) Il s'agit de vérifier les conditions de la définition 1.73.

Vu que $(A, +)$ est un groupe abélien de neutre 0, nous avons

$$[a] + [b] = [a + b] = [b + a] = [b + a], \tag{1.76}$$

ainsi que

$$[a] + [0] = [a + 0] = [a] \tag{1.77}$$

et

$$[a] + ([b] + [c]) = [a] + [b + c] = [a + b + c] = [a + b] + [c]. \tag{1.78}$$

Donc $(A/I, +)$ est un groupe abélien de neutre $[0]$.

L'associativité de A donne l'associativité dans A/I :

$$\big([a][b]\big)[c] = [ab][c] = [abc] = [a][bc] = [a]\big([b][c]\big). \tag{1.79}$$

Et enfin pour la distributivité,

$$[a]\big([b] + [c]\big) = [a][b + c] = [a(b + c)] = [ab + ac] = [ab] + [ac] = [a][b] + [a][c]. \tag{1.80}$$

Nous avons prouvé que A/I est un anneau de neutre $[0]$ et d'unité $[1]$.

Pour (3) Nous devons vérifier les trois conditions de la définition 1.74. Cela est immédiat parce que $\pi(x) = [x]$.

□

Définition 1.81.
*Soient A un anneau commutatif et $S \subset A$. Nous disons que $\delta \in A$ est un **PGCD** de S si*

(1) δ divise tous les éléments de S.

(2) si d divise également tous les éléments de S, alors d divise δ.

*Nous disons que $\mu \in A$ est un **PPCM** de S si*

(1) $S \mid \mu$,

(2) si $S \mid m$, alors $\mu \mid m$.

Remarque 1.82.
Au sens de la définition 1.81, le pgcd n'est pas unique. Dans $\mathbb{Z}$ par exemple les nombres 4 et -4 sont tout deux pgcd de $\{4, 16\}$.

Dans $\mathbb{Z}$ cependant, nous modifions implicitement la définition et nous n'acceptons que les positifs, de telle sorte à ce que l'unique pgcd soit effectivement le plus grand pour l'ordre usuel sur $\mathbb{Z}$.

Pour l'unicité dans $\mathbb{Z}$, voir 3.30.

1.6 Les entiers

Lemme 1.83.
Toute partie bornée de $\mathbb{Z}$ possède un plus grand élément.

Proposition 1.84.
Soit $a, b \in \mathbb{Z}$ tels que a divise b. Alors $a \leqslant b$.

1.6.1 Quelques autres résultats

Nous supposons ici connaitre, et avoir démontré les rudiments du calcul dans $\mathbb{N}$ et $\mathbb{Z}$, en particulier les produits remarquables.

1.6.2 Anneau intègre

Définition 1.85 (Diviseurs dans un anneau)**.**
*Soient $a, b \in A$. On dit que a divise b, ou que a est un **diviseur (à gauche)** de b s'il existe $c \in A$ tel que $ac = b$. On dit que c'est un diviseur de b à droite si $ca = b$ pour un certain $c \in A$.*

Un cas particulier est le cas des diviseurs de zéro. L'absence de tels diviseurs dans un anneau est une propriété intéressante : on dit dans ce cas que l'anneau est intègre. Nous étudions ces anneaux plus en détail en section 3.9.

Un élément $a \in A$ est **régulier à droite** si $ba = 0$ implique $b = 0$. Il est régulier à gauche si $ab = 0$ implique $b = 0$.

Définition 1.86 (Éléments nilpotents, unipotents et inversibles)**.**
*On dit que $a \in A$ est **nilpotent** s'il existe $n \in \mathbb{N}$ tel que $a^n = 0$. Il est dit **unipotent** si $a - 1$ est nilpotent, c'est-à-dire si $(a-1)^n = 0$ pour un certain $n \in \mathbb{N}$.*

*Un élément $a \in A$ est dit **inversible** s'il existe $b \in A$ tel que $ab = 1$.*

L'ensemble $U(A)$ des éléments inversibles de A est un groupe pour la multiplication. Nous notons $A^* = A \backslash \{0\}$.

Conformément à la définition 1.85 de diviseur, nous posons la définition suivante pour les diviseurs de zéro.

Définition 1.87 ([22])**.**
*Un élément $a \neq 0$ est un **diviseur de zéro à gauche** s'il existe $x \neq 0$ tel que $xa = 0$. L'élément a est un diviseur de zéro **à droite** s'il existe $y \neq 0$ tel que $ay = 0$.*

*Nous disons que a est un **diviseur de zéro** si il est un diviseur de zéro à gauche ou à droite.*

Proposition-définition 1.88 ([1])**.**
Soit A un anneau non réduit à $\{0\}$. Les assertions suivantes sont équivalentes :

(1) A ne possède pas de diviseurs de zéro.

(2) La règle du produit nul s'applique dans A : pour tous $a, b \in A$, si $ab = 0$, alors $a = 0$ ou $b = 0$.

(3) On peut simplifier par un même élément non-nul, deux expressions produit dans A qui sont égales : pour tous $a, b, c \in A$ avec $a \neq 0$, si $ab = ac$, alors $b = c$.

*Un anneau non réduit à $\{0\}$ qui vérifie ces propriétés est dit **intègre**.*

Démonstration. En trois implications.

(1) implique (2) Si $ab = 0$ avec $a, b \neq 0$ alors a est un diviseur de zéro. Vu que nous supposons que A n'a pas de diviseurs de zéros, soit a soit b est nul.

(2) implique (3) Si $ab = ac$, alors $a(b - c) = 0$ et l'hypothèse dit que soit $a = 0$, soit $b - c = 0$. Donc si $a \neq 0$, alors $b - c = 0$.

(3) implique (1) Si $A = \{0\}$, le point (3) n'est pas applicable.

Si $a \neq 0$ et $xa = 0$, alors nous avons aussi $xa = 0 \times a$. Par propriété de simplification, $x = 0$. Donc a n'est pas un diviseur de zéro à gauche. Nous prouvons de la même façon qu'il n'y a pas de diviseurs de zéro à droite.

□

1.7 Somme à valeurs dans un groupe abélien

1.7.1 Somme à valeurs dans un anneau

Définition 1.89.
Si $f \colon \{0, \ldots, N\} \to A$ est une application vers un anneau A, alors nous définissons la notation $\sum_{i=0}^{N} f(i)$ par récurrence de la façon suivante :

(1) $\sum_{i=0}^{0} f(i) = f(0)$,
(2) $\sum_{i=0}^{k} f(i) = \sum_{i=0}^{k-1} f(i) + f(k)$.

1.7.2 Somme à valeurs dans un groupe abélien

Si S est un ensemble fini, nous savons de la proposition 1.32 qu'il existe un unique $N \in \mathbb{N}$ pour lequel il existe une bijection $\varphi\colon \{0,\dots,N\} \to S$. Cette bijection n'est à priori pas unique.

Lemme-définition 1.90 ([1]).
Soient un groupe abélien $(G,+)$ ainsi qu'un ensemble fini I contenant n éléments. Soit une application $f\colon I \to G$. Si $\sigma_1,\sigma_2\colon \{1,\dots,n\} \to I$ sont deux bijections, alors[18]

$$\sum_{i=1}^{n} f\big(\sigma_1(i)\big) = \sum_{i=1}^{n} f\big(\sigma_2(i)\big). \tag{1.81}$$

La valeur commune est notée

$$\sum_{i\in I} f(i) \tag{1.82}$$

La définition 1.90 donne lieu à un certain nombre de remarques.

(1) Elle donne la somme sur un ensemble fini. Un problème avec les ensembles infinis (outre la convergence) est l'ordre de sommation. Si vous voulez sommer sur $\mathbb{Z}$, dans quel ordre le faire ?

(2) Pour aller plus loin, et sommer sur des ensembles infinis, il faut regarder la définition 10.91.

Proposition 1.91.
Soient un groupe abélien $(G,+)$, un ensemble fini I est un ensemble fini, une application $f\colon I \to G$ et une bijection $\sigma\colon I \to I$. Alors

$$\sum_{i\in I} f(i) = \sum_{i\in I} f\big(\sigma(i)\big). \tag{1.83}$$

Si nous avons une application $L\colon S \to S$, nous notons

$$\sum_{s\in S} f\big(L(s)\big) = \sum_{s\in S} (f\circ L)(s). \tag{1.84}$$

Cette façon d'écrire donne une interprétation pour la notation $\sum_{g\in G} f(hg)$ qui arrive dans la proposition 1.92. Il s'agit de considérer l'application L_h du lemme 1.72, de considérer[19]

$$\sum_{g\in G} f(hg) = \sum_{g\in G} (f\circ L_h)(g) \tag{1.85}$$

et de faire tourner la définition 1.90. La même chose tient pour définir $\sum_{g\in G}(gh)$ à l'aide de R_h.

Proposition 1.92 ([1]).
Soient un groupe fini G et une fonction $f\colon G \to A$ où A est un anneau commutatif. Alors

$$\sum_{g\in G} f(g) = \sum_{g\in G} f(gh) = \sum_{g\in G} f(hg) \tag{1.86}$$

pour tout $h \in G$.

Démonstration. Nous avons une bijection $\varphi\colon \{0,\dots,N\} \to G$ garantie par la proposition 1.32. La définition est que

$$\sum_{g\in G} f(g) = \sum_{i=0}^{N} f\big(\varphi(i)\big). \tag{1.87}$$

18. Pour rappel, le symbole $\sum_{i=1}^{n}$ est défini par 1.89.
19. Le fait que L_h soit une bijection n'a pas d'importance ici.

Par ailleurs, le lemme 1.72 donne une bijection $L_h\colon G \to G$ et permet de considérer la composée

$$\begin{aligned} \varphi'\colon \{0,\dots,N\} &\to G \\ \varphi' &= L_h \circ \varphi. \end{aligned} \tag{1.88}$$

La proposition 1.90 nous permet d'utiliser la bijection φ' au lieu de φ pour exprimer la somme $\sum_{g\in G}$. Ensuite un jeu de notation utilisant (1.85) donne

$$\begin{aligned} \sum_{g\in G} f(g) &= \sum_{i=0}^{N} f\big(\varphi(i)\big) = \sum_{i=0}^{N} f\big(\varphi'(i)\big) = \sum_{i=0}^{N} (f\circ L_h\circ\varphi)(i) \\ &= \sum_{i=0}^{N} (f\circ L_h)\big(\varphi(i)\big) = \sum_{g\in G} (f\circ L_h)(g) = \sum_{g\in G} f(hg). \end{aligned} \tag{1.89}$$

En ce qui concerne $\sum_{g\in G} f(gh)$, c'est la même chose, en utilisant R_h au lieu de L_h. □

Tout cela nous permet de faire une somme sympathique et bien connue.

Lemme 1.93.
Soit $n \in \mathbb{N}$. Nous avons

$$\sum_{k=0}^{n} k = \frac{n(n+1)}{2}. \tag{1.90}$$

Démonstration. La preuve est pratiquement immédiate par récurrence. Nous allons donner une preuve plus « constructive », qui formalise l'idée classique d'écrire la somme à l'endroit et à l'envers.

Nous notons S la somme $\sum_{k=0}^{n} k$. Le lemme 1.90 dit que si $\sigma_i\colon \{0,\dots,n\} \to \{0,\dots,n\}$ sont deux bijections, alors $\sum_{k=0}^{n} f\big(\sigma_1(k)\big) = \sum_{k=0}^{n} f\big(\sigma_2(k)\big)$. Nous sommes intéressé au cas $f(i) = i$.

En prenant $\sigma_1(k) = k$ et $\sigma_2(k) = n-k$, nous avons

$$S = \sum_{k=0}^{n} k = \sum_{k=0}^{n} (n-k). \tag{1.91}$$

Donc

$$2S = \sum_{k=0}^{n} \big(k + (n-k)\big) = \sum_{k=0}^{n} n = n\sum_{k=0}^{n} 1 = n(n+1). \tag{1.92}$$

En divisant par deux, nous obtenons le résultat annoncé. □

Dans le même ordre d'idée, pour le produit.

Proposition 1.94.
Si E est un ensemble fini et si G est un groupe abélien, alors pour toute fonction $f\colon E \to G$ et pour toute permutation σ de E,

$$\prod_{i\in E} f(i) = \prod_{i\in E} f\big(\sigma(i)\big) \tag{1.93}$$

1.7.3 Fonction puissance

Voici une première définition de la fonction puissance. Il y en aura d'autres, de plus en plus générales. Voir le thème 33.

Définition 1.95.
Si A est un anneau, si $a \in A$ et si $n \in \mathbb{N}$, nous définissons a^n par récurrence :

(1) $a^0 = 1$ *(l'unité pour la multiplication dans A),*

(2) $a^{k+1} = a \cdot a^k$.

Le lemme suivant dit que le point (2) de la définition 1.95 aurait pu être écrit $a^k \cdot a$ au lieu de $a \cdot a^k$.

Lemme 1.96 ([1]).
Si A est un anneau, si $a \in A$ et si $n \in \mathbb{N}$, alors

$$a^n = a \cdot a^{n-1} = a^{n-1} \cdot a. \tag{1.94}$$

1.8 Corps

1.8.1 Définitions, morphismes

Définition 1.97 ([23]).
Un anneau[20] $(A, +, \times)$ *est un* ***corps*** *dans lequel tout élément non nul est inversible pour l'opération* $\times$ *(pour l'opération* $+$*, tous les éléments sont inversibles parce que* $(A, +)$ *est un groupe).*

Remarque 1.98.
Un anneau est ce qu'on appelle « *ring* » en anglais. Un corps est en anglais « *field* ». De plus le mot « *field* » comprend la commutativité. Donc certains utilisent le mot « corps » pour dire « corps commutatif » et parlent alors d'anneau *à division* pour parler de corps non commutatifs.

La proposition suivante donne une caractérisation d'un corps, en disant un tout petit peu plus que la définition 1.97.

Proposition 1.99.
L'anneau A est un corps si et seulement si $U(A) = A^*$.

Démonstration. En deux parties.

Sens direct Nous supposons que A est un corps. D'une part tous les éléments non nuls sont inversibles, c'est-à-dire $A^* \subset U(A)$.

Pour l'inclusion inverse, nous montrons qu'une élément inversible ne peut pas être nul. Cela n'est autre que le lemme 1.75 couplé à la proposition 1.76 : $a \cdot 0 = 0 \neq 1$ pour tout a.

Sens inverse Si $U(A) = A^*$, nous avons immédiatement que tous les éléments non nuls sont inversibles et donc que A est un corps.

□

Lemme 1.100.
Un corps non nul est un anneau intègre[21].

Démonstration. Soit un produit nul $ab = 0$. Si $a \neq 0$, alors il est inversible et nous multiplions $ab = 0$ par a^{-1}. Nous trouvons $b = 0$ parce que $0a^{-1} = 0$. □

Conséquence : dans un corps nous avons toujours la règle du produit nul, et l'élément nul n'est jamais inversible.

Définition 1.101 (Morphisme de corps).
Un corps étant un anneau sans plus de structure, un ***morphisme de corps*** *n'est qu'un morphisme des anneaux.*

Le lemme suivant montre que définir un morphisme de corps comme étant simplement un morphisme des anneaux est une bonne idée.

Lemme 1.102.
Si $\phi \colon \mathbb{K} \to \mathbb{K}'$ est un morphisme de corps, alors

(1) pour tout $a \in \mathbb{K}$ nous avons $\varphi(a^{-1}) = \varphi(a)^{-1}$;

(2) le morphisme φ est injectif.

20. Définition 1.73.
21. Définition 1.88.

Démonstration. Vu que $\varphi(1) = 1$, nous avons aussi

$$1 = \varphi(aa^{-1}) = \varphi(a)\varphi(a^{-1}). \tag{1.95}$$

Donc, par unicité de l'inverse[22], $\varphi(a^{-1}) = \varphi(a)^{-1}$.

Pour l'injectivité nous supposons $\varphi(a) = \varphi(b)$. Étant donné que $\mathbb{K}'$ est un corps, nous pouvons multiplier par $\varphi(b)^{-1}$:

$$\varphi(a)\varphi(b)^{-1} = 1. \tag{1.96}$$

En utilisant le premier point nous avons $1 = \varphi(a)\varphi(b^{-1})$, puis le morphisme d'anneaux : $1 = \varphi(ab^{-1})$, et encore le morphisme d'anneaux nous permet de déduire $ab^{-1} = 1$ et donc $a = b$. □

1.8.2 Corps des fractions

Définition 1.103 ([24]).
Soit un anneau commutatif et intègre[23] *A et $E = A \times A \backslash \{0\}$. Nous y définissons les deux opérations suivantes :*

(1) $(a, b) + (c, d) = (ad + cb, bd)$ *;*

(2) $(a, b)(c, d) = (ac, bd)$.

Et aussi la relation d'équivalence $(a, b) \sim (c, d)$ si et seulement si $ad = bc$.

*Le **corps des fractions** de A est le quotient*

$$\mathrm{Frac}(A) = \big(A \times A \backslash \{0\}\big)/\sim . \tag{1.97}$$

Nous notons a/b la classe de (a, b).

Les éléments de $\mathrm{Frac}(A)$ *sont des **fractions rationnelles** de A.*

Le fait que A soit intègre est important pour être certain que $bd \neq 0$ sous l'hypothèse que $b, d \neq 0$.

La proposition suivante dit que $\mathrm{Frac}(A)$ est le plus petit corps contenant A.

Proposition 1.104 ([1]).
Si $\mathbb{L}$ est un corps contenant A en tant que sous-anneau[24]*, alors il existe un morphisme de corps injectif $\epsilon\colon \mathrm{Frac}(A) \to \mathbb{L}$.*

Démonstration. Il suffit de vérifier que la formule

$$\epsilon(a/b) = ab^{-1} \tag{1.98}$$

vérifie toutes les conditions. Notons que dans le membre de droite de (1.98), l'inverse et le produit sont calculés dans $\mathbb{L}$.

Le fait que ϵ soit bien défini provient du fait que A soit commutatif :

$$\epsilon(ax/bx) = (ax)(bx)^{-1} = ab^{-1}xx^{-1} = ab^{-1} = \epsilon(a/b). \tag{1.99}$$

Le fait que ϵ soit un morphisme est une vérification de routine, par exemple ceci pour la somme :

$$\epsilon\big(a/b + c/d\big) = \epsilon\big((ad + cb)/bd)\big) = (ad + cb)(bd)^{-1} = ab^{-1} + cd^{-1}, \tag{1.100}$$

tandis que

$$\epsilon(a, b) + \epsilon(c, d) = ab^{-1} + cd^{-1}, \tag{1.101}$$

qui est égal (il faut aussi vérifier pour le produit).

Enfin ϵ est injective parce que si $\epsilon(a/b) = \epsilon(c/d)$, alors $ab^{-1} = cd^{-1}$, d'où il est facilement vu que $ad = cb$, c'est-à-dire $a/b = c/d$. □

22. Lemme 1.70 (2).
23. Définition 1.88.
24. Bien entendu, nous pouvons trouver plein de corps contenant A en tant que sous-ensemble sans pour autant étendre A en tant qu'anneau ; il suffit d'y mettre une loi de composition farfelue.

Notons que si A est un anneau qui n'est pas un corps, le corps $\mathrm{Frac}(A)$ existe, mais si $R \in \mathrm{Frac}(A)$, il n'a pas de sens de vouloir calculer $R(\alpha)$ pour $\alpha \in A$.

Définition 1.105 (Évaluation d'une fraction rationnelle)**.**
Soit un corps $\mathbb{K}$ contenant l'anneau A. Si $R = P/Q \in \mathrm{Frac}(A)$ et si $\alpha \in \mathbb{K}$ nous définissons

$$R(\alpha) = (P/Q)(\alpha) = P(\alpha)Q^{-1}(\alpha). \tag{1.102}$$

Dans cette formule, les polynômes, l'inverse et le produit sont calculés dans $\mathbb{K}$ et non dans A.

Théorème-définition 1.106.
Soit $\mathbb{A}$ un anneau commutatif intègre.

(1) Il existe un corps commutatif $\mathbb{K}$ et un morphisme d'anneaux injectif $\epsilon\colon \mathbb{A} \to \mathbb{K}$ tels que pour tout $\lambda \in \mathbb{K}$, il existe $(a, b) \in \mathbb{A} \times \mathbb{A}^*$ tels que

$$\lambda = \epsilon(a)\big(\epsilon(b)\big)^{-1} \tag{1.103}$$

(2) Si $(\mathbb{K}', \epsilon')$ est un autre couple qui vérifie la propriété, les corps $\mathbb{K}$ et $\mathbb{K}'$ sont isomorphes.

*Le corps $\mathbb{K}$ associé à l'anneau $\mathbb{A}$ est le **corps des fractions** de $\mathbb{A}$, et sera noté $\mathrm{Frac}(\mathbb{A})$.*

Lemme 1.107 (Simplification de fraction)**.**
L'application $\mathbb{A} \times \mathbb{A}^* \to \mathbb{K}$ donnée par $(a, b) \mapsto \epsilon(a)\big(\epsilon(b)\big)^{-1}$ envoie (xa, xb) sur le même que (a, b).

La proposition suivante montre encore que le corps des fractions est le plus petit corps que l'on puisse imaginer à partir d'un anneau.

Proposition 1.108.
Soit un anneau A. Tout corps contenant un sous-anneau isomorphe à A contient un sous-corps isomorphe à $\mathrm{Frac}(A)$.

Démonstration. Soit un corps $\mathbb{K}$ contenant un sous-anneau A' isomorphe à A. Nous considérons la partie suivante de $\mathbb{K}$:

$$S = \{ab^{-1} \text{ tel que } a, b \in A\}. \tag{1.104}$$

Ensuite nous montrons que

$$\begin{aligned} \varphi\colon S &\to \mathrm{Frac}(A) \\ ab^{-1} &\mapsto a/b \end{aligned} \tag{1.105}$$

est un isomorphisme de corps.

Bien définie Si $ab^{-1} = xy^{-1}$ alors $ay = xb$ et donc $a/b = x/y$ par définition des classes de $\mathrm{Frac}(A)$.

Surjectif Tout élément de $\mathrm{Frac}(A)$ est de la forme a/b avec $a, b \in A$. Un tel élément est l'image par φ de $ab^{-1} \in S$.

Injectif Si $\varphi(ab^{-1}) = \varphi(xy^{-1})$ alors $a/b = x/y$, et par définition des classes nous avons $ay = bx$ qui donne immédiatement $ab^{-1} = xy^{-1}$.

□

1.8.3 Suites de Cauchy dans un corps totalement ordonné

Définition 1.109.
Ordre et choses reliées dans un corps.

*(1) Un corps $\mathbb{K}$ est **totalement ordonné** s'il existe une relation d'ordre total*[25] *tel que*

(1a) $x \leqslant y$ implique $x + z \leqslant y + z$ pour tout $x, y, z \in \mathbb{K}$

25. Définition 1.8.

(1b) $x \geqslant 0$ *et* $y \geqslant 0$ *implique* $xy \geqslant 0$.

(2) Si $\mathbb{K}$ *est un corps totalement ordonné, nous y définissons la valeur absolue par*

$$|x| = \begin{cases} x & \text{si } x \geqslant 0 \\ -x & \text{si } x \leqslant 0. \end{cases} \tag{1.106}$$

(3) La suite (x_n) *dans le corps totalement ordonné* $\mathbb{K}$ *est* ***de Cauchy*** *si pour tout* $\epsilon \in \mathbb{K}^+$, *il existe* $N \in \mathbb{N}$ *tel que si* $p, q \geqslant N$ *alors* $|x_p - x_q| \leqslant \epsilon$.

(4) La suite (x_n) *dans le corps totalement ordonné* $\mathbb{K}$ *est* ***convergente*** *s'il existe* $q \in \mathbb{K}$ *tel que pour tout* $\epsilon \in \mathbb{K}^+$, *il existe* N *tel que si* $k \geqslant N$ *alors* $|x_k - q| \leqslant \epsilon$.

(5) Un corps $\mathbb{K}$ *est* ***archimédien*** *s'il est totalement ordonné et si pour tout* $x, y \in \mathbb{K}$ *avec* $x > 0$, *il existe* $n \in \mathbb{N}$ *tel que* $nx \geqslant y$.

(6) Un corps totalement ordonné est ***complet*** *si toute suite de Cauchy y est convergente.*

(7) Si $a, \epsilon \in \mathbb{K}$ *alors nous définissons la* ***boule ouverte*** *de centre* a *et de rayon* ϵ *par*

$$B(a, \epsilon) = \{x \in \mathbb{K} \text{ tel que } |a - x| < \epsilon\}, \tag{1.107}$$

et la ***boule fermée*** *par*

$$\overline{B(a, \epsilon)} = \{x \in \mathbb{K} \text{ tel que } |a - x| \leqslant \epsilon\}. \tag{1.108}$$

Remarque 1.110.
Nous étudierons plus tard la notion de caractéristique d'un anneau [26] ou d'un corps. Tout anneau totalement ordonné est nécessairement de caractéristique nulle, lemme 3.63.

Remarque 1.111.
En mettant côte à côte les points (4) et (7) nous pouvons dire que (x_k) converge vers q si et seulement si pour tout $\epsilon > 0$, nous avons $x_k \in B(q, \epsilon)$ à partir d'un certain indice N.

Ces boules prendront une nouvelle force avec le super-théorème 7.94.

Parmi ces définitions, celles de suite convergente, de Cauchy et de corps complet seront utilisées dans le cas de $\mathbb{Q}$ (et de $\mathbb{R}$ pour la complétude). Elles seront prouvées être équivalentes aux définitions topologiques dans le cas particulier de $\mathbb{R}$ et $\mathbb{Q}$ lorsque la topologie métrique sera définie. Dans cet état d'esprit nous n'allons pas démontrer tout de suite que $\mathbb{R}$ est un corps complet. Nous allons directement démontrer que c'est un espace topologique complet.

Lemme 1.112 (Propriétés de la valeur absolue)**.**
Soit $\mathbb{K}$ *un corps totalement ordonné. Si* $x, y \in \mathbb{K}$ *alors*

(1) Si $x \geqslant 0$ *alors* $-x \leqslant 0$.

(2) $|x| \geqslant 0$

(3) $|x| = 0$ *si et seulement si* $x = 0$

(4) $|x + y| \leqslant |x| + |y|$.

Démonstration. Point par point

(1) Nous partons de $x \geqslant 0$ et nous ajoutons $-x$ des deux côtés en profitant de la définition d'un corps totalement ordonné : $x - x \geqslant -x$ et donc $0 \geqslant -x$, c'est-à-dire $-x \leqslant 0$.

(2) Si $x \geqslant 0$ alors c'est vrai. Sinon, $x \leqslant 0$ et $|x| = -x \geqslant 0$ par le point (1).

(3) Si $x = 0$ alors $x = -x$ et $|x| = 0$. Au contraire si $x \neq 0$ alors $-x \neq 0$ et que x soit positif ou négatif, nous aurons toujours $\pm x \neq 0$.

(4) Nous supposons que $x \leqslant y$ et nous distinguons divers cas suivant la positivité de x et y.

(4a) Si $x, y \geqslant 0$. Dans ce cas, $x + y \geqslant y \geqslant 0$, donc $|x + y| = x + y = |x| + |y|$.

26. définition 3.59

(4b) Si $x, y \leqslant 0$. Dans ce cas, $x + y \leqslant 0$ et nous avons $|x + y| = -x - y = |x| + |y|$.

(4c) Si $x \leqslant 0$ et $y \geqslant 0$. Nous subdivisons encore en deux cas suivant que $x + y$ est positif ou négatif. Si $x + y \geqslant 0$, alors nous écrivons successivement

$$x \leqslant 0 \tag{1.109a}$$

$$x + y \leqslant y \leqslant y + |x| = |x| + |y| \tag{1.109b}$$

et donc $|x + y| = x + y \leqslant |x| + |y|$.

Nous supposons à présent que $x \leqslant 0$, $y \geqslant 0$ et $x + y \leqslant 0$. Dans ce cas il suffit d'écrire $|x + y| = |(-x) + (-y)|$ pour retomber dans le cas précédent à inversion près de x et y.

□

Remarque 1.113.
La partie (4) est très importante parce que c'est elle qui fera presque toutes les majorations dont nous aurons besoin en analyse. En effet elle donne l'inégalité triangulaire de la façon suivante : si $x, y, z \in \mathbb{K}$ nous avons

$$|x - y| = |(x - z) + (z - y)| \leqslant |x - z| + |z - y|. \tag{1.110}$$

Lemme 1.114 (À propos de boules)**.**
Soient un corps totalement ordonné $\mathbb{K}$ et des éléments $x, y, \epsilon \in \mathbb{K}$.

(1) Nous avons $y \in B(x, \epsilon)$ si et seulement si $x - \epsilon < y < x + \epsilon$.

(2) Si $y \in \overline{B(x, \epsilon)}$ alors $y \in B(x, \epsilon')$ pour tout $\epsilon' > \epsilon$.

Démonstration. Pour rappel,

$$|x - y| = \begin{cases} x - y & \text{si } x - y \geqslant 0 \\ y - x & \text{si } x - y \leqslant 0. \end{cases} \tag{1.111}$$

Nous pouvons maintenant démontrer nos choses.

(1) Des inégalités $x - \epsilon < y$ et $y < x + \epsilon$ nous tirons $x - y < \epsilon$ et $y - x < \epsilon$. Donc quel que soit le signe de $x - y$ nous avons toujours $|x - y| < \epsilon$.

Dans l'autre sens, nous supposons que $|x - y| < \epsilon$.

Si $x - y \geqslant 0$ alors l'hypothèse signifie $x - y < \epsilon$, ce qui donne $y > x - \epsilon$. Mais l'inégalité $x - y \geqslant 0$ donne également $x \geqslant y$ et donc $x + \epsilon \geqslant y + \epsilon > y$. Notez le jeu de l'inégalité non stricte qui se change en inégalité stricte.

Si $x - y \leqslant 0$ nous pouvons faire le même raisonnement.

(2) C'est immédiat parce que

$$|x - y| \leqslant \epsilon < \epsilon'. \tag{1.112}$$

□

Proposition 1.115.
Toute suite convergente dans un corps totalement ordonné est de Cauchy.

Démonstration. Soit un corps totalement ordonné $\mathbb{K}$ et une suite $x_n \xrightarrow{\mathbb{K}} x$. Soit $\epsilon > 0$. Il est important de se rendre compte que $\epsilon \in \mathbb{K}$ et que l'inégalité est au sens de l'ordre dans $\mathbb{K}$; en particulier ce n'est pas $\epsilon \in \mathbb{R}$ ni $\epsilon \in \mathbb{Q}$. D'ailleurs nous n'avons encore pas défini ni $\mathbb{R}$ ni $\mathbb{Q}$.

Vu que (x_n) converge vers x, il existe $N \in \mathbb{N}$ tel que pour tout $k > N$,

$$|x_k - x| < \epsilon. \tag{1.113}$$

Soient $p, q > N$. Alors en utilisant la majoration du lemme 1.112(4),

$$|x_p - x_q| = \big|(x_p - x) + (x - x_q)\big| \leqslant |x_p - x| + |x - x_q| \leqslant 2\epsilon. \tag{1.114}$$

Donc la suite (x_n) est de Cauchy. □

1.9 Les rationnels

Une construction très explicite est faite dans [8]. Ici nous allons prendre plus court :

Définition 1.116.
Le corps des fractions de $\mathbb{Z}$ (définition 1.103) est noté $\mathbb{Q}$ et ses éléments sont les ***rationnels****.*

Les résultats énoncés ici sont utilisés plus bas et servent de guide à un éventuel contributeur qui voudrait écrire une partie dédiée à $\mathbb{Q}$ et ses propriétés de base[27]. Nous espérons que des preuves se trouvent dans [8]. En tout cas, le lecteur est invité à ne rien prendre comme évident.

Lemme 1.117.
Tout rationnel est majoré par un naturel.

Proposition 1.118.
L'ensemble des rationnels est dénombrable.

Proposition 1.119.
Si $q < 1$, alors $qx < x$ pour tout $x \in \mathbb{Q}^+$.

Proposition 1.120.
Le corps $\mathbb{Q}$ est archimédien[28]*.*

1.9.1 Suites de Cauchy dans les rationnels

Proposition 1.121 ([8])**.**
Principales propriétés des suites de Cauchy dans $\mathbb{Q}$.

(1) Toute suite convergente est de Cauchy[29]*.*

(2) Toute suite de Cauchy est bornée.

(3) Si $x_n \to 0$ et si (y_n) est bornée, alors $x_n y_n \to 0$

(4) Si (x_n) et (y_n) sont de Cauchy alors $(x_n + y_n)$, $(x_n - y_n)$ et $(x_n y_n)$ sont également de Cauchy.

(5) Si il existe $a, b \in \mathbb{Q}$ tels que $x_n \to a$ et $y_n \to b$ alors $x_n + y_n \to a + b$, $x_n - y_n \to a - b$ et $x_n y_n \to ab$.

(6) Soit (x_n) une suite de Cauchy qui ne converge pas vers zéro. Alors il existe n_0 tel que la suite $\left(\frac{1}{x_n}\right)_{n \geqslant n_0}$ soit de Cauchy.

Démonstration. Point par point.

(1) C'est la proposition 1.115.

(2) Soit (x_n) une suite de Cauchy dans $\mathbb{Q}$. Avec $\epsilon = 1$ dans la définition, si $q > N_1$, nous avons

$$|x_q - x_{N_1}| \leqslant 1. \tag{1.115}$$

Et donc pour tout q plus grand que N_1, $x_N - 1 \leqslant x_q \leqslant x_N + 1$, ou encore, pour tout n :

$$|x_n| \leqslant \max\{|x_1|, |x_2|, \dots, |x_N|, |x_N + 1|\}. \tag{1.116}$$

La suite est donc bornée.

(3) Soit $\epsilon > 0$. Les hypothèses disent qu'il existe un N tel que $|x_n| \leqslant \epsilon$ dès que $n \geqslant N$. Et il existe aussi $M \geqslant 0$ tel que $|y_n| \leqslant M$ pour tout n. Du coup, lorsque $n \geqslant N$ nous avons $|x_n y_n| \leqslant M\epsilon$.

27. Par exemple, définir une relation d'ordre sur $\mathbb{Q}$ et expliciter l'inclusion de $\mathbb{Z}$ dans $\mathbb{Q}$.
28. Définition 1.109(5).
29. Et non la réciproque, qui sera justement la grande innovation des nombres réels.

(4) En ce qui concerne la somme,

$$|x_p + y_p - x_q - y_q| \leqslant |x_p - x_q| + |y_p - y_q|. \tag{1.117}$$

Soit N_1 tel que si $p, q \geqslant N_1$ alors $|x_p - x_q| \leqslant \epsilon$ et N_2 de même pour la suite (y_n). En prenant $N = \max\{N_1, N_2\}$, la somme (1.117) est plus petite que 2ϵ dès que $p, q \geqslant N$.

Passons à la démonstration du fait que le produit de deux suites de Cauchy est de Cauchy. Les suites (x_n) et (y_n) sont bornées et quitte à prendre le maximum, nous disons qu'elles sont toutes les deux bornées par le nombre M : pour tout n nous avons $|x_n| \leqslant M$ et $|y_n| \leqslant M$. Nous avons :

$$|x_p y_p - x_q y_q| \leqslant |x_p y_p - x_q y_p| + |x_q y_p - x_q y_q| \leqslant |y_p||x_p - x_q| + |x_q||y_p - y_q|. \tag{1.118}$$

Vu que (x_n) et (y_n) sont de Cauchy, si p et q sont assez grands, les deux différences sont majorées par ϵ et nous avons

$$|x_p y_p - x_q y_q| \leqslant M\epsilon + M\epsilon = 2M\epsilon, \tag{1.119}$$

ce qui prouve que $(x_n y_n)$ est de Cauchy.

(5) En ce qui concerne la somme, nous pouvons tout de suite calculer

$$|x_n + y_n - (a + b)| \leqslant |x_n - a| + |y_n - b|. \tag{1.120}$$

Il existe une valeur de n à partir de laquelle le premier terme est plus petit que ϵ et une à partir de laquelle le second terme est plus petit que ϵ. En prenant le maximum des deux, la somme est plus petite que 2ϵ.

En ce qui concerne le produit,

$$|x_n y_n - ab| \leqslant |x_n y_n - a y_n| + |a y_n - ab| \leqslant |y_n||x_n - a| + |a||y_n - b|. \tag{1.121}$$

Les suites $|x_n - a|$ et $|y_n - b|$ convergent vers zéro ; la suite (y_n) est bornée parce que convergente (combinaison des points (1) et (2)) et a (la suite constante) est également bornée. Donc par le point (3), nous avons

$$y_n|x_n - a| + a|y_n - b| \to 0. \tag{1.122}$$

Au passage nous avons également utilisé la propriété de la somme que nous venons de démontrer.

(6) Soit (x_n) une suite de Cauchy dans $\mathbb{Q}$ ne convergeant pas vers zéro : il existe $\alpha > 0$ tel que pour tout $N \in \mathbb{N}$, il existe $n \geqslant N$ tel que $|x_n| > \alpha$. Mais notre suite est de Cauchy, donc il existe $n_0 \in \mathbb{N}$ tel que si $p, q \geqslant n_0$ alors

$$|x_p - x_q| \leqslant \frac{\alpha}{2}. \tag{1.123}$$

En fixant $N = n_0$, on obtient un naturel $n \geqslant n_0$ tel que $|x_n| \geqslant \alpha$. De plus, comme la suite est de Cauchy, si $p > n$ nous avons aussi $|x_n - x_p| \leqslant \frac{\alpha}{2}$. Cela implique $|x_p| \geqslant \frac{\alpha}{2}$ et en particulier $x_p \neq 0$.

Nous venons de prouver que la suite ne s'annule plus à partir de l'indice n, et même que $|x_k| \geqslant \alpha/2$ pour tout $k \geqslant n$. La suite $(1/x_k)_{k \geqslant n}$ est donc bien définie.

Soit $\epsilon > 0$. Soit n_0 tel que $|x_p - x_q| < \epsilon$ pour tout $p, q > n_0$. Soit K plus grand que n_0 et que n. En prenant $p, q \geqslant K$, nous avons $|x_p| > \frac{\alpha}{2}$ et $|x_q| > \frac{\alpha}{2}$. Nous en déduisons que

$$\left|\frac{1}{x_p} - \frac{1}{x_q}\right| \leqslant \frac{|x_q - x_p|}{|x_p x_q|} \leqslant \frac{4}{\alpha^2}|x_q - x_p| \leqslant \frac{4}{\alpha^2}\epsilon. \tag{1.124}$$

Donc $\left(\frac{1}{x_n}\right)$ est de Cauchy.

□

1.9.2 Insuffisance des rationnels

Nous allons voir qu'il n'existe pas de nombres rationnels x tels que $x^2 = 2$, mais que pourtant il existe une infinité de suites de rationnels (x_n) tels que $x_n^2 \to 2$.

Lemme 1.122.
Un entier x est pair si et seulement si l'entier x^2 est pair.

Démonstration. Si x est un nombre pair, alors il existe un entier a tel que $x = 2a$ alors $x^2 = 4a^2$ est pair.

Inversement, si x est impair alors il existe un entier a tel que $x = 2a + 1$ et alors $x^2 = 4a^2 + 4a + 1 = 2(2a^2 + 2a) + 1$ est impair. □

Le théorème 3.43 nous dira que tous les $\sqrt{n}$ sont irrationnels dès que n n'est pas un carré parfait. Voici déjà le résultat pour $n = 2$. Le fait que $\sqrt{2}$ existe dans $\mathbb{R}$ sera la proposition 1.166.

Proposition 1.123 (Irrationalité de $\sqrt{2}$)**.**
Il n'existe pas de fractions d'entiers dont le carré soit égal à 2.

Démonstration. Nous supposons que la fraction d'entiers a/b est telle que $a^2/b^2 = 2$, et nous allons construire une suite d'entiers strictement décroissante et strictement positive, ce qui est impossible.

Grâce au lemme 1.122 nous avons successivement les affirmations suivantes :

— $\frac{a^2}{b^2} = 2$ avec $a \neq 0$ et $b \neq 0$.

— $a^2 = 2b^2$, donc a^2 est pair.

— a est alors pair et a^2 est divisible par 4. Soit $a^2 = 4k$.

— $4k/b^2 = 2$, donc $4k = 2b^2$, donc $b^2 = 2k$ et b^2 est pair.

— Nous déduisons que b est pair.

La fraction $\frac{a/2}{b/2}$ est alors une nouvelle fraction d'entiers dont le carré vaut 2. En procédant de la même façon, en remplaçant a par $a/2$ et b par $b/2$, on obtient que la fraction d'entiers $\frac{a/4}{b/4}$ a la même propriété.

En particulier, tous les nombres de la forme $a/2^n$ sont des entiers. Ils forment une suite strictement décroissante d'entiers strictement positifs. Impossible, me diriez-vous ? Et vous auriez bien raison : toute partie non vide de $\mathbb{N}$ admet un plus petit élément [30]. Il n'y a donc pas de fractions d'entiers dont le carré vaut 2. □

Lemme 1.124 (Série géométrique, voir aussi l'exemple 10.67)**.**
Si $q \neq 1 \in \mathbb{Q}$ et $p \in \mathbb{N}$ nous avons

$$\sum_{k=0}^{p} q^k = \frac{1 - q^{p+1}}{1 - q}. \tag{1.125}$$

Démonstration. En posant $S_p = 1 + q + q^2 + \cdots + q^p$, nous avons $S_p - qS_p = 1 - q^{p+1}$ et donc

$$S_p = \sum_{k=0}^{p} q^k = \frac{1 - q^{p+1}}{1 - q}. \tag{1.126}$$

□

Proposition 1.125.
La suite donnée par

$$x_n = 1 + \frac{1}{1!} + \cdots + \frac{1}{n!} \tag{1.127}$$

est de Cauchy et ne converge pas dans $\mathbb{Q}$.

30. Voir [8], et attention : ce n'est pas tout à fait évident.

Démonstration. Si $p > q > 0$ nous avons

$$x_p - x_q = \sum_{k=q+1}^{p} \frac{1}{k!} \tag{1.128a}$$

$$\leqslant \sum_{k=q+1}^{p} \frac{1}{(q+1)!} \frac{1}{(q+1)^{k-q-1}} \tag{1.128b}$$

$$\leqslant \frac{1}{(q+1)!} \lim_{p\to\infty} \sum_{k=0}^{p} \frac{1}{(q+1)^k} \tag{1.128c}$$

$$= \frac{1}{(q+1)!} \frac{1}{1-\frac{1}{q+1}} \tag{1.128d}$$

$$= \frac{1}{(q+1)!} \frac{q+1}{q} \tag{1.128e}$$

$$= \frac{1}{q!q}. \tag{1.128f}$$

Justifications :

— Pour (1.128b), il s'agit de remplacer dans $k!$ tous les facteurs plus grands que $(q+1)$ par $q+1$. Cela rend le dénominateur plus petit.

— Pour (1.128c), il y a une inégalité parce que la suite $p \mapsto \sum_{k=0}^{p} 1/(q+1)^k$ est une suite strictement croissante.

— Pour (1.128d), le lemme 1.124 donne la valeur de la somme finie. En ce qui concerne la limite, nous avons demandé $p > q > 0$ et donc $q + 1 > 1$. Dans ce cas la limite fonctionne.

Cette inégalité une fois établie nous permet de prouver les assertions. La suite (x_n) est de Cauchy car, pour tout $\epsilon \in \mathbb{Q}$ s'écrivant $\epsilon = \frac{a}{b}$ avec $a, b \in \mathbb{N}$, en prenant $p, q > b$, nous avons

$$x_p - x_q \leqslant \frac{1}{b!b} < \frac{1}{b} < \frac{a}{b} = \epsilon. \tag{1.129}$$

Montrons par l'absurde que cette suite ne converge pas dans $\mathbb{Q}$. Pour cela, nous supposons que $\lim_{n\to\infty} x_n = \frac{a}{b} \in \mathbb{Q}$. Pour tout $p > q$ nous avons établi

$$0 < x_p - x_q < \frac{1}{qq!}. \tag{1.130}$$

Prenons la limite $p \to \infty$; par stricte croissance de la suite, les inégalités restent strictes :

$$0 < \frac{a}{b} - x_q < \frac{1}{qq!}. \tag{1.131}$$

Si $n > b$ alors nous pouvons écrire

$$\frac{a}{b} - x_n = \frac{\alpha}{n!} \tag{1.132}$$

avec $\alpha \in \mathbb{Z}$ parce que le dénominateur commun entre $\frac{a}{b}$ et x_n est dans $n!$. En prenant donc $q > n$ dans (1.131) nous pouvons écrire

$$0 < \frac{\alpha}{q!} < \frac{1}{qq!}, \tag{1.133}$$

c'est-à-dire $0 < \alpha < \frac{1}{q}$, ce qui est impossible pour $\alpha \in \mathbb{Z}$. □

Lemme 1.126.
Soit $A > 0$ dans $\mathbb{Q}$. Il existe un rationnel $q > 0$ tel que $q^2 < A$.

Démonstration. Vu que $\mathbb{Q}$ est archimédien (proposition 1.120), il existe $n \in \mathbb{N}$ tel que $1 < nA$. Pour ce n, nous avons

$$\left(\frac{1}{n}\right)^2 < \frac{1}{n} < A. \tag{1.134}$$

□

La proposition suivante donne une suite de rationnels qui convergerait dans $\mathbb{R}$ vers $\sqrt{A}$ (non encore défini à ce stade). Il est expliqué dans [3] que la suite est motivée par la méthode de Newton 33.53.

Proposition 1.127 ([3]).
Soient $A > 0$ dans $\mathbb{Q}$ et $x_0 \in \mathbb{Q}$. La suite (x_k) définie par

$$x_{k+1} = \frac{1}{2}\left(x_k + \frac{A}{x_k}\right) \tag{1.135}$$

a les propriétés suivantes :

(1) La suite $y_k = x_k^2$ converge dans $\mathbb{Q}$ vers A.

(2) La suite (x_k) est de Cauchy dans $\mathbb{Q}$.

(3) La suite (x_k) ne converge pas dans $\mathbb{Q}$ dans le cas de $A = 2$.

Démonstration. En plusieurs points.

La suite s_k En posant $y_k = x_k^2$ nous calculons que

$$y_{k+1} - A = \frac{(y_k - A)^2}{4y_k}. \tag{1.136}$$

Autrement dit, la suite $s_k = y_k - A$ vérifie

$$s_{k+1} = \frac{s_k^2}{4(A + s_k)}. \tag{1.137}$$

Quelle que soit la valeur de $s_0 = x_0^2 - A$, nous avons

$$s_1 = \frac{s_0^2}{4(A + s_0)} = \frac{(x_0^2 - A)^2}{4(A + x_0^2 - A)} = \frac{(x_0^2 - A)^2}{4x_0^2} > 0. \tag{1.138}$$

Donc à partir de s_1, tous les éléments sont positifs. Vu que $A > 0$ nous avons aussi

$$s_{k+1} < \frac{s_k^2}{4s_k} = \frac{s_k}{4} \tag{1.139}$$

et donc $s_k < s_0/4^k$. Donc $s_k \to 0$.

La suite (y_k) Nous venons de prouver que si $y_k = A + s_k$, alors $s_k \to 0$. Autrement dit, la suite y_k converge vers A dans $\mathbb{Q}$.

La suite (y_k) est donc de Cauchy par la proposition 1.121(1).

La suite (x_k) est de Cauchy Soit $\epsilon > 0$ dans $\mathbb{Q}$. Vu que (y_k) est de Cauchy, il existe $n_0 \in \mathbb{N}$ tel que

$$|x_r^2 - x_s^2| < \epsilon \tag{1.140}$$

pour tout $r, s \geqslant n_0$.

Soit par ailleurs $q \neq 0$ dans $\mathbb{Q}$ tel que $q^2 < A$, assuré par le lemme 1.126. Quitte augmenter la valeur de n_0, nous supposons que $x_r, x_s > q$, et en particulier que $x_r + x_s \neq 0$. Cela permet d'écrire d'abord

$$x_r^2 - x_s^2 = (x_r + x_s)(x_r - x_s) \tag{1.141}$$

et ensuite de prendre la valeur absolue et de diviser par $|x_r + x_s|$:

$$|x_r - x_s| = \frac{|x_r^2 - x_s^2|}{|x_r + x_s|} < \frac{\epsilon}{2q}. \tag{1.142}$$

Donc (x_k) est une suite de Cauchy.

Pas de convergence pour $A = 2$ Supposons que $x_k \to a \in \mathbb{Q}$. Dans ce cas nous aurions $x_k^2 \to a^2 = A = 2$ (proposition 1.121(5)). Mais nous savons par la proposition 1.123 que $a^2 = 2$ est impossible dans $\mathbb{Q}$.

□

Notons que cette proposition ne présume en rien de l'existence ou de la non-existence dans $\mathbb{Q}$ d'un élément qui pourrait décemment être nommé $\sqrt{A}$. Il se fait que le théorème 3.43 dira que $\sqrt{n}$ est soit entier soit irrationnel.

1.128.
Un petit programme en python explorer la suite de la proposition 1.127.

```python
#!/usr/bin/python3

def rec(A,x):
    return ((x**2+A)/x)/2

A = 3        # Compute square root of 3
x = 1000     # Initial guess: 1000

for i in range(1,100):
    print(i, x, x**2, x**2-A)
    x = rec(A, x)
```

tex/frido/codeSnip_4.py

1.10 Les réels

Une construction des réels via les coupures de Dedekind est donnée dans [25].

1.129.
La construction des réels va nécessiter un petit « bootstrap » au niveau de la topologie. En effet la notion de suite de Cauchy est une notion topologique (définition 7.274) alors que la topologie métrique (celle entre autres de $\mathbb{Q}$) ne sera définie que par le théorème 7.94. Nous avons donc dû définir en la définition 1.109 *ex nihilo* les notions de

— suite de Cauchy
— suite convergente
— complétude

Nous allons ensuite construire $\mathbb{R}$ comme ensemble de classes d'équivalence de suites de Cauchy dans $\mathbb{Q}$. Ce ne sera que plus tard, après avoir défini la notion d'espace métrique que nous allons voir que sur $\mathbb{R}$, ces trois notions coïncident avec celles topologiques [31]. Et par conséquent que $\mathbb{R}$ sera un espace métrique complet [32].

Dans cette optique, il est intéressant de lire ce que dit Wikipédia à propos des suites de Cauchy dans l'article consacré à la construction des nombres réels[26].

1.10.1 L'ensemble

Soit $\mathcal{E}$ l'ensemble des suites de Cauchy [33] dans $\mathbb{Q}$. Soit aussi l'ensemble $\mathcal{E}_0$ constituée des suites qui convergent vers zéro [34].

31. Proposition 7.281.
32. Théorème 1.153 pour la complétude en tant que corps et théorème 1.115 pour la complétude en tant que espace métrique.
33. Définition 1.109(3)
34. Nous rappelons qu'à ce niveau nous n'avons pas encore prouvé que toutes les suites de Cauchy convergent.

En posant

$$x + y = (x_n + y_n) \tag{1.143}$$

et

$$xy = (x_n y_n), \tag{1.144}$$

l'ensemble $\mathcal{E}$ devient un anneau[35] commutatif dont le neutre de l'addition est la suite constante $x_n = 0$ et le neutre pour la multiplication est la suite constante $x_n = 1$.

Proposition 1.130.
La partie $\mathcal{E}_0$ est un idéal[36] de l'anneau $\mathcal{E}$.

Démonstration. Nous savons par la proposition 1.121(1) que les suites convergentes sont de Cauchy ; par conséquent $\mathcal{E}_0 \subset \mathcal{E}$.

L'ensemble structuré $(\mathcal{E}_0, +)$ est un sous-groupe de $\mathcal{E}$ par les propriétés de la proposition 1.121 (il s'agit du fait que la somme de deux suites convergent vers zéro est une suite convergente vers zéro).

En ce qui concerne la propriété fondamentale des idéaux, si $x \in \mathcal{E}_0$ et $y \in \mathcal{E}$ nous devons prouver que $xy \in \mathcal{E}_0$. Vu que $(\mathcal{E}_0, \cdot)$ est commutatif, cela suffira pour être un idéal bilatère. Vu que y est une suite de Cauchy, elle est bornée ; et étant donné que $x \to 0$ nous avons alors $xy \to 0$ (par la proposition 1.121(3)). □

Théorème-définition 1.131 (L'anneau des réels[8])**.**
Sur l'ensemble quotient $\mathcal{E}/\mathcal{E}_0$, les opérations

(1) $\bar{u} + \bar{v} = \overline{u + v}$

(2) $\bar{u} \cdot \bar{v} = \overline{uv}$

sont bien définies et donnent à $\mathcal{E}/\mathcal{E}_0$ une structure de corps commutatif appelé ***corps des réels*** *et noté $\mathbb{R}$*

Démonstration. Nous divisons la preuve en plusieurs parties.

Les opérations sont bien définies La partie $\mathcal{E}_0$ est un idéal par la proposition 1.130. Le quotient est donc bien défini et est un anneau par la proposition 1.80(2).

Caractérisation des classes Soit $q \in \mathbb{Q}$ et une suite x convergente vers q. Cette suite est de Cauchy comme toute suite convergente. Montrons que

$$\bar{x} = \{\text{suites qui convergent vers } q\}. \tag{1.145}$$

Si $y \in \bar{x}$ alors $y = x + h$ avec $h \in \mathcal{E}_0$, et comme $h_n \to 0$, on a $y_n \to q$. Réciproquement, si $y_n \to q$ alors pour chaque n nous avons

$$y_n = x_n + (y_n - x_n), \tag{1.146}$$

mais $y_n - x_n \to 0$. Donc la suite $y - x \in \mathcal{E}_0$ ce qui signifie que $y \in \bar{x}$.

Neutre et unité Il est vite vérifié que $\bar{0}$, la classe de la suite constante égale à zéro est neutre pour l'addition. De même, $\bar{1}$, est un neutre pour la multiplication.

Corps Nous devons prouver que tout élément non nul est inversible. C'est-à-dire que si $x \in \mathcal{E}$ ne converge pas vers zéro[37] alors il existe $y \in \mathcal{E}$ tel que $xy \in \bar{1}$.

Nous savons par la proposition 1.121(6) que x étant une suite de Cauchy dans $\mathbb{Q}$, il existe $n_0 \in \mathbb{N}$ tel que $\left(\frac{1}{x_n}\right)_{n \geqslant n_0}$ est une suite de Cauchy. Nous posons alors

$$y_n = \begin{cases} 0 & \text{si } n \leqslant n_0 \\ \frac{1}{x_n} & \text{si } n > n_0. \end{cases} \tag{1.147}$$

35. Définition 1.73.
36. Définition 1.79.
37. $x \in \mathcal{E}$ peut soit ne pas converger du tout, soit converger vers autre chose que zéro.

Nous avons alors

$$(xy)_n = \begin{cases} 0 & \text{si } n \leqslant n_0 \\ 1 & \text{si } n > n_0 \end{cases} \tag{1.148}$$

et donc $xy \in \bar{1}$.

□

1.132 (Quelques notations entre $\mathbb{Q}$ et $\mathbb{R}$)**.**
Si $k \mapsto x_k$ est une suite, nous notons (x_k) avec des parenthèses la suite elle-même. Le k dans (x_k) est un indice muet, et dans les cas où il peut y avoir une ambiguïté, nous pouvons noter $(x_k)_{k\in\mathbb{N}}$. Cette dernière notation est plus lourde, mais plus exacte.

Le mieux est d'écrire simplement x la suite, mais alors il faut être prudent et ne pas noter x la limite. Nous éviterons donc d'écrire $x_k \to x$.

Si (x_k) est une suite de Cauchy dans $\mathbb{Q}$, nous notons $\bar{x}$ l'élément de $\mathbb{R}$ qui lui correspond. En fait $\bar{x} = (x_k)$: $\bar{x}$ est la suite-elle même, mais pour nous souvenir de l'origine nous allons adopter cette notation.

D'autre part nous définissons

$$\begin{aligned} \varphi\colon \mathbb{Q} &\to \mathbb{R} \\ q &\mapsto \overline{[k \mapsto q]}, \end{aligned} \tag{1.149}$$

c'est-à-dire que $\varphi(q)$ est la classe de la suite constante $x_k = q$.

Proposition 1.133.
Soit l'application

$$\begin{aligned} \varphi\colon \mathbb{Q} &\to \mathbb{R} \\ q &\mapsto \bar{q}. \end{aligned} \tag{1.150}$$

où par $\bar{q}$ nous entendons la classe de la suite constante égale à q (qui est de Cauchy).

(1) C'est un homomorphisme de corps injectif.

(2) Image(φ) *est un sous-corps de* $\mathbb{R}$

(3) $\varphi\colon \mathbb{Q} \to \text{Image}(\varphi)$ *est un isomorphisme de corps.*

Démonstration. Le fait que ce soit un homomorphisme est simplement

— $\varphi(q + q') = \overline{q + q'} = \bar{q} + \overline{q'}$

— $\varphi(qq') = \overline{qq'} = \bar{q}\overline{q'}$.

En ce qui concerne l'injectivité, si q est tel que $\varphi(q) = \bar{0} = \mathcal{E}_0$, c'est que

$$\varphi(q) = \{\text{suites de Cauchy qui convergent vers zéro}\} \tag{1.151}$$

Mais nous savons aussi que[38]

$$\varphi(q) = \bar{q} = \{\text{suites de Cauchy qui convergent vers } q\} \tag{1.152}$$

Nous en déduisons que $q = 0$. □

Lorsque dans la suite nous parlerons d'un élément de $\mathbb{Q}$ comme étant un réel, nous aurons en tête l'image de cet élément par φ.

38. Voir dans la démonstration du théorème 1.131.

1.10.2 Relation d'ordre

Nous définissons les parties $\mathcal{E}^+$ et $\mathcal{E}^-$ de $\mathcal{E}$ par

(1) $x \in \mathcal{E}^+$ si et seulement si pour tout $\epsilon > 0$, il existe N_ϵ tel que $n > N_\epsilon$ implique $x_n > -\epsilon$.

(2) $x \in \mathcal{E}^-$ si et seulement si pour tout $\epsilon > 0$, il existe N_ϵ tel que $n > N_\epsilon$ implique $x_n < \epsilon$.

Nous notons aussi $\mathcal{E}^{++} = \mathcal{E}^+ \backslash \mathcal{E}_0$.

Lemme 1.134.
Les parties $\mathcal{E}^+$ et $\mathcal{E}^-$ partitionnent $\mathcal{E}$ de la façon suivante :

(1) $\mathcal{E}^+ \cap \mathcal{E}^- = \mathcal{E}_0$

(2) $\mathcal{E}^+ \cup \mathcal{E}^- = \mathcal{E}$

Démonstration. On prouve d'abord que $\mathcal{E}^+ \cap \mathcal{E}^- \subset \mathcal{E}_0$, l'inclusion inverse est évidente. Soit $\epsilon > 0$ et $x \in \mathcal{E}^+ \cap \mathcal{E}^-$. Il existe un $N \in \mathbb{N}$ tel que $x_n > -\epsilon$ et $x_n < \epsilon$ pour tout $n \geqslant N$. Par conséquent, $|x_n| \leqslant \epsilon$ pour tout $n \geqslant N$ et la suite x converge vers zéro, c'est-à-dire $x \in \mathcal{E}_0$.

Pour prouver le second point, soit $x \in \mathcal{E}\backslash\mathcal{E}^-$, et prouvons que $x \in \mathcal{E}^+$. La condition $x \notin \mathcal{E}^-$ donne qu'il existe un $\alpha > 0$ (dans $\mathbb{Q}$) tel que pour tout n, il existe $p > n$ avec $x_p > \alpha$. Mais x est une suite de Cauchy, donc nous avons un n_0 tel que si $n, p \geqslant n_0$ alors $|x_n - x_p| \leqslant \frac{\alpha}{2}$. En particulier, si $n \geqslant n_0$, et si $p > n$ est tel que $x_p > \alpha$, on obtient

$$x_n > \frac{\alpha}{2} > 0 \tag{1.153}$$

Par conséquent $x \in \mathcal{E}^+$ parce que $x \in \mathcal{E}$ et les x_n sont tous positifs à partir d'un certain rang. □

Lemme 1.135 ([8])**.**
Quelques propriétés du partitionnement.

(1) $x \in \mathcal{E}^-$ *si et seulement si* $(-x) \in \mathcal{E}^+$

(2) $x \in \mathcal{E}^+$ *et* $y \in \mathcal{E}^+$ *implique* $x + y \in \mathcal{E}^+$

(3) $x \in \mathcal{E}^+$ *et* $y \in \mathcal{E}^+$ *implique* $xy \in \mathcal{E}^+$

(4) Si $x, y \in \mathcal{E}$ *sont tels que* $x - y \in \mathcal{E}_0$ *alors soit* $x, y \in \mathcal{E}^+$ *soit* $x, y \in \mathcal{E}^-$.

Démonstration. Point par point.

(1) Définition de $\mathcal{E}^+$ et $\mathcal{E}^-$.

(2) Pour $n \geqslant N_{\epsilon/2}$ nous avons $x_n > -\epsilon/2$ et $y_n > -\epsilon/2$. Donc $x_n + y_n > -\epsilon$.

(3) Si x ou y est dans $\mathcal{E}_0$ alors $xy \in \mathcal{E}_0$ et c'est bon. Si par contre $x, y \in \mathcal{E}^{++}$ alors pour n suffisamment grand, $x_n > 0$ et $y_n > 0$. Et dans ce cas, $(xy)_n > 0$, c'est-à-dire $xy \in \mathcal{E}^+$.

(4) Supposons que $x - y \in \mathcal{E}_0$ avec $x \in \mathcal{E}^+$ et prouvons qu'alors $y \in \mathcal{E}^+$. Soit donc $\epsilon > 0$; il existe n_1 tel que $x_n > -\frac{\epsilon}{2}$ dès que $n \geqslant n_1$. Mais $x - y \in \mathcal{E}_0$, donc il existe n_2 tel que $|x_n - y_n| < \frac{\epsilon}{2}$ dès que $n \geqslant n_2$. En prenant n plus grand que n_1 et n_2, nous avons en même temps

$$\begin{cases} x_n > -\dfrac{\epsilon}{2} & (1.154a) \\ |x_n - y_n| < \dfrac{\epsilon}{2}. & (1.154b) \end{cases}$$

Cela implique que $y_n > -\epsilon$ et donc que $y \in \mathcal{E}^+$.

Nous pouvons de même prouver que si $x \in \mathcal{E}^-$ alors $y \in \mathcal{E}^-$.

□

Définition 1.136 (Positivité dans $\mathbb{R}$)**.**
Vocabulaire et notations.

(1) Nous notons $\mathbb{R} = \mathcal{E}/\mathcal{E}_0$.

(2) Nous notons $\mathbb{R}^+ = \mathcal{E}^+$.

(3) *Nous notons* $\mathbb{R}^- = \mathcal{E}^-$.

(4) *Un élément de* $\mathbb{R}$ *est* ***positif*** *s'il est la classe d'une suite de Cauchy appartenant à* $\mathcal{E}^+$.

(5) *Un élément de* $\mathbb{R}$ *est* ***négatif*** *s'il est la classe d'une suite de Cauchy appartenant à* $\mathcal{E}^-$.

(6) *Lorsque nous parlons de nombres réels, le symbole « 0 » signifie* $\mathcal{E}_0$ *ou plus précisément la classe d'un élément de* $\mathcal{E}_0$ *modulo* $\mathcal{E}_0$.

1.137.
Avec les conventions de la définition 1.136, et en anticipant sur nos connaissances à propos des réels,

(1) zéro est positif et négatif.

(2) L'intersection entre $\mathbb{R}^+$ et $\mathbb{R}^-$ est le singleton $\{0\}$.

(3) L'ensemble des nombres *strictement* positifs est noté $(\mathbb{R}^+)^*$ ou $\mathbb{R}^+\backslash\{0\}$.

(4) Le mot « positif » signifie « positif ou nul » ; le mot « négatif » signifie « négatif ou nul ».

Cela vient des conventions de la remarque 1.137 qui sont également celles de Wikipédia[27].

Lemme-définition 1.138 (Ordre dans $\mathbb{R}$)**.**
Si $a, b \in \mathbb{R}$ *nous notons* $a \leqslant b$ *si et seulement si* $b - a$ *est positif. Nous notons aussi* $a > b$ *si et seulement si* $b - a \in \mathbb{R}^+\backslash\{0\}$, *etc.*

Avec ces définitions,

(1) *l'ensemble* $(\mathbb{R}, \leqslant)$ *est un corps totalement ordonné (définitions 1.8 et 1.109) ;*

(2) *l'application*

$$\begin{aligned} \varphi\colon \mathbb{Q} &\to \mathbb{R} \\ q &\mapsto \bar{q} \end{aligned} \tag{1.155}$$

dont nous avons déjà parlé dans la proposition 1.133 est strictement croissante.

Démonstration. Nous prouvons la stricte croissance de φ. Si $q < l$ alors $\varphi(q) - \varphi(l) = \overline{q - l}$ est la classe de la suite constante $q - l$ qui est un élément strictement positif de $\mathbb{Q}$. Nous avons donc $\overline{q - l} \in \mathbb{R}^+$, et donc $\varphi(q) < \varphi(l)$. □

Remarque 1.139.
Comme déjà mentionné plus haut, à chaque fois que nous parlerons d'un élément de $\mathbb{Q}$ comme étant un élément de $\mathbb{R}$, nous considérons la classe de la suite constante.

Lemme 1.140.
Si $x, y, z \in \mathbb{R}$ *avec* $x > 0$ *sont tels que* $z > y/x$ *alors* $zx > y$.

Démonstration. Nous savons que

$$z - \frac{y}{x} \in \mathcal{E}^+\backslash\{0\} = \mathcal{E}^{++}. \tag{1.156}$$

Vu que $x \in \mathcal{E}^{++}$, multiplier par x fait rester dans $\mathcal{E}^{++}$:

$$zx - x\frac{y}{x} \in \mathcal{E}^{++}. \tag{1.157}$$

Un représentant de $x\frac{y}{x}$ est la suite $n \mapsto x_n \frac{y_n}{x_n} = y_n$. Donc $x\frac{y}{x} = y$. Cela signifie que $zx - y \in \mathcal{E}^{++}$ et donc que $zx > y$. □

Lemme 1.141.
Pour tout $a \in \mathbb{R}$, *il existe* $p \in \mathbb{N}$ *tel que* $p > a$.

Démonstration. Nous allons donner deux preuves différentes de ce lemme.

Première façon L'élément a de $\mathbb{R}$ admet un représentant (a_n) qui est une suite de Cauchy dans $\mathbb{Q}$. C'est donc une suite bornée, c'est-à-dire qu'il existe $m, q \in \mathbb{N}$ tels que $|a_n| \leqslant m/q$ pour tout n (proposition 1.121(2)). Soit M un naturel strictement plus grand que m/q[39].

La suite de Cauchy $(M - a_n)_{n\in\mathbb{N}}$ est constituée de rationnels positifs et est donc dans $\mathcal{E}^+$. La classe de $M - a$ est donc un réel positif[40]. Par définition de la relation d'ordre, $M \geqslant a$.

Seconde façon La suite (a_n) est majorée par $\frac{m}{q}$, donc on a dans $\mathbb{Q}$ et pour tout n :

$$a_n \leqslant \frac{m}{q} = M \leqslant qM. \tag{1.158}$$

L'application $\varphi\colon \mathbb{Q} \to \mathbb{R}$ est croissante, donc

$$\varphi\big((a_n)\big) \leqslant \varphi(qM). \tag{1.159}$$

□

En corolaire, nous avons

Lemme 1.142.
Pour tout $x \in \mathbb{R}$, il existe $q \in \mathbb{Z}$ tel que $q < x$.

Démonstration. Utilisation du lemme précédent avec $a = -x$: on prend $q = -p$. □

Théorème 1.143 ([8]).
Le corps $\mathbb{R}$ est archimédien[41].

Démonstration. Le lemme 1.138 dit que $\mathbb{R}$ est totalement ordonné. Soient $x, y \in \mathbb{R}$ avec $x > 0$; posons $a = \frac{y}{x}$. Le lemme 1.141 nous donne un $p \in \mathbb{N}$ tel que $p > a$.Nous concluons alors avec le lemme 1.140 :

$$px > ax = \frac{y}{x}x = y. \tag{1.160}$$

□

Le lemme suivant n'est pas loin de dire que $\mathbb{Q}$ est dense dans $\mathbb{R}$, à part que nous n'avons pas encore donné de topologie sur $\mathbb{R}$.

Lemme 1.144.
Si $x, y \in \mathbb{R}$ sont tels que $x < y$, alors il existe $s \in \mathbb{Q}$ tel que $x < s < y$.

Démonstration. Nous avons par hypothèse que $y - x > 0$ et donc le fait que $\mathbb{R}$ soit archimédien (théorème 1.143) nous donne $q \in \mathbb{N}$ tel que $q(y - x) > 1$. Soit

$$E = \{n \in \mathbb{Z} \text{ tel que } \frac{n}{q} \leqslant x\}. \tag{1.161}$$

Cet ensemble n'est pas vide à cause du lemme 1.142 ; de plus, comme $|x|q \leqslant n_0$ pour un certain n_0 (à cause du lemme 1.141), l'ensemble E est majoré par n_0. Donc E possède un plus grand élément[42] p qui vérifie

$$\frac{p}{q} \leqslant x < \frac{p+1}{q}. \tag{1.162}$$

De plus $(p + 1)/q < y$. En effet nous avons

$$\frac{p+1}{q} = \frac{p}{q} + \frac{1}{q} \leqslant x + \frac{1}{q} < x + y - x = y \tag{1.163}$$

39. Lemme 1.117.
40. Et nous allons d'ailleurs arrêter de toujours préciser « la classe de » lorsque ce n'est pas nécessaire.
41. Définition 1.109(5).
42. Lemme 1.83.

où nous avons utilisé l'inégalité stricte $y - x > \frac{1}{q}$.

Nous avons donc

$$x < \frac{p+1}{q} < y, \tag{1.164}$$

et le nombre $(p+1)/q$ convient comme s. □

Remarque 1.145.
Le lemme 1.144 a également pour conséquence que des ensembles comme $[-1, 1]$ ne sont pas bien ordonnés (définition 1.8). En effet la partie $]0, 1[$ ne possède pas de minimum parce que si $x \in]0, 1[$ alors $0 < x$ et il existe $s \in \mathbb{Q}$ (a fortiori $s \in \mathbb{R}$) tel que $0 < s < x$, c'est-à-dire que x n'est pas un minimum de $]0, 1[$.

Tant que nous y sommes dans les encadrements de réels...

1.146.
Soit $q_0 \in \mathbb{Q}$ tel que $0 \leqslant q_0 < 1$. On définit alors $d_1 \in \{0, 1\}$ comme valant 1 si $2q_0 \geqslant 1$ et 0 sinon. Puis on pose $q_1 = 2q_0 - d_1$.

Poursuivant de la sorte, on crée une suite $(d_n)_{n\geqslant 1}$: c'est le **développement dyadique** de q_0.

Lemme 1.147 ([1])**.**
Soit q, q' deux rationnels tels que $0 \leqslant q < q' < 1$. Il existe deux entiers naturels a et N tels que $q < \frac{a}{2^N} < q'$.

Démonstration. On crée les développements dyadiques de q et q', que l'on note respectivement $(d_n)_{n\geqslant 1}$ et $(d'_n)_{n\geqslant 1}$. Notons

$$E = \{n \in \mathbb{N} \text{ tel que } d_n \neq d'_n\}. \tag{1.165}$$

Comme $q \neq q'$, les développements dyadiques sont différents [43], l'ensemble E est non-vide, et il admet un plus petit élément N. Or, $q < q'$, et donc nécessairement $d_N < d'_N$. On construit alors $a = \sum_{i=1}^{N} d_i 2^i$. □

Corolaire 1.148.
Pour tous réels x, y tels que $0 \leqslant x < y \leqslant 1$, il existe un nombre de la forme $d = a/2^n$, avec $n \in \mathbb{N}$ et $a \in \mathbb{N}$, $a \leqslant 2^n$, tel que $x < d < y$.

Lemme 1.149 ([1])**.**
Soient des réels a, b, x, y tels que

$$a \leqslant x \leqslant b \tag{1.166}$$

et

$$a \leqslant y \leqslant b, \tag{1.167}$$

alors $|x - y| \leqslant |b - a|$.

1.10.3 Complétude

Le théorème 16.125 donne une complétion de tout espace métrique en un espace complet. Il serait tentant de l'utiliser ici pour définir $\mathbb{R}$ à partir de $\mathbb{Q}$. Cette méthode ne fonctionne cependant pas parce que la démonstration de 16.125 utilise le fait que $\mathbb{R}$ est complet.

Lemme 1.150 ([8])**.**
Toute suite de Cauchy dans $\mathbb{Q}$ converge dans $\mathbb{R}$ vers le réel qu'elle représente.

Plus précisément, en suivant les notations de 1.132, si (x_k) est une suite de Cauchy dans $\mathbb{Q}$, alors la suite $\varphi(x_k)$ dans $\mathbb{R}$

(1) est de Cauchy dans $\mathbb{R}$,

(2) converge dans $\mathbb{R}$ vers $\bar{x}$.

43. À vérifier tout de même...

Démonstration. Soit (x_n) une suite de Cauchy de $\mathbb{Q}$, c'est-à-dire que $x_k \in \mathbb{Q}$ pour tout k et qu'elle est de Cauchy. Elle représente un réel $\bar{x} \in \mathbb{R}$, et nous voulons prouver que pour la topologie de $\mathbb{R}$ nous avons $\lim_{n\to\infty} x_n = \bar{x}$. Dans cette dernière limite, chacun des x_n est vu dans $\mathbb{R}$.

Si $\epsilon \in \mathbb{Q}$ est donné, il existe N_ϵ tel que si $p, q \geqslant N_\epsilon$ alors $|x_p - x_q| < \epsilon$, c'est-à-dire

$$x_p - \epsilon < x_q < x_p + \epsilon. \tag{1.168}$$

Soit $p \geqslant N_\epsilon$ fixé. Pour tout $q \geqslant N_\epsilon$ nous avons $x_p - \epsilon < x_q < x_p + \epsilon$. Par conséquent, la suite $n \mapsto (x_p - \epsilon) - q_n$ est un élément de $\mathcal{E}^-$ et au niveau des classes nous pouvons écrire

$$\overline{n \mapsto (x_p - \epsilon) - x_n} \leqslant 0. \tag{1.169}$$

Vu que $x_p - \epsilon$ représente la suite constante nous avons l'inégalité suivante dans $\mathbb{R}$:

$$x_p - \epsilon - \bar{x} \leqslant 0 \tag{1.170}$$

ou encore : $\bar{x} \geqslant x_p - \epsilon$. En faisant de même avec l'autre partie de l'inégalité, $x_p - \epsilon \leqslant \bar{x} \leqslant x_p + \epsilon$, ce qui implique que

$$x_p \in B(\bar{x}, \epsilon) \tag{1.171}$$

dès que $p \geqslant N_\epsilon$. Cela signifie que $x_p \to \bar{x}$ dans $\mathbb{R}$. □

Proposition 1.151.
Soit une suite convergente $x_k \xrightarrow{\mathbb{Q}} q$. Alors

$$\varphi(x_k) \xrightarrow{\mathbb{R}} \varphi(q) \tag{1.172}$$

où φ est la fonction qui à un rationnel fait correspondre la classe de la suite constante correspondante[44].

Démonstration. Le fait d'avoir une convergence $x_k \to q$ dans $\mathbb{Q}$ implique que la suite (x_k) est de Cauchy, par la proposition 1.121(1).

Le lemme 1.150 nous indique que $\varphi(x_k)$ est une suite dans $\mathbb{R}$ qui converge vers $\bar{q}$, la classe de la suite (x_k).

À prouver : $\varphi(x) = \bar{q}$. Autrement dit, nous devons prouver que la classe de la suite constante $a_k = q$ et la classe de la suite x sont les mêmes.

La suite $(x_k - q)$ est de Cauchy dans $\mathbb{Q}$ et converge vers zéro par hypothèse. Donc les suites x et (q) sont dans la même classe. □

Proposition 1.152 ([1])**.**
Deux choses à propos de suites de rationnels convergeant vers un réel.

(1) Soit un réel x. Il existe une suite de rationnels strictement croissante qui converge vers x.

(2) Si de plus $x > 0$, alors la suite (toujours strictement croissante) peut être choisie parmi les rationnels strictement positifs.

Démonstration. Le lemme 1.144 nous sera d'une grande aide. Soit $x \in \mathbb{R}$. Il existe $q_0 \in \mathbb{Q}$ tel que $x - 1 < q_0 < x$. Ensuite nous construisons la suite par récurrence : q_k est choisi tel que $q_{k-1} < q_k < x$. Cela règle le point (1).

Pour (2). Il suffit de faire la même chose, en partant de $0 < q_0 < x$. □

Théorème 1.153 (Complétude de $\mathbb{R}$, critère de Cauchy[8])**.**
Nous avons :

(1) Le corps $\mathbb{R}$ est un corps complet (définition 1.109(6))

(2) Une suite dans $\mathbb{R}$ est convergente (définition 1.109(4)) si et seulement si elle est de Cauchy (définition 1.109(3)).

44. Voir les notations en 1.132.

Notez la grande similitude entre ce théorème et le théorème 7.294. Ils ne sont pas équivalents, ne parlent pas exactement du même objet « $\mathbb{R}$ », ni des mêmes notions de suites de Cauchy et de complétude.

Démonstration. Soit (x_n) une suite de Cauchy dans $\mathbb{R}$. Pour chaque n, il existe par le lemme 1.144 un $y_n \in \mathbb{Q}$ tel que

$$x_n - \frac{1}{n} < y_n < x_n + \frac{1}{n}. \tag{1.173}$$

(y_n) est une suite de Cauchy dans $\mathbb{Q}$ Nous prouvons que (y_n) est une suite de Cauchy dans $\mathbb{Q}$ (définition 1.109(3)). Vu que (x_n) est de Cauchy pour le corps $\mathbb{R}$, si $\epsilon > 0$ dans $\mathbb{R}$ est donné, il existe n_ϵ tel que si $p, q \geqslant n_\epsilon$, alors $|x_p - x_q| < \epsilon$.

Nous avons :

$$|y_p - y_q| \leqslant |y_p - x_p| + |x_p - x_q| + |x_q - y_q| < \frac{1}{p} + \epsilon + \frac{1}{q}. \tag{1.174}$$

En choisissant $N_\epsilon > \max\{n_\epsilon, \frac{1}{\epsilon}\}$ (ce qui est possible par le lemme 1.141), et en prenant $p, q > N_\epsilon$, nous avons

$$|y_p - y_q| \leqslant 3\epsilon, \tag{1.175}$$

ce qui prouve que (y_p) est une suite de Cauchy dans $\mathbb{Q}$, pour la notion de suite de Cauchy dans $\mathbb{Q}$.

Le réel représenté Vu que (y_p) est de Cauchy dans $\mathbb{Q}$, elle représente un réel que nous notons $\bar{y}$.

Convergence de (x_n) Nous prouvons que $x_n \xrightarrow{\mathbb{R}} \bar{y}$.

Nous savons qu'une suite de Cauchy de rationnels converge dans $\mathbb{R}$ vers le réel qu'elle représente, c'est-à-dire : $y_n \xrightarrow{\mathbb{R}} \bar{y}$ où chaque $y_n \in \mathbb{Q}$ est vu comme la suite constante (cela est le lemme 1.150). Autrement dit, pour $\epsilon > 0$, il existe un $N_\epsilon \in \mathbb{N}$ tel que si $p > N_\epsilon$ alors $|\bar{y} - y_p| < \epsilon$. Pour un tel p nous avons

$$|\bar{y} - x_p| \leqslant |\bar{y} - y_p| + |y_p - x_p| \leqslant \epsilon + \frac{1}{p}. \tag{1.176}$$

Donc dès que p est plus grand que $\max\{N_\epsilon, \frac{1}{\epsilon}\}$, nous avons $|\bar{y} - x_p| < 2\epsilon$, ce qui signifie que la suite (x_n) converge vers $\bar{y}$ dans $\mathbb{R}$.

Ceci achève de prouver que $\mathbb{R}$ est un corps complet.

En ce qui concerne l'équivalence entre les suites convergentes et de Cauchy, nous venons de prouver que toute suite de Cauchy dans $\mathbb{R}$ est convergente. La réciproque est la proposition 1.115. □

Nous avons terminé avec la construction des réels. Les propriétés topologiques arrivent en la section 7.7. En particulier le théorème 7.294 pour la complétude de $\mathbb{R}$ en tant qu'espace métrique.

1.10.4 Intervalles

Nous avons déjà défini la notion d'intervalle pour un espace totalement ordonné en 1.13. Nous posons quelques notations dans $\mathbb{R}$.

Définition 1.154.
Soient $a \neq b$ dans $\mathbb{R}$. Nous définissons les parties suivantes de $\mathbb{R}$:

(1) $]a, b[= \{x \in \mathbb{R}$ *tel que* $a < x < b\}$

(2) $[a, b[= \{x \in \mathbb{R}$ *tel que* $a \leqslant x < b\}$

(3) $]a, b] = \{x \in \mathbb{R}$ *tel que* $a < x \leqslant b\}$

(4) $[a, b] = \{x \in \mathbb{R}$ *tel que* $a \leqslant x \leqslant b\}$

(5) $]-\infty, a] = \{x \in \mathbb{R}$ *tel que* $x \leqslant a\}$

(6) $]-\infty, a[= \{x \in \mathbb{R} \text{ tel que } x < a\}$

(7) $]a, \infty[= \{x \in \mathbb{R} \text{ tel que } x > a\}$

(8) $[a, \infty[= \{x \in \mathbb{R} \text{ tel que } x \geqslant a\}$.

(9) $]-\infty, \infty[= \mathbb{R}$.

La proposition 1.158 nous dira que tous les intervalles de $\mathbb{R}$ sont d'une de ces formes.

1.10.5 Maximum, supremum et compagnie

Ce n'est un secret pour personne que $\mathbb{R}$ est un ensemble totalement ordonné[45] : il y a des éléments plus grands que d'autres, et mieux : à chaque fois que je prends deux éléments différents dans $\mathbb{R}$, il y en a un des deux qui est plus grand que l'autre. Il n'y a pas d'*ex æquo* dans $\mathbb{R}$.

Définition 1.155.
Soit A, une partie de $\mathbb{R}$.

*(1) Un nombre M est un **majorant** de A si M est plus grand que tous les éléments de A : pour tout $x \in A$, $M \geqslant x$.*

*(2) Un nombre m est un **minorant** de A si m est plus petit que tous les éléments de A : pour tout $x \in A$, $m \leqslant x$.*

Nous parlons de majorant ou de minorants stricts *lorsque les inégalités sont strictes.*

Nous insistons sur le fait que l'inégalité n'est pas stricte. Ainsi, 1 est un majorant de $[0, 1]$. Dès qu'un ensemble a un majorant, il en a plein. Si s majore l'ensemble A, alors $s + 1$, $s + 4$, et $s + \frac{3}{7}$ majorent également A.

Exemple 1.156.
Une petite galerie d'exemples de majorants.

— L'intervalle fermé $[4, 8]$ admet entre autres 8 et 130 comme majorants,

— l'intervalle ouvert $]4, 8[$ admet également 8 et 130 comme majorants,

— 7 n'est pas un majorant de $[1, 5] \cup]8, 32]$,

— 10/10 majore les notes qu'on peut obtenir à un devoir.

— l'intervalle $[4, \infty[$ n'a pas de majorants.

△

Proposition-définition 1.157 (Least-upper-bound property[28])**.**
Soit A une partie majorée de $\mathbb{R}$. Il existe un unique élément $M \in \mathbb{R}$ tel que

(1) $M \geqslant x$ pour tout $x \in A$,

(2) pour tout ε, le nombre $M - \varepsilon$ n'est pas un majorant de a, c'est-à-dire qu'il existe un élément $x \in A$ tel que $x > M - \varepsilon$.

*Cet élément est nommé **supremum** de A et est noté* $\sup(A)$*. De la même façon, **l'infimum** de A, noté* $\inf A$*, est le plus grand de ses minorants.*

Par convention, si la partie n'est pas bornée vers le haut, nous dirons que son supremum n'existe pas, ou bien qu'il est égal à $+\infty$, suivant les contextes. Pour votre culture générale, sachez toutefois que $\infty \notin \mathbb{R}$.

Démonstration. Nous faisons la preuve pour l'infimum.

Unicité En ce qui concerne l'unicité, soient m_1 et m_2, deux infimums de A. Supposons $m_1 > m_2$. Alors il existe $\epsilon > 0$ tel que $m_2 < m_2 + \epsilon < m_1$ (c'est le lemme 1.144). Cela prouve que $m_2 + \epsilon$ est un minorant de A et donc que m_2 n'est pas un infimum.

45. Lemme 1.138(1).

Existence Soit A, une partie de $\mathbb{R}$. Nous allons trouver son infimum en suivant une méthode de dichotomie. Pour cela nous allons construire trois suites en même temps de la façon suivante. D'abord nous choisissons un point x_0 de A et un point x_1 qui minore A (qui existe par hypothèse) :

$$\begin{aligned} &x_0 \text{ est un élément de } A,\\ &x_1 \text{ est un minorant de } A,\\ &a_0 = x_0\\ &b_0 = x_1\\ &b_1 = x_1. \end{aligned} \tag{1.177}$$

Ensuite, nous faisons la récurrence suivante :

$$\begin{aligned} x_{n+1} &= \frac{a_n + b_n}{2},\\ a_{n+1} &= \begin{cases} a_n & \text{si } x_{n+1} \text{ minore } A\\ x_{n+1} & \text{sinon,} \end{cases}\\ b_{n+1} &= \begin{cases} x_{n+1} & \text{si } x_{n+1} \text{ minore } A\\ b_n & \text{sinon.} \end{cases} \end{aligned} \tag{1.178}$$

Nous allons montrer que (a_n) et (b_n) sont des suites convergentes de même limite et que cette limite est l'infimum de A.

Soit $n \in \mathbb{N}$; il y a deux possibilités. Soit $a_n = a_{n-1}$ et $b_n = x_n$, soit $a_n = x_n$ et $b_n = b_{n-1}$. Supposons que nous soyons dans le premier cas (le second se traite de façon similaire). Alors nous avons

$$\begin{aligned} |a_n - b_n| &= |a_{n-1} - x_n|\\ &= \left|a_{n-1} - \frac{a_{n-1} + b_{n-1}}{2}\right|\\ &= \frac{1}{2}|a_{n-1} - b_{n-1}|, \end{aligned} \tag{1.179}$$

ce qui prouve que $|a_n - b_n| \to 0$. Nous montrons maintenant que la suite (a_n) est de Cauchy. En effet nous avons

$$|a_n - a_{n-1}| = \begin{cases} 0 \\ \left|\frac{a_n - b_n}{2}\right| \end{cases} \leqslant \frac{1}{2n}. \tag{1.180}$$

Il en est de même pour la suite (b_n). Ce sont deux suites de Cauchy (donc convergentes par la proposition 1.115) qui convergent vers la même limite. Soit ℓ cette limite.

Le nombre ℓ minore A. En effet si $a \in A$ est plus petit que ℓ, les éléments b_n tels que $|b_n - \ell| < |a - \ell|$ ne peuvent pas minorer A. D'autre part, pour tout ϵ, le nombre $\ell + \epsilon$ ne peut pas minorer A. En effet, ℓ est la limite de la suite décroissante (a_n), donc il existe a_n entre ℓ et $\ell + \epsilon$. Mais a_n ne minore pas A, donc $\ell + \epsilon$ ne minore pas non plus A.

Nous avons prouvé que toute partie minorée de $\mathbb{R}$ possède un infimum.

La preuve que toute partie majorée possède un supremum se fait de la même façon. □

1.10.5.1 Intervalles

Proposition 1.158.
Tous les intervalles de $\mathbb{R}$ sont d'une des formes listées dans la définition 1.154.

1.10.5.2 Quelques exemples

En matière de notations, le maximum de l'ensemble A est noté $\max A$, le supremum est noté $\sup A$. Le minimum et l'infimum sont notés $\min A$ et $\inf A$.

Exemple 1.159.
Exemples de différence entre majorant, supremum et maximum.

— Le nombre 10 est un supremum, majorant et maximum de l'intervalle fermé $[0, 10]$,

— Le nombre 10 est un majorant et un supremum, mais pas un maximum de l'intervalle ouvert $]0, 10[$,

— Le nombre 136 est un majorant, mais ni un maximum ni un supremum de l'intervalle $[0, 10]$.

△

En utilisant les notations concises, ces différents cas s'écrivent ainsi :

$$10 = \max[0, 10] = \sup[0, 10] \quad 10 = \sup[0, 10[\tag{1.181}$$

Exemple 1.160.
Si on dit que un pont s'effondre à partir d'une charge de 10 tonnes, alors 10 tonnes est un *supremum* des charges que le pont peut supporter : si on met $9,999999$ tonnes dessus, il tient encore le coup, mais si on ajoute un gramme, alors il s'effondre (on sort de l'ensemble des charges acceptables). △

Exemple 1.161.
Si on dit qu'un pont résiste jusqu'à 10 tonnes, alors 10 tonnes est un *maximum* de la charge acceptable. Sur ce pont-ci, on peut ajouter le dernier gramme. Mais à partir de là, le moindre truc qu'on ajoute, il s'effondre. △

Lemme 1.162.
À propos de bornes d'un intervalle.

(1) La borne inférieure d'un intervalle est son infimum,

(2) la borme supérieure est le supremum.

(3) Si de plus l'intervalle est fermé, l'infimum est un minimum et le supremum est un maximum.

Exemple 1.163.
Quelques exemples dans les intervalles.

(1) $A = [1, 2]$. Tous les nombres plus petits ou égaux à 1 sont minorants, 1 est infimum et minimum. Le nombre 2 est un majorant, le maximum et le supremum.

(2) $B =]3, \pi[$. Le nombre π est le supremum et est un majorant, mais n'est pas le maximum (parce que $\pi \notin B$). L'ensemble B n'a pas de maximum. Bien entendu, -1000 est un minorant.

Dans les deux cas, le nombre 53 est un majorant. △

Il existe évidemment de nombreux exemples plus vicieux.

Exemple 1.164.
Prenons $E = \{\frac{1}{n} \text{ tel que } n \in \mathbb{N}_0\}$, dont les premiers points sont indiqués sur la figure 1.1. Cet ensemble est constitué des nombres $1, \frac{1}{2}, \frac{1}{3}, \ldots$ Le plus grand d'entre eux est 1 parce que tous les nombres de la forme $\frac{1}{n}$ avec $n \geqslant 1$ sont plus petits ou égaux à 1. Le nombre 1 est donc maximum de E.

L'ensemble E n'a par contre pas de minimum parce que tout élément de E s'écrit $\frac{1}{n}$ pour un certain n et est plus grand que $\frac{1}{n+1}$ qui est également dans E.

Prouvons que zéro est l'infimum de E. D'abord, tous les éléments de E sont strictement positifs, donc zéro est certainement un minorant de E. Ensuite, nous savons que pour tout $\varepsilon > 0$, il existe un n tel que $\frac{1}{n}$ est plus petit que ε. L'ensemble E possède donc un élément plus petit que $0 + \varepsilon$, et zéro est bien l'infimum. △

L'exemple suivant est une source classique d'erreurs en ce qui concerne l'infimum. Il sera à relire après avoir vu la définition de limite (définition 7.129).

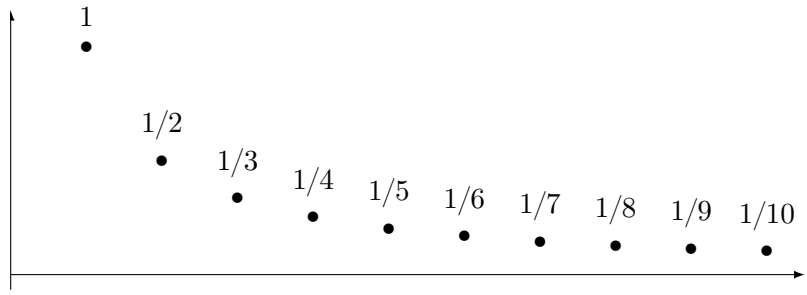


FIGURE 1.1 – Les premiers points du type $x_n = 1/n$.

Exemple 1.165.
Les premiers points de l'ensemble $F = \{\frac{(-1)^n}{n}$ tel que $n \in \mathbb{N}_0\}$ sont représentés à la figure 1.2. Bien que (comme nous le verrons plus tard) la limite de la suite $x_n = (-1)^n/n$ soit zéro, il n'est pas correct de dire que zéro est l'infimum de l'ensemble F. Le dessin, au contraire, montre bien que -1 est le minium (aucun point est plus bas que -1), tandis que le maximum est $1/2$.

Nous reviendrons avec cet exemple dans la suite. Pour l'instant, ayez bien en tête que zéro n'est rien de spécial pour l'ensemble F en ce qui concerne les notions de maximum, minimum et compagnie. △

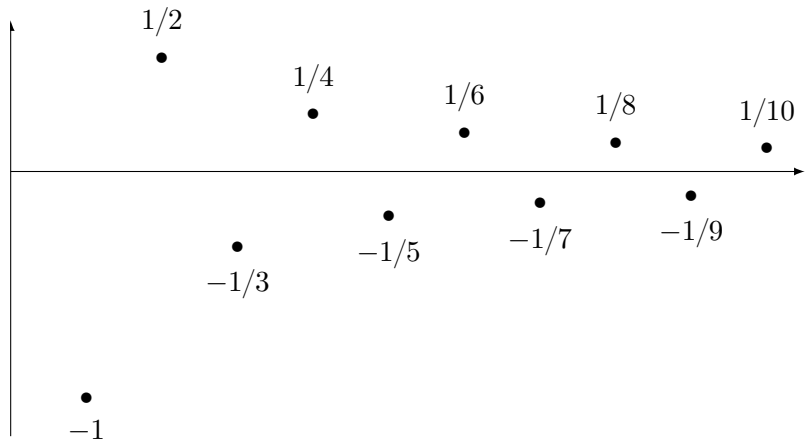


FIGURE 1.2 – Les quelques premiers points du type $(-1)^n/n$.

1.10.6 Racines

Dans cette section, nous définissons $\sqrt{x}$ pour $x \in \mathbb{R}^+$. Vous notez que c'est fait de façon assez algébrique[46], ou en tout cas, en restant proche des définitions. Des définitions plus technologiques utilisant la continuité de $x \mapsto x^n$ et qui prouvent que c'est bijectif sur un domaine choisi avec prudence existent, et c'est fait dans la définition 11.325. Il est même expliqué dans [3] que la méthode décrite ici permet de définir $\sqrt[n]{x}$ pour tout n entier, et pas seulement pour $n = 2$.

Proposition 1.166.
Soit $q \in \mathbb{Q}^+$. Il existe un unique $r \in \mathbb{R}$ tel que $r^2 = q$.

Plus précisément, en termes des notations de 1.132, pour tout $q \in \mathbb{Q}^+$, il existe un unique $r \in \mathbb{R}^2$ tel que $r^2 = \varphi(q)$.

Démonstration. En deux parties : d'abord l'existence et ensuite l'unicité.

Existence Si $q = 0$, c'est $r = 0$. Nous supposons $q > 0$. La suite (x_k) de la proposition 1.127 a la propriété d'être de Cauchy dans $\mathbb{Q}$. Donc il existe un réel r qui est la classe de cette suite.

46. Discutable parce que des limites sont utilisées.

Nous posons donc

$$r = \bar{x}. \tag{1.182}$$

En ce qui concerne r^2, nous avons, par définition du produit dans $\mathbb{R}$,

$$r^2 = \bar{x}^2 = \overline{(x_k^2)}, \tag{1.183}$$

c'est la classe de la suite de Cauchy donnée par les x_k^2. Posons $y_k = x_k$; la relation (1.183) s'écrit

$$r^2 = \bar{y}. \tag{1.184}$$

La proposition 1.127 nous dit également que y est une suite de Cauchy et que

$$y_k \xrightarrow{\mathbb{Q}} q \tag{1.185}$$

La proposition 1.151 donne alors $\bar{y} = \bar{q}$, et finalement

$$r^2 = \bar{q} = \varphi(q). \tag{1.186}$$

Ici tout n'est pas encore terminé avec l'existence parce qu'il faut nous assurer que $r \geqslant 0$. Ce n'est pas très compliqué : si $r < 0$, alors nous pouvons faire le choix $-r$ qui convient tout aussi bien : $(-r)^2 = r^2$.

Unicité Supposons $r_1, r_2 \in \mathbb{R}$ tels que $r_1^2 = r_2^2$. Le lemme 1.138 dit que $\mathbb{R}$ est totalement ordonné ; disons pour fixer les idées que $r1 \leqslant r_2$. Cela signifie, par définition de l'ordre sur $\mathbb{R}$, que $r_2 - r_1 \geqslant 0$. En posant $s = r_2 - r_1$ nous avons $r_2 = r_1 + s$. Passons au carré ; la distribution dans le calcul suivant provient du fait que $\mathbb{R}$ est un corps :

$$r_2^2 = (r_1 + s)^2 = r_1^2 + 2r_1 s + s^2. \tag{1.187}$$

Vu que $r_1^2 = q = r_2^2$, nous avons $2r_1 s + s^2 = 0$ ou encore

$$s(2r_1 + s) = 0. \tag{1.188}$$

Vu que $\mathbb{R}$ est un corps, il est un anneau intègre[47] et la règle du produit nul s'applique : soit $s = 0$, soit $2r_2 + s = 0$. Vu que $r_2 > 0$ et que $s \geqslant 0$, nous avons $2r_2 + s > 0$ et donc $s = 0$.

Nous en déduisons que $r_1 = r_2$.

□

1.11 Les complexes

Problèmes et choses à faire

Encore une fois, cette section n'est pas du tout faite. Le but de cette section serait de

- Construire $\mathbb{C}$ en tant qu'ensemble
- Construire les opérations courantes.
- Démontrer les propriétés de base.

Attention : il est sans espoir de parler de forme trigonométrique ici parce que les exponentielles et fonctions trigonométriques ne sont définies qu'avec les séries.

Nous donnons ici en vrac quelques propriétés et définitions que se doivent d'être présentes dans la version finale de cette section, si elle existe un jour.

Comme partout dans ce chapitre sur la construction des ensembles de nombres, certaines propriétés qui ont l'air toute simples peuvent, en fonction des définitions prises, s'avérer pas du tout évidentes.

47. Lemme 1.100.

1.11.1 Définitions

Un nombre complexe s'écrit sous la forme $z = a + bi$, où a et b sont des nombres réels appelés (et notés) respectivement partie réelle ($a = \Re(z)$) et partie imaginaire ($b = \Im(z)$) de z. L'ensemble des nombres de cette forme s'appelle l'ensemble des nombres complexes ; cet ensemble porte une structure de corps et est noté $\mathbb{C}$. Le nombre complexe $i = 0 + 1i$ est un nombre imaginaire qui a la particularité que $i^2 = -1$.

Deux nombres complexes $a+bi$ et $c+di$ sont égaux si et seulement si $a = c$ et $b = d$, c'est-à-dire leurs parties réelles sont égales, et leurs parties imaginaires sont égales.

Pour $z = a + bi$ un nombre complexe, on note $\bar{z} = a - bi$ le *complexe conjugué* de z. Dans le plan de Gauss, il s'agit du symétrique de z par rapport à la droite réelle (généralement dessinée horizontalement).

On définit le module du complexe z par $|z| = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}$. Dans le plan de Gauss, il s'agit de la distance entre 0 et z.

Proposition 1.167.
Si z_1 et z_2 sont des nombres complexes, alors

$$|z_1 z_2| = |z_1||z_2|. \tag{1.189}$$

Nous avons aussi, pour tout $n \in \mathbb{N}$,

$$|z^n| = |z|^n. \tag{1.190}$$

Démonstration. D'abord $(a + bi)(c + di) = ac - db + (ad + bc)i$, de telle sorte que

$$|(a + bi)(c + di)|^2 = (ac - bd)^2 + (ad + bc)^2. \tag{1.191}$$

Mais en calculant d'autre part $|a + bi|^2|c + di|^2$, nous tombons sur la même valeur.

Une simple récurrence permet de conclure que $|z^n| = |z|^n$. □

Voila. Vous êtes déjà content d'apprendre que l'on peut démontrer $|z^n| = |z|^n$ sans faire appel à la forme trigonométrique des nombres complexes.

Proposition 1.168.
Pour tout nombres complexes $z = a + bi$ et z', nous avons

(1) $z\bar{z} = a^2 + b^2$;
(2) $\bar{\bar{z}} = z$;
(3) $|z| = |\bar{z}|$;
(4) $|zz'| = |z|\,|z'|$;
(5) $|z + z'| \leqslant |z| + |z'|$.

Lemme 1.169.
Pour tout $z \in \mathbb{C}$ nous avons $z\bar{z} = \bar{z}z = |z|^2$.

Chapitre 2

Théorie des groupes

Pour rappel, la notion de groupe est définie en 1.69.

2.1 Groupes

Définition 2.1 ([29]).
Soit G un groupe. Le ***centralisateur*** *de $H \subset G$ est*

$$\mathcal{Z}_G(H) = \{g \in G \text{ tel que } hg = gh \, \forall h \in H\}; \tag{2.1}$$

il contient donc tous les éléments de G qui commutent avec ceux de H.

Si H est un sous-groupe, son ***normalisateur*** *est*

$$N_G(H) = \{g \in G \text{ tel que } gH = Hg\}. \tag{2.2}$$

Définition 2.2.
Le ***centre*** *d'un groupe G est l'ensemble des éléments de G qui commutent avec tous les autres :*

$$Z_G = \{z \in G \text{ tel que } gz = zg \, \forall g \in G\}. \tag{2.3}$$

Si $g \in G$ nous notons $Z_G(g)$ le ***centralisateur*** *de g dans G :*

$$Z_G(g) = \{h \in G \text{ tel que } hg = gh\}. \tag{2.4}$$

C'est l'ensemble des éléments de G qui commutent avec g.

Définition 2.3.
Un sous-groupe N de G est ***normal*** *ou* ***distingué*** *si pour tout $g \in G$ et pour tout $n \in N$, $gng^{-1} \in N$. Autrement dit lorsque $gNg^{-1} \subset N$.*

Lorsque N est normal dans G il est parfois noté $N \lhd G$.

Définition 2.4.
Un sous-groupe H de G est un sous-groupe ***caractéristique*** *si $\alpha(H) = H$ pour tout $\alpha \in \mathrm{Aut}(G)$.*

Définition 2.5 (Groupe simple)**.**
Un groupe est dit ***simple*** *si il est non trivial et si les seuls sous-groupes normaux qu'il admet sont lui-même et le sous-groupe réduit à l'identité.*

Définition 2.6 (Sous-groupe engendré)**.**
Soit A une partie du groupe G. Le sous-groupe ***engendré*** *par A est l'intersection de tous les sous-groupes de G contenant A. Nous notons ce groupe $\mathrm{gr}(A)$.*

Lorsque A est fini (disons $A = \{a_1, \dots, a_n\}$), on note aussi le sous-groupe engendré $\langle a_1, \dots, a_n \rangle$.

En particulier, $\mathrm{gr}(\emptyset) = \{e\}$ [1].

1. Vu que l'ensemble vide est une partie de tous les ensemble, la définition $\mathrm{gr}(\emptyset) = \{e\}$ est ambigüe : c'est l'identité de quel groupe ? C'est le contexte qui dira.

Un sous-groupe engendré n'est jamais vide parce qu'il contient toujours au moins le neutre (parce que c'est un sous-groupe). Si G est un groupe, le sous-groupe $\mathrm{gr}(\emptyset)$ lui-même contient e [2].

Lemme 2.7.
Si G est un groupe et A une partie de G, alors $\mathrm{gr}(A)$ est un sous-groupe de G.

Le sous-groupe engendré par A est le plus petit (pour l'inclusion) groupe de G contenant A. Plus formellement, nous avons le résultat suivant.

Lemme 2.8.
Tout sous-groupe de G contenant A contient $\mathrm{gr}(A)$.

Démonstration. Si H est un sous-groupe de G contenant A, alors $\mathrm{gr}(A)$ est l'intersection de H avec tous les autres sous-groupes de G contenant A. Il contient donc $\mathrm{gr}(A)$. □

Lemme 2.9 ([30])**.**
Si A est une partie du groupe G, alors le sous-groupe $\mathrm{gr}(A)$ engendré [3] par A est l'ensemble de tous les produits finis d'éléments de A et de A^{-1} (l'identité est le produit à zéro éléments).

C'est-à-dire que tout élément de $\mathrm{gr}(A)$ peut être écrit sous la forme [4]

$$\prod_{i=1}^{n} g_i^{a_i} \tag{2.5}$$

où $a_i \in \mathbb{Z}$ et $g\colon \mathbb{N} \to A$ n'est pas spécialement injective : il peut arriver que $g_i = g_j$.

Démonstration. Vu qu'un produit vide est égal à l'identité, le lemme est vrai (un peu trivialement) dans le cas où $A = \emptyset$. À partir de maintenant nous supposons que A est non vide.

Nous nommons $\mathrm{gr}(A)$ le groupe engendré par A et par H l'ensemble

$$H = \{g_1 \dots g_n \text{ tel que } g_i \in A \cup A^{-1}\}. \tag{2.6}$$

Nous commençons par prouver que H est un groupe.

— Vu que A est non vide, nous considérons $a \in A$. Dans ce cas, $e = aa^{-1} \in H$. Donc $e \in H$.
— L'inverse de $g_1 \dots g_n$ est $g_n^{-1} \dots g_1^{-1}$ qui est également dans H.
— Le produit de $g_1 \dots g_n$ par $h_1 \dots h_n$ est également dans H [5].

Vu que H est un groupe contenant A, nous avons $\mathrm{gr}(A) \subset H$ parce que $\mathrm{gr}(A)$ est une intersection dont un des éléments est H.

Par ailleurs tout groupe contenant A doit contenir les inverses et les produits finis, donc $H \subset \mathrm{gr}(A)$.

Au final, $H = \mathrm{gr}(A)$, ce qu'il fallait. □

Lemme 2.10.
Soit un groupe G et un sous-groupe $H = \mathrm{gr}(h_1, \dots, h_n)$. Si $\alpha \in G$, alors

$$\alpha H \alpha^{-1} = \mathrm{gr}(\alpha h_1 \alpha^{-1}, \dots, \alpha h_n \alpha^{-1}). \tag{2.7}$$

Démonstration. Il s'agit d'une conséquence du lemme 2.9. Un élément de $\mathrm{gr}(\alpha h_1 \alpha^{-1}, \dots, \alpha h_n \alpha^{-1})$ est un produit d'éléments de G de la forme $\alpha h_i \alpha^{-1}$ ou $(\alpha h_j \alpha^{-1})^{-1} = \alpha h_j^{-1} \alpha^{-1}$. Or nous avons

$$\alpha h_i \alpha^{-1} \alpha h_j \alpha^{-1} = \alpha h_i h_j \alpha^{-1} \in \alpha H \alpha^{-1}. \tag{2.8}$$

Donc

$$\mathrm{gr}(\alpha h_1 \alpha^{-1}, \dots, \alpha h_n \alpha^{-1}) \subset \alpha H \alpha^{-1}. \tag{2.9}$$

L'inclusion dans l'autre sens est du même tonneau. □

2. Demandez-vous si il est possible que $\mathrm{gr}(\emptyset)$ contienne d'autres éléments que e.
3. Définition 2.6.
4. Les a_i négatifs correspondent aux inverses. Notons que si $g \in A$, il n'y a pas de garanties que g^{-1} soit également dans A.
5. Et c'est ici qu'on se rend compte que la décomposition n'est probablement que rarement unique.

Définition 2.11 (Partie génératrice, groupe monogène)**.**
*Soit G, un groupe et A une partie de G. Si $\mathrm{gr}(A) = G$, alors nous disons que A est une **partie génératrice** le groupe G.*

*Un groupe est **monogène** s'il a une partie génératrice réduite à un seul élément.*

Définition 2.12 (Groupe cyclique)**.**
*Un élément $a \in G$ est un **générateur** de G si tous les éléments de G s'écrivent sous la forme a^n pour un certain $n \in \mathbb{Z}$. Un groupe fini et monogène est dit **cyclique**.*

Exemple 2.13.
Soit le groupe $\big(\mathbb{Z}/10\mathbb{Z}, +\big)$. L'élément $[2]_{10}$ n'est pas générateur parce que ses puissances[6] sont

$$\mathrm{gr}([2]_{10}) = \{[2]_{10}, [4]_{10}, [6]_{10}, [8]_{10}, [0]_{10}\}. \tag{2.10}$$

Par contré l'élément $[3]_{10}$ est générateur : ses puissances sont dans l'ordre

$$[3]_{10}, [6]_{10}, [9]_{10}, [2]_{10}, [5]_{10}, [8]_{10}, [1]_{10}, [4]_{10}, [7]_{10}, [0]_{10}. \tag{2.11}$$

△

Un exemple presque identique, mais un peu masqué sera l'exemple 17.151.

2.2 Sous-groupe normal

Proposition 2.14.
Soit N un sous-groupe de G. Les propriétés suivantes sont équivalentes :

(1) $gNg^{-1} \subseteq N$ pour tout $g \in G$,

(2) $gNg^{-1} = N$ pour tout $g \in G$,

(3) $gN = Ng$ pour tout $g \in G$,

(4) N est une union de classes de conjugaison de G,

(5) N est normal[7].

Définition 2.15.
Soit $g \in G$ et $n \in \mathbb{Z}$. Nous définissons g^n par

(1) $g^0 = e$ et $g^n = gg^{n-1}$ si n est positif.

(2) si $n < 0$, nous posons $g^n = (g^{-1})^{-n}$.

Définition 2.16 (Ordre d'un groupe et d'un élément)**.**
Ce sont deux choses différentes.

*(1) Si G est un groupe, l'**ordre** est la cardinalité de G et est noté $|G|$.*

*(2) L'**ordre** d'un élément g de G est le naturel*

$$\min\{n \in \mathbb{N} \text{ tel que } g^n = e\}, \tag{2.12}$$

s'il existe ; dans le cas contraire, nous disons que l'ordre de g est infini.

Nous verrons que le corolaire 2.35 au théorème de Lagrange dira que l'ordre d'un élément divise l'ordre du groupe.

Lemme 2.17 ([31])**.**
Si H et K sont normaux dans le groupe G et si $H \cap K = \{e\}$ alors $HK \simeq H \times K$.

6. Attention aux notations ; en général on écrit la loi de groupe de façon multiplicative et on parle des puissances d'un élément, mais ici on écrit la loi de groupe additivement, donc les « puissances » sont en réalité les multiples.
7. Définition 2.3.

Définition 2.18.
*L'**exposant** du groupe G est le plus petit entier non nul n tel que $g^n = e$ pour tout $g \in G$. S'il n'existe pas, nous disons que l'exposant du groupe est infini.*

Proposition 2.19.
Si l'ensemble des ordres de tous les éléments d'un groupe est majoré, alors l'exposant du groupe est le plus petit commun multiple des ordres des éléments du groupe

Pour un groupe fini, l'exposant est le ppcm *des ordres des éléments du groupe.*

Le théorème de Burnside 9.242 nous donnera un bon paquet d'exemples de groupes d'exposant fini dans $\mathrm{GL}(n, \mathbb{C})$.

Proposition 2.20.
Soit un groupe G. Nous considérons un sous-groupe normal H de G ainsi qu'un homomorphisme $\psi\colon G \to H$. Alors

(1) $\psi(H)$ est normal dans $\psi(G)$

(2) Si G/H est abélien alors $\psi(G)/\psi(H)$ est abélien.

Démonstration. Soient $h \in H$ et $g \in G$. Alors $\psi(g)\psi(h)\psi(g)^{-1} = \psi(ghg^{-1}) \in \psi(H)$. Donc $\psi(H)$ est normal dans $\psi(G)$.

Pour la seconde partie nous notons $[\ldots]$ les classes par rapport à $\psi(H)$ et $\overline{\ldots}$ celles par rapport à H. Nous avons

$$\begin{aligned}
[\psi(g_1)][\psi(g_2)] &= \big[\psi(g_1)\psi(g_2)\big] && (2.13a)\\
&= \big[\psi(g_1 g_2)\big] && (2.13b)\\
&= \{\psi(g_1 g_2)\psi(h) \text{ tel que } h \in H\} && (2.13c)\\
&= \{\psi(g_1 g_2 h) \text{ tel que } h \in H\} && (2.13d)\\
&= \psi\Big(\{g_1 g_2 h \text{ tel que } h \in H\}\Big) && (2.13e)\\
&= \psi\big(\overline{g_1 g_2}\big) && (2.13f)\\
&= \psi\big(\overline{g_2 g_1}\big) && (2.13g)\\
&= \text{refaire à l'envers} && (2.13h)\\
&= [\psi(g_2)][\psi(g_1)]. && (2.13i)
\end{aligned}$$

Par conséquent $\psi(G)/\psi(H)$ est abélien. □

2.2.1 Classes de conjugaison

Définition 2.21.
*Soit un groupe G et un élément $g \in G$. La **classe de conjugaison** de g est la partie*

$$C_g = \{kgk^{-1} \text{ tel que } k \in G\}. \tag{2.14}$$

Lemme 2.22.
Un groupe est abélien si et seulement si ses classes de conjugaison sont des singletons.

Démonstration. Supposons que G soit abélien. Alors

$$C_g = \{kgk^{-1} \text{ tel que } k \in G\} = \{g\}. \tag{2.15}$$

Donc les classes de conjugaison sont des singletons.

Dans l'autre sens, si les classes sont des singletons, on a $kgk^{-1} = g$ pour tous $k, g \in G$. Cela signifie immédiatement que G est abélien. □

2.3 Groupe dérivé

Définition 2.23.
Si G est un groupe et si $g, h \in G$, nous notons $[g, h] = ghg^{-1}h^{-1}$ le ***commutateur*** *de g et h.*

L'élément neutre est toujours un commutateur : pour $g = h$, $[g, g] = ggg^{-1}g^{-1} = e$.

Définition 2.24.
Le ***groupe dérivé*** *de G est le sous-groupe noté $D(G)$ ou $[G, G]$ engendré*[8] *par les commutateurs.*

Autrement dit, $D(G)$ est l'intersection de tous les sous-groupes de G contenant tous les commutateurs. Le groupe $D(G)$ contient toujours au moins le neutre parce que c'est un groupe.

En vertu du lemme 2.9, le groupe dérivé de G est l'ensemble des produits finis de commutateurs. C'est-à-dire que si S_m est l'ensemble des produits de m commutateurs, alors

$$D(G) = \bigcup_{m=1}^{\infty} S_m. \tag{2.16}$$

Lemme 2.25.
Le groupe dérivé est un sous-groupe caractéristique[9]*, et un sous-groupe normal*[10]*.*

Démonstration. Il est évident que si $\alpha \in \mathrm{Aut}(G)$ alors

$$\alpha\big([g, h]\big) = \big[\alpha(g), \alpha(h)\big], \tag{2.17}$$

c'est-à-dire que $D(G)$ est un sous-groupe caractéristique. En particulier si c est un commutateur, alors xcx^{-1} en est encore un, ce qui montre que $D(G)$ est normal dans G. Plus spécifiquement,

$$x(ghg^{-1}h^{-1})x^{-1} = (xgx^{-1})(xhx^{-1})(xg^{-1}x^{-1})(xh^{-1}x^{-1}) \tag{2.18a}$$
$$= (xgx^{-1})(xhx^{-1})(xgx^{-1})^{-1}(xhx^{-1})^{-1}. \tag{2.18b}$$

□

Proposition 2.26.
Le groupe quotient $G/D(G)$ est abélien.

Démonstration. En ce qui concerne le fait que $G/D(G)$ soit abélien, nous savons que pour tout $g, h \in G$ nous avons $h^{-1}g^{-1}hg \in D(G)$ et donc

$$[g][h] = [gh] = [ghh^{-1}g^{-1}hg] = [hg] = [h][g]. \tag{2.19}$$

□

Le groupe quotient $G/D(G)$ est appelé l'**abélianisé** de G et est parfois noté G^{ab}.

Si $f\colon G \to A$ est un homomorphisme entre le groupe G et un groupe abélien A, alors $f\big(D(G)\big) = \{0\}$. Du coup f passe au quotient de G par $D(G)$, et il existe une unique application $\bar{f}\colon G/D(G) \to A$ telle que $f = \bar{f} \circ \pi$ où $\pi\colon G \to G/D(G)$ est la projection canonique.

2.4 Théorèmes d'isomorphismes

Définition 2.27.
Soient un groupe G, un ensemble X et une application $f\colon X \to G$. Le ***noyau*** *de f est la partie*

$$\ker(f) = \{x \in X \text{ tel que } f(x) = e\} \tag{2.20}$$

où e est l'unité de G.

8. Définition 2.6.
9. Définition 2.4.
10. Définition 2.3.

Si G est un groupe et si N est un sous-groupe normal, alors l'ensemble G/N a une structure de groupe et la projection canonique $\pi\colon G \to G/N$ est un homomorphisme surjectif de noyau N.

Théorème 2.28 (Premier théorème d'isomorphisme)**.**
Soit $\theta\colon G \to H$ un homomorphisme de groupe. Alors

(1) $\operatorname{Ker}\theta$ *est normal dans G,*

(2) $\operatorname{Image}\theta$ *est un sous-groupe de H*

(3) nous avons un isomorphisme naturel

$$G/\operatorname{Ker}\theta \simeq \operatorname{Image}\theta \tag{2.21}$$

Démonstration. Point par point.

(1) Le fait que $\operatorname{Ker}\theta$ soit un sous-groupe de G est clair ; montrons qu'il est normal. Si $g \in G$ et $u \in \operatorname{Ker}\theta$, alors $\theta(g^{-1}ug) = \theta(g^{-1})\theta(u)\theta(g) = \big(\theta(g)\big)^{-1}\theta(g) = 1_H$, et donc $g^{-1}ug \in \operatorname{Ker}\theta$.

(2) Il suffit de remarquer que si $h = \theta(g)$ et $h' = \theta(g')$, alors $h^{-1}h' = \theta(g^{-1}g')$.

(3) Si $[g]$ représente la classe de g dans $G/\operatorname{Ker}\theta$, l'isomorphisme est donné par $\varphi[g] = \theta(g)$.

□

Théorème 2.29 (Deuxième théorème d'isomorphisme)**.**
Soient H et N deux sous-groupes de G et supposons que N soit normal. Alors

(1) $NH = HN$ est un sous-groupe.

(2) Le groupe N est normal dans NH.

(3) Le groupe $N \cap H$ est normal dans H.

(4) Nous avons l'isomorphisme

$$\frac{HN}{N} \simeq \frac{H}{H \cap N}. \tag{2.22}$$

(5) L'isomorphisme du point (4) est encore valable si N n'est pas normal mais si seulement H normalise N, c'est-à-dire si $hNh^{-1} \in N$ pour tout $h \in H$.

Démonstration. Point par point.

(1) Il est clair que $1_G \in NH$. Soient nh et $n'h'$ deux éléments de NH ; alors en tenant compte du fait que N est normal,

$$nhn'h' = n\underbrace{hn'h^{-1}}_{\in N}hh' \in NH. \tag{2.23}$$

Cela prouve que NH est un groupe.

De la même façon, nous prouvons que HN est un groupe par

$$hnh'n' = hh'\underbrace{h'^{-1}nh'}_{\in N}n' \in HN \tag{2.24}$$

Nous devons encore prouver que $HN = NH$. Pour cela, $nh \in HN$, car $nh = hh^{-1}nh$, les trois derniers facteurs formant un élément de N par normalité ; de même $hn \in NH$, montrant que $NH = HN$. Enfin, comme $(nh)^{-1} = h^{-1}n^{-1})$, les inverses de NH sont dans $HN = NH$.

(2) N est normal dans G, a fortiori dans l'un de ses sous-groupes.

(3) Il suffit de voir que, si $h \in H$ et $n \in N \cap H$, alors $hnh^{-1} \in N \cap H$. Or, $hnh^{-1} \in H$ puisque H est un sous-groupe ; et $hnh^{-1} \in N$ car N est un sous-groupe normal de G.

(4) Il faut d'abord remarquer que H et N étant des groupes et le produit NH étant un groupe, nous avons $NH = HN$. Soit le morphisme injectif

$$\begin{aligned} j\colon H &\to HN \\ h &\mapsto h \end{aligned} \tag{2.25}$$

et la surjection canonique

$$\sigma\colon HN \to HN/N \tag{2.26}$$

Nous considérons ensuite l'application composée

$$\begin{aligned} f\colon H &\to HN/N \\ h &\mapsto hN. \end{aligned} \tag{2.27}$$

<u>f **est surjective**</u> L'application f est surjective parce que l'élément $hnN \in HN/N$ est l'image de h, étant donné que $hnN = hN$.

<u>$\mathrm{Ker}(f) = H \cap N$</u> Si $a \in H \cap N$, nous avons $f(a) = aN = N$, et donc $H \cap N \subset \mathrm{Ker}\, f$. D'autre part, si $h \in H$ vérifie $h \in \mathrm{Ker}\, f$, alors $f(h) = hN = N$, ce qui est uniquement possible si $h \in N$.

Le premier théorème d'isomorphisme implique alors que $H/\mathrm{Ker}\, f \simeq \mathrm{Image}\, f$, c'est-à-dire

$$H/N \cap H \simeq HN/N. \tag{2.28}$$

□

Théorème 2.30 (Troisième théorème d'isomorphisme)**.**
Soient N et M deux sous-groupes normaux de G avec $M \subset N$. Alors N/M est normal dans G/M et

$$(G/M)/(N/M) \simeq G/N. \tag{2.29}$$

Démonstration. Afin de montrer que N/M est normal dans G/M, nous considérons $g \in G$, $nM \in N/M$ et nous calculons

$$gnMg^{-1} = gng^{-1}\underbrace{gMg^{-1}}_{=M} = \underbrace{gng^{-1}}_{\in N} M \in N/M. \tag{2.30}$$

Pour prouver l'isomorphisme nous considérons le morphisme

$$\begin{aligned} \varphi\colon G/M &\to G/N \\ gM &\mapsto gN. \end{aligned} \tag{2.31}$$

C'est surjectif et le noyau est N/M parce que $\varphi(gM) = N$ uniquement si $g \in N$. Nous pouvons appliquer le premier théorème d'isomorphisme à φ en écrivant

$$(G/M)/\mathrm{Ker}\, \varphi \simeq \mathrm{Image}\, \varphi, \tag{2.32}$$

c'est-à-dire

$$(G/M)/(N/M) \simeq G/N. \tag{2.33}$$

□

2.5 Indice d'un sous-groupe et ordre des éléments

Lemme 2.31.
Lorsque H est normal dans G, alors la définition

$$[a] \cdot [b] = [ab] \tag{2.34}$$

définit une loi de groupe sur l'ensemble G/H.

Démonstration. Le neutre est $[e]$ et l'associativité ne pose pas plus de problèmes que l'existence d'un inverse. Le point à vérifier est que la formule (2.34) est une bonne définition : $[ah] \cdot [bh'] = [ab]$ pour tout $h, h' \in H$. Nous avons :

$$[ah] \cdot [ah'] = [ahah'] = [ahb]. \tag{2.35}$$

Pour montrer que cela est $[ab]$, l'astuce est d'introduire bb^{-1} à côté du a :

$$[ahb] = [abb^{-1}hb] = [ab] \tag{2.36}$$

parce que $b^{-1}hb \in H$ du fait que H soit normal dans G. □

Exemple 2.32 ([32])**.**
Il ne faudrait pas croire que le groupe quotient G/H est forcément un sous-groupe de G. Par exemple le quotient $\mathbb{Z}/2\mathbb{Z}$ est l'ensemble $\{0, 1\}$ muni de l'addition. En particulier $1 + 1 = 0$, ce qui est évidemment faux dans $\mathbb{Z}$. Le groupe $(\mathbb{Z}, +)$ ne possède aucun élément d'ordre 2.

Il n'en est pas moins vrai que l'application

$$\begin{aligned} f\colon G &\to G/H \\ g &\mapsto [g] \end{aligned} \tag{2.37}$$

est un morphisme de groupes. △

Définition 2.33.
*Si H est un sous-groupe d'un groupe fini l'**indice** de H dans G est le nombre $|G|/|H|$, souvent noté $|G : H|$.*

Le théorème de Lagrange dira en particulier que l'indice est toujours un nombre entier. C'est à ne pas confondre avec le degré d'une extension de corps (définition 6.54).

Théorème 2.34 (Théorème de Lagrange)**.**
Soit H un sous-groupe du groupe fini G. Alors

(1) L'ordre de H divise l'ordre de G.

(2) Les trois nombres suivants sont égaux :

- *le nombre de classes de H à gauche,*
- *le nombre de classes de H à droite,*
- *l'indice de H dans G.*

En particulier si H est distingué dans G nous avons

$$|G/H| = \frac{|G|}{|H|}. \tag{2.38}$$

Démonstration. Nous commençons par montrer que les classes de H ont toutes le même nombre d'éléments que H. En effet pour chaque $g \in G$ nous avons la bijection

$$\begin{aligned} \varphi\colon H &\to gH \\ h &\mapsto gh. \end{aligned} \tag{2.39}$$

L'injectivité de φ est le fait que $gh = gh'$ implique $h = h'$. La surjectivité est par définition de la classe.

Les classes à gauche formant une partition de G, le cardinal de G est le produit de la taille des classes par le nombre de classes :

$$|G| = |H| \cdot \text{nombre de classes.} \tag{2.40}$$

En particulier nous voyons que $|H|$ divise $|G|$.

La dernière formule exprime simplement que G/H est par définition le nombre de classes de H à gauche (ou à droite) dans G. □

Corolaire 2.35.
L'ordre d'un élément d'un groupe fini divise l'ordre du groupe. En particulier dans un groupe d'ordre n tous les éléments vérifient $q^n = e$.

Démonstration. Soit G un groupe fini et considérons, à $g \in G$ fixé, le sous-groupe

$$H = \{g^k \text{ tel que } k \in \mathbb{N}\}. \tag{2.41}$$

Par le théorème de Lagrange 2.34, l'ordre de H divise $|G|$, mais l'ordre de H est le plus petit k tel que $g^k = e$, c'est-à-dire l'ordre de g. □

D'autres résultats à propos d'ordres et d'indices de groupes finis dans la proposition 3.38 et le lemme 3.40. En particulier le théorème de Cauchy 3.34 qui dit si p divise l'ordre du groupe G, alors G contient au moins un élément d'ordre p.

2.6 Suite de composition

Définition 2.36.
*Une **suite de composition** pour un groupe G est une suite finie de sous-groupes $(G_i)_{i=0,\dots,n}$ telle que*

$$\{e\} = G_n \subseteq G_{n-1} \subseteq \dots \subseteq G_1 \subseteq G_0 = G \tag{2.42}$$

et telle que G_{i+1} est normal[11] *dans G_i. Les groupes G_i/G_{i+1} sont les **quotients** de la suite de composition.*

*Une suite de **Jordan-Hölder** est une suite de composition dont tous les quotients sont simples.*

L'objet de nos prochaines pérégrinations mathématiques est de montrer que tout groupe fini admet une suite de Jordan-Hölder (théorème 2.41).

Lemme 2.37 (du papillon[33])**.**
Soit G un groupe et des sous-groupes A et B. Soit A' normal dans A et B' normal dans B. Alors

(1) $A'(A \cap B')$ est normal dans $A'(A \cap B)$

(2) $(A' \cap B)B'$ est normal dans $(A \cap B)B'$

(3) Nous avons les isomorphismes de groupes

$$\frac{A'(A \cap B)}{A'(A \cap B')} \simeq \frac{(A \cap B)B'}{(A' \cap B)B'} \simeq \frac{B'(A \cap B)}{B'(A' \cap B)}. \tag{2.43}$$

Démonstration. Nous n'allons pas démontrer chacun des points ; pour plus de détails, nous dirons simplement que « la preuve est très similaire dans les autres cas ».

Commençons par montrer que $A'(A \cap B')$ est un groupe. Si $a, b \in A'$ et $x, y \in A \cap B'$,

$$axby = xx^{-1}axbx^{-1}xy \tag{2.44}$$

11. Nous rappelons au cas où que « normal » signifie « distingué ».

En utilisant la normalité, $x^{-1}ax \in A'$, donc $xx^{-1}axbx^{-1} \in A'$ et donc le tout est dans $A'(A \cap B')$. L'ensemble $A'(A \cap B')$ est également stable pour l'inverse parce que

$$x^{-1}a^{-1} = \underbrace{x^{-1}a^{-1}x}_{\in A'} x^{-1}. \tag{2.45}$$

Nous montrons maintenant que $A'(A \cap B')$ est normal dans $A'(A \cap B)$. Soient $a, b \in A'$, $x \in A \cap B'$ et $f \in A \cap B$. Alors

$$\begin{aligned}
(bf)^{-1}(ax)(bf) &= (bf)^{-1}(a \underbrace{xbx^{-1}}_{=c\in A'} xf) && \text{(2.46a)}\\
&= f^{-1}b^{-1}acxf && \text{(2.46b)}\\
&= f^{-1}b^{-1}acf \underbrace{f^{-1}xf}_{=y\in A\cap B'} && \text{(2.46c)}\\
&= \underbrace{f^{-1}b^{-1}acf}_{\in A'} y && \text{(2.46d)}\\
&\in A'(A \cap B'). && \text{(2.46e)}
\end{aligned}$$

Pour prouver l'isomorphisme

$$\frac{A'(A \cap B)}{A'(A \cap B')} = \frac{(A \cap B)B'}{(A' \cap B)B'}, \tag{2.47}$$

nous allons utiliser le deuxième théorème d'isomorphisme (2.30(5)). Que nous appliquons à $H = A \cap B$ et $N = A'(A \cap B')$. La vérification que H normalise N est usuelle. Nous commençons par écrire

$$\frac{A'(A \cap B')(A \cap B)}{A'(A \cap B')} \simeq \frac{A \cap B}{A \cap B \cap A'(A \cap B')}. \tag{2.48}$$

Pour simplifier un peu cette expression nous prouvons d'abord que

$$(A \cap B) \cap A'(A \cap B') = (A' \cap B)(A \cap B'). \tag{2.49}$$

L'inclusion $\supset$ est facile. Pour l'autre sens, étant donné que $A'(A \cap B') \subset A$ nous avons

$$A \cap B \cap A'(A \cap B) = B \cap A'(A \cap B). \tag{2.50}$$

Un élément de $B \cap A'(A \cap B)$ est un élément de B qui s'écrit sous la forme $s = ax$ avec $a \in A'$ et $x \in A \cap B'$. Nous avons alors $a = sx^{-1}$ avec $s \in B$ et $x^{-1} \in B'$. Par conséquent $a \in B$ et donc $a \in A' \cap B$. Nous avons donc

$$(A \cap B) \cap A'(A \cap B') = B \cap A'(A \cap B) \subset (A' \cap B)(A \cap B'), \tag{2.51}$$

et donc l'égalité (2.49). Toujours dans l'idée de simplifier (2.48) nous remarquons que $A \cap B'$ est un sous-ensemble de $A \cap B'$, donc $A'(A \cap B')(A \cap B) = A'(A \cap B)$. Il reste donc

$$\frac{A'(A \cap B)}{A'(A \cap B')} = \frac{A \cap B}{(A' \cap B)(A \cap B')}. \tag{2.52}$$

Étant donné que les hypothèses sur A et B sont symétriques, le membre de droite peut aussi s'écrire en inversant A et B. Nous en sommes à

$$\frac{B'(A \cap B)}{B'(A' \cap B)} = \frac{A'(A \cap B)}{A'(A \cap B')}. \tag{2.53}$$

Nous devons encore justifier $B'(A \cap B) = (A \cap B)B'$ et $B'(A' \cap B) = (A' \cap B)B'$. Faisons le premier et laissons le second au lecteur. Si $b \in B'$ et $x \in A \cap B$, alors

$$bx = x \underbrace{x^{-1}bx}_{\in B'} \in (A \cap B)B'. \tag{2.54}$$

□

Proposition 2.38.
Si G est un groupe fini et que (G_i) est une suite de composition pour G, alors l'ordre de G est le produit des ordres de ses quotients.

Démonstration. Étant donné que G_{i+1} est toujours normal dans G_i, le théorème de Lagrange (2.34) s'applique et nous avons à chaque pas de la suite de composition nous avons

$$|\frac{G_i}{G_{i+1}}| = \frac{|G_i|}{|G_{i+1}|} \tag{2.55}$$

et il suffit d'écrire $|G|$ de façon télescopique :

$$|G| = \prod_{0\leqslant i\leqslant n-1} \frac{|G_i|}{|G_{i+1}|} \tag{2.56}$$

□

Nous disons que les deux suites de composition $(G_i)_{0\leqslant i\leqslant r}$ et $(G_j)_{0\leqslant j\leqslant s}$ sont **équivalentes** si $r = s$ et s'il existe une permutation $\sigma \in S_{r-1}$ telle que

$$\frac{G_i}{G_{i+1}} \simeq \frac{H_{\sigma(i)}}{H_{\sigma(i)+1}}. \tag{2.57}$$

Proposition 2.39 (Schreider)**.**
Deux suites de composition d'un même groupe admettent des raffinements équivalents.

Démonstration. Soient les suites de composition

$$\{e\} = G_m \subseteq \ldots \subseteq G_1 \subseteq G_0 = G \tag{2.58a}$$
$$\{e\} = H_m \subseteq \ldots \subseteq H_1 \subseteq H_0 = G \tag{2.58b}$$

Nous raffinons la suite (G_i) en remplaçant $G_{i+1} \subseteq G_i$ par

$$G_{i+1} = G_{i+1}(G_i \cap H_n) \subset G_{i+1}(G_i \cap H_{n-1}) \subseteq \ldots \subseteq G_{i+1}(G_i \cap H_0) = G_i, \tag{2.59}$$

et de même pour (H_j). Le groupe $G_{i+1}(G_i \cap H_k)$ est normal dans $G_{i+1}(G_i \cap H_{k-1})$ parce que G_{i+1} étant normal dans G_i et H_k dans H_{k-1}, le lemme 2.37 s'applique. Nous avons donc bien défini un raffinement.

Nous devons maintenant prouver que les deux raffinements ainsi construits sont des suites de composition équivalentes. D'abord elles ont la même longueur mn parce que chacun des m éléments de la suite (G_i) a été remplacé par n éléments et inversement, chacun de n éléments de la suite (H_j) a été remplacé par m éléments.

Par ailleurs, les quotients du raffinement de (G_i) sont de la forme

$$\frac{G_{i+1}(G_i \cap H_k)}{G_{i+1}(G_i \cap H_{k+1})} \simeq \frac{H_{k+1}(H_k \cap G_i)}{H_{k+1}(H_k \cap G_{i+1})} \tag{2.60}$$

en vertu du lemme du papillon (2.37). Le membre de droite de (2.60) est un des quotients du raffinement de (H_j). □

Lemme 2.40 (Schreider strictement décroissant)**.**
Soient Σ_1 et Σ_2, deux suites de composition strictement décroissantes du groupe G. Alors elles admettent des raffinements équivalents strictement décroissants.

Démonstration. Par hypothèse, Σ_1 et Σ_2 n'ont pas de répétitions. Soient Σ_1'' et Σ_2'', des raffinements équivalents donnés par le lemme de Schreider. Étant donné que ce sont des suites de composition équivalentes, elles ont le même nombre de quotients réduits à $\{e\}$, c'est-à-dire le même nombre de répétitions.

Les suites Σ_1' et Σ_2' obtenues en retirant les répétitions de Σ_1'' et Σ_2'' sont des raffinements équivalents de Σ_1 et Σ_2 et strictement décroissants. □

Théorème 2.41 (Jordan-Hölder)**.**
Tout groupe fini admet une suite de Jordan-Hölder.

Deux suites de Jordan-Hölder sont équivalentes.

Démonstration. Nous ne prouvons que le second point.

Par définition, une suite de Jordan-Hölder n'a pas de raffinement strictement décroissant (à part elle-même) parce que G_{i+1} est normal maximum dans G_i. Si Σ_1 et Σ_2 sont des suites de Jordan-Hölder nous pouvons considérer les raffinements équivalents strictement décroissants Σ_1' et Σ_2' du lemme de Schreider 2.40. Nous avons $\Sigma_1' \sim \Sigma_2'$, mais par ce que nous venons de dire à propos de la maximalité, $\Sigma_1' = \Sigma_1$ et $\Sigma_2' = \Sigma_2$. D'où le résultat. □

2.7 Groupes résolubles

Définition 2.42.
*Le groupe G est **résoluble** s'il existe une suite finie de sous-groupes G_i*

$$\{e\} = G_n \subset G_{n-1} \subset \ldots \subset G_1 \subset G_0 = G \tag{2.61}$$

avec G_i normal dans G_{i+1} et G_i/G_{i+1} abélien.

Il s'agit d'un groupe qui admet une suite de composition [12] dont les quotients sont abéliens.

Lemme 2.43 ([34])**.**
Soit G un groupe et H un sous-groupe normal. Le groupe G/H est abélien si et seulement si $D(G) \subset H$.

Démonstration. Les propositions suivantes sont équivalentes :

— Le groupe G/H est abélien
— pour tout $x, y \in G$, $\bar{x}\bar{y} = \bar{y}\bar{x}$
— $\bar{x}\bar{y}\bar{x}^{-1}\bar{y}^{-1} = \bar{e}$
— $\overline{xyx^{-1}y^{-1}} = \bar{e}$
— $[x, y] \in H$
— $D(G) \subset H$.

□

Proposition 2.44 ([34])**.**
Un groupe est résoluble si et seulement si sa suite dérivée termine sur $\{e\}$.

Démonstration. Grâce au lemme 2.25 et à la proposition 2.26, si la suite dérivée termine sur $\{e\}$ alors la suite dérivée est une suite qui répond aux conditions de la définition 2.42 de groupe résoluble.

Il faut donc encore montrer le sens direct. Nous supposons que G est un groupe résoluble et nous étudions sa suite dérivée. Nous avons une suite

$$\{e\} = G_n \subset G_{n-1} \subset \ldots \subset G_1 \subset G_0 = G \tag{2.62}$$

avec G_i/G_{i+1} abélien et G_{i+1} normal dans G_i. Nous allons prouver par récurrence que $D^i(G) \subset G_i$.

Pour $i = 0$ nous avons bien $G \subset G_0$. Notre hypothèse de récurrence est :

$$D^i(G) \subset G_i \tag{2.63}$$

Par le lemme 2.43 nous avons aussi

$$D(G_i) \subset G_{i+1}. \tag{2.64}$$

En dérivant (2.63) et en tenant compte de (2.64), $D^{i+1}(G) \subset D(G_i) \subset G_{i+1}$. Donc par récurrence nous avons bien $D^k(G) \subset G_k$ pour tout k. Mais $G_r = \{e\}$ pour un certain r, donc pour ce r nous avons $D^r(G) = \{e\}$, ce qu'il fallait. □

12. Voir définition 2.36.

Proposition 2.45.
Soient des groupes G et H. Nous supposons que G est résoluble et nous considérons un homomorphisme $\psi\colon G \to H$. Alors $\psi(G)$ est résoluble.

Démonstration. Vu que G est résoluble, il existe une suite de sous-groupes G_i tels que

$$\{e\} = G_n \subset G_{n-1} \subset \ldots \subset G_1 \subset G_0 = G \tag{2.65}$$

avec G_i normal dans G_{i+1} et G_i/G_{i+1} abélien. Nous posons $\psi(G)_i = \psi(G_i)$ et nous avons $\psi(G)_n = \psi\big(\{e\}\big) = \{e\}$ ainsi que $\psi(G)_0 = \psi(G)$; donc

$$\{e\} = \psi(G)_n \subset \psi(G)_{n-1} \subset \ldots \subset \psi(G)_1 \subset \psi(G)_0 = \psi(G). \tag{2.66}$$

Les faits que $\psi(G)_i$ soit normal dans $\psi(G)_{i+1}$ et que $\psi(G)_i/\psi(G)_{i+1}$ soit abélien est directement la proposition 2.20.

□

2.8 Action de groupes

Définition 2.46 (Thème 62)**.**
*Une **action de groupe** G sur un ensemble E est la donnée, pour chaque élément $g \in G$, d'une fonction $\phi_g : E \to E$, de telle sorte que :*

$$\begin{aligned} \phi_e(x) &= x, & \forall x \in E; \\ \phi_{gh}(x) &= \phi_g(\phi_h(x)), & \forall g, h \in G, \forall x \in E. \end{aligned}$$

*On dit dans ce cas que G **agit** sur E.*

Lemme 2.47.
Pour tout $g \in G$,

(1) L'application $\phi_g\colon E \to E$ est injective,

(2) Pour l'inverse : $(\phi_g)^{-1} = \phi_{g^{-1}}$.

Démonstration. Si $x, y \in E$ sont tels que $\phi_g(x) = \phi_g(y)$ alors en appliquant $\phi_{g^{-1}}$ aux deux membres nous trouvons

$$(\phi_{g^{-1}}\phi_g)(x) = (\phi_{g^{-1}}\phi_g)(y), \tag{2.67}$$

ce qui donne $x = y$ parce que $\phi_{g^{-1}}\phi_g = \phi_{g^{-1}g} = \phi_e = \mathrm{Id}$.

Les dernières trois égalités écrites disent que $\phi_{g^{-1}}$ est l'inverse[13] de ϕ_g. □

Pour alléger les notations, on convient d'écrire $g \cdot x$, voire plus simplement gx au lieu de $\phi_g(x)$. Le deuxième axiome d'action de groupe dit que la notation ghx ne souffre d'aucune ambiguïté.

Définition 2.48 (Quelques notions autour de l'action)**.**
*Si G agit sur un ensemble E, nous notons $G \cdot x$ l'**orbite** de $x \in E$ sous l'action de G :*

$$G \cdot x = \{gx \text{ tel que } g \in G\}.$$

*Nous notons G_x ou $\mathrm{Fix}(x)$ le **stabilisateur** de x :*

$$\mathrm{Fix}(x) = G_x = \{g \in G \text{ tel que } g \cdot x = x\}. \tag{2.68}$$

*Pour $g \in G$, nous notons enfin $\mathrm{Fix}(g)$ le **fixateur** de g :*

$$\mathrm{Fix}(g) = \{x \in E \text{ tel que } g \cdot x = x\}. \tag{2.69}$$

13. Si vous décidez de dire ça a un jury dans un concours, soyez prêts à préciser les domaines.

Définition 2.49.
*L'action de G sur E est **fidèle** si l'identité est le seul élément de G à fixer tous les points de E, c'est-à-dire si $gx = x\,\forall x \in E \Rightarrow g = e$.*

Un exemple d'action fidèle tout à fait non trivial sera donné avec l'action du groupe modulaire sur le plan de Poincaré dans le théorème 22.93.

Le groupe G agit toujours sur lui même à gauche et à droite. L'action à gauche est $g \cdot h = gh$; celle à droite est $g \cdot h = hg^{-1}$.

Définition 2.50.
*L'action **adjointe** définie par $g \cdot h = ghg^{-1}$ est une manière pour un groupe d'agir sur lui-même par automorphismes. Cela est souvent noté $\mathbf{Ad}(g)h = ghg^{-1}$.*

En effet pour tout $g \in G$, l'application $\mathbf{Ad}(g)\colon G \to G$ est un automorphisme de G.

Si H est un sous-groupe de G, nous notons G/H le quotient de G par la relation $g \sim gh$ pour tout $h \in H$. Lorsque la distinction est importante, nous noterons $(G/H)_g$ pour les classes à gauche et $(G/H)_d$ pour les classes à droite.

Nous avons une relation d'équivalence à gauche et une à droite. D'abord

$$x \sim_g y \Leftrightarrow xh = y \tag{2.70}$$

pour un certain $h \in H$. Ensuite

$$x \sim_d y \Leftrightarrow hx = y \tag{2.71}$$

pour un certain $h \in H$.

Le lemme suivant est une généralisation du théorème de Lagrange 2.34.

Lemme 2.51.
L'ensemble $(G/H)_g$ est fini si et seulement si l'ensemble $(G/H)_d$ est fini. S'il en est ainsi, alors $(G/H)_g$ et $(G/H)_d$ ont même cardinal qui vaut l'indice de H dans G.

Démonstration. L'application

$$\begin{aligned} f\colon (G/H)_g &\to (G/H)_d \\ [x]_g &\mapsto [x^{-1}]_d \end{aligned} \tag{2.72}$$

est une bijection bien définie. En effet si $x \sim_g y$, nous avons $h \in H$ tel que $y^{-1}h = x^{-1}$, c'est-à-dire que $x^{-1} \sim_d y^{-1}$ et f est bien définie. Le fait que f soit surjective est évident. Pour l'injectivité, soient $x, y \in G$ tels que

$$f([x]_g) = f([y]_g). \tag{2.73}$$

Alors $x^{-1} \sim_d y^{-1}$, ce qui implique l'existence de $h \in H$ tel que $hx^{-1} = y^{-1}$, ou encore que $xh^{-1} = y$, ce qui signifie que $x \sim_g y$.

Pour l'énoncé à propos de l'indice, nous procédons en plusieurs étapes simples.

(1) Les classes (les éléments de $(G/H)_g$) forment une partition de G.

(2) Toutes les classes ont le même nombre d'éléments par la bijection

$$\begin{aligned} f\colon [x]_g &\to [y]_g \\ xh &\mapsto yh. \end{aligned} \tag{2.74}$$

(3) Le nombre d'éléments dans une classe est égal à $|H|$ par la bijection

$$\begin{aligned} g\colon [x]_g &\to H \\ xh &\mapsto h. \end{aligned} \tag{2.75}$$

Par conséquent

$$|G| = |H| \cdot \text{nombre de classes} = |H| \cdot \text{cardinal de } (G/H)_g, \tag{2.76}$$

et nous avons bien

$$\text{cardinal de } (G/H)_g = \frac{|G|}{|H|} = |G : H|. \tag{2.77}$$

□

Proposition 2.52 (Orbite-stabilisateur[35])**.**
Soit G un groupe agissant sur un ensemble E et $x \in E$.

(1) Les ensembles $G \cdot x$ et G/G_x sont équipotents.

(2) L'orbite de $\mathrm{Fix}(x)$ est finie si et seulement si $\mathrm{Fix}(x)$ est d'indice fini dans G. Dans ce cas nous avons

$$\mathrm{Card}(G \cdot x) = |G : \mathrm{Fix}(x)|. \tag{2.78}$$

Une autre façon d'écrire la même formule :

$$|G| = |\mathrm{Fix}(x)||\mathcal{O}_x|. \tag{2.79}$$

C'est la formule (2.78) qui est nommée formule des classes sur wikipédia.

Démonstration. (1) Soit l'application

$$\begin{aligned} \psi\colon G \cdot x &\to G/G_x \\ a \cdot x &\mapsto [a]. \end{aligned} \tag{2.80}$$

Cette application est bien définie parce que si $a \cdot x = b \cdot x$, alors il existe $h \in G_x$ tel que $b = ah$, et par conséquent $[a] = [b]$. Cette application est une bijection et par conséquent $G \cdot x$ est équipotent à G/G_x.

(2) Soit $y \in \mathcal{O}_x$ et $A_y = \{g \in G \text{ tel que } g \cdot x = y\}$. L'ensemble A_y est une classe à gauche de $\mathrm{Fix}(x)$, par conséquent $|A_y| = |\mathrm{Fix}(x)|$ pour tout $y \in \mathcal{O}_x$. Les A_y pour différents y sont disjoints et nous avons de plus

$$\bigcup_{y \in \mathcal{O}_x} A_y = G. \tag{2.81}$$

Les ensembles A_y divisent donc G en $|\mathcal{O}_x|$ paquets de $|\mathrm{Fix}(x)|$ éléments. D'où la formule (2.79).

□

Corolaire 2.53.
Soit C_g la classe de conjugaison d'un élément g du groupe fini G. Alors

$$\mathrm{Card}(C_g) = |G : Z_G(g)| \tag{2.82}$$

où $Z_G(g)$ est le centralisateur de g dans G[14] *de G.*

Démonstration. Cela est une application de la proposition 2.52 (formule (2.78)) dans le cas de l'action adjointe de G sur lui-même.

En effet, si nous considérons l'action adjointe, l'orbite est la classe de conjugaison : $C_g = G \cdot g$. Et le stabilisateur de g pour l'action adjointe n'est autre que le centralisateur de g :

$$\begin{aligned} \mathrm{Fix}(g) &= \{h \in G \text{ tel que } h \cdot g = g\} && (2.83a) \\ &= \{h \in G \text{ tel que } hgh^{-1} = g\} && (2.83b) \\ &= \{h \in G \text{ tel que } gh = hg\} && (2.83c) \\ &= Z_G(g). && (2.83d) \end{aligned}$$

Donc la formule $\mathrm{Card}(G \cdot g) = |G : G_g|$ devient, dans le cas de l'action adjointe de G sur lui-même : $\mathrm{Card}(C_g) = |G : Z_G(g)|$. □

14. Définition 2.2.

Lemme 2.54.
Soit G un groupe agissant sur l'ensemble E. On définit $x \sim x'$ si et seulement s'il existe $g \in G$ tel que $g \cdot x = x'$. Alors

(1) la relation $\sim$ est une relation d'équivalence.

(2) la classe $[x]$ est l'orbite $\mathcal{O}_x$ de x sous G.

Corolaire 2.55 (Équation des orbites)**.**
Soit G un groupe agissant sur l'ensemble E et $\mathcal{O}_1, \ldots, \mathcal{O}_k$ la liste des orbites (distinctes). Alors

(1) $E = \bigcup_i \mathcal{O}_i$, l'union est disjointe,

(2) $\mathrm{Card}(E) = \sum_i \mathrm{Card}(\mathcal{O}_i)$.

Définition 2.56.
Soit G un groupe agissant sur l'ensemble E. Un ***domaine fondamental*** *ou une* ***transversale*** *est une partie de E contenant un et un seul élément de chaque orbite.*

Autrement dit, les images des éléments d'un domaine fondamental forment une partition de l'ensemble :

$$E = \bigsqcup_{g \in G} g(F) \tag{2.84}$$

où $g(F) = \phi_g(F) = \{\phi_g(x) \text{ tel que } x \in F\}$. L'union est disjointe, c'est-à-dire que si $g \neq g'$, alors $g(F) \cap g'(F) = \varnothing$.

Proposition 2.57 (Équation des classes[36])**.**
Soit G, un groupe fini opérant sur un ensemble E. Si E'' est un ensemble contenant exactement un élément de chaque orbite dans $E \backslash \mathrm{Fix}_G(E)$, alors

$$|G| = |\mathrm{Fix}_G(E)| + \sum_{x \in E''} \frac{|G|}{|\mathrm{Fix}_G(x)|}. \tag{2.85}$$

Si de plus G est un p-groupe, alors

$$|E| = |\mathrm{Fix}_G(E)| \mod p. \tag{2.86}$$

Démonstration. Par le corolaire 2.55, nous avons $|G| = \sum_{x \in E'} |\mathcal{O}_x|$ où E' est une transversale. En séparant la somme entre les orbites à un élément et les autres,

$$|G| = \mathrm{Card}(\mathrm{Fix}_G(E)) + \sum_{x \in E''} \frac{|G|}{|\mathrm{Fix}_G(x)|} \tag{2.87}$$

où nous avons utilisé le fait que $|G| = |\mathrm{Fix}_G(x)||\mathcal{O}_x|$.

Si G est un p-groupe alors si $x \in E''$, $\mathrm{Fix}_G(x)$ est un sous-groupe propre de G et donc $|\mathrm{Fix}_G(x)|$ est un diviseur propre de $|G|$. Du coup la fraction $|G|/|\mathrm{Fix}_G(x)|$ est une puissance non nulle de p et l'équation (2.85) devient immédiatement (2.86). □

Corolaire 2.58 (Équation des classes)**.**
Soit G, un groupe et $C_1, \ldots, C_l$ la liste de ses classes de conjugaison contenant plus de un élément. Alors

$$\mathrm{Card}(G) = \mathrm{Card}\big(Z(G)\big) + \sum_i |G : Z_{g_i}| = \mathrm{Card}\big(Z(G)\big) + \sum_i \frac{\mathrm{Card}(G)}{\mathrm{Card}\big(\mathrm{Fix}(g_i)\big)} \tag{2.88}$$

si $g_i \in C_i$.

Démonstration. Étant donné que les classes de conjugaison sont disjointes, le cardinal de G est bien la somme des cardinaux de ses classes. Les classes ne contenant que un seul élément sont celles des éléments de $Z(G)$. En ce qui concerne les autres orbites, $\mathrm{Card}(C_{g_i}) = |G : Z_{g_i}|$ par le théorème orbite-stabilisateur (proposition 2.52). □

Théorème 2.59 (Formule de Burnside)**.**
Si G est un groupe fini agissant sur l'ensemble fini E et si Ω est l'ensemble des orbites, alors

$$\mathrm{Card}(\Omega) = \frac{1}{|G|}\sum_{g\in G}\mathrm{Card}\big(\mathrm{Fix}(g)\big). \tag{2.89}$$

Démonstration. Nous considérons l'ensemble

$$A = \{(g,x)\in G\times E \text{ tel que } gx = x\}, \tag{2.90}$$

et nous en calculons le cardinal de deux façons. D'abord

$$\begin{aligned}
\mathrm{Card}(A) &= \sum_{x\in E}\mathrm{Card}\{g\in g \text{ tel que } gx = x\} && (2.91a)\\
&= \sum_{x\in E}\mathrm{Card}(\mathrm{Fix}(x)) && (2.91b)\\
&= \sum_{\omega\in\Omega}\sum_{x\in\omega}\mathrm{Card}(\mathrm{Fix}(x)) && (2.91c)\\
&= \sum_{\omega\in\Omega}\frac{|G|}{\mathrm{Card}(\omega)} && (2.91d)\\
&= |G|. && (2.91e)
\end{aligned}$$

Pour obtenir (2.91d) nous avons utilisé l'équation des classes (2.79). L'autre façon de calculer $\mathrm{Card}(A)$ est de regrouper ainsi :

$$\mathrm{Card}(A) = \sum_{g\in G}\mathrm{Card}\{x\in E \text{ tel que } gx = x\} = \sum_{g\in G}\mathrm{Card}(\mathrm{Fix}(g)). \tag{2.92}$$

En égalisant les deux expressions de $\mathrm{Card}(A)$ nous trouvons

$$|G|\,\mathrm{Card}(\Omega) = \sum_{g\in G}\mathrm{Card}(\mathrm{Fix}(g)). \tag{2.93}$$

□

Proposition 2.60.
Soit G un groupe et H, un sous-groupe du centre de G.

(1) H est normal dans G.

(2) Si G/H est monogène, alors G est abélien.

(3) Si G est fini de centre Z, alors $|G:H|$ n'est pas premier.

Théorème 2.61.
Soit G un groupe cyclique[15] *d'ordre n.*

(1) Tout sous-groupe de G est cyclique.

(2) Pour chaque diviseur d de n, il existe un unique sous-groupe H_d de G d'ordre d.

Si a est un générateur de G, alors H_d peut être décrit des façons suivantes :

$$H_d = \{x\in G \text{ tel que } x^d = e\} = \{x\in G \text{ tel que } \exists y\in G \text{ tel que } y^{n/d} = x\} = \langle a^{n/d}\rangle. \tag{2.94}$$

Définition 2.62.
Soit G un groupe agissant sur un ensemble E. Nous disons que l'action est ***transitive*** *si elle possède une seule orbite. L'action est* ***libre*** *si $g\cdot x = g'\cdot x$ implique $g = g'$.*

15. Définition 2.12.

2.9 Permutations, groupe symétrique

Nous donnons ici quelques éléments à propos du groupe symétrique. Beaucoup de choses supplémentaires sont reportées à la section 5.6. Voir aussi le thème 61.

Définition 2.63.
*Soit un ensemble E. Une **permutation** de l'ensemble E est une bijection $E \to E$. Le **groupe symétrique** de E le groupe des bijections $E \mapsto E$; il est noté S_E.*

*Le **groupe symétrique** S_n est le groupe des permutations de l'ensemble $\{1, \dots, n\}$. C'est donc l'ensemble des bijections $\{1, \dots, n\} \to \{1, \dots, n\}$.*

Définition 2.64.
*Une **transposition** est une permutation qui inverse deux éléments. Plus précisément, une bijection $\sigma\colon E \to E$ est une transposition si il existe $a, b \in E$ tels que*

$$\sigma(x) = \begin{cases} a & si\ x = b \\ b & si\ x = a \\ x & sinon. \end{cases} \tag{2.95}$$

Lemme 2.65 ([37])**.**
Le groupe symétrique S_n est un ensemble fini contenant $n!$ éléments.

Lemme 2.66 ([38])**.**
Deux résultats.

(1) Tout groupe est isomorphe à un sous-groupe d'un groupe symétrique.

(2) Tout groupe fini d'ordre n est isomorphe à un sous-groupe de S_n.

Démonstration. Soit, pour $g \in G$ donné, l'application

$$\begin{aligned} \tau_g\colon G &\to G \\ x &\mapsto gx. \end{aligned} \tag{2.96}$$

Cela est une bijection de G. En effet, d'une part, $\tau_g(x) = y$ pour $x = g^{-1}y$ (surjection) et, d'autre part, $\tau_g(x) = \tau_g(y)$ implique $gx = gy$ et donc $x = y$ (injection).

Nous avons donc $\tau_g \in S_G$. De plus l'application

$$\begin{aligned} \varphi\colon G &\to S_G \\ g &\mapsto \tau_g \end{aligned} \tag{2.97}$$

est un morphisme de groupe. Il est injectif parce que si $\tau_g = \tau_h$ alors $gx = hx$ pour tout x. En particulier $g = h$.

Donc $\varphi\colon G \to \mathrm{Image}(\varphi)$ est un isomorphisme entre G et un sous-groupe de S_G.

Un groupe fini de cardinal n est isomorphe à un sous-groupe de S_G ; or S_G est isomorphe à un des S_n. □

2.9.1 Décomposition en cycles

Définition 2.67.
*Le **support** d'une permutation σ est l'ensemble constitué des éléments modifiés par σ :*

$$\operatorname{supp}\sigma = \{i \in \{1, \dots, n\} \text{ tel que } \sigma(i) \neq i\}.$$

Définition 2.68.
*Nous disons qu'un élément $\sigma \in S_n$ **inverse** les nombres $i < j$ si $\sigma(i) > \sigma(j)$. Soit N_σ le nombre d'inversions que $\sigma \in S_n$ possède (c'est le nombre de couples (i, j) avec $i < j$ tels que $\sigma(i) > \sigma(j)$). L'entier*

$$\epsilon(\sigma) = (-1)^{N_\sigma} \tag{2.98}$$

*est la **signature** de σ.*

Un élément du groupe symétrique S_n peut être décomposé en produit de cycles de supports disjoints de la façon suivante. Pour $\sigma \in S_n$, nous écrivons d'abord le cycle qui correspond à l'orbite de 1. Ce sera le cycle

$$(1, \sigma(1), \sigma^2(1), \dots, \sigma^k(1)) \tag{2.99}$$

avec $\sigma^{k+1}(1) = 1$. Ensuite nous recommençons avec le plus petit élément de $\{1, \dots, n\}$ à ne pas être dans ce cycle, et puis le suivant, etc. La *structure* d'une telle décomposition est la donnée des nombres k_i donnant le nombre de cycles de longueur i.

Lemme 2.69 ([35])**.**
Soit $\sigma = (i_1, \dots, i_k) \in S_n$, un cycle de longueur k et $\theta \in S_n$. Alors

$$\theta\sigma\theta^{-1} = \big(\theta(i_1), \dots, \theta(i_k)\big). \tag{2.100}$$

Tous les cycles de longueur k sont conjugués entre eux.

Proposition 2.70 (Classes de conjugaison et structure en cycles[39])**.**
Une classe de conjugaison dans S_n est formée des permutations ayant une décomposition en cycles disjoints de même structure. Autrement dit, deux permutations σ et σ' sont conjuguées si et seulement si le nombre k_i de cycles de longueur i dans σ est le même que le nombre k'_i de cycles de longueur i dans σ'.

Démonstration. Soit $\sigma = c_1 \dots c_m$ la décomposition de σ en cycles de supports disjoints. Les c_i sont des cycles de supports disjoints. Si τ est une permutation, alors

$$\sigma' = \tau\sigma\tau^{-1} = (\tau c_1 \tau^{-1}) \dots (\tau c_m \tau^{-1}), \tag{2.101}$$

mais $\tau c_i \tau^{-1}$ est un cycle de même longueur que c, puisque le lemme 2.69 nous dit que si $\sigma = (a_1, \dots, a_k)$, alors $\tau c \tau^{-1} = \big(\tau(a_1), \dots, \tau(a_k)\big)$. Notons encore que les cycles $\tau c_i \tau^{-1}$ restent à support disjoints.

Donc tous les éléments de la classe de conjugaison de σ sont des permutations de même structure de σ.

Réciproquement, si $\sigma' = c'_1 \dots c'_m$ est une décomposition de σ' en cycles disjoints tels que la longueur de c_i est la même que la longueur de c'_i, alors il suffit de construire des permutations τ_i telles que $\tau_i c_i \tau_i^{-1} = c'_i$, à travers le lemme 2.69. Comme les supports des c_i et des c'_i sont disjoints, la permutation $\tau_1 \dots \tau_m$ conjugue σ et σ'. □

Exemple 2.71.
Voyons les classes de conjugaison de S_3. Étant donné que ce groupe agit par définition sur un ensemble à 3 éléments, aucun élément de S_3 ne possède un cycle de plus de 3 éléments. Il y a donc seulement des cycles de longueur deux ou trois (à part les triviaux). Aucun élément de S_3 n'a une décomposition en cycles disjoints contenant deux cycles de deux ou un cycle de deux et un de trois.

En résumé il y a trois classes de conjugaison dans S_3. La première est celle contenant seulement l'identité. La seconde est celle contenant les cycles de longueur deux et la troisième contient les cycles de longueur 3.

Ce sont donc

$$C_1 = \{\mathrm{Id}\} \tag{2.102a}$$
$$C_2 = \{(1,2), (1,3), (2,3)\} \tag{2.102b}$$
$$C_3 = \{(1,2,3), (2,1,3)\}. \tag{2.102c}$$

△

Exemple 2.72.
Les classes de conjugaison de S_4. Nous savons que les classes de conjugaison dans S_4 sont caractérisées par la structure des décompositions en cycles (proposition 2.70). Le groupe symétrique S_4 possède dont les classes de conjugaison suivantes.

(1) Le cycle vide qui représente la classe constituée de l'identité seule.

(2) Les transpositions (de type (a, b)) qui sont au nombre de 6.

(3) Les 3-cycles. Pour savoir quel est leur nombre nous commençons par remarquer qu'il y a 4 façons de prendre 3 nombres parmi 4 et ensuite 2 façons de les arranger. Il y a donc 8 éléments dans cette classe de conjugaison.

(4) Les 4-cycles. Le premier est arbitraire (parce que c'est cyclique). Pour le second il y a 3 possibilités, et deux possibilités pour le troisième ; le quatrième est alors automatique. Cette classe de conjugaison contient donc 6 éléments.

(5) Les doubles transpositions, du type $(a, b)(c, d)$. Dans ce cas, tous les nombres sont permutés, et l'image de 1 détermine la double transposition. Il y a 3 images possibles, et donc 3 éléments dans cette classe.

△

Proposition 2.73.
Tout élément de S_n peut être écrit sous la forme d'un produit fini de transpositions.

Cette décomposition n'est pas à confondre avec celle en cycles de support disjoints. Par exemple $(1, 2, 3) = (1, 3)(1, 2)$.

Proposition-définition 2.74.
Si une permutation peut être écrire sous forme d'un produit d'un nombre pair de permutations, alors toute décomposition en permutations sera en quantité paire.

Une telle permutation est une ***permutation paire****.*

Lemme 2.75 ([31])**.**
Un k-cycle est une permutation impaire si k est pair et paire si k est impair.

Proposition 2.76 ([35])**.**
Soit S_n le groupe symétrique.

(1) L'application $\epsilon\colon S_n \to \{1, -1\}$ est l'unique homomorphisme surjectif de S_n sur $\{-1, 1\}$.

(2) Si $s = t_1 \cdots t_k$ est le produit de k transpositions, alors $\epsilon(s) = (-1)^k$.

Démonstration. Soit $\sigma, \theta \in S_n$. Afin de montrer que $\epsilon(\sigma\theta) = \epsilon(\sigma)\epsilon(\theta)$, nous divisons les couples (i, j) tels que $i \leqslant j$ en 4 groupes suivant que $\theta(i) \gtrless \theta(j)$ et $\sigma\big(\theta(i)\big) \gtrless \sigma\big(\theta(j)\big)$. Nous notons N_1, N_2, N_3 et N_4 le nombre de couples dans chacun des quatre groupes :

(i, j)	$\sigma\big(\theta(i)\big) < \sigma\big(\theta(j)\big)$	$\sigma\big(\theta(i)\big) > \sigma\big(\theta(j)\big)$
$\theta(i) < \theta(j)$	N_1	N_2
$\theta(i) > \theta(j)$	N_3	N_4

Nous avons immédiatement $N_\theta = N_3 + N_4$ et $N_{\sigma\theta} = N_2 + N_4$. Les éléments qui participent à N_σ sont ceux où $\theta(i)$ et $\theta(j)$ sont dans l'ordre inverse de $\sigma\big(\theta(i)\big)$ et $\sigma\big(\theta(j)\big)$ (parce que θ est une bijection). Donc $N_\sigma = N_2 + N_3$. Par conséquent nous avons

$$\epsilon(\sigma)\epsilon(\theta) = (-1)^{N_2+N_3}(-1)^{N_3+N_4} = (-1)^{N_2+N_4} = (-1)^{N_{\sigma\theta}} = \epsilon(\sigma\theta). \tag{2.103}$$

Nous avons prouvé que ϵ est un homomorphisme. Pour montrer que ϵ est surjectif sur $\{-1, 1\}$ nous devons trouver un élément $\tau \in S_n$ tel que $\epsilon(\tau) = -1$. Si τ est la transposition $1 \leftrightarrow 2$ alors le couple $(1, 2)$ est le seul à être inversé par τ et nous avons $\epsilon(\tau) = -1$.

Avant de montrer l'unicité, nous montrons que si $\sigma = t_1 \dots t_k$ alors $\epsilon(\sigma) = (-1)^k$. Pour cela il faut montrer que $\epsilon(\tau) = -1$ dès que τ est une transposition. Soit τ_{ij}, la transposition (i, j) et $\theta = (i, i + 1, \dots, j - 1)$ alors le lemme 2.69 dit que

$$\tau_{ij} = \theta\tau_{j-1,j}\theta^{-1}. \tag{2.104}$$

La signature étant un homomorphisme,

$$\epsilon(\tau_{ij}) = \epsilon(\theta)\epsilon(\tau_{j-1,j})\epsilon(\theta)^{-1} = \epsilon(\tau_{j-1,j}) = -1. \tag{2.105}$$

Nous passons maintenant à la partie unicité de la proposition. Soit un homomorphisme surjectif $\varphi\colon S_n \to \{-1,1\}$ et τ, une transposition telle que $\varphi(\tau) = -1$ (qui existe parce que sinon φ ne serait pas surjectif[16]). Si τ' est une autre transposition, il existe $\sigma \in S_n$ tel que $\tau' = \sigma\tau\sigma^{-1}$ (lemme 2.69). Dans ce cas, $\varphi(\tau') = \varphi(\tau) = -1$, et si $\sigma = \tau_1 \dots \tau_k$),

$$\varphi(\sigma) = (-1)^k = \epsilon(\sigma). \tag{2.106}$$

□

Corolaire 2.77.
Si $\sigma \in S_n$, alors

$$\epsilon(\sigma) = \epsilon(\sigma^{-1}). \tag{2.107}$$

Démonstration. Comme dit par la proposition 2.76, ϵ est un homomorphisme, donc

$$\epsilon(\sigma)\epsilon(\sigma^{-1}) = \epsilon(\sigma\sigma^{-1}) = \epsilon(\mathrm{Id}) = 1. \tag{2.108}$$

Vu que $\epsilon(\sigma)$ et $\epsilon(\sigma^{-1})$ ne peuvent être que ± 1, ils doivent être tous les deux 1 ou tous les deux -1 pour que le produit soit 1. □

2.10 Produit semi-direct de groupes

Définition 2.78.
*Une **suite exacte** est une suite d'applications comme suit :*

$$\dots \xrightarrow{f_i} A_i \xrightarrow{f_{i+1}} A_{i+1} \xrightarrow{f_{i+2}} \dots \tag{2.109}$$

où pour chaque i, les applications f_i et f_{i+1} vérifient $\ker(f_{i+1}) = \mathrm{Image}(f_i)$. *Lorsque les ensembles A_i sont des groupes, alors nous demandons de plus que les f_i soient de homomorphismes.*

Très souvent nous sommes confrontés à des suites exactes de la forme

$$1 \longrightarrow A \xrightarrow{f} G \xrightarrow{g} B \longrightarrow 1 \tag{2.110}$$

où G, A et B sont des groupes, 1 est l'identité. La première flèche est l'application $\{1\} \to A$ qui à 1 fait correspondre 1. La dernière est l'application $B \to 1$ qui à tous les éléments de B fait correspondre 1. Le noyau de f étant l'image de la première flèche (c'est-à-dire $\{1\}$), l'application f est injective. L'image de g étant le noyau de la dernière flèche (c'est-à-dire B en entier), l'application g est surjective.

Définition 2.79.
*Soient N et H deux groupes et un morphisme de groupes $\phi\colon H \to \mathrm{Aut}(N)$. Le **produit semi-direct** de N et H relativement à ϕ, noté $N \times_\phi H$ est l'ensemble $N \times H$ muni de la loi (que l'on vérifiera être de groupe)*

$$(n,h) \cdot (n',h') = (n\phi_h(n'), hh'). \tag{2.111}$$

Attention à l'ordre quelque peu contre-intuitif. Lorsque nous notons $N \times_\phi H$, c'est bien $\phi\colon H \to \mathrm{Aut}(N)$, c'est-à-dire H qui agit sur N et non le contraire.

Lorsque N et H sont des sous-groupes d'un même groupe, le plus souvent ϕ est l'action adjointe définie en 2.50.

Le théorème suivant permet de reconnaitre un produit semi-direct lorsqu'on en voit un.

16. Nous utilisons ici le fait que tous les éléments de S_n sont des produits de transpositions, proposition 2.73.

Théorème 2.80 ([6]).
Soit une suite exacte de groupes

$$1 \longrightarrow N \xrightarrow{i} G \xrightarrow{s} H \longrightarrow 1 \tag{2.112}$$

S'il existe un sous-groupe $\tilde{H}$ de G à partir duquel s est un isomorphisme, alors

$$G \simeq i(N) \times_\sigma \tilde{H} \tag{2.113}$$

où σ est l'action adjointe[17] *de $\tilde{H}$ sur $i(N)$.*

Démonstration. Nous posons $\tilde{N} = i(N)$ et nous allons subdiviser la preuve en petits pas.

(1) $\tilde{N}$ est normal dans G. En effet étant donné que la suite est exacte nous avons $\tilde{N} = \ker(s)$. Le noyau d'un morphisme est toujours un sous-groupe normal.

(2) $\tilde{N} \cap \tilde{H} = \{e\}$. L'application s étant un isomorphisme depuis $\tilde{H}$, il n'y a pas d'éléments de $\tilde{H}$ dans $\ker(s)$ autre que e.

(3) $G = \tilde{N}\tilde{H}$. Nous considérons $g \in G$ et $h \in \tilde{H}$ tel que $s(g) = s(h)$. L'existence d'un tel h est assurée par le fait que s est surjective depuis $\tilde{H}$. Du coup nous avons $e = s(gh^{-1})$, c'est-à-dire $gh^{-1} \in \ker(s) = \tilde{N}$. Nous avons donc bien la décomposition $g = (gh^{-1})h$, et donc $G = \tilde{N}\tilde{H}$.

(4) L'écriture $g = nh$ avec $n \in \tilde{N}$ et $h \in \tilde{H}$ est unique. Si $nh = n'h'$, alors $n = n'h'h^{-1}$, ce qui signifierait que $h'h^{-1} \in \tilde{N}$. Mais étant donné que $\tilde{H} \cap \tilde{N} = \{e\}$, nous obtenons $h = h'$ et par suite $n = n'$.

(5) L'application

$$\begin{aligned} \phi\colon G &\to \tilde{N} \times \tilde{H} \\ nh &\mapsto (n, h) \end{aligned} \tag{2.114}$$

est une bijection. C'est une conséquence des points (3) et (4).

(6) Si sur $\tilde{N} \times \tilde{H}$ nous mettons le produit

$$(n, h) \cdot (n', h') = (n\sigma_h n', hh') \tag{2.115}$$

où σ est l'action adjointe du groupe sur lui-même, c'est-à-dire $\sigma_x(y) = xyx^{-1}$, alors ϕ est un isomorphisme. Si $g, g' \in G$ s'écrivent (de façon unique par le point (5)) $g = nh$ et $g' = n'h'$ alors

$$\begin{aligned} \phi(nhn'h') &= \phi(n \underbrace{hn'h^{-1}}_{\in \tilde{N}} hh') && \text{(2.116a)} \\ &= \phi\big((nhn'h^{-1})(hh')\big) && \text{(2.116b)} \\ &= (nhn'h^{-1}, hh') && \text{(2.116c)} \\ &= (n, h) \cdot (n', h') && \text{(2.116d)} \\ &= \phi(nh)\phi(n'h'). && \text{(2.116e)} \end{aligned}$$

□

Corolaire 2.81.
Soit G un groupe, et N, H des sous-groupes de G tels que

(1) H normalise N (c'est-à-dire que $hnh^{-1} \in N$ pour tout $h \in H$ et $n \in N$[18]*),*

(2) $H \cap N = \{e\}$,

(3) $HN = G$.

17. Le fait que H agisse sur $i(N)$ fait partie du théorème.
18. Ou encore que H agit sur N par automorphismes internes.

Alors l'application

$$\begin{aligned} \psi\colon N \times_\sigma H &\to G \\ (n,h) &\mapsto nh \end{aligned} \tag{2.117}$$

est un isomorphisme de groupes.

Dans les hypothèses, l'ordre entre N et H est important lorsqu'on dit que c'est N qui agit sur H ; mais l'hypothèse $NH = G$ est équivalente à $HN = G$ (passer à l'inverse pour s'en assurer).

Insistons encore un peu sur la notation : dans $N \times_\sigma H$, c'est H qui agit sur N par σ.

2.11 Groupe de torsion

Soit G un groupe. Un élément $g \in G$ est un **élément de torsion** s'il est d'ordre fini. La **torsion** de G est l'ensemble de ses éléments de torsion. Nous disons qu'un groupe est un **groupe de torsion** si tous ses éléments sont de torsion.

Exemple 2.82.
Le groupe additif $\mathbb{Q}/\mathbb{Z}$ est un groupe de torsion parce que si $[x] = [p/q]$, alors $q[x] = [p] = [0]$. $\triangle$

2.12 Famille presque nulle

Soit $(G,+)$ un groupe abélien et $\mathcal{F} = \{g_i\}_{i\in I}$ une famille d'éléments de G indicés par un ensemble I. Le **support** de $\mathcal{F}$ est l'ensemble $\{i \in I \text{ tel que } g_i \neq 0\}$. La famille est dite **presque nulle** si le support est fini.

Nous disons que $\mathcal{F}$ est une **suite** si $I = \mathbb{N}$.

Chapitre 3

Anneaux

Attention aux conventions. Dans le Frido, un corps peut être réduit à $\{0\}$ et un idéal premier ne peut pas être $\{0\}$. Ces conventions ont une série de conséquences un peu partout, par exemple dans la proposition 3.121 où nous parlons d'idéal maximum propre. Comparez par exemple avec [40]. Soyez attentifs.

En cas de doutes, nous suivons les conventions de Wikipédia.

3.1 Inversibles et nilpotents

Le concept d'anneau est la définition 1.73.

Lemme 3.1.
Si a et b commutent, nous avons, pour tout $r \in \mathbb{N}$ et $r > 0$, la formule

$$a^{r+1} - b^{r+1} = (a-b)\left(\sum_{k=0}^{r} a^{r-k}b^k\right). \tag{3.1}$$

Démonstration. Démontrons cela par récurrence. Le cas $r = 0$ est évident. Pour un r donné, si (3.1) est vraie, alors

$$\begin{aligned}
a^{r+2} - b^{r+2} &= a^{r+1}a - a^{r+1}b + a^{r+1}b - b^{r+1}b \\
&= a^{r+1}(a-b) + (a^{r+1} - b^{r+1})b \\
&= a^{r+1}(a-b) + (a-b)\left(\sum_{k=0}^{r} a^{r-k}b^k\right)b \\
&= (a-b)\left(a^{r+1} + \left(\sum_{k=0}^{r} a^{r-k}b^k\right)b\right) \\
&= (a-b)\left(a^{r+1} + \sum_{k=0}^{r} a^{r-k}b^{k+1}\right) \\
&= (a-b)\left(a^{r+1} + \sum_{k'=1}^{r+1} a^{(r+1)-k'}b^{k'}\right) \\
&= (a-b)\left(\sum_{k'=0}^{r+1} a^{(r+1)-k'}b^{k'}\right).
\end{aligned}$$

□

Proposition 3.2.
Si a est un élément nilpotent de l'anneau A, alors $1-a$ est inversible. Si a est nilpotent non nul, alors il est diviseur de zéro.

Démonstration. Soit n le minimum tel que $a^n = 0$. En vertu de la formule (3.1) nous avons

$$1 = 1 - a^n = (1-a)(1 + a + \cdots + a^{n-1}) = (1 + a + \cdots + a^{n-1})(1-a). \tag{3.2}$$

La somme $1 + a + \cdots + a^{n-1}$ est donc un inverse de $(1-a)$. □

3.2 PGCD, PPCM et éléments inversibles

La définition de pgcd et ppcm dans un anneau commutatif est la définition 1.81. Dans la plus grande tradition, elle a été introduite sans motivations, et utilisée par-ci par-là. Nous revenons maintenant dessus.

Commençons par donner une autre vision de la divisibilité dans les anneaux intègres.

Proposition 3.3.
Dans un anneau intègre[1] *A, on a l'équivalence suivante concernant deux éléments $a, b \in A$:*

$$a \mid b \Leftrightarrow (b) \subset (a). \tag{3.3}$$

Donc la divisibilité devient en réalité une relation d'ordre dont nous pouvons chercher un maximum et un minimum. Si S est une partie de A, nous notons $a \mid S$ pour exprimer que $a \mid x$ pour tout $x \in S$; de la même façon, $S \mid b$ signifie que $x \mid b$ pour tout $x \in S$.

Nous rappelons également la définition 1.74 de morphisme d'anneaux. Remarquons que si f est un morphisme, nous avons $f(0) = 0$ et $f(x)^{-1} = f(x^{-1})$.

Lemme 3.4 ([41])**.**
Les éléments inversibles d'un anneau sont diviseurs de tous les éléments.

Démonstration. Soit k inversible d'inverse k' : $kk' = 1$; soit aussi $a \in A$. Alors $a = k(k'a)$, ce qui montre que k divise a. □

Lemme 3.5 ([41])**.**
Dans un anneau, 1 est un pgcd de a et b si et seulement si les seuls diviseurs communs sont les inversibles.

Démonstration. Supposons pour commencer que 1 est un pgcd de a et b. Un diviseur commun de a et b doit donc diviser 1. Or un diviseur de 1 est forcément inversible.

Dans l'autre sens, les diviseurs communs de a et b sont tous inversibles et donc diviseurs de 1. Donc 1 est un pgcd de a et b. □

3.3 Le groupe et anneau des entiers

Certes $(\mathbb{Z}, +)$ est un groupe mais en ajoutant la multiplication, $(\mathbb{Z}, +, \times)$ devient un anneau[2]

3.3.1 Division euclidienne

Théorème 3.6 (Division euclidienne[42])**.**
Soient $a \in \mathbb{Z}$ et $b \in \mathbb{N}^*$. Il existe un unique couple $(q, r) \in \mathbb{Z} \times \mathbb{N}$, avec $0 \leqslant r < b$, tel que

$$a = bq + r. \tag{3.4}$$

Démonstration. Remarquons que $r = a - bq$, et donc, une fois l'existence et l'unicité de q établie, celle de r suivra.

1. Définition 1.88.
2. Définition 1.73.

Unicité Nous supposons avoir $(q, r) \in \mathbb{Z} \times \mathbb{N}$ tels que

$$\begin{cases} 0 \leqslant r < b & (3.5a) \\ a = qb + r. & (3.5b) \end{cases}$$

Ce système implique que

$$0 \leqslant a - qb < b. \tag{3.6}$$

En ajoutant qb dans les trois membres de cette inégalité,

$$qb \leqslant a < (q+1)b. \tag{3.7}$$

Cela implique que

$$q = \max\{k \in \mathbb{Z} \text{ tel que } kb \leqslant a\}. \tag{3.8}$$

Donc q est unique et la relation $a = bq + r$ implique que r est également unique.

$$E = \{q \in \mathbb{Z} | bq \leqslant a\}.$$

C'est un sous-ensemble d'entiers non-vide (il contient $-|a|$) et admet $|a|$ comme majorant ; il admet donc un maximum q par le lemme 1.83. Ce maximum vérifie

$$bq \leqslant a < b(q+1). \tag{3.9}$$

Cela donne $0 \leqslant a - bq < b$ et le résultat en posant $r = a - qb$.

□

Définition 3.7.
*L'opération $(a, b) \mapsto (q, r)$ donnée par le théorème 3.6 est la **division euclidienne**. Le nombre q est le **quotient** et r est le **reste** de la division de a par b.*

3.3.2 Sous-groupes de $(\mathbb{Z}, +)$

Proposition 3.8.
Une partie H du groupe $(\mathbb{Z}, +)$ est un sous-groupe si et seulement s'il existe $n \in \mathbb{N}$ tel que $H = n\mathbb{Z}$.

Démonstration. Soit $H \neq \{0\}$ un sous-groupe de $\mathbb{Z}$. L'ensemble $H \cap \mathbb{N}^*$ contient un élément minimum que nous notons n. Nous avons certainement $n\mathbb{Z} \subset H$ parce que H est un groupe (donc $n + n$ et $-n$ sont dans H dès que n est dans H). Nous devons prouver que $H \subset n\mathbb{Z}$.

Si $x \in H$, par le théorème de division euclidienne 3.6, il existe $q \in \mathbb{Z}$ et $r \in \mathbb{N}$, uniques, tels que $x = nq + r$ et $0 \leqslant r < n$. Nous savons déjà que $nq \in H$, donc $r = x - nq \in H$. Le nombre r est donc un élément de H strictement plus petit que n. Mais nous avions décidé que n serait le plus petit élément de $H \cap \mathbb{N}^*$. Par conséquent $r = 0$ et $x = nq \in n\mathbb{Z}$. □

Notons que si un sous-groupe H de $\mathbb{Z}$ est donné, alors le nombre n tel que $H = n\mathbb{Z}$ est unique. En effet si $n\mathbb{Z} = m\mathbb{Z}$ nous avons que n divise m (parce que $m \in m\mathbb{Z} \subset n\mathbb{Z}$) et que m divise n parce que $n \in m\mathbb{Z}$. Par conséquent $n = m$.

3.3.3 PGCD, PPCM et Bézout

Vu que $\mathbb{Z}$ est un anneau intègre, nous avons la définition 1.81 de pgcd et de ppcm.

Proposition 3.9 (PPCM et PGCD)**.**
Soient $p, q \in \mathbb{Z}^*$.

(1) Le pgcd de p et q est le plus grand diviseur commun de p et q.

(2) Le ppcm de p et q est leur plus petit multiple commun.

Démonstration. Démontrons le premier point. Notons δ le pgcd de p et q. Si d est un diviseur commun de p et q, alors d divise δ. Dans $\mathbb{Z}$, $d \mid \delta$ implique $d \leqslant \delta$ (proposition 1.84). □

Lemme 3.10.
Soient $p, q \in \mathbb{Z}^*$. Les entiers $\operatorname{ppcm}(p,q)$ *et* $\operatorname{pgcd}(p,q)$ *fournissent les isomorphismes de groupes suivants :*

$$p\mathbb{Z} \cap q\mathbb{Z} = \operatorname{ppcm}(p,q)\mathbb{Z} \tag{3.10a}$$
$$p\mathbb{Z} + q\mathbb{Z} = \operatorname{pgcd}(p,q)\mathbb{Z}. \tag{3.10b}$$

Définition 3.11.
Si $\operatorname{pgcd}(p,q) = 1$*, nous disons que p et q sont* ***premiers entre eux****. Si nous avons un ensemble d'entiers a_i, nous disons qu'ils sont premiers* ***dans leur ensemble*** *si 1 est le PGCD de tous les a_i ensemble.*

Les nombres 2, 4 et 7 ne sont pas premiers deux à deux (à cause de 2 et 4), mais ils sont premiers dans leur ensemble parce qu'il n'y a pas de diviseurs communs à tout le monde.

Définition 3.12.
Un ***nombre premier*** *est un naturel acceptant exactement deux diviseurs distincts.*

Avec cette définition, 0 n'est pas premier, 1 n'est pas premier et 2 est premier.

Théorème 3.13 (Théorème de Bézout [3][43], thème 50)**.**
Deux entiers non nuls $a, b \in \mathbb{Z}^*$ sont premiers entre eux si et seulement s'il existe $u, v \in \mathbb{Z}$ tels que

$$au + bv = 1 \tag{3.11}$$

Démonstration. Soit $d = \operatorname{pgcd}(a,b)$ et des nombres u, v tels que $au + bv = 1$. Le PGCD d divise à la fois a et b, et donc divise $au + bv$. Nous en déduisons que d divise 1 et est par conséquent égal à 1.

Nous supposons maintenant que $\operatorname{pgcd}(a,b) = 1$ et nous considérons l'ensemble

$$E = \{au + bv \text{ tel que } u, v \in \mathbb{Z}\} \cap \mathbb{N}^*. \tag{3.12}$$

C'est-à-dire l'ensemble des nombres strictement positifs pouvant s'écrire sous la forme $au+bv$. Cet ensemble est non vide parce qu'il contient par exemple soit a soit $-a$. Soit m le plus petit élément de E et écrivons

$$m = au_1 + bv_1. \tag{3.13}$$

Par le théorème de division euclidienne [4] (avec a et m), il existe des entiers uniques q et r tels que

$$a = mq + r \tag{3.14}$$

avec $0 \leqslant r < m$. En remplaçant m par sa valeur (3.13), $a = (au_1 + bv_1)q + r$ et

$$r = a(1 - u_1 q) - bv_1 q, \tag{3.15}$$

c'est-à-dire que $r \in \mathbb{Z}a + \mathbb{Z}b$ en même temps que $0 \leqslant r < m$. Si r était strictement positif, il serait dans E. Mais cela est impossible par minimalité de m. Donc $r = 0$ et a est divisible par m.

De la même façon nous prouvons que b est divisible par m. Vu que m divise à la fois a et b nous avons $m = 1$. □

Corolaire 3.14.
Soient p et q deux entiers premiers entre eux. Alors

$$p\mathbb{Z} + q\mathbb{Z} = \mathbb{Z}; \tag{3.16}$$

en particulier, pour tout $x \in \mathbb{Z}$, il existe u_x, v_x entiers tels que $u_x p + v_x q = x$.

3. Il y a une super application ici : https://perso.univ-rennes1.fr/matthieu.romagny/agreg/dvt/mauvais_prix.pdf.
4. Théorème 3.6.

Notons que l'application $p\mathbb{Z} + q\mathbb{Z}$ vers $\mathbb{Z}$ n'est évidemment pas injective : les u_x et v_x ne sont pas uniques à x fixé.

Démonstration. Soit $x \in \mathbb{Z}$. Le théorème de Bézout nous donne k et l tels que $kp + lq = 1$. Du coup, $(xk)p + (xl)q = x$. □

La proposition suivante établit que si x est assez grand, alors il peut même être écrit comme une combinaison de p et q à coefficients positifs. Elle sera utilisée pour démontrer que les états apériodiques d'une chaine de Markov peuvent être atteints à tout moment (assez grand), voir la définition 37.47 et ce qui suit.

Proposition 3.15.
Soient a et b deux éléments de $\mathbb{N}$ premiers entre eux. Il existe $N > 0$ tel que tout $x > N$ appartient à $a\mathbb{N} + b\mathbb{N}$.

Démonstration. Soient a et b, premiers entre eux, et $x \in \mathbb{N}$. Disons tout de suite, pour éviter les cas triviaux et pénibles, que x, a et b sont strictement positifs.

Une décomposition pour x On applique le théorème 3.6 de division euclidienne à x et $a + b$: il existe des entiers p_x, r_x, uniques, tels que

$$\begin{cases} x = (p_x - 1)(a + b) + r_x & (3.17a) \\ 0 \leqslant r_x < a + b. & (3.17b) \end{cases}$$

En d'autres termes, $p_x(a + b)$ est le premier multiple de $a + b$ supérieur ou égal à x. De plus, p_x est strictement positif car x l'est. Il existe alors des entiers u et v tels que

$$ua + vb = p_x(a + b) - x \tag{3.18}$$

par le corolaire 3.14. Ainsi, x peut s'écrire

$$x = (p_x - u)a + (p_x - v)b. \tag{3.19}$$

Des maximums Il s'agit maintenant de savoir si nous pouvons être assuré d'avoir $p_x > u$ et $p_x > v$ dès que x est assez grand. Pour cela, grâce au corolaire 3.14, nous considérons les nombres u_i et v_i définis par

$$u_i a + v_i b = i \tag{3.20}$$

pour $i = 1, \dots, a + b$. Nous posons $u^* = \max\{u_i\}$, $v^* = \max\{v_i\}$, et $p^* = \max\{u^*, v^*\}$. Nous posons alors $N = p^*(a + b)$, et considérons $x > N$.

Nouvelle décomposition pour x Nous voulons écrire

$$x = (p_x - u_k)a + (p_x - v_k)b \tag{3.21}$$

pour un certain k. Cela demande $u_k a + v_k b = ua + vb = p_x(a + b) - x$ par l'équation (3.18). Vu que $p_x(a+b) - x > 0$, les nombres u_k et v_k existent : il suffit de prendre $k = p_x(a+b) - x$.

Conclusion Avec tous ces choix, nous avons d'abord $x > p^*(a + b)$ et donc

$$x = (p_x - 1)(a + b) + r_x > p^*(a + b), \tag{3.22}$$

ce qui donne

$$(p_x - 1)(a + b) > p^*(a + b) - r_x > (p - 1)(a + b). \tag{3.23}$$

ou encore $p_x > p^*$. Nous avons finalement

$$p_x \geqslant p^* \geqslant u^* \geqslant u_k \tag{3.24}$$

et

$$p_x \geqslant p^* \geqslant v^* \geqslant v_k. \tag{3.25}$$

De ce fait, la décomposition (3.21) est celle que nous voulions.

□

3.16.
Une méthode pour obtenir les entiers naturels u et v qui permettent la décomposition $x = au + bv$ est d'abord de choisir u_0 et v_0 tels que au_0 et bv_0 soient les plus proches possibles de $x/2$, puis de décomposer le nombre (relativement petit) $x - au_0 - bv_0$ en $au_1 + bv_1$. Deux nombres u et v qui fonctionnent sont alors $u = u_0 + u_1$ et $v = v_0 + v_1$.

Exemple 3.17.
Écrivons $1000 = u \cdot 7 + v \cdot 5$ avec $u, v \in \mathbb{N}$. D'abord $72 \cdot 7 = 504$ et $100 \cdot 5 = 500$. Nous avons donc

$$1004 = 72 \cdot 7 + 100 \cdot 5. \tag{3.26}$$

Ensuite $4 = 25 - 21 = -3 \cdot 7 + 5 \cdot 5$. Au final,

$$1000 = 75 \cdot 7 + 95 \cdot 5. \tag{3.27}$$

△

3.3.4 Calcul effectif du PGCD et de Bézout

Soient a et b, deux entiers que nous allons prendre positifs. Nous allons voir maintenant l'algorithme de **Euclide étendu** qui est capable, pour a et b donnés, de calculer le PGCD de a et b, et un couple de Bézout (u, v) tel que $ua + vb = \operatorname{pgcd}(a, b)$. Ce calcul est indispensable si on veut implémenter RSA (18.2).

Cela se base sur le lemme suivant.

Lemme 3.18.
Soient $a, b \in \mathbb{N}$ et des nombres q et r tels que $a = qb + r$. Alors $\operatorname{pgcd}(a, b) = \operatorname{pgcd}(r, b)$.

Démonstration. Il suffit de voir que les diviseurs communs de a et b sont diviseurs de r et que les diviseurs communs de r et b divisent a.

Si s divise a et b, alors dans l'équation

$$\frac{a}{s} = \frac{qb}{s} + \frac{r}{s}$$

les termes a/s et qb/s sont entiers, donc r/s est aussi entier, et s divise r.

Inversement, si s divise r et b, alors il divise $qb + r$ et donc a. □

Remarque 3.19.
Ce lemme est surtout intéressant lorsque $a = qb + r$ est la division euclidienne de a par b : en effet, dans ce cas $r < b$, et le calcul du PGCD de deux nombres (a et b) est ramené à un calcul de PGCD de deux nombres plus petits (b et r).

L'algorithme pour calculer $\operatorname{pgcd}(a, b)$ consiste à écrire des divisions euclidiennes successives de la manière suivante :

$$a = q_2 b + r_2 \qquad r_2 < b \tag{3.28a}$$

$$b = q_3 r_2 + r_3 \qquad r_3 < r_2 \tag{3.28b}$$

$$\vdots \tag{3.28c}$$

en remarquant que $\operatorname{pgcd}(a, b) = \operatorname{pgcd}(b, r_2) = \operatorname{pgcd}(r_2, r_3)$. Étant donné que les inégalités $r_2 < b$ et $r_3 < r_2$ sont strictes, en continuant ainsi nous finissons sur zéro, c'est-à-dire qu'il existera un n pour lequel $r_{n+1} = 0$; et donc

$$r_{n-1} = q_{n+1} r_n,$$

et à ce moment nous avons $\operatorname{pgcd}(a, b) = \operatorname{pgcd}(r_{n-1}, r_n) = r_n$.

3.3.4.1 Algorithme d'Euclide pour le PGCD

Écrivons l'algorithme[44] en détail (parce que Bézout, ça va être la même chose en cinq fois plus compliqué). On pose

$$r_0 = a \tag{3.29a}$$
$$r_1 = b \tag{3.29b}$$

(ce qui explique que nous n'ayons pas utilisé r_0 et r_1 précédemment). Ensuite on écrit la division euclidienne $a = q_2 b + r_2$, c'est-à-dire $r_0 = q_2 r_1 + r_2$. Cela donne r_2 et q_2 en termes de r_0 et r_1 :

$$r_2 = r_0 - q_2 r_1. \tag{3.30}$$

Ensuite, sachant r_2 nous pouvons continuer :

$$r_1 = q_3 r_2 + r_3 \tag{3.31}$$

donne q_3 et $r_3 = r_1 - q_3 r_2$. On continue avec $r_2 = q_4 r_3 + r_4$. Tout cela pour poser la suite

$$\begin{aligned} r_0 &= a \\ r_1 &= b \\ r_k &= q_{k+2} r_{k+1} + r_{k+2} \end{aligned} \tag{3.32}$$

où la troisième définit r_{k+2} et q_{k+2} en fonction de r_k et r_{k+1}, à l'aide du théorème de la division euclidienne. La suite (r_k) ainsi construite est strictement décroissante et à chaque étape le lemme 3.18 et le principe de l'algorithme d'Euclide nous donnent

$$\begin{cases} \operatorname{pgcd}(r_k, r_{k+1}) = \operatorname{pgcd}(r_{k+1}, r_{k+2}) = \operatorname{pgcd}(a, b) & (3.33a) \\ 0 \leqslant r_{k+1} < r_k. & (3.33b) \end{cases}$$

La suite étant strictement décroissante, nous prenons r_n, le dernier non nul : $r_{n+1} = 0$. Dans ce cas la dernière équation sera

$$r_{n-1} = q_n r_n \tag{3.34}$$

avec $\operatorname{pgcd}(a, b) = \operatorname{pgcd}(r_n, r_{n-1}) = r_n$.

Exemple 3.20.
Calculons le PGCD de 18 et 231. Pour cela nous écrivons les divisions euclidiennes en chaine :

$$231 = 18 \cdot 12 + 15 \tag{3.35a}$$
$$18 = 1 \cdot 15 + 3 \tag{3.35b}$$
$$15 = 5 \cdot 5 + 0. \tag{3.35c}$$

Donc le PGCD est 3 (le dernier reste non nul). △

3.3.4.2 Algorithme étendu : calcul effectif des coefficients de Bézout

La difficulté est de construire la suite qui donne des coefficients de Bézout. Elle va être construite à l'envers. Nous supposons déjà connaitre la liste complète des r_k jusqu'à $r_n = \operatorname{pgcd}(a, b)$, ainsi que la liste complète des divisions euclidiennes

$$r_k = q_{k+2} r_{k+1} + r_{k+2}. \tag{3.36}$$

Nous voulons trouver les couples (u_k, v_k) de telle façon à avoir à chaque étape

$$r_n = u_k r_k + v_k r_{k-1}. \tag{3.37}$$

Notons que c'est à chaque fois r_n que nous construisons. La première équation de type Bézout à résoudre est

$$r_n = u_n r_n + v_n r_{n-1}, \tag{3.38}$$

sachant que $r_{n-1} = q_n r_n$. On pose $v_n = 0$ et $u_n = 1$ et c'est bon. Pour la récurrence, supposons les coefficients u_k et v_k connus, et déterminons les coefficients u_{k-1} et v_{k-1}. Pour ce faire, nous égalons les deux expressions pour r_n :

$$r_n = u_k r_k + v_k r_{k-1} = u_{k-1} r_{k-1} + v_{k-1} r_{k-2}; \tag{3.39}$$

dans laquelle nous substituons $r_{k-2} = q_k - r_{k-1} + r_k$:

$$\begin{aligned} u_k r_k + v_k r_{k-1} &= u_{k-1} r_{k-1} + v_{k-1}(q_k r_{k-1} + r_k) && (3.40)\\ &= (u_{k-1} + q_k v_{k-1}) r_{k-1} + v_{k-1} r_k && (3.41) \end{aligned}$$

et nous égalons les coefficients de r_k et r_{k-1} pour obtenir

$$\begin{cases} v_{k-1} = u_k & (3.42a)\\ u_{k-1} = v_k - v_{k-1} q_k. & (3.42b) \end{cases}$$

Dès que u_k et v_k ainsi que q_k sont connus, on peut calculer u_{k-1} et v_{k-1}.

La dernière équation, celle avec $k = 1$, est

$$r_n = u_1 r_1 + v_1 r_0, \tag{3.43}$$

c'est-à-dire

$$\operatorname{pgcd}(a, b) = u_1 b + v_1 a. \tag{3.44}$$

Nous avons ainsi trouvé des coefficients de Bézout pour a et b.

3.3.5 Générateurs

Proposition 3.21 ([45]).
Soit $n \geqslant 2$.

(1) L'anneau $\mathbb{Z}/n\mathbb{Z}$ contient n éléments

(2) Nous avons $\mathbb{Z}/n\mathbb{Z} = \{[k]_n\}_{k=0,\dots,n-1}$.

Proposition 3.22 ([45]).
Soient $n \geqslant 2$, a premier avec n, $b \in \mathbb{N}$, et $m_1, \dots, m_n \in \mathbb{N}$. Alors

(1) Nous avons $\mathbb{Z}/n\mathbb{Z} = \{[am_i + b]_n\}_{i=1,\dots,}$,

(2) Nous avons $\mathbb{Z}/n\mathbb{Z} = \{[b + k]_n\}_{k=0,\dots,n-1}$

(3) Nous avons $\mathbb{Z}/n\mathbb{Z} = \{[ka]_{k=0,\dots,n-1}\}$.

Proposition 3.23 ([45]).
Soit $n \geqslant 2$ et $m \in \mathbb{Z}$. Nous avons équivalence entre

(1) $\operatorname{pgcd}(n, m) = 1$,

(2) $[m]_n$ engendre le groupe $(\mathbb{Z}/n\mathbb{Z}, +)$

(3) $[m]_n$ est inversible dans $\big((\mathbb{Z}/n\mathbb{Z})^*, \ \cdot\ \big)$.

3.3.6 Générateurs pour le groupe multiplicatif

Cette section est presque certainement à déplacer.

Proposition 3.24 ([46]).
Le groupe $\big((\mathbb{Z}/n\mathbb{Z})^*, \ \cdot\ \big)$ est cyclique.

La proposition suivante est un pas important dans l'algorithme de Shor permettant aux ordinateurs quantiques de factoriser rapidement des grands nombres.

Proposition 3.25 ([47]).
Soient $A, B \in \mathbb{N}$ tels que $\operatorname{pgcd}(A, B) = 1$. Alors il existe $p, m \in \mathbb{N}$ tels que $A^p = mB + 1$.

3.3.7 Décomposition en facteurs premiers

Lemme 3.26 (Lemme de Gauss)**.**
Soient $a, b, c \in \mathbb{Z}$ tels que a divise bc. Si a est premier avec c, alors a divise b.

Démonstration. Vu que a est premier avec c, nous avons $\operatorname{pgcd}(a, c) = 1$ et le théorème de Bézout 3.13 nous donne donc $s, t \in \mathbb{Z}$ tels que $sa + tc = 1$. En multipliant par b, nous avons $sab + tbc = b$. Mais les deux termes du membre de gauche sont multiples de a parce que a divise bc. Par conséquent b est somme de deux multiples de a et donc est multiple de a. □

Lemme 3.27 (Lemme d'Euclide[48])**.**
Si un nombre premier p divise le produit de deux nombres entiers b et c, alors p divise b ou c.

Démonstration. Vu que p est premier, s'il ne divise pas a c'est que $\operatorname{pgcd}(a, p) = 1$. Dans ce cas le lemme de Gauss 3.26 implique que p divise b. □

Le théorème fondamental de l'arithmétique permet de décomposer des nombres en facteurs premiers.

Théorème 3.28 ([49])**.**
Tout entier strictement positif peut être écrit comme un produit de nombres premiers d'une unique façon, à l'ordre près des facteurs.

En d'autres termes, pour tout entier $n > 1$, il existe une suite finie unique $(p_1, k_1), \dots (p_r, k_r)$ telle que :

(1) les p_i sont des nombres premiers tels que, si $i < j$, alors $p_i < p_j$;

(2) les k_i sont des entiers naturels non nuls ;

(3) $n = \prod_{i=1}^{r} p_i^{k_i}$.

Démonstration. Soit n un entier positif. Nous prouvons l'existence d'une décomposition en facteurs premiers par récurrence. Le nombre $n = 1$ est le produit d'une famille finie de nombres premiers : la famille vide.

Supposons que tout entier strictement inférieur à un certain entier $n > 1$ est produit de nombres premiers. Deux possibilités apparaissent pour n : il est premier ou non. Si n est premier, et donc produit d'un unique entier premier, à savoir lui-même, le résultat est vrai. Si n n'est pas premier, il se décompose sous la forme kl avec k et l strictement inférieurs à n. Dans ce cas, l'hypothèse de récurrence implique que les entiers k et l peuvent s'écrire comme produits de nombres premiers. Leur produit aussi, ce qui fournit une décomposition de n en produit de nombres premiers. Par application du principe de récurrence, tous les entiers naturels peuvent s'écrire comme produit de nombres premiers.

Nous prouvons maintenant l'unicité. Prenons deux produits de nombres premiers qui sont égaux. Prenons n'importe quel nombre premier p du premier produit. Il divise le premier produit, et, de là, aussi le second. Par le lemme d'Euclide 3.27, p doit alors diviser au moins un facteur dans le second produit. Mais les facteurs sont tous des nombres premiers eux-mêmes, donc p doit être égal à un des facteurs du second produit. Nous pouvons donc simplifier par p les deux produits. En continuant de cette manière, nous voyons que les facteurs premiers des deux produits coïncident précisément. □

Lemme 3.29 ([1])**.**
Nous notons $\mathcal{P}$ l'ensemble des nombres premiers dans $\mathbb{N}$. Soient des suites finies $(a_p)_{p\in\mathcal{P}}$ et $(b_p)_{p\in\mathcal{P}}$. Nous posons

$$a = \prod_{p\mathcal{P}} p^{a_p} \quad \text{et} \quad b = \prod_{p\in\mathcal{P}} p^{b_p}. \tag{3.45}$$

Alors $a \mid b$ si et seulement si $a_p \leqslant b_p$ pour tout p.

Démonstration. Dire que $a \mid b$ signifie qu'il existe $s \in \mathbb{N}$ tel que $as = b$; le théorème 3.28 nous permet de décomposer s en $s = \prod_{p\in\mathcal{P}} p^{s_p}$. Vu que le produit dans $\mathbb{N}$ est commutatif et associatif,

$$b = as = \prod_{p\in\mathcal{P}} p^{s_p+a_p}. \tag{3.46}$$

Par unicité de la décomposition de b (toujours le théorème 3.28), nous en déduisons que $b_p = s_p + a_p \geqslant a_p$.

Dans l'autre sens, l'hypothèse $a_p \leqslant b_p$ implique l'existence de $s_p \geqslant 0$ tels que $b + p = a_p + s_p$. En posant $s = \prod_{p\in\mathcal{P}} p^{s_p}$, nous avons

$$as = \prod_{p\in\mathcal{P}} p^{s_p+a_p} = \prod_{p\in\mathcal{P}} p^{b_p} = b. \tag{3.47}$$

Donc $a \mid b$. □

Lemme 3.30 ([1])**.**
Dans $\mathbb{N}$, le pgcd[5] *et le ppcm sont uniques.*

Démonstration. Supposons que δ_1 et δ_2 soient des pgcd de la partie S. Vu que $\delta_1 \mid S$, nous avons $\delta_1 \mid \delta_2$ parce que δ_2 est un pgcd. Le même raisonnement, inversant δ_1 et δ_2 montre que $\delta_2 \mid \delta_1$. Si (a_p) sont les éléments de la décomposition de δ_1 et (b_p) ceux de δ_2, alors le lemme 3.29 nous indique que $a_p \leqslant b_p$ et $b_p \leqslant a_p$, ce qui implique que $a_p = b_p$.

Le ppcm se fait de même. □

Proposition 3.31.
Soient $a, b \in \mathbb{Z}\backslash\{0\}$ décomposés en $a = \prod_{p\in\mathcal{P}} p^{a_p}$ et $b = \prod_{p\in\mathcal{P}} p^{b_p}$. En posant

$$m_p = \min\{a_p, b_p\} \tag{3.48a}$$
$$M_p = \max\{a_p, b_p\}, \tag{3.48b}$$

nous avons

$$\operatorname{pgcd}(a, b) = \prod_{p\in\mathcal{P}} p^{m_p} \tag{3.49a}$$
$$\operatorname{ppcm}(a, b) = \prod_{p\in\mathcal{P}} p^{M_p}. \tag{3.49b}$$

Démonstration. Nous notons $\delta = \prod_{p\in\mathcal{P}} p^{m_p}$ et $\mu = \prod_{p\in\mathcal{P}} p^{M_p}$.

Nous commençons par montrer que δ est le pgcd de a et b. Vu que $\delta_p = \min\{a_p, b_p\}$, nous avons $\delta_p \leqslant a_p$ et $\delta_p \leqslant b_p$. Le lemme 3.29 nous dit alors que $\delta \mid a$ et $\delta \mid b$. De même, si $s \mid a$ et $s \mid b$, nous avons $s_p \leqslant a_p$ et $s_p \leqslant b_p$, ce qui montre que $s_p \leqslant m_p$ et donc que $s \mid \delta$.

Le fait que μ soit le ppcm de a et b se montre de même. □

Corolaire 3.32 ([1])**.**
Un élément $m \in \mathbb{Z}^*$ vérifie $m \leqslant p^n$ et $\operatorname{pgcd}(m, p^n) \neq 1$ si et seulement si $m = qp$ pour un certain $q \leqslant p^{n-1}$.

Problèmes et choses à faire

Il faut vérifier si le corolaire 3.32 est correct. Et puis rédiger des démonstrations de tout ce petit monde.

Lemme 3.33.
Un entier $n \geqslant 1$ se décompose de façon unique en produit de la forme $n = qm^2$ où q est un entier sans facteurs carrés et m, un entier.

5. Le pgcd et ppcm sont définis dans 1.81.

Démonstration. Pour $n = 1$, c'est évident. Nous supposons $n \geqslant 2$.

En ce qui concerne l'existence, nous décomposons n en facteurs premiers[6] et nous séparons les puissances paires des puissances impaires :

$$n = \prod_{i=1}^{r} p_p^{2\alpha_i} \prod_{j=1}^{s} q_j^{2\beta_j+1} \tag{3.50a}$$

$$= \underbrace{\left(\prod_{i=1}^{r} p_i^{2\alpha_i} \prod_{j=1}^{s} q^{2\beta_j}\right)}_{m^2} \underbrace{\prod_{j=1}^{s} q_j}_{q} . \tag{3.50b}$$

Nous passons à l'unicité. Supposons que $n = q_1 m_1^2 = q_2 m_2^2$ avec q_1 et q_2 sans facteurs carrés (dans leur décomposition en facteurs premiers). Soit $d = \operatorname{pgcd}(m_1, m_2)$ et k_1, k_2 définis par $m_1 = dk_1$, $m_2 = dk_2$. Par construction, $\operatorname{pgcd}(k_1, k_2) = 1$. Étant donné que

$$n = q_1 d^2 k_1^2 = q_2 d^2 k_2^2, \tag{3.51}$$

nous avons $q_1 k_1^2 = q_2 k_2^2$ et donc k_1^2 divise $q_2 k_2^2$. Mais k_1 et k_2 n'ont pas de facteurs premiers en commun, donc k_1^2 divise q_2, ce qui n'est possible que si $k_1 = 1$ (parce que k_1^2 n'a que des facteurs premiers alors que q_2 n'en a pas). Dans ce cas, $d = m_1$ et m_1 divise m_2. Si $m_2 = lm_1$ alors l'équation (3.51) se réduit à $n = q_1 m_1^2 = q_2 l^2 m_1^2$ et donc

$$q_1 = q_2 l^2, \tag{3.52}$$

ce qui signifie $l = 1$ et donc $m_1 = m_2$. □

Dans $\mathbb{N}$, il y a assez bien de nombres premiers. Nous allons voir maintenant que la somme des inverses des nombres premiers diverge. Pour comparaison, la somme des inverses des carrés converge. Il y a donc plus de nombres premiers que de carrés.

3.3.8 Ordre d'un élément dans un groupe fini

Il y a des informations en plus dans la partie sur les groupes monogènes, 5.2.

Théorème 3.34 (Théorème de Cauchy[50])**.**
Soit un groupe fini d'ordre n. Pour tout diviseur premier p de n, le groupe G possède au moins un élément d'ordre p.

Le lemme suivant indique que sous hypothèse de commutativité, l'ordre d'un élément est une notion multiplicative.

Lemme 3.35 ([51])**.**
Soit G un groupe et $a, b \in G$ tels que $ab = ba$ d'ordres respectivement r et s, deux nombres premiers entre eux. Alors l'élément ab est d'ordre rs.

Démonstration. Étant donné que $(ab)^{rs} = a^{rs} b^{rs} = 1$, l'ordre de ab divise rs. Et vu que r et s sont premiers entre eux, l'ordre de ab s'écrit sous la forme $r_1 s_1$ avec $r_1 \mid r$ et $s_1 \mid s$. Nous avons

$$a^{r_1 s_1} b^{r_1 s_1} = (ab)^{r_1 s_1} = 1, \tag{3.53}$$

que nous élevons à la puissance r_2 où r_2 est définit en posant $r = r_1 r_2$:

$$a^{r s_1} b^{r s_1} = 1. \tag{3.54}$$

Et comme $a^{r s_1} = 1$, nous concluons que $b^{r s_1} = 1$. Donc $s \mid r s_1$. Par le lemme de Gauss 3.26, nous avons en fait $s \mid s_1$. Vu qu'on a aussi $s_1 \mid s$, nous avons $s = s_1$.

Le même type d'argument donne $r = r_1$, et finalement l'ordre de ab est $r_1 s_1 = rs$. □

6. Théorème 3.28.

Lemme 3.36 ([35]).
Un sous-groupe d'indice 2 *est un sous-groupe normal.*

Démonstration. Si H est un tel sous-groupe d'un groupe G, alors G possède exactement deux classes à gauche par rapport à H (théorème de Lagrange 2.34) et se partitionne donc en deux parties : $G = H \cup xH$ avec $x \notin H$. De même pour les classes à droite : $G = H \cup Hx$. Puisque la classe à droite Hx n'est pas H, on a $xH = Hx$, et H est normal dans G par la proposition 2.14. □

Lemme 3.37 ([52]).
Soit H, un sous-groupe normal d'indice m de G. Alors pour tout $a \in G$ nous avons $a^m \in H$.

Démonstration. Par définition de l'indice, le groupe G/H est d'ordre m. Donc si $[a] \in G/H$, nous avons $[a]^m = [e]$, ce qui signifie $[a^m] = [e]$, ou encore $a^m \in H$. □

Proposition 3.38 ([52]).
Soit un groupe fini G et H, un sous-groupe normal d'ordre n et d'indice m avec m et n premiers entre eux. Alors H est l'unique sous-groupe de G à être d'ordre n.

Démonstration. Soit H' un sous-groupe d'ordre n. Si $h \in H'$ alors $h^n = 1$ et $h^m \in H$ par le lemme 3.37. Étant donné que m et n sont premiers entre eux, par le théorème de Bézout 3.13, il existe $a, b \in \mathbb{Z}$ tels que

$$am + bn = 1. \tag{3.55}$$

Du coup $h = h^1 = (h^m)^a(h^n)^b$. En tenant compte du fait que $h^n = 1$ et $h^m \in H$, nous avons $h \in H$. Ce que nous venons de prouver est que $H' \subset H$ et donc que $H = H'$ parce que $|H'| = |H| = |G|/m$. □

3.39.
Notons que cette proposition ne dit pas qu'il existe un sous-groupe d'ordre n et d'indice m. Il dit juste que s'il y en a un et s'il est normal, alors il n'y en a pas d'autres.

Lemme 3.40.
L'ensemble des ordres[7] *d'un groupe commutatif est stable par PPCM*[8].

Autrement dit, si $x \in G$ est d'ordre r et si $y \in G$ est d'ordre s, alors il existe un élément d'ordre $\operatorname{ppcm}(r, s)$.

Démonstration. Soit $m = \operatorname{ppcm}(r, s)$. Afin d'écrire m sous une forme pratique, nous considérons les décompositions en facteurs premiers de r et s :

$$r = \prod_{i=1}^{k} p_i^{\alpha_i} \tag{3.56a}$$

$$s = \prod_{i=1}^{k} p_i^{\beta_i} \tag{3.56b}$$

où $\{p_i\}_{i=1\ldots k}$ est l'ensemble des nombres premiers arrivant dans les décompositions de r et de s. Si nous posons

$$r' = \prod_{\substack{i=1 \\ \alpha_1 > \beta_i}}^{k} p_i^{\alpha_i} \tag{3.57a}$$

$$s' = \prod_{\substack{i=1 \\ \alpha_i \leqslant \beta_i}}^{k} p_i^{\beta_i}, \tag{3.57b}$$

7. Définition 2.16.
8. Définition 1.81.

alors $\text{ppcm}(r,s) = r's'$ et r' et s' sont premiers entre eux. L'élément $x^{r/r'}$ est d'ordre r' et l'élément $y^{s/s'}$ est d'ordre s'. Maintenant nous utilisons le fait que G soit commutatif et le lemme 3.35 pour conclure que l'ordre de $x^{r/r'}y^{s/s'}$ est $r's' = m$. □

3.3.9 Écriture des fractions

Théorème 3.41.
Tout élément de $\mathbb{Q}^+$ s'écrit de façon unique comme quotient de deux entiers premiers entre eux.

Démonstration. En deux parties[9]

Unicité Supposons avoir $\frac{a}{b} = \frac{c}{d}$ avec $\text{pgcd}(a,b) = \text{pgcd}(c,d) = 1$. Nous avons

$$ad = bc \tag{3.58}$$

donc

(1) a divise bc mais est premier avec b donc a divise c par le lemme de Gauss 3.26.

(2) c divise ad mais est premier avec d donc c divise a par le lemme de Gauss 3.26.

En conclusion a divise c et c divise a, ergo $a = c$. L'égalité $b = d$ est alors immédiate.

Existence Soit le quotient $\frac{a}{b}$. Nous avons

$$\frac{a}{b} = \frac{a/\text{pgcd}(a,b)}{b/\text{pgcd}(a,b)}, \tag{3.59}$$

qui est encore un quotient d'entiers parce que $\text{pgcd}(a,b)$ divise aussi bien a que b. Il faut montrer que les nombres $a/\text{pgcd}(a,b)$ et $b/\text{pgcd}(a,b)$ sont premiers entre eux. Pour cela nous supposons que k est un diviseur commun. En particulier, les nombres $a/k\,\text{pgcd}(a,b)$ et $b/k\,\text{pgcd}(a,b)$ sont des entiers, ce qui fait que $k\,\text{pgcd}(a,b)$ est un diviseur commun de a et b. Étant donné que $\text{pgcd}(a,b)$ est le plus grand tel diviseur, nous devons avoir $k\,\text{pgcd}(a,b) = \text{pgcd}(a,b)$ c'est-à-dire que $k = 1$. Donc les nombres $a/\text{pgcd}(a,b)$ et $b/\text{pgcd}(a,b)$ sont premiers entre eux.

□

Proposition 3.42.
Les entiers p et q sont premiers entre eux si et seulement si p^2 et q^2 sont premiers entre eux.

Démonstration. Si p^2 et q^2 sont premiers entre eux, par le théorème de Bézout 3.13 il existe $a, b \in \mathbb{Z}$ tels que

$$ap^2 + bq^2 = 1 \tag{3.60}$$

Dans ce cas, $(ap)p + (bq)q = 1$, ce qui montre (par encore Bézout, mais l'autre sens) que p et q sont premiers entre eux.

Réciproquement, supposons que p et q ne sont pas premiers entre eux. Alors $\text{pgcd}(p,q) = k \neq 1$. L'entier k divise p et donc p^2 ; et l'entier k divise q et donc q^2. Au final, k divise p^2 et q^2, ce qui montre que p^2 et q^2 ne sont pas premiers entre eux. □

Une des conséquences de ces résultats sera le fait que $\sqrt{n}$ est irrationnelle dès que n n'est pas un carré parfait, théorème 3.43.

Nous avons déjà vu dans la proposition 1.123 que $\sqrt{2}$ était irrationnel. En fait le théorème suivant va nous montrer que le nombre $\sqrt{n}$ est soit entier, soit irrationnel.

Théorème 3.43.
Soit $n \in \mathbb{N}$. Le nombre $\sqrt{n}$ est rationnel si et seulement si n est un carré parfait.

9. Définitions des pgcd et ppcm en 1.81.

Démonstration. Supposons que $\sqrt{n}$ soit rationnel. Le théorème 3.41 nous donne $p, q \in \mathbb{N}$ premiers entre eux tels que $\sqrt{n} = p/q$. La proposition 3.42 nous enseigne de plus que p^2 et q^2 sont premiers entre eux. Nous avons

$$p^2 = nq^2. \tag{3.61}$$

Le nombre q est alors un diviseur commun de q^2 et de p. Donc $q = 1$ et $n = p^2$ est un carré parfait. □

3.3.10 Équation diophantienne linéaire à deux inconnues

Soient a, b et c des entiers naturels donnés. Nous considérons l'équation

$$ax + by = c \tag{3.62}$$

à résoudre[53] pour $(x, y) \in \mathbb{N}^2$.

Si a ou b est nul, c'est facile ; nous supposons donc que a et b sont tout deux non nuls. Nous commençons par simplifier l'équation en cherchant les diviseurs communs. Soit $d = \text{pgcd}(a, b)$ et notons $a = da'$, $b = db'$. Nous avons déjà l'équation

$$d(a'x + b'y) = c, \tag{3.63}$$

et donc si c n'est pas un multiple de d, il n'y a pas de solutions [10]. Si par contre c est un multiple de d alors nous écrivons $c = c'd$ et l'équation devient

$$a'x + b'y = c' \tag{3.64}$$

C'est maintenant que nous utilisons le théorème de Bézout 3.13 : vu que a' et b' sont premiers entre eux, nous avons la relation $a'u + b'v = 1$ pour certains [11] nombres entiers u et v. Nous récrivons notre équation sous la forme $a'x + b'y = c'(a'u + b'v)$ et rassemblons les termes :

$$a'(x - c'u) = b'(c'v - y). \tag{3.65}$$

C'est-à-dire que si (x, y) est une solution, alors a' divise $b'(c'v - y)$. Mais comme a' et b' sont premiers entre eux, le nombre a' doit forcément [12] diviser $c'v - y$. Disons $c'v - y = ka'$. Alors $a'(x - c'u) = b'ka'$ et donc

$$x = b'k + c'u. \tag{3.66}$$

Nous trouvons alors une expression pour y en injectant cela dans $a'x + b'y = c'$ et en utilisant Bézout : $a'c'u = (1 - b'v)c'$. Au final nous avons prouvé que toutes les solutions sont de la forme

$$\begin{cases} x = b'k + c'u & \text{(3.67a)} \\ y = vc' - a'k & \text{(3.67b)} \end{cases}$$

avec $k \in \mathbb{Z}$. Si nous voulons réellement seulement des solutions dans $\mathbb{N}$ et non dans $\mathbb{Z}$, il faut seulement un peu restreindre les valeurs de k. Il en reste un nombre fini parce que l'équation pour x borne k vers le bas tandis que celle pour y borne k vers le haut.

Par ailleurs, il est très vite vérifié que tous les couples (x, y) de la forme (3.67) sont solutions.

Exemple 3.44.
Résoudre l'équation $2x + 6y = 52$.

Nous pouvons factoriser 2 dans le membre de gauche et simplifier alors toute l'équation par 2 :

$$x + 3y = 26. \tag{3.68}$$

10. Exemple : $8x + 2y = 9$. Le membre de gauche est certainement un nombre pair et il n'y a donc pas de solutions.
11. Nous avons décrit un algorithme pour les trouver en démontrant l'équation 3.44.
12. C'est Gauss 3.26 qui le dit, et vous savez que lorsque Gauss dit, c'est *forcément* vrai.

Nous cherchons une relation de Bézout pour $u+3v=1$. Ce n'est heureusement pas très compliqué : $u=-5$, $v=2$. Nous pouvons alors écrire

$$x+3y=26\times(-5+3\times 2), \tag{3.69}$$

et donc $x+5\times 26=3(y-26\times 6)$, et en posant $k=y-26\times 6$ nous avons

$$x=3k-130. \tag{3.70}$$

En injectant nous trouvons l'équation $3k-130+3y=26$ et donc

$$y=52-k. \tag{3.71}$$

△

3.3.11 Quotients

Nous écrivons $a=b \mod p$ essentiellement s'il existe $k\in\mathbb{Z}$ tel que $b+kp=a$. Plus généralement nous notons $[a]_p=\{a+kp|k\in\mathbb{Z}\}$ et l'écriture « $a=n \mod p$ » peut tout autant signifier $a=[b]_p$ que $a\in[b]_p$. La différence entre les deux est subtile mais peut de temps en temps valoir son pesant d'or.

Proposition 3.45.
Soit $n\in\mathbb{N}$. Le groupe $(\mathbb{Z}/n\mathbb{Z},+)$ est monogène. Si $n\neq 0$, il est cyclique d'ordre n.

Démonstration. Nous considérons la surjection canonique $\mu\colon\mathbb{Z}\to\mathbb{Z}/n\mathbb{Z}$. Si $a\in\mathbb{Z}$, alors $\mu(a)=a\mu(1)$. Par conséquent $\mathbb{Z}/n\mathbb{Z}=\mathrm{gr}\big(\mu(1)\big)$ parce que tout groupe contenant $\mu(1)$ contient tous les multiples de $\mu(1)$, et par conséquent contient $\mu(\mathbb{Z})=\mathbb{Z}/n\mathbb{Z}$.

Soit $x\in\mathbb{Z}/n\mathbb{Z}$ et considérons x_0, le plus petit naturel représentant x. Nous notons $x=[x_0]$. Le théorème de la division euclidienne 3.6 donne l'existence de q et r avec $0\leqslant r<n$ et $q\geqslant 0$ tels que

$$x_0=nq+r. \tag{3.72}$$

Nous avons $[x_0]=[r]=\mu(r)$ parce que x_0-r est un multiple de n. Nous avons donc $[x_0]\in\mu(\mathbb{N}_{n-1})$. Par conséquent

$$\mathbb{Z}/n\mathbb{Z}=\mu(\mathbb{Z})=\mu(\mathbb{N}_{n-1}). \tag{3.73}$$

La restriction $\mu\colon\mathbb{N}_{n-1}\to\mathbb{Z}/n\mathbb{Z}$ est donc surjective. Montrons qu'elle est également injective. Si $\mu(x_0)=\mu(x_1)$, alors $x_1=x_0+kn$. Si nous supposons que $x_1>x_0$, alors $k>0$ et si $x_0\in\mathbb{N}_{n-1}$, alors $x_1>n-1$.

L'ordre de $\mathbb{Z}/n\mathbb{Z}$ est donc le même que le cardinal de $\mathbb{N}_{n-1}$, c'est-à-dire n. Le groupe $\mathbb{Z}/n\mathbb{Z}$ est donc fini, d'ordre n et monogène parce que $\mathbb{Z}/n\mathbb{Z}=\mathrm{gr}(\mu(1))$. Il est donc cyclique. □

Lemme 3.46 ([54])**.**
Soit $q\in\mathbb{N}$ avec $q\geqslant 2$. Soient $n,d\in\mathbb{N}$ tels que $q^d-1\mid q^n-1$. Alors $d\mid n$.

Démonstration. Par le théorème de division euclidienne 3.6, il existe $a,b\in\mathbb{Z}$ tels que $n=ad+b$ avec $0\leqslant b<d$. En remarquant que $q^d\in[1]_{q^d-1}$ nous avons

$$q^n=(q^d)^aq^b\in[1]_{q^d-1}q^b=[q^b]_{q^d-1}. \tag{3.74}$$

Pour cela nous avons utilisé d'abord le fait que si $a\in[z]_k$, alors $a^n\in[z^n]_k$ et ensuite le fait que $[1]_kx=[x]_k$. D'autre part l'hypothèse $q^d-1\mid q^n-1$ implique

$$q^n\in[1]_{q^d-1}. \tag{3.75}$$

Par conséquent le nombre q^n est à la fois dans $[q^b]_{q^d-1}$ et dans $[1]_{q^d-1}$. Cela implique que

$$[1]_{q^d-1}=[q^b]_{q^d-1}, \tag{3.76}$$

ou encore que $q^b\in[1]_{q^d-1}$ ou encore que $q^d-1\mid q^b-1$.

Étant donné que $b<d$ et que $q\geqslant 2$, nous avons que $q^b-1<q^d-1$; donc pour que q^d-1 divise q^b-1, il faut que q^b-1 soit zéro, c'est-à-dire $b=0$.

Mais dire $b=0$ revient à dire que $d\mid n$, ce qu'il fallait démontrer. □

3.4 Binôme de Newton et morphisme de Frobenius

Proposition 3.47 ([55]).
Pour tout x, $y \in \mathbb{R}$ et $n \in \mathbb{N}$, nous avons

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \tag{3.77}$$

où

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{3.78}$$

sont les ***coefficients binomiaux****.*

Démonstration. La preuve se fait par récurrence. La vérification pour $n = 0$ et $n = 1$ se fait aisément pour peu que l'on se rappelle que $x^0 = 1$ et que $0! = 1$, ce qui donne entre autres $\binom{0}{0} = 1$.

Supposons que la formule (3.77) soit vraie pour $n \geqslant 1$, et prouvons la pour $n+1$. Nous avons

$$\begin{aligned}(x+y)^{n+1} &= (x+y) \cdot \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k \\ &= \sum_{k=0}^{n} \binom{n}{k} x^{n-k+1} y^k + \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^{k+1} \\ &= x^{n+1} + \sum_{k=1}^{n} \binom{n}{k} x^{n-k+1} y^k + \sum_{k=0}^{n-1} \binom{n}{k} x^{n-k} y^{k+1} + y^{n+1}.\end{aligned} \tag{3.79}$$

La seconde grande somme peut être transformée en posant $i = k+1$:

$$\sum_{k=0}^{n-1} \binom{n}{k} x^{n-k} y^{k+1} = \sum_{i=1}^{n} \binom{n}{i-1} x^{n-(i-1)} y^{i-1+1}, \tag{3.80}$$

dans lequel nous pouvons immédiatement renommer i par k. En remplaçant dans la dernière expression de (3.79), nous trouvons

$$(x+y)^{n+1} = x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \left[\binom{n}{k} + \binom{n}{k-1} \right] x^{n-k+1} y^k. \tag{3.81}$$

La thèse découle maintenant de la formule

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k} \tag{3.82}$$

qui est vraie parce que

$$\frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)(n-k+1)!} = \frac{n!(n-k+1) + n!k}{k!(n-k+1)!} = \frac{n!(n+1)}{k!(n-k+1)!}, \tag{3.83}$$

par simple mise au même dénominateur. □

3.5 Idéal dans un anneau

La définition d'un idéal dans un anneau est la définition 1.79.

Définition 3.48 (Idéal engendré par un élément).
Si p est un élément d'un anneau A alors nous notons (p) l'idéal dans A ***engendré*** *par p, c'est-à-dire pA.*

Définition 3.49.
Un sous-ensemble $B \subset A$ d'un anneau est un ***sous anneau*** *si*

(1) $1 \in B$

(2) *B est un sous-groupe pour l'addition*

(3) *B est stable pour la multiplication.*

Remarque 3.50.
Un idéal n'est pas toujours un anneau parce que l'identité pourrait manquer. Un idéal qui contient l'identité est l'anneau complet.

Exemple 3.51.
L'ensemble $2\mathbb{Z}$ est un idéal de $\mathbb{Z}$. On peut aussi le noter (2). △

Proposition 3.52 (Premier théorème d'isomorphisme pour les anneaux)**.**
Soient A et B des anneaux et un homomorphisme $f\colon A \to B$. Nous considérons l'injection canonique $j\colon f(A) \to B$ et la surjection canonique $\phi\colon A \to A/\ker f$. Alors il existe un unique isomorphisme

$$\tilde{f}\colon A/\ker f \to f(A) \tag{3.84}$$

tel que $f = j \circ \tilde{f} \circ \phi$.

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "\phi"'] & B \arrow[d, "j"] \\ A/\ker f \arrow[r, "\tilde{f}"'] & f(A) \subset B \end{tikzcd} \tag{3.85}$$

Proposition 3.53.
Soient I un idéal de A et la projection canonique

$$\phi\colon A \to A/I. \tag{3.86}$$

Elle est une bijection entre les idéaux de A contenant I et les idéaux de A/I.

Dit de façon imagée :

$$\{\text{idéaux de } A \text{ contenant } I\} \simeq \{\text{idéaux de } A/I\}. \tag{3.87}$$

Démonstration. Si $I \subset J$ et si J est un idéal de A, alors $\phi(J)$ est un idéal dans A/I. En effet un élément de $\phi(J)$ est de la forme $\phi(j)$ et un élément de A/I est de la forme $\phi(i)$. Leur produit vaut

$$\phi(i)\phi(j) = \phi(ij) \in \phi(J). \tag{3.88}$$

Soit maintenant K un idéal dans A/I et soit $J = \phi^{-1}(K)$. Étant donné qu'un idéal doit contenir 0 (parce qu'un idéal est un groupe pour l'addition), $[0] \in K$ et par conséquent $I \subset \phi^{-1}(K)$. □

Proposition 3.54 ([1])**.**
Si A est un anneau, nous avons les équivalences

(1) *A est un corps*[13].

(2) *A est non nul et ses seuls*[14] *seuls idéaux à gauche sont $\{0\}$ et A.*

(3) *A est non nul et ses seuls idéaux à droite sont $\{0\}$ et A.*

Démonstration. Nous allons montrer que le point (1) est équivalent aux deux autres.

(3) implique (2) Si I est un idéal à gauche différent de $\{0\}$, alors il contient un certain $a \neq 0$. Vu que A est un corps, il contient un inverse a^{-1}, et comme I est un idéal, $a^{-1}I \subset I$. En particulier $a^{-1}a \in I$. Donc $1 \in I$ et $I = A$.

13. Définition 1.97.
14. Je vous laisse vous poser de grandes questions sur le fait que le vide est un idéal ou non.

(2) implique (1) Supposons que les seuls idéaux de A soient $\{0\}$ et A. Soit $a \in A$. Si a est non nul, alors $aA = A$, en particulier, $1 \in aA$, c'est-à-dire qu'il existe $b \in A$ tel que $ab = 1$. L'élément a est donc inversible.

□

Définition 3.55.
*Un idéal I dans un anneau A est dit **idéal maximal** ou idéal maximum si tout idéal J vérifiant $I \subset J \subset A$ est soit I, soit A.*

Proposition 3.56 (Thème 46)**.**
Un idéal I dans un anneau A est maximum si et seulement si A/I est un corps.

Démonstration. Soit un idéal maximum $I \subset A$. Alors les idéaux contenant I sont A et I. L'application ϕ de la proposition 3.53 est une bijection, donc l'ensemble des idéaux de A/I ne contient que deux éléments. Les seuls idéaux de A/I sont donc $\{0\}$ et A/I ; donc A/I est un corps par la proposition 3.54.

Dans l'autre sens, c'est la même chose : si A/I est un corps, il possède exactement deux idéaux, donc A ne contient que deux idéaux contenant I. Donc I est un idéal maximum. □

3.5.1 Résultats supplémentaires sur l'anneau des entiers

Corolaire 3.57.
Les quotients de $\mathbb{Z}$ sont $\mathbb{Z}/n\mathbb{Z}$.

Démonstration. Tous les idéaux de $\mathbb{Z}$ sont de la forme $n\mathbb{Z}$. En effet en vertu de la proposition 3.8, les seuls sous-groupes de $\mathbb{Z}$ (en tant que groupe additif) sont les $n\mathbb{Z}$. Tous les idéaux sont donc de cette forme. De plus les $n\mathbb{Z}$ sont effectivement tous des idéaux : si $a \in n\mathbb{Z}$ et si $k \in \mathbb{Z}$ alors $ak \in n\mathbb{Z}$. Cela est donc un idéal. □

Proposition 3.58.
Soient $n \geqslant 2$ un entier et $\phi\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ la surjection canonique. Nous noterons $\tilde{a} = \phi(a)$. Alors l'ensemble des inversibles de $\mathbb{Z}/n\mathbb{Z}$ est donné par

$$U(\mathbb{Z}/n\mathbb{Z}) = \phi(P_n) = \{\tilde{x} \text{ tel que } 0 \leqslant x \leqslant n \text{ tel que } \operatorname{pgcd}(x,n) = 1\}. \tag{3.89}$$

où P_n est l'ensemble $P_n = \{x \in \{0,\dots,n-1\} \text{ tel que } \operatorname{pgcd}(x,n) = 1\}$.

De plus,

$$\operatorname{Card}\big(U(\mathbb{Z}/n\mathbb{Z})\big) = \varphi(n). \tag{3.90}$$

Démonstration. Soit $0 \leqslant x \leqslant n$ tel que $\operatorname{pgcd}(x,n) = 1$. Il existe donc [15] $u, v \in \mathbb{Z}$ tels que $ux + vn = 1$. En passant aux classes,

$$\tilde{u}\tilde{x} = \tilde{1}, \tag{3.91}$$

donc $\tilde{u}$ est l'inverse de $\tilde{x}$. Cela prouve que $\phi(P_n) \subset U(\mathbb{Z}/n\mathbb{Z})$.

Nous prouvons maintenant l'inclusion inverse. Soient $\tilde{x}$ et $\tilde{y}$ inverses l'un de l'autre : $\tilde{x}\tilde{y} = \tilde{1}$. Il existe donc $q \in \mathbb{Z}$ tel que $xy - qn = 1$, ce qui prouve [16] que $\operatorname{pgcd}(x,n) = 1$. □

3.6 Caractéristique

Lemme-définition 3.59.
Soit l'application

$$\begin{aligned} \mu\colon \mathbb{Z} &\to A \\ n &\mapsto n \cdot 1_A. \end{aligned} \tag{3.92}$$

(1) C'est un morphisme d'anneaux.

15. Théorème de Bézout 3.13
16. À nouveau avec le Théorème de Bézout.

(2) Le noyau est un sous-groupe de $\mathbb{Z}$

(3) Il existe un unique $p \in \mathbb{Z}$ tel que $\ker(\mu) = p\mathbb{Z}$.

*Ce p est la **caractéristique** de A.*

Par exemple la caractéristique que $\mathbb{Q}$ est zéro parce qu'aucun multiple de l'unité n'est nul. À propos de diagonalisation en caractéristique 2, voir l'exemple 9.132.

Lemme 3.60.
Si A est de caractéristique nulle, alors A est infini.

Démonstration. En effet, $\ker \mu = \{0\}$ implique que $n1_A \neq m1_A$ dès que $n \neq m$ et par conséquent A contient $\mathbb{Z}1_A$, et est infini. □

Lemme 3.61.
Si p est la caractéristique de l'anneau A, alors nous avons l'isomorphisme d'anneaux

$$\mathbb{Z}1_A \simeq \mathbb{Z}/p\mathbb{Z}. \tag{3.93}$$

Démonstration. L'isomorphisme est donné par l'application $n1_A \mapsto \phi(n)$ si ϕ est la projection canonique $\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$. □

Proposition 3.62.
La caractéristique d'un anneau fini divise son cardinal.

Démonstration. Si A est un anneau, le groupe $\mathbb{Z}$ agit sur A par

$$n \cdot a = a + n1_A. \tag{3.94}$$

Chaque orbite de cette action est de la forme

$$\mathcal{O}_a = \{a + n1_A \text{ tel que } n = 0, \dots, p-1\} \tag{3.95}$$

où p est la caractéristique de A. Les orbites ont p éléments et forment une partition de A, donc le cardinal de A est un multiple de p. □

Lemme 3.63 ([56])**.**
Un anneau totalement ordonné est de caractéristique nulle.

Démonstration. Le morphisme $\mu\colon \mathbb{Z} \to A$, $n \mapsto n1_A$ est strictement croissant, en particulier $\mu(x) \neq \mu(y)$ dès que $x \neq y$. Donc $\ker(\mu) = \{0\}$. □

L'ensemble typique de caractéristique p est $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$.

Proposition 3.64.
Soit A un anneau commutatif de caractéristique première p. Alors $\sigma(x) = x^p$ est un automorphisme de l'anneau A. Nous avons la formule

$$(a+b)^p = a^p + b^p \tag{3.96}$$

pour tout $a, b \in A$.

Démonstration. Nous utilisons la formule du binôme de la proposition 3.47 et le fait que les coefficients binomiaux non extrêmes sont divisibles par p et donc nuls. □

Proposition 3.65.
Soit A un anneau commutatif unitaire de caractéristique p. L'application

$$\begin{aligned} \mathrm{Frob}_A\colon A &\to A \\ x &\mapsto x^p \end{aligned} \tag{3.97}$$

est un automorphisme d'anneau unitaire.

Nous le nommons le **morphisme de Frobenius**. Nous utiliserons aussi les itérés du morphisme de Frobenius : $\mathrm{Frob}^k : x \mapsto x^{p^k}$.

Exemple 3.66.
Soit à factoriser $X^p - 1$ dans $\mathbb{F}_p$. Grâce au morphisme de Frobenius, nous avons immédiatement

$$X^p - 1 = (X - 1)^p. \tag{3.98}$$

△

3.7 Module sur un anneau

Définition 3.67 (module sur un anneau[57])**.**
*Soit un anneau A. Un **module à gauche** sur A est la donnée d'un triple $(M, +, \cdot)$ où*

(1) $+$ est une loi de composition interne à M, c'est-à-dire $+ : M \times M \to M$,

(2) $\cdot$ est une loi de composition externe, c'est-à-dire $\cdot : A \times M \to M$

telles que

(1) $(M, +)$ est un groupe[17]*.*

(2) $a \cdot (x + y) = a \cdot x + a \cdot y$,

(3) $(a + b) \cdot x = a \cdot x + b \cdot x$,

(4) $(ab) \cdot x = a \cdot (b \cdot x)$

(5) $1 \cdot x = x$.

pour tout $a, b \in A$ et $x, y \in M$.

Proposition 3.68.
Si M est un module sur un anneau, alors $(M, +)$ est un groupe commutatif.

Démonstration. Il suffit de calculer $(1 + 1) \cdot (x + y)$ de deux façons différentes :

$$(1 + 1) \cdot (x + y) = 1 \cdot (x + y) + 1 \cdot (x + y) = x + y + x + y \tag{3.99}$$

d'une part et

$$(1 + 1) \cdot (x + y) = (1 + 1) \cdot x + (1 + 1) \cdot y = x + x + y + y, \tag{3.100}$$

d'autre part. En égalant les deux expressions, il vient

$$x + y + x + y = x + x + y + y, \tag{3.101}$$

qui se simplifie (nous sommes dans un groupe) en $y + x = x + y$. □

Définition 3.69.
*Un **espace vectoriel** est un module sur un corps commutatif*[18]*.*

Définition 3.70 ([58])**.**
*Soient un A-module M et un ensemble I. Une famille $(m_i)_{i \in I}$ est **libre** si ils sont **linéairement indépendants**, c'est-à-dire si pour tout choix d'une partie finie J dans I et d'éléments $(a_j)_{j \in J}$ dans A, si nous avons*

$$\sum_{j \in J} a_j m_j = 0, \tag{3.102}$$

alors $a_j = 0$ pour tout j.

17. Nous verrons dans la proposition 3.68 qu'il est forcément commutatif.
18. La condition de commutativité n'est pas indispensable, mais comme nous ne parlerons que de corps commutatifs...

Définition 3.71 ([59]).
Soit S, une partie du A-module M. Le ***sous-module engendré*** *par S est l'ensemble des éléments de M qui sont des combinaisons linéaires finies d'éléments de S, c'est-à-dire de sommes de la forme*

$$\sum_{t\in T} a_t t \tag{3.103}$$

où T est fini dans S et $a_t \in A$.

3.7.1 Module produit

Lemme-définition 3.72 ([58]).
*Soient un anneau A et un ensemble I. Le A-**module produit** A^I est l'ensemble des applications $I \to A$.*

En termes de notations, nous écrivons ceci :

$$A^I = \{(a_i)_{i\in I}, a_i \in A\}. \tag{3.104}$$

L'ensemble A^I devient un module par les définition, pour $x, y \in A^I$ et $a \in A$:

$$ax = (ax_i)_{i\in I} \tag{3.105a}$$
$$x + y = (x_i + y_i)_{i\in I}. \tag{3.105b}$$

En d'autres termes, $A^I = \mathrm{Fun}(I, A)$.

Lemme 3.73.
Pour chaque $i \in I$ nous considérons l'élément $e_i \in A^I$ donné par

$$e_i = (\delta_{ij})_{j\in I}. \tag{3.106}$$

La famille $\{e_i\}_{i\in I}$ est libre[19] *dans A^I.*

Démonstration. Soient J fini dans I ainsi que des éléments $a_j \in A$ ($j \in J$). Nous supposons que[20] $\sum_{j\in J} a_j e_j = 0$. Calculons un peu :

$$\sum_{j\in J} a_j e_j = \sum_{j\in J} (a_j \delta_{ji})_{i\in I} = \left(\sum_{j\in J} a_j \delta_{ji}\right)_{i\in I}. \tag{3.107}$$

Pour que le tout soit nul dans A^I, il faut que

$$\sum_{j\in J} a_j \delta_{ji} \tag{3.108}$$

soit nul pour tout $i \in I$. Si nous fixons $i \in I$, la somme sur j possède un seul terme non annulé par δ_{ji}, et c'est le terme $j = i$. Nous avons donc $a_i = 0$. □

Définition 3.74.
Nous notons $A^{(I)}$ le sous-module de A^I engendré[21] *par les e_i.*

Théorème 3.75 (Propriété universelle de $A^{(I)}$[58]).
Soient un anneau A ainsi qu'un A-module P. Pour $\phi \in \mathrm{Hom}_A(A^{(I)}, P)$, nous considérons

$$\begin{aligned} \phi|_I \colon I &\to P \\ i &\mapsto \phi(e_i). \end{aligned} \tag{3.109}$$

19. Définition 3.70.
20. Pour rappel, les sommes finies sont définies par 1.90.
21. Définition 3.71.

(1) L'application

$$\begin{aligned} f\colon \operatorname{Hom}_A(A^{(I)}, P) &\to \operatorname{Fun}(I, P) \\ \phi &\mapsto \phi|_I \end{aligned} \tag{3.110}$$

est une bijection.

(2) L'application inverse est $g\colon \operatorname{Fun}(I, P) \to \operatorname{Hom}_A(A^{(I)}, P)$ *donnée par*

$$g(\psi)\big(\sum_{j\in J} a_j e_j\big) = \sum_{j\in J} a_j \psi(j) \tag{3.111}$$

pour tout J *fini dans* I *et choix de* $a_j \in A$.

Démonstration. Nous allons montrer que $g\big(f(\phi)\big) = \phi$ et que $f\big(g(\psi)\big) = \psi$ pour tout $\phi \in \operatorname{Hom}_A(A^{(I)}, P)$ et pour tout $\psi \in \operatorname{Fun}(I, P)$.

Dans un premier sens nous avons :

$$g\big(f(\phi)\big)\big(\sum_j a_j e_j\big) = \sum_j a_j f(\phi)(j) \tag{3.112a}$$

$$= \sum_j a_j \phi(e_j) \tag{3.112b}$$

$$= \phi(\sum_j a_j e_j). \tag{3.112c}$$

Justifications :

— Pour (3.112b), nous avons utilisé le fait que $f(\phi)(i) = \phi|_I(i) = \phi(e_i)$.

— Pour (3.112c), nous utilisons le fait que ϕ est un morphisme de modules.

Et pour l'autre sens,

$$f\big(g(\psi)\big)(i) = g(\psi)(e_i) = \psi(i). \tag{3.113}$$

□

3.7.2 Sous-module

Soient M un A-module et $x = (x_i)_{i\in I}$ une famille d'éléments de M paramétrée par l'ensemble I. Nous considérons l'application

$$\begin{aligned} \mu_x\colon A^{(I)} &\to M \\ (a_i)_{i\in I} &\mapsto \sum_{i\in I} a_i x_i. \end{aligned} \tag{3.114}$$

Ici $A^{(I)}$ désigne l'ensemble de toutes les applications $I \to A$ de support fini.

Définition 3.76.
À l'instar des espaces vectoriels, les modules ont une notion de partie libre, génératrice et de bases :

(1) Si μ_x *est surjective, nous disons que* x *est une partie* ***génératrice****.*

(2) Si μ_x *est injective, nous disons que la partie* x *est* ***libre****.*

(3) Si μ_x *est bijective, nous disons que la partie* x *est une* ***base****.*

Définition 3.77.
Un sous-ensemble $N \subset M$ *est un* ***sous-module*** *si* $(N, +)$ *est un sous-groupe de* $(M, +)$ *et si* $a \cdot x \in N$ *pour tout* $x \in N$ *et pour tout* $a \in A$.

Exemple 3.78.
Un anneau A est lui-même un A-module et ses sous-modules sont les idéaux. △

Définition 3.79.
Soit M un module sur un anneau commutatif A. Un ***projecteur*** *est une application linéaire $p\colon M \to M$ telle que $p^2 = p$.*

Une famille $(p_i)_{i\in I}$ sur M est ***orthogonale*** *si $p_i \circ p_j = 0$ pour tout $i \neq j$. La famille est* ***complète*** *si $\sum_{i\in I} p_i = \mathbb{1}$.*

Théorème 3.80.
Soient des sous modules $M_1, \ldots, M_n$ du module M tels que $M = M_1 \oplus \ldots \oplus M_n$. Les applications p_i définies par

$$p_i(x_1 + \ldots + x_n) = x_i \tag{3.115}$$

forment une famille orthogonale de projecteurs et $p_1 + \cdots + p_n = \mathrm{Id}$.

Inversement, si $(p_1, \ldots, p_n)$ est une famille orthogonale de projecteurs dans un module $\mathcal{E}$ tel que $\sum_{i=1}^n p_i = \mathrm{Id}$, alors

$$M = \bigoplus_{i=1}^{n} p_i(M). \tag{3.116}$$

Définition 3.81.
Un module est ***simple*** *ou* ***irréductible*** *s'il n'a pas d'autres sous-modules que $\{0\}$ et lui-même. Un module est* ***indécomposable*** *s'il ne peut pas être écrit comme somme directe de sous-modules.*

Un module simple est a fortiori indécomposable. L'inverse n'est pas vrai comme le montre l'exemple suivant.

Exemple 3.82.
Soit $\mathcal{E} = \mathbb{C}[X]/(X^2)$ vu comme $\mathbb{C}[X]$-module. C'est le $\mathbb{C}[X]$-module des polynômes de la forme $aX + b$ avec $a, b \in \mathbb{C}$. L'ensemble des polynômes de la forme aX est un sous-module. Le module $\mathcal{E}$ n'est donc pas simple. Il est cependant indécomposable parce que $\{aX\}$ est le seul sous-module non trivial. En effet si $\mathcal{F}$ est un sous-module de $\mathcal{E}$ contenant $aX + b$ avec $b \neq 0$, alors $\mathcal{F}$ contient $X(aX + b) = bX$ et donc contient tout $\mathcal{E}$. △

Définition 3.83 (Algèbre[60])**.**
Si $\mathbb{K}$ est un corps commutatif[22]*, une $\mathbb{K}$-****algèbre*** *A est un espace vectoriel*[23] *muni d'une opération bilinéaire $\times\colon A \times A \to A$, c'est-à-dire telle que pour tout $x, y, z \in A$ et pour tout $\alpha, \beta \in \mathbb{K}$,*

(1) $(x + y) \times z = x \times z + y \times z$
(2) $x \times (y + z) = x \times y + x \times z$
(3) $(\alpha x) \times (\beta y) = (\alpha\beta)(x \times y)$.

Si A et B sont deux $\mathbb{K}$-algèbres, une application $f\colon A \to B$ est un ***morphisme d'algèbres*** *entre A et B si pour tout $x, y \in A$ et pour tout $\alpha \in \mathbb{K}$,*

(1) $f(xy) = f(x)f(y)$
(2) $f(x + \alpha y) = f(x) + \alpha f(y)$

où nous avons noté xy pour $x \times y$.

Lemme 3.84 ([1])**.**
Soient une algèbre A et une famille $(X_i)_{i\in I}$ de sous-algèbres de A (ici I est un ensemble quelconque). Alors la partie $X = \bigcap_{i\in I} X_i$ est une sous-algèbre de A.

Démonstration. Nous devons prouver que si x et y sont dans X et $\lambda \in \mathbb{K}$, alors xy, $x + y$ et λx sont dans X. Pour tout $i \in I$ nous avons $x, y \in X_i$ et donc $xy \in X_i$, $x + y \in X_i$ et $\lambda x \in X_i$ (parce que X_i est une algèbre). Donc xy,$x + y$ et λx sont dans X_i pour tout I, et donc dans X. □

Définition 3.85.
*L'****algèbre engendrée*** *par X est l'intersection de toutes les sous-algèbres de A contenant X (qui est une algèbre par le lemme 3.84).*

22. Définition 1.97
23. Définition 3.69.

3.8 Polynômes

3.8.1 Polynômes d'une variable

Et voila la définition que tout le monde attendait ; la définition des anneaux de polynômes. Pour ne pas taper trop fort du premier coup, nous commençons par les polynômes d'une seule variable.

Nous allons définir et étudier ici l'anneau des polynômes sur un anneau A, c'est-à-dire ce qui sera noté $A[X]$. Pour $\mathbb{K}(X)$ lorsque $\mathbb{K}$ est un corps, voir 6.74.

L'ensemble des polynômes sur A sera simplement $A^{(\mathbb{N})}$. Vu que $\mathbb{N}$ est un ensemble bien particulier possédant plein de structure, nous allons pouvoir mettre sur $A^{(\mathbb{N})}$ une structure non seulement de A-module (ça c'est déjà fait), mais en plus d'anneau ainsi qu'une évaluation.

Définition 3.86.
L'ensemble des ***polynômes*** *en une indéterminée sur l'anneau A est l'anneau $A^{(\mathbb{N})}$ que nous avons défini en 3.74.*

Notez que nous n'avons pas encore donné la notation $A[X]$; nous verrons plus tard comment elle arrive.

Vu que $A^{(\mathbb{N})}$ est engendré par les e_i, tout polynôme sur A s'écrit $P = \sum_{i=1}^{n} a_i e_i$.

Définition 3.87.
Nous ajoutons deux structures à $A^{(\mathbb{N})}$.

L'évaluation *Si $\alpha \in A$ et si $P \in A^{(\mathbb{N})}$, nous définissons $P(\alpha)$ par*

$$P(\alpha) = (\sum_{i=0}^{n} a_i e_i)(\alpha) = \sum_{i=0}^{n} a_i \alpha^i, \tag{3.117}$$

étant entendu que $\alpha^0 = 1$ dans A.

Cette définition s'étend immédiatement au cas où B est un anneau qui étend A. Dans ce cas nous pouvons définir $P(b)$ pour tout $P \in \mathbb{A}^{(\mathbb{N})}$ et $b \in B$ avec la même formule (3.117).

Le produit *C'est ici que la structure particulière de $\mathbb{N}$ est utilisée. Nous définissons le produit $A^{\mathbb{N}} \times A^{(\mathbb{N})} \to A^{(\mathbb{N})}$ de la façon suivante. Si $(P_k)_{k\in\mathbb{N}}$ est la suite (presque partout nulle) d'éléments de A qui définit P et si $(Q_k)_{k\in\mathbb{N}}$ est celle de Q, nous notons*

$$(PQ)_n = \sum_{k=0}^{n} P_k Q_{n-k}, \tag{3.118}$$

et donc $PQ = \sum_i (PQ)_i e_i$. Plus explicitement,

$$(\sum_{i=0}^{n} a_i e_i)(\sum_{j=0}^{m} b_j e_j) = \sum_{k=0}^{\infty} \Big(\sum_{\substack{(i,j)\in\mathbb{N}^2 \\ i+j=k}} a_i b_j \Big) e_k. \tag{3.119}$$

Notons qu'à droite, la somme sur k est une somme finie.

Proposition 3.88.
Soit un anneau A. À propos de structure sur $A^{(\mathbb{N})}$.

(1) Avec le produit, l'ensemble $A^{(\mathbb{N})}$ devient un anneau.

(2) L'application

$$\begin{aligned} g\colon A^{(\mathbb{N})} &\to A \\ P &\mapsto P(\alpha) \end{aligned} \tag{3.120}$$

est un morphisme d'anneaux[24]*. En particulier, $(PQ)(\alpha) = P(\alpha)Q(\alpha)$.*

24. Définition 1.74.

Démonstration. En plusieurs points

Anneau L'identité pour le produit dans $A^{(\mathbb{N})}$ est le polynôme donné par $a_0 = 1$ et $a_i = 0$ pour $i \neq 0$. Cela se vérifie en utilisant directement la définition (3.119). La distributivité aussi [25].

Le morphisme Nous notons P_k les éléments de la suite définissant P et Q_k ceux de Q. Alors nous avons

$$(P+Q)(\alpha) = \sum_k (P_k + Q_k)\alpha^k = \sum_k P_k\alpha^k + \sum_k Q_k\alpha^k = P(\alpha) + Q(\alpha). \tag{3.121}$$

Vous aurez noté que la première égalité était la définition (3.105b). De même,

$$P(\alpha)Q(\alpha) = \big(\sum_n P_n\alpha^n\big)\big(\sum_k Q_k\alpha^k\big) = \sum_k Q_k\big(\sum_n P_n\alpha^n\big)\alpha^k = \sum_k\sum_n Q_kP_n\alpha^{n+k} \tag{3.122a}$$

$$= \sum_m \big(\sum_{l=0}^{m} P_lQ_{m-l}\big)\alpha^m = \sum_m (PQ)_m\alpha^m = (PQ)(\alpha). \tag{3.122b}$$

□

Définition 3.89.
*Soit $P \in \mathcal{P}$, $P \neq 0$. On appelle **degré** de P le plus grand nombre naturel n pour lequel le coefficient correspondant est non-nul. Ce naturel est noté* $\deg(P)$.

3.8.2 La notation $A[X]$

Si A est un anneau, nous avons déjà défini les polynômes en une indéterminée sur A comme étant le module $A^{(\mathbb{N})}$ qui est devenu un anneau par la proposition 3.88.

Le polynôme donné par la suite $(a_n)_{n\in\mathbb{N}}$ est souvent notée

$$\sum_k a_kX^k. \tag{3.123}$$

Par exemple avec $a = (4, 2, 8)$ nous avons $a = 8X^2 + 2X + 4$. Nous utiliserons souvent cette notation, qui est très pratique parce qu'elle s'adapte bien aux règles de multiplication et d'addition, en particulier la distributivité.

Il y a (au moins) deux façons de comprendre ce que signifie réellement « X » dans cette notation.

3.8.2.1 Première façon (qui botte en touche)

La première est de dire qu'il n'a pas de significations, et que X^2 est un simple abus de notations pour écrire $(0, 0, 1, 0, \cdots)$. Avec cette façon de voir, nous notons l'anneau des polynômes sur A par « $A[X]$ » où le X n'a pas d'autres raisons d'être que d'avertir le lecteur que nous réservons la lettre « X » pour utiliser la notation pratique des polynômes.

3.8.2.2 Seconde façon (la bonne)

La seconde façon de voir le « X » est de nous rappeler que $A^{(\mathbb{N})}$ a une base en tant de que module : les e_k dont nous avons parlé plus haut. Nous posons $X = e_1$, et nous prenons la convention $X^0 = 1$. Alors nous avons $e_k = X^k$ et nous notons $A[X]$ l'anneau $A^{(\mathbb{N})}$ exprimé avec X.

Dans les deux cas, il n'est pas vraiment légitime d'écrire des égalités comme « $P(X) = X^2 + 2X - 3$ », et encore moins de dire « Le polynôme P, *évalué* en X vaut $X^2 + 2X - 3$ » : il est plus correct d'écrire « $P = X^2 + 2X - 3$ ».

Le lemme suivant montre que ces notations tombent vraiment à point. La véritable difficulté de l'énoncé est de comprendre qu'il n'est pas trivial.

Nous avons vu dans la définition 3.87 que si B est un anneau qui étant A, et si $P \in A[X]$, alors nous avons une définition de $P(b)$ pour tout $b \in B$. Nous appliquons cela à $B = A[X]$, qui est un

25. Je n'ai pas fait les calculs, écrivez-moi pour me dire si ça va facilement.

anneau qui étend A. Autrement dit, si P et Q sont des polynômes, ça a un sens d'écrire $P(Q)$ et le résultat sera un élément de $A[X]$.

Dans le cas particulier $Q = X$, nous avons une chouette formule.

Lemme 3.90.
Nous avons

$$P(X) = P \tag{3.124}$$

pour tout $P \in A[X]$.

Démonstration. Si $P = (a_k)_{k\in\mathbb{N}}$ alors par définition $P(\alpha) = \sum_k a_k\alpha^k$ dès que α est dans un anneau B qui étend A. Nous considérons le cas particulier $B = \mathbb{A}[X]$ et $\alpha = X$, c'est-à-dire $Q = (0, 1, 0, \ldots)$, l'élément $P(X)$ de $A[X]$ vaut

$$\sum_k a_k X^k, \tag{3.125}$$

qui est exactement P lui-même. □

Mais il faut bien comprendre que si P est le polynôme $(-3, 2, 1, 0, \ldots)$, noté $X^2 + 2X - 3$, écrire $P(X) = X^2 + 2X - 3$ est une pirouette de notations que rien ne justifie par rapport à simplement écrire $P = X^2 + 2X - 3$.

3.8.3 Polynômes de plusieurs variables

Définition 3.91.
L'ensemble des ***polynôme de*** n ***variables*** *sur l'anneau* A *est* $A^{(\mathbb{N}^n)}$*, c'est-à-dire l'ensemble des suites indexées par* $\mathbb{N}^n$ *et dont seulement une quantité finie de coefficients sont non nuls.*

Le produit sur $A[X_1, \ldots, X_n]$ *est défini par*

$$(PQ)(k_1, \ldots, k_n) = \sum_{\substack{(l_1,\ldots,l_n),(m_1,\ldots,m_n)\in\mathbb{N}^n\times\mathbb{N}^n \\ l_i+m_i=k_i}} P_{l_1,\ldots,l_n} Q_{m_1,\ldots,m_n}. \tag{3.126}$$

3.92.
Dans $A[X_1, \ldots, X_n]$, la multiplication n'est pas la multiplication de fonctions $\mathbb{N}^n \to \mathbb{K}$, parce que le but est d'obtenir la multiplication usuelle au niveau des évaluations.

Définition 3.93.
Si P *est un polynôme de* n *variables sur* A*, et si* $(x_1, \ldots, x_n) \in A^n$*,* ***l'évaluation*** *de* P *sur* $(x_1, \ldots, x_n)$ *est*

$$P(x_1, \ldots, x_n) = \sum_{(k_1,\ldots,k_n)\in\mathbb{N}^n} P_{k_1,\ldots,k_n} x_1^{k_1} \ldots x_n^{k_n}. \tag{3.127}$$

Notez que la somme, bien que sur $\mathbb{N}^n$*, est une somme finie.*

3.94.
Comme dans le cas des polynômes d'une seule variable, les X_i dans la notation $A[X_1, \ldots, X_n]$ sont à prendre à la légère. L'anneau des polynômes de n variables sur A aurait mieux fait d'être noté par exemple par $\mathcal{P}_n(A)$.

Le fait est que nous avons les polynômes élémentaires définis par

$$X_1(k_1, \ldots, k_n) = \begin{cases} 1 & \text{si } (k_1, \ldots, k_n) = (1, 0 \ldots, 0) \\ 0 & \text{sinon.} \end{cases} \tag{3.128}$$

et que l'anneau des polynômes peut être vu comme A (les polynômes constants) étendus par les X_i.

Quoi qu'il en soit, les X_1 dans la notation $A[X_1, \ldots, X_n]$ sont des indices muets. L'anneau $A[X_1, \ldots, X_n]$ est exactement le même que $A[T_1, \ldots, T_n]$.

3.8.4 Action du groupe symétrique

Par souci de notations, nous notons $\mathcal{P}_n(A)$ l'anneau des polynômes de n variables sur A. La propriété universelle de $\mathcal{P}_n(A) = A^{(\mathbb{N}^n)}$ du théorème 3.75 nous donne une application

$$g\colon \operatorname{Fun}\big(\mathbb{N}^n, \mathcal{P}_n(A)\big) \to \operatorname{Hom}_A\big(\mathcal{P}_n(A), \mathcal{P}_n(A)\big) \tag{3.129}$$

Avec cela nous pouvons énoncer et démontrer le lemme qui donne l'action de S_n [26] sur $\mathcal{P}_n(A)$.

Lemme 3.95 ([61]).
Pour $\sigma \in S_n$ nous définissons

$$\begin{aligned} \phi_\sigma\colon \mathbb{N}^n &\to \mathcal{P}_n(A) \\ m &\mapsto e_{\sigma(m)}. \end{aligned} \tag{3.130}$$

Alors l'application

$$\begin{aligned} \rho\colon S_n &\to \operatorname{Hom}_A\big(\mathcal{P}_n(A), \mathcal{P}_n(A)\big) \\ \sigma &\mapsto g(\phi\sigma) \end{aligned} \tag{3.131}$$

est une action [27].

Démonstration. Nous commençons par donner une expression à notre ρ. Un élément de $\mathcal{P}_n(A)$ est de la forme $\sum_{m\in\mathbb{N}^n} a_m e_m$, et nous avons [28]

$$\rho(\sigma)\big(\sum_{m\in\mathbb{N}^n} a_m e_m\big) = \sum_m a_m \phi_\sigma(m) = \sum_m a_m e_{\sigma(m)}. \tag{3.132}$$

Nous avons tout de suite $\rho(\mathrm{Id}) = \mathrm{Id}$.

En ce qui concerne la composition, nous avons d'une part

$$\rho(\sigma_1\sigma_2)\big(\sum_m a_m e_m\big) = g(\phi_{\sigma_1\sigma_2})\big(\sum_m a_m e_m\big) = \sum_m a_m e_{\sigma_1\sigma_2(m)}, \tag{3.133}$$

et d'autre part,

$$\begin{aligned} \rho(\sigma_1)\rho(\sigma_2)\big(\sum_m a_m e_m\big) &= \rho(\sigma_1)\big(\sum_m a_m e_{\sigma_2(m)}\big) && \text{(3.134a)} \\ &= \rho(\sigma_1)\big(\sum_m a_{\sigma_2^{-1}(m)} e_m\big) && \text{(3.134b)} \\ &= \sum_m a_{\sigma_2^{-1}(m)} e_{\sigma_1(m)} && \text{(3.134c)} \\ &= \sum_m a_m e_{\sigma_1\sigma_2(m)} && \text{(3.134d)} \end{aligned}$$

La proposition 1.91 est utilisée pour (3.134b) et pour (3.134d). □

3.9 Anneau intègre

La définition d'un anneau intègre est la définition 1.88.

Exemple 3.96.
Un corps [29] est toujours un anneau intègre. En effet, soient un corps $\mathbb{K}$ et deux éléments $x, y \in \mathbb{K}$ tels que $xy = 0$. Si y est inversible, alors nous pouvons multiplier par y^{-1} pour trouver $x = 0$. Cela prouve que $\mathbb{K}$ est un anneau intègre. △

26. Définition du groupe symétrique S_n en 2.63.
27. Définition 2.46.
28. La somme est définie par 1.90, et ça va être important. Ah oui, en réalité partout, les sommes sont finies parce que les a_m ($m \in \mathbb{N}^n$) sont presque tous nuls. Il faudrait écrire sur la somme sur $\{m \in \mathbb{N}^2$ tel que $a_m \neq 0\}$, mais vous vous imaginez la complication dans la notation.
29. Définition 1.97.

Exemple 3.97.
L'ensemble $\mathbb{Z}$ avec les opérations usuelles est un anneau intègre. △

Exemple 3.98.
L'anneau $\mathbb{Z}/6\mathbb{Z}$ n'est pas intègre parce que $3 \cdot 2 = 0$ alors que ni 3 ni 2 ne sont nuls. △

Nous verrons au théorème 3.172 que l'anneau A est intègre si et seulement si $A[X]$ est intègre.

Corolaire 3.99.
L'anneau $\mathbb{Z}/n\mathbb{Z}$ est intègre si et seulement si n est premier.

Démonstration. Supposons que n soit premier. La proposition 3.58 donne les inversibles de $\mathbb{Z}/n\mathbb{Z}$ par

$$U(\mathbb{Z}/n\mathbb{Z}) = \{\tilde{x} \text{ tel que } 0 \leqslant x \leqslant n \text{ tel que } \operatorname{pgcd}(x, n) = 1\}. \tag{3.135}$$

Mais comme n est premier, $\operatorname{pgcd}(x, n) = 1$ pour tout x, et donc tous les éléments de $\mathbb{Z}/n\mathbb{Z}$ sont inversibles. Donc $\mathbb{Z}/n\mathbb{Z}$ est intègre.

Si n n'est pas premier, alors $n = pq$ avec $1 < p \leqslant q < n$. Alors

$$[p]_n[q]_n = [pq]_n = [0]_n. \tag{3.136}$$

Donc lorsque n n'est pas premier, l'anneau $\mathbb{Z}/n\mathbb{Z}$ possède des diviseurs de zéro et n'est alors pas intègre. □

3.9.1 Caractéristique d'un anneau intègre

Lemme 3.100.
La caractéristique[30] *d'un anneau intègre est zéro ou un nombre premier.*

Démonstration. Si A est intègre, alors $\mathbb{Z}1_A$ est a fortiori intègre. Notons p la caractéristique de A. Si $p = 0$, la preuve est finie ; supposons donc que $p \neq 0$. Alors, l'anneau $\mathbb{Z}/p\mathbb{Z}$ est isomorphe à $\mathbb{Z}1_A$, et est donc intègre. Or, la proposition 3.99 dit que $\mathbb{Z}/p\mathbb{Z}$ est intègre si et seulement si p est premier, ce qui conclut la preuve. □

Exemple 3.101.
Il existe des corps dont la caractéristique n'est pas égale au cardinal (contrairement à ce que laisserait penser l'exemple des $\mathbb{Z}/p\mathbb{Z}$). En effet les matrices $n \times n$ inversibles sur $\mathbb{F}_3$ forment un corps qui n'est pas de cardinal trois alors que la caractéristique est 3 :

$$\begin{pmatrix} 1 & \\ & 1 \end{pmatrix} + \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} + \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} = 0. \tag{3.137}$$

△

Exemple 3.102.
Si $\mathbb{K}$ est un corps de caractéristique 2, alors l'égalité $x = -x$ n'implique pas $x = 0$, vu que $2x = 0$ est vérifiée pour tout x. Cela se répercute sur un certain nombre de résultats. Par exemple, en caractéristique deux, une forme antisymétrique n'est pas toujours alternée : voir le lemme 9.2. △

3.9.2 Divisibilité et classes d'association

Lemme 3.103.
Si A est un anneau intègre et si $a, b \in A$ sont tels que $a \mid b$ et $b \mid a$, alors il existe un inversible $u \in A$ tel que $a = ub$.

30. Définition 3.59.

Démonstration. Les hypothèses à propos de la divisibilité nous indiquent que $a = xb$ et $b = ya$ pour certains $x, y \in A$. Du coup,

$$b(1 - yx) = 0. \tag{3.138}$$

Étant donné que $\mathbb{A}$ est intègre, cela montre que $b = 0$ ou $1 - yx = 0$. Si $b = 0$ nous avons immédiatement $a = 0$ et le lemme est prouvé. Si au contraire $yx = 1$, c'est que y et x sont inversibles et inverses l'un de l'autre. □

Définition 3.104.
On dit de deux éléments $a, b \in A$ qu'ils sont ***associés*** *si ils vérifient les hypothèses du lemme 3.103 ; en d'autres termes, a et b sont associés s'il existe un inversible $u \in A$ tel que $a = ub$.*

La ***classe d'association*** *d'un élément $a \in A$ est l'ensemble des éléments qui lui sont associés ; en d'autres termes, c'est $aU(A)$.*

Exemple 3.105.
Dans $\mathbb{Z}[i]$, les inversibles sont ± 1 et $\pm i$. Donc les éléments associés à z sont z, $-z$, iz et $-iz$.

Notons au passage que la notion de divisibilité dans $\mathbb{Z}[i]$ n'est pas immédiatement intuitive. En effet bien que 7 ne soit pas divisible par 2 (ni dans $\mathbb{Z}$ ni dans $\mathbb{Z}[i]$), le nombre $7 + 6i$ est divisible par $2 + i$ dans $\mathbb{Z}[i]$. En effet :

$$(2 + i)(4 + i) = 7 + 6i. \tag{3.139}$$

△

Problèmes et choses à faire
Est-ce que quelqu'un connaît un anneau contenant $\mathbb{Z}$ dans lequel 7 est divisible en 2 ?

Peut-être $\mathbb{Z}$ étendu par tous les $1/2^n$?

3.9.3 PGCD et PPCM

Pour le théorème de Bézout et autres opérations avec des modulo, voir le thème 50. Le pgcd et le ppcm sont définis en 1.81.

Lemme 3.106.
Soient A un anneau intègre et $S \subset A$. Si δ est un PGCD de S, alors l'ensemble des PGCD de S est la classe d'association de δ.

De la même façon si μ est un PPCM de S, alors l'ensemble des PPCM de S est la classe d'association de μ.

Démonstration. Soient δ un PGCD de S et u un inversible dans A. Si $x \in S$ nous avons $\delta \mid x$ et donc $x = a\delta$. Par conséquent $x = au^{-1}u\delta$ et donc $u\delta$ divise x. De la même manière, si d divise x pour tout $x \in S$, alors d divise δ et donc $\delta = ad$ et $u\delta = aud$, ce qui signifie que d divise $u\delta$.

Dans l'autre sens nous devons prouver que si δ' est un autre PGCD de S, alors il existe un inversible $u \in \mathbb{A}$ tel que $\delta' = u\delta$. Vu que δ' divise x pour tout $x \in S$, nous avons $\delta' \mid \delta$, et symétriquement nous trouvons $\delta \mid \delta'$. Par conséquent (lemme 3.103), il existe un inversible u tel que $\delta = u\delta'$.

Le même type de raisonnement tient pour le PPCM. □

Si δ est un PGCD de S, nous dirons *par abus de langage* que δ est *le* PGCD de S, gardant en tête qu'en réalité toute sa classe d'association est PGCD. Nous noterons aussi, toujours par abus que $\delta = \text{pgcd}(S)$.

Remarque 3.107.
La classe d'association d'un élément n'est pas toujours très grande. Les inversibles dans $\mathbb{Z}$ étant seulement ± 1, nous pouvons obtenir l'unicité du PGCD et du PPCM en imposant qu'ils soient positifs.

Pour les polynômes, nous obtenons l'unicité en demandant que le PGCD soit unitaire.

Dans les cas pratiques, il y a donc en réalité peu d'ambiguïté à parler du PGCD ou du PPCM d'un ensemble.

3.9.4 Anneaux intègres et corps

Le fait d'être intègre pour un anneau n'assure pas le fait d'être un corps. Nous avons cependant ce résultat pour les anneaux finis.

Proposition 3.108.
Un anneau fini intègre est un corps.

Démonstration. Soit A un tel anneau. Soit $a \neq 0$. Les applications

$$l_a \colon x \to ax \tag{3.140a}$$
$$r_a \colon x \to xa \tag{3.140b}$$

sont injectives. En tant que applications injectives entre ensembles finis, elles sont surjectives. Il existe donc b et c tels que $1 = ba = ac$. Il se fait que b et c sont égaux parce que[31]

$$b = b(ac) = (ba)c = c. \tag{3.141}$$

Par conséquent b est un inverse de a. □

Proposition 3.109.
Soit $n \in \mathbb{N}^*$. Les conditions suivantes sont équivalentes :

(1) n est premier.

(2) $\mathbb{Z}/n\mathbb{Z}$ est un anneau intègre.

(3) $\mathbb{Z}/n\mathbb{Z}$ est un corps.

Démonstration. L'équivalence entre les deux premiers points est le contenu du corolaire 3.99. Le fait que $\mathbb{Z}/n\mathbb{Z}$ soit un corps lorsque $\mathbb{Z}/n\mathbb{Z}$ est intègre est la proposition 3.108. Le fait que $\mathbb{Z}/n\mathbb{Z}$ soit intègre lorsque $\mathbb{Z}/n\mathbb{Z}$ est un corps est une propriété générale des corps : ce sont en particulier des anneaux intègres (lemme 1.100). □

3.9.5 Élément irréductible

Définition 3.110 (Élément irréductible[62])**.**
*Un élément d'un anneau commutatif est **irréductible** si il n'est ni inversible, ni le produit de deux éléments non inversibles.*

3.111.
Nous allons voir dans la section 3.12 que le concept d'élément irréductible n'est vraiment utile que dans le cas des anneaux intègres.

Exemple 3.112.
Un corps n'a pas d'éléments irréductibles parce qu'à part zéro tous les éléments sont inversibles. Mais 0 n'est pas irréductible parce qu'il peut être écrit comme produit d'éléments non inversibles : $0 = 0 \cdot 0$. △

Exemple 3.113.
Les éléments irréductibles de l'anneau $\mathbb{Z}$ sont les nombres premiers. En effet les seuls inversibles de $\mathbb{Z}$ sont ± 1. Si p est premier et $p = ab$ avec $a, b \in \mathbb{Z}$, alors nous avons soit $a = \pm 1$ soit $b = \pm 1$. △

3.10 Anneau factoriel

Définition 3.114 (Anneau factoriel)**.**
*Un anneau commutatif A est **factoriel** s'il vérifie les propriétés suivantes.*

31. Il faut être un peu souple avec les notations communément employées dans les ouvrages mathématiques, et que nous reprenons telles quelles. Dans l'équation qui suit, $b(ac)$ est le produit de b par l'élément ac, et non quelque chose comme le produit de b avec l'idéal (ac) par exemple.

(1) L'anneau A est intègre (pas de diviseurs de zéro).

(2) Si $a \in A$ est non nul et non inversible, alors il admet une décomposition en facteurs irréductibles : $a = p_1 \dots p_k$ où les p_i sont irréductibles.

(3) Si $a = q_1 \dots q_m$ est une autre décomposition de a en irréductibles, alors $m = k$ et il existe une permutation[32] $\sigma \in S_k$ telle que p_i et $q_{\sigma(i)}$ soient associés[33].

Un anneau factoriel permet de caractériser le pgcd et le ppcm de nombres.

Proposition 3.115.
Soit une famille $\{a_n\}$ d'éléments de A qui se décomposent en irréductibles comme

$$a_i = \prod_k p_k^{\alpha_{k,i}}. \tag{3.142}$$

Alors

$$\text{pgcd}\{a_n\} = \prod_k p_k^{\min_i\{\alpha_{k,i}\}}. \tag{3.143}$$

De plus le PGCD est :

(1) Un multiple de tous les diviseurs communs des a_i.

(2) Unique pour cette propriété à multiple près par un inversible[34].

De la même manière,

$$\text{ppcm}\{a_n\} = \prod_k p_k^{\max_i\{\alpha_{k,i}\}}. \tag{3.144}$$

Un anneau factoriel a une relation de préordre partiel donnée par $a < b$ si a divise b. En termes d'idéaux, cela donne l'ordre inverse de celui de l'inclusion[35] : $a < b$ si et seulement si $(b) \subset (a)$.

Exemple 3.116.
L'anneau $\mathbb{Z}[i\sqrt{3}]$ n'est pas factoriel parce que 4 a au moins deux décompositions distinctes en irréductibles :

$$4 = 2 \cdot 2, \tag{3.145}$$

et

$$4 = (1 + i\sqrt{3})(1 - i\sqrt{3}). \tag{3.146}$$

△

Nous allons voir dans l'exemple 3.156 que $\mathbb{Z}[i\sqrt{2}]$ est factoriel parce qu'il sera euclidien.

3.11 Anneau principal et idéal premier

Définition 3.117.
*Un idéal I dans A est **principal à gauche** s'il existe $a \in I$ tel que $I = Aa$. Il est **principal à droite** s'il existe $a \in I$ tel que $I = aA$. Nous disons qu'il est **principal** s'il est principal à gauche et à droite.*

Définition 3.118.
*Un anneau est **principal** si*

(1) il est commutatif et intègre

(2) tous ses idéaux sont principaux.

32. Définition 2.63.
33. Définition 3.104.
34. Soyez prudent avec cette affirmation : je n'en n'ai pas de démonstrations sous la main et ne suis pas certain que ce soit vrai.
35. Voir proposition 3.3.

Souvent pour prouver qu'un anneau est principal, nous prouvons qu'il est euclidien (définition 3.151) et nous utilisons la proposition 3.153 qui dit qu'un anneau euclidien est principal.

Une manière de prouver qu'un anneau n'est pas principal est de prouver qu'il n'est pas factoriel, théorème 3.144.

Définition 3.119.
Nous disons qu'un idéal I dans A est ***premier*** *si I est strictement inclus dans A et si pour tout $a, b \in A$ tels que $ab \in I$ nous avons $a \in I$ ou $b \in I$.*

Lemme 3.120.
L'idéal nul (réduit à $\{0\}$) est premier si et seulement si A est intègre.

Démonstration. En deux sens.

Si $\{0\}$ est premier Alors $A \neq \{0\}$ parce que $I = \{0\}$ est propre (définition d'idéal premier).
De plus, si $ab = 0$, alors $ab \in I$ qui est un idéal premier. Donc soit a soit b est dans I, c'est-à-dire que soit a soit b est nul. Donc A est intègre.

Si A est intègre Alors $A \neq \{0\}$ et l'idéal $I = \{0\}$ est strictement inclus dans A. Si $ab \in I$, alors $ab = 0$ et comme A est intègre, soit a soit b est nul, c'est-à-dire appartient à I.

□

Proposition 3.121 ([40]).
Soit un anneau commutatif[36] *et un idéal I dans A.*

(1) I est un idéal premier si et seulement si A/I est un anneau intègre.

(2) I est un idéal maximal si et seulement si A/I est un corps.

(3) Tout idéal maximal propre est premier.

Démonstration. En plein d'étapes.

I premier implique A/I intègre Évacuons le cas trivial pour être sûr. Si $I = \{0\}$ alors A est intègre par le lemme 3.120. Donc $A/I = A/\{0\} = A$ est intègre également.
Soient $a, b \in A$ tels que $[a][b] = [0]$. Donc $[ab] = [0]$, c'est-à-dire $ab \in I$. Vu que I est un idéal premier nous avons $a \in I$ ou $b \in I$, c'est-à-dire $[a] = 0$ ou $[b] = 0$; nous en déduisons que A/I est un anneau intègre.

A/I intègre implique I premier Soit $ab \in I$. Alors $[ab] = 0$, ce qui signifie que $[a][b] = 0$ donc que $[a] = 0$ ou que $[b] = 0$ parce que A/I est intègre. Mais la condition $[a] = 0$ signifie $a \in I$, et $[b] = 0$ signifie $b \in I$. Nous avons donc prouvé que soit a soit b est dans I, c'est-à-dire que I est premier.

Si I est un idéal maximum Nous devons montrer que tout élément non nul de A/I est inversible. Un élément non nul de A/I est $[x]$ avec $x \in A \backslash I$.
Nous considérons $J = Ax + I$, qui est un idéal parce que pour tout $a \in A$, $aAx + aI \in Ax + I$. Mais comme I est maximal, $J = I$ ou $J = A$.
Si $J = I$, nous aurions que pour tout $a \in A$ et pour tout $i \in I$, $ax + i \in I$. En particulier pour $a = 1$ et $i = 0$ nous aurions $x \in I$, ce qui est contraire à l'hypothèse faite sur x.
Donc $J = A$. En particulier, $1 \in J$, c'est-à-dire qu'il existe $a \in A$ et $i \in I$ tels que $ax + i = 1$. En passant aux classes, $[ax] = 1$, c'est-à-dire $[a][x] = 1$ qui signifie que $[a]$ est un inverse de $[x]$ dans A/I.

Si A/I est un corps Si $x \in A \backslash I$, il faut prouver que tout idéal contenant I et x est A.
Un idéal contenant I et x doit contenir l'idéal $J = Ax + I$. Vu que $x \notin I$, nous avons $[x] \neq 0$ dans A/I. Donc $[x]$ est inversible et il existe $a \in A$ tel que $[ax] = [A]$. C'est-à-dire que $ax - 1 \in I$. Nous avons alors

$$1 = ax + \underbrace{(1 - ax)}_{\in I}. \tag{3.147}$$

36. Tous les anneaux du Frido sont commutatifs

C'est-à-dire que $1 \in Ax + I$ et donc $Ax + I = A$.

Enfin nous prouvons que tout idéal maximal propre est premier.

Si I est maximal, A/I est un corps par le point (2), et vu que I est propre, le corps A/I n'est pas réduit à $\{0\}$. Donc le lemme 1.100 dit que A/I est un anneau intègre. Le point (1) dit alors que I est un idéal premier. □

Remarque 3.122.
Vu qu'un corps peut être réduit à $\{0\}$, dans (2), l'idéal peut être A. Mais pas dans (3), parce qu'un idéal premier est propre, ça fait partie de la définition 3.119.

Proposition 3.123 ([63])**.**
Si A est un anneau commutatif intègre, alors un idéal I dans A est premier si et seulement si A/I est intègre.

Démonstration. Supposons que I soit un idéal premier. Si $\bar{a}, \bar{b} \in A/I$ sont tels que $\bar{a}\bar{b} = 0$, alors $\overline{ab} = 0$, ce qui signifie que $ab \in I$. Mais alors, vu que I est premier, soit a soit b est dans I. Cela signifie que soit $\bar{a}$ soit $\bar{b}$ est nul dans A/I. Cela prouve que A/I est un anneau intègre.

Dans l'autre sens, nous supposons que A/I est intègre. Cela implique immédiatement que $I \neq A$ parce que A/A n'est pas un anneau intègre (tout le monde est évidemment diviseur de zéro).

Soient donc $a, b \in A$ tels que $ab \in I$. Alors $\bar{a}\bar{b} = \overline{ab} = 0$ dans A/I, mais comme A/I est intègre, cela implique que soit $\bar{a}$ soit $\bar{b}$ est nul. Autrement dit, soit a soit b est dans I. □

Proposition 3.124 (Thème 46, [1])**.**
Soit A un anneau principal[37] *qui n'est pas un corps. Pour un idéal propre I de A, les conditions suivantes sont équivalentes :*

(1) I est un idéal maximal[38] *;*

(2) I est un idéal premier non nul[39] *;*

(3) il existe p irréductible[40] *dans A tel que $I = (p)$.*

Démonstration. En plusieurs implications.

(1) implique (2) Par hypothèse, I est un idéal propre, de plus il n'est pas égal à $\{0\}$, parce que lorsque A et $\{0\}$ sont les seuls idéaux, nous avons un corps (proposition 3.54). Étant donné que I est un idéal maximal, le quotient A/I est un corps par la proposition 3.56.

Soient maintenant, pour entrer dans le vif du sujet, des éléments $a, b \in A$ tels que $ab \in I$. Dans le corps A/I nous avons $\overline{ab} = 0$, et par définition du produit dans le quotient, $\bar{a}\bar{b} = 0$. Par intégrité de l'anneau A/I (un corps est un anneau intègre, exemple 3.96) nous avons soit $\bar{a} = 0$, soit $\bar{b} = 0$, soit les deux en même temps. Dans tous les cas, soit a soit b est dans I.

(2) implique (3) Maintenant I est un idéal premier non réduit à $\{0\}$. Vu que A est un anneau principal, il existe $x \in A$ tel que $I = (x)$. Nous devons prouver que x peut être choisi irréductible ; et nous allons faire plus : nous allons prouver que x ne peut être que irréductible[41].

Supposons que x ne soit pas irréductible. Alors il existe $a, b \in A$ non inversibles tels que $x = ab$. Si $a \in (x)$ alors il existe $k \in A$ tel que $a = xk$, et en particulier, $a = abk$, c'est-à-dire $1 = bk$ (parce que A est principal et donc intègre). Cela signifie que b est inversible alors que nous avions dit qu'il ne l'était pas. Nous en déduisons que a n'est pas dans (x). On montre de manière similaire que b n'est pas dans (x) non plus.

Nous nous retrouvons donc avec $a, b \in A$ tel que $ab \in I$ sans que ni a ni b sont soient dans I. Cela contredit le fait que I soit un idéal premier. En conclusion, x est irréductible.

37. Définition 3.118.
38. Définition 3.55.
39. Définition 3.119.
40. Définition 3.110.
41. ça me semble un peu trop facile. Lisez attentivement, et écrivez-moi pour dire si vous êtes d'accord ou pas.

(3) implique (1) Nous avons $I = (p)$ avec p irréductible dans A. Supposons que J est un idéal différent de A contenant I. Vu que A est principal, il existe $y \in A$ tels que $J = (y)$. En particulier $p \in J$, donc $p = ay$ pour un certain $a \in A$. Mais p est irréductible, donc soit a est inversible, soit y est inversible. Si y est inversible, alors $J = A$, ce qui est exclu. Si a est inversible, alors $(y) = (p)$, et $I = J$.

□

3.125.
Dans la proposition 3.124, l'hypothèse d'idéal propre est importante. En effet dans le cas $I = A$, nous avons évidemment que I est un idéal maximum. Mais A n'est d'abord pas un idéal premier parce qu'un idéal premier doit être strictement inclus dans l'anneau. Et ensuite, A est en général loin d'être garanti d'être égal à (p) pour un de ses éléments p.

Proposition 3.126.
Soit A un anneau principal, et soit $p \in A$ un élément irréductible. Alors

(1) (p) est un idéal maximum.

(2) $A/(p)$ est un corps.

Démonstration. Nous notons $I = (p)$. Soit un idéal J contenant I. Vu que A est principal, J aussi est monogène : $J = (q)$. Mais comme p est dans I qui est dans J, il existe $a \in A$ tel que $p = qa$.

Vu que p est irréductible, soit q soit a est inversible. Si q est inversible, alors $J = A$. Si a est inversible, alors nous avons $p = qa$, donc $q = pa^{-1}$, ce qui signifie que $q \in (p)$ et donc que $J = I$.

Cela prouve que (p) est un idéal maximum.

Le fait que $A/(p)$ soit un corps est maintenant la proposition 3.56. □

Exemple 3.127.
L'anneau $\mathbb{Z}$ est principal parce qu'il est intègre et que ses seuls idéaux sont les $n\mathbb{Z}$ qui sont principaux : $n\mathbb{Z}$ est engendré par n. △

Exemple 3.128 (Les idéaux de $\mathbb{Z}/n\mathbb{Z}$)**.**
Les idéaux de $\mathbb{Z}/n\mathbb{Z}$ sont principaux, mais l'anneau $\mathbb{Z}/n\mathbb{Z}$ n'est pas principal lorsque n n'est pas premier. Nous allons voir ça.

Les idéaux de $\mathbb{Z}/n\mathbb{Z}$ sont principaux Soit un idéal S dans $\mathbb{Z}/n\mathbb{Z}$. Nous considérons la projection canonique $\phi\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$. La proposition 3.53 dit que $S = \phi(J)$ où J est un idéal de $\mathbb{Z}$ contenant $n\mathbb{Z}$. Mais le corolaire 3.57 nous dit qu'alors $J = m\mathbb{Z}$ pour un certain m. Pour que $m\mathbb{Z}$ contienne $n\mathbb{Z}$, il faut que m divise n.

Bref, $S = \phi(m\mathbb{Z})$ avec $m \mid n$. Nous montrons maintenant que S est engendré par $[m]_n$. D'abord, l'élément $[m]_n$ est bien dans $\phi(m\mathbb{Z})$. Ensuite un élément de $\phi(m\mathbb{Z})$ est de la forme

$$[km]_n = k[m]_n \in ([m]_n). \tag{3.148}$$

Donc $S \subset ([m]_n)$. Et l'inclusion dans l'autre sens est tout aussi immédiate : un élément de $([m]_n)$ est de la forme

$$k[m]_n = [km]_n = \phi(km) \in \phi(m\mathbb{Z}). \tag{3.149}$$

Si n n'est pas premier, $\mathbb{Z}/n\mathbb{Z}$ n'est pas principal Le fait est que lorsque n n'est pas premier, $\mathbb{Z}/n\mathbb{Z}$ n'est pas intègre (corolaire 3.99).

Moralité Un anneau comme $\mathbb{Z}/6\mathbb{Z}$ est un anneau dont tous les idéaux sont principaux, mais qui n'est pas principal.

△

Exemple 3.129.
Nous verrons dans la proposition 25.20 que l'anneau des fonctions holomorphes sur un compact de $\mathbb{C}$ est principal. △

Définition 3.130.
Nous disons que deux éléments d'un anneau principal sont ***premiers entre eux*** *si leur PGCD est* 1.

Théorème 3.131.
Si A est un anneau principal et si p et q sont premiers entre eux dans A, alors on a l'isomorphisme d'anneaux

$$A/pqA \simeq A/pA \times A/qA. \tag{3.150}$$

3.11.1 Bézout

Théorème 3.132 ([64])**.**
Toute partie S d'un anneau principal admet un PGCD et un PPCM. De plus

$$\delta = \operatorname{pgcd}(S) \Leftrightarrow (\delta) = \sum_{s\in S}(s)\mu = \operatorname{ppcm}(S) \Leftrightarrow (\mu) = \bigcap_{s\in S}(s) \tag{3.151}$$

Démonstration. Vu que l'anneau A est principal, tous ses idéaux sont principaux et donc engendrés par un seul élément. En particulier il existe $\delta, \mu \in A$ tels que

$$(\delta) = \sum_{s\in S}(s) \tag{3.152a}$$

$$(\mu) = \bigcap_{s\in S}(s) \tag{3.152b}$$

PGCD Montrons ce que δ est un PGCD de S. Pour tout $x \in S$, nous avons $(x) \subset (\delta)$, et donc $\delta \mid x$. Par ailleurs si $d \mid x$ pour tout $x \in S$, nous avons $(x) \subset (d)$ et donc

$$\sum_{x\in S}(x) \subset (d), \tag{3.153}$$

puis $(\delta) \subset (d)$ et finalement $d \mid \delta$.

PPCM Si $x \in S$ nous avons $(\mu) \subset (x)$ et donc $x \mid \mu$. D'autre part si $x \mid m$ pour tout $x \in S$, alors $(m) \subset (x)$ et donc $(m) \subset (\mu)$, finalement $\mu \mid m$.

□

Corolaire 3.133 (Théorème de Bézout[64])**.**
Soit un anneau principal A. Deux éléments $a, b \in A$ sont premiers entre eux si et seulement s'il existe un couple $(u, v) \in A^2$ tel que

$$ua + vb = 1. \tag{3.154}$$

À la place de 1 *on aurait pu écrire n'importe quel inversible.*

Démonstration. Pour cette preuve, nous allons écrire $\operatorname{pgcd}(a, b)$ l'ensemble de PGCD de a et b, c'est-à-dire la classe d'association d'un PGCD.

Si a et b sont premiers entre eux, alors

$$1 \in \operatorname{pgcd}(a, b) = \sum_{x=a,b}(x) = (a) + (b). \tag{3.155}$$

À l'inverse, si nous avons $ua + vb = 1$, alors $1 \in (a) + (b)$, mais vu que $(a) + (b)$ est un idéal principal, $(1) = (a) + (b)$ et donc $1 \in \operatorname{pgcd}(a, b)$. □

Le lemme de Gauss est une application immédiate de Bézout. Il y aura aussi un lemme de Gauss à propos de polynômes (lemme 6.45), et une généralisation directe au théorème de Gauss, théorème 6.44.

Lemme 3.134 (lemme de Gauss).
Soit A un anneau principal et $a, b, c \in A$ tels que a divise bc. Si a est premier avec c, alors a divise b.

Démonstration. Vu que a est premier avec c, nous avons $\text{pgcd}(a, c) = 1$ et Bézout (3.13) nous donne donc $s, t \in \mathbb{A}$ tels que $sa + tc = 1$. En multipliant par b,

$$sab + tbc = b. \tag{3.156}$$

Mais les deux termes du membre de gauche sont multiples de a parce que a divise bc. Par conséquent b est somme de deux multiples de a et donc est multiple de a. □

Un cas usuel d'utilisation est le cas de $A = \mathbb{N}^*$.

3.11.2 Élément premier

Définition 3.135 ([65]).
*Soit un anneau commutatif A. Un élément $p \in A$ est **premier** si il est*

(1) non nul,
(2) non inversible,
(3) si p divise un produit ab, alors il divise soit a soit b (ou le deux).

Proposition 3.136 ([66]).
Dans un anneau intègre[42] *tout élément premier est irréductible*[43].

Démonstration. Soit p, un élément premier dans un anneau intègre A.

p n'est pas inversible Cela fait partie de la définition d'un élément premier.

p n'est pas un produit d'inversibles Soient $a, b \in A$ tels que $p = ab$. Par le point (3) de la définition 3.135, p divise soit a soit b. Supposons que p divise a. Alors il existe $x \in A$ tel que $a = px$. En remettant dans $p = ab$ nous avons :

$$p = pxb. \tag{3.157}$$

Mais l'anneau est intègre et permet donc des simplifications par tout élément non nul. La relation 3.157 donne donc

$$1 = xb, \tag{3.158}$$

ce qui signifie que b est inversible.
Un travail similaire montre que a est inversible si p divise b.

□

Exemple 3.137.
Si nous avons l'égalité $7 = ab$ dans $\mathbb{Z}$, alors soit a soit b vaut 1 et est donc inversible. △

Sur un anneau non intègre, la notion d'élément premier n'est pas aussi intéressante que sur un anneau intègre. Par exemple la proposition 3.136 devient fausse.

Exemple 3.138.
Soit l'anneau $\mathbb{Z}^2$. L'élément $(1, 0)$ est premier mais pas irréductible.

$(1, 0)$ est premier L'élément $(1, 0)$ est non nul ; ça c'est pas cher. Pour qu'il soit inversible, il faudrait $(1, 0)(x, y) = (1, 1)$. Entre autres, $0 \times y = 1$, ce qui est impossible. Donc il n'est pas inversible.
Supposons que $(1, 0)$ divise le produit $(a, b)(c, d) = (ac, b)$. Alors il existe (x, y) tel que $(1, 0)(x, y) = (ac, bd)$. Cela signifie que $x = ac$ et $0 \times y = bd$. En particulier, soit $b = 0$ soit $d = 0$. Si $b = 0$, nous avons $(a, b) = (a, 0)$ et effectivement, $(1, 0)$ le divise.

42. Si pas intègre, voir l'exemple 3.138.
43. Toutes les définitions dans le thème 52.

$(1,0)$ **n'est pas irréductible** Nous avons $(1,0) = (1,0)(1,0)$. Donc l'élément $(1,0)$ est le produit de deux éléments non inversibles.

△

Exemple 3.139.
Si $\mathbb{K}$ est un corps, l'élément XY de $\mathbb{K}[X,Y]$ n'est pas premier parce que $XY \mid X^2Y^2$ alors que XY ne divise ni X^2 ni Y^2. △

Proposition 3.140 ([67, 1], thème 52)**.**
Soit un anneau principal A et un élément $p \neq 0$ dans A. Nous avons équivalence de :

(1) (p) est un idéal premier,

(2) p est un élément premier,

(3) p est un élément irréductible,

(4) (p) est un idéal maximum propre[44].

Démonstration. En plusieurs implications.

(1) implique (2) En plusieurs points.

— La condition $p \neq 0$ est dans les hypothèses de la proposition.

— Si p était inversible, nous aurions $(p) = A$ et donc pas que (p) est un idéal premier.

— Soient $a, b \in A$ tels que $p \mid ab$. En particulier, $(ab) \in (p)$. Mais comme (p) est un idéal premier, cela implique soit $a \in (p)$ soit $b \in (p)$. Donc soit p divise a soit p divise b.

(2) implique (3) Un anneau principal est intègre ; c'est dans la définition 3.118. Dans un anneau intègre, tout élément premier est irréductible, c'est la proposition 3.136.

(3) implique (4) Soit un idéal I contenant (p). Vu que A est principal, I est engendré par un seul élément ; soit $I = (a)$. Vu que $p \in I$, l'élément a divise p. Mais comme p est un élément premier, $a \mid p$ implique $a = p$ ou $a = 1$. Dans le premier cas, $I = (a) = (p)$, et dans le second cas, $I = (a) = (1) = A$. Donc (p) est bien un idéal maximum.

De plus l'idéal (p) est propre. En effet avoir $(p) = A$ dirait en particulier que $1 \in (p)$, c'est-à-dire qu'il existe $x \in A$ tel que $xp = 1$. Or p étant irréductible, il est non inversible.

(4) implique (1) C'est la proposition 3.121(3).

□

Un exemple d'élément premier non irréductible est $[4]_6$ dans l'anneau non principal $\mathbb{Z}/6\mathbb{Z}$. Voir 3.149 et le lemme 3.150.

3.11.3 Anneau noethérien

Définition 3.141.
*Un anneau est dit **noethérien** si toute suite croissante d'idéaux est stationnaire (à partir d'un certain rang).*

Montrer que tout anneau principal est noethérien est le premier pas pour montrer que tout anneau principal est factoriel.

Lemme 3.142.
Tout anneau principal[45] *est noethérien.*

44. Ce « propre » n'est pas dans l'énoncé sur Wikipédia. Je ne comprends pas pourquoi, et j'ai posé la question sur la page de discussion.
https://fr.wikipedia.org/wiki/Discussion:Idéal_premier

45. Définition 3.118.

Démonstration. Soit (J_n) une suite croissante d'idéaux et J la réunion. L'ensemble J est encore un idéal parce que les J_i sont emboités. Étant donné que l'idéal est principal nous pouvons prendre $a \in J$ tel que $J = (a)$. Il existe N tel que $a \in J_N$. Alors pour tout $n \geqslant N$ nous avons

$$J \subset J_N \subset J_n \subset J. \tag{3.159}$$

La première inclusion est le fait que $J = (a)$ et $a \in J_N$. La seconde est la croissance des idéaux et la troisième est le fait que J est une union. Par conséquent pour tout $n \geqslant N$ nous avons $J_N = J_n = J$. La suite est par conséquent stationnaire. □

Exemple 3.143.
Il y a moyen d'avoir un anneau noetherien non principal. C'est le cas de $\mathbb{Z}/6\mathbb{Z}$ dont nous parlerons dans 3.150. △

Théorème 3.144 ([68])**.**
Tout anneau principal est factoriel.

Exemple 3.145 ($\mathbb{Z}[i\sqrt{5}]$ n'est ni factoriel ni principal)**.**
Vu que $(i\sqrt{5})^2 = -5$, les éléments de $\mathbb{Z}[i\sqrt{5}]$ sont les éléments de $\mathbb{C}$ de la forme $a + bi\sqrt{5}$ avec $a, b \in \mathbb{Z}$. Nous définissons la **norme** sur $\mathbb{Z}[i\sqrt{5}]$ par [46]

$$\begin{aligned} N\colon \mathbb{Z}[i\sqrt{5}] &\to \mathbb{N} \\ z &\mapsto z\bar{z}. \end{aligned} \tag{3.160}$$

Le fait que ce soit à valeurs dans $\mathbb{N}$ est un simple calcul :

$$N(x + iy\sqrt{5}) = (x + iy\sqrt{5})(x - iy\sqrt{5}) = x^2 + 5y^2. \tag{3.161}$$

De plus N est multiplicative : $N(z_1z_2) = N(z_1)N(z_2)$.

Nous pouvons maintenant déterminer les inversibles de $\mathbb{Z}[i\sqrt{5}]$. Si α est inversible, alors il existe β tel que $\alpha\beta = 1$. Au niveau de la norme,

$$N(\alpha)N(\beta) = 1, \tag{3.162}$$

ce qui implique que $N(\alpha) = 1$. Or l'équation $x^2 + 5y^2 = 1$ dans $\mathbb{N}$ donne $y = 0$, $x = \pm 1$.

Au final, les inversibles de $\mathbb{Z}[i\sqrt{5}]$ sont ± 1.

L'anneau $\mathbb{Z}[i\sqrt{5}]$ n'est alors pas factoriel (définition 3.114) parce que

$$2 \times 3 = (1 + i\sqrt{5})(1 - i\sqrt{5}). \tag{3.163}$$

Cela donne deux décompositions du nombre 6 en produit d'éléments non associés [47] (2 n'est associé qu'à 2 et -2) parce que les inversibles sont 1 et -1.

Le fait que $\mathbb{Z}[i\sqrt{5}]$ ne soit pas factoriel implique qu'il ne soit pas principal, théorème 3.144. △

3.12 Anneau $\mathbb{Z}/6\mathbb{Z}$

Nous allons donner quelques propriétés de cet anneau, et en particulier voir que dans cet anneau non intègre, la notion d'élément irréductible n'est pas très intéressante.

Voici pour commencer un calcul la table de multiplication de $A = \mathbb{Z}/6\mathbb{Z}$. Pour les multiples de (par exemple) $[4]_6$ nous écrivons

$$1 \times [4]_6 = [4_6] \tag{3.164}$$

et ensuite

$$2 \times [4]_6 = [8]_6 = [2]_6, \tag{3.165}$$

46. C'est le carré de la norme usuelle, mais c'est l'usage dans le milieu.
47. Définition 3.104.

puis

$$3 \times [4]_6 = [2+4]_6 = [6]_6 = [0]_6, \tag{3.166}$$

et caetera. Le résultat est :

$\times$	$[0]_6$	$[1]_6$	$[2]_6$	$[3]_6$	$[4]_6$	$[5]_6$
$[0]_6$	0	0	0	0	0	0
$[1]_6$	0	1	2	3	4	5
$[2]_6$	0	2	4	0	2	4
$[3]_6$	0	3	0	3	0	3
$[4]_6$	0	4	2	0	4	2
$[5]_6$	0	5	4	3	2	1

(3.167)

Pour ne pas alourdir, nous n'avons pas écrit $[x]_6$ partout au lieu de x.

3.146 (Inversibles)**.**
Les éléments inversibles de $\mathbb{Z}/6\mathbb{Z}$ sont ceux qui ont un $[1]_6$ dans leur table de multiplication. Ce sont donc

$$U(\mathbb{Z}/6\mathbb{Z}) = \big\{[1]_6, [5]_6\big\}. \tag{3.168}$$

3.147 (Diviseurs de zéro)**.**
Les diviseurs de zéro sont ceux qui ont un $[0]_6$ dans leur table de multiplication, c'est-à-dire

$$\big\{[2]_6, [3]_6, [4]_6\big\}. \tag{3.169}$$

3.148 (Irréductibles)**.**
Les irréductibles sont ceux qui ne sont ni inversibles ni produit de deux éléments non inversibles. Les non inversibles sont :

$$\big\{[0]_6, [2]_6, [3]_6, [4]_6\big\}. \tag{3.170}$$

Ils sont candidats à être irréductibles. Les produits de ces éléments (on oublie les crochets) sont :

$$2 \times 2 = 4 \tag{3.171a}$$
$$2 \times 3 = 0 \tag{3.171b}$$
$$2 \times 4 = 2 \tag{3.171c}$$
$$3 \times 3 = 3 \tag{3.171d}$$
$$3 \times 4 = 0 \tag{3.171e}$$
$$4 \times 4 = 4. \tag{3.171f}$$

Donc $[0]_6$, $[2]_6$, $[3]_6$ et $[4]_6$ ne sont plus candidats à être irréductible. Bref, il ne reste aucun candidats.

L'anneau $\mathbb{Z}/6\mathbb{Z}$ n'a aucun élément irréductible.

3.149 (Éléments premiers)**.**
Les éléments non nuls et non inversibles sont 2, 3 et 4.

Pour 2 L'élément $[2]_6$ divise 2, 4 et 0.

- Les (a, b) tels que $ab = 2$ sont : $(1, 2)$, $(2, 4)$ et $(5, 4)$. L'élément 2 divise donc toujours a ou b.
- Les (a, b) tels que $ab = 4$ sont : $(1, 4)$, $(2, 5)$ et $(4, 4)$. L'élément 2 divise donc toujours a ou b.
- Les (a, b) tels que $ab = 0$ sont : $(0, x)$, $(3, 2)$ et $(4, 3)$. L'élément 2 divise donc toujours a ou b. En particulier, $[2]_6$ divise $[0]_6$; c'est important.

Donc $[2]_6$ est un élément premier.

Pour 3 L'élément $[3]_6$ divise 3 et 0.

— Les (a, b) tels que $ab = 3$ sont : $(1, 3)$ et $(3, 5)$. L'élément 3 divise donc toujours a ou b.

— Les (a, b) tels que $ab = 0$ sont : $(0, x)$, $(3, 2)$ et $(4, 3)$. L'élément 3 divise donc toujours a ou b.

Donc $[3]_6$ est un élément premier. L'élément $[4]_6$ divise 4, 2 et 0.

— Les (a, b) tels que $ab = 4$ sont : $(1, 4)$, $(2, 5)$ et $(4, 4)$. L'élément 4 divise donc toujours a ou b.

— Les (a, b) tels que $ab = 2$ sont : $(1, 2)$, $(2, 4)$ et $(5, 4)$. L'élément 4 divise donc toujours a ou b.

— Les (a, b) tels que $ab = 0$ sont : $(0, x)$, $(3, 2)$ et $(4, 3)$. L'élément 4 divise donc toujours a ou b.

Donc $[4]_6$ est un élément premier.

Au final, les éléments premiers dans $\mathbb{Z}/6\mathbb{Z}$ sont

$$\{[2]_6, [3]_6, [4]_6\}. \tag{3.172}$$

Vous noterez que $\mathbb{Z}/6\mathbb{Z}$ a des éléments premiers non irréductibles. Cela est un contre-exemple à la proposition 3.140 dans le cas d'un anneau non-intègre.

Lemme 3.150 ([1])**.**
L'anneau $\mathbb{Z}/6\mathbb{Z}$ est noetherien, mais ni intègre ni principal[48].

Démonstration. Vu que c'est un anneau fini, toute suite croissante de quoi que ce soit devient stationnaire ; donc $\mathbb{Z}/6\mathbb{Z}$ est noetherien.

Vu que $\mathbb{Z}/6\mathbb{Z}$ a des diviseurs de zéro, il n'est pas intègre. Et vu qu'il n'est pas intègre, il n'est pas factoriel non plus. □

3.13 Anneau euclidien

Définition 3.151 (Wikipédia)**.**
Soit A un anneau intègre. Un ***stathme euclidien*** *sur A est une application $\alpha\colon A\backslash\{0\} \to \mathbb{N}$ tel que*

(1) $\forall a, b \in A\backslash\{0\}$, il existe $q, r \in A$ tel que

$$a = qb + r \tag{3.173}$$

et $\alpha(r) < \alpha(b)$.

(2) Pour tout $a, b \in A\backslash\{0\}$, $\alpha(b) \leqslant \alpha(ab)$.

Un anneau est ***euclidien*** *s'il accepte un stathme euclidien.*

Le stathme est la fonction qui donne le « degré » à utiliser dans la division euclidienne. La contrainte est que le degré du reste soit plus petit que le degré du dividende.

Exemple 3.152.
Le stathme de $\mathbb{N}$ pour la division euclidienne usuelle est $\alpha(n) = n$. Si $a, b \in \mathbb{N}$ nous écrivons

$$a = qb + r \tag{3.174}$$

où q est l'entier le plus proche *inférieur* à a/b (on veut que le reste soit positif) et $r = a - qb$. Nous avons donc

$$r - b = a - b(q + 1) < a - b\frac{a}{b} = 0, \tag{3.175}$$

ce qui montre que $r < b$. △

48. Toutes les définitions dans le thème 52.

Cet exemple ne fonctionne pas avec $\mathbb{Z}$ au lieu de $\mathbb{N}$ parce que le stathme doit avoir des valeurs dans $\mathbb{N}$. Cela ne veut cependant pas dire qu'il n'existe pas de stathme sur $\mathbb{Z}$; cela veut seulement dire que $\alpha(x) = x$ ne fonctionne pas.

Proposition 3.153 ([69]).
Un anneau euclidien est principal.

Démonstration. Soit A un anneau principal et α un stathme sur A. Nous considérons un idéal I non nul de A. Nous devons montrer que I est généré par un élément. En l'occurrence nous allons montrer qu'un élément $a \in I\backslash\{0\}$ qui minimise $\alpha(a)$ va générer[49]. Soit $x \in I$. Par construction, il existe $q, r \in A$ tels que $x = aq + r$ avec $r = 0$ ou $\alpha(r) < \alpha(a)$. Étant donné que $x, a \in I$, $r \in I$. Si $r \neq 0$, alors r contredirait la minimalité de $\alpha(a)$. Donc $r = 0$ et $x = aq$, ce qui signifie que I est principal. □

Proposition 3.154.
L'anneau $\mathbb{Z}$ est principal et euclidien.

Démonstration. Nous allons seulement montrer que $\alpha(x) = |x|$ est un stathme euclidien. Ainsi $\mathbb{Z}$ sera euclidien et donc principal par la proposition 3.153.

D'abord $\mathbb{Z}$ est intègre, c'est l'exemple 3.97.

La condition $\alpha(b) \leqslant \alpha(ab)$ est immédiate : $|a| \leqslant |ab|$ pour tout $a, b \in \mathbb{Z}$.

Soient maintenant $a, b \in \mathbb{Z}$. Nous définissons $q_0, r_0 \in \mathbb{N}$ tels que

$$|a| = q_0|b| + r_0 \tag{3.176}$$

avec $r_0 < |b|$. Cela existe parce que $\alpha(x) = x$ est un stathme sur $\mathbb{N}$ par l'exemple 3.152.

Si $a > 0$ et $b > 0$ Alors $a = q_0 b + r_0$ et le couple (q_0, r_0) vérifie les conditions de la définition 3.151(1).

Si $a > 0$ et $b < 0$ Alors $a = -q_0 b + r_0$, et le couple $(-q_0, r_0)$ vérifie les conditions de la définition 3.151(1).

Si $a < 0$ et $b > 0$ Alors $a = -q_0 b - r_0$, et le couple $(-q_0, -r_0)$ vérifie les conditions de la définition 3.151(1) parce que

$$\alpha(-r_0) = r_0 < |b| = \alpha(b). \tag{3.177}$$

Si $a < 0$ et $b < 0$ Alors $a = q_0 b - r_0$, et le couple $(q_0, -r_0)$ vérifie les conditions de la définition 3.151(1).

□

Nous venons de voir que $\mathbb{Z}$ est principal ; le lemme suivant nous dit que $\mathbb{Z}[X]$ n'est pas principal, lui.

Lemme 3.155 ([70]).
Si A est un anneau intègre qui n'est pas un corps, alors $A[X]$ n'est pas principal.

Démonstration. Soit un élément non nul $a \in A$.

Un idéal principal contenant a et X est $\mathbf{A[X]}$ Soit (P) un idéal principal contenant a et X. Vu que $a \in (P)$, il existe Q tel que $a = QP$. Donc P divise a dans $\mathbb{Z}[X]$. Les degrés font que P est un polynôme constant, c'est-à-dire en réalité un élément de A. Soit $P = k \in A$.

Vu que P divise X, nous avons aussi $X = kQ$ pour un certain $Q \in \mathbb{Z}[X]$. Les degrés disent qu'il existe $k' \in A$ tel que $Q = k'X$ et donc $Q = k'X = k'kQ$, ce qui implique que $kk' = 1$. L'idéal engendré par k contient donc en particulier $kk' = 1$ et donc contient $A[X]$ en entier :

$$1 = k'k \in k'(P) = (P). \tag{3.178}$$

49. Un tel élément existe...

Si $(a, X) = A[X]$ alors a est inversible Si $(a, X) = A[X]$, en particulier, $1 \in (a, X)$, ce qui signifie qu'il existe des polynômes $U, V \in A[X]$ tels que

$$1 = UX + Va. \tag{3.179}$$

Nous évaluons cette égalité en 0 : vu que $(UX)(0) = 0$ nous avons $1 = V(0)a$, ce qui signifie que $V(0)$ est un inverse de A. Donc a est inversible.

Si a n'est pas inversible alors (a, X) n'est pas principal Si (a, X) était principal, alors nous aurions, par ce qui est dit plus haut, $(a, X) = A[X]$. Mais cette dernière égalité impliquerait que a est inversible.

En conclusion, si A n'est pas un corps, il possède un élément ni nul ni inversible. Dans ce cas, l'idéal (a, X) n'est pas principal dans $A[X]$ et nous en déduisons que $A[X]$ n'est pas un anneau principal. □

Nous verrons dans le lemme 3.178 que si $\mathbb{K}$ est un corps, alors $\mathbb{K}[X]$ est principal.

Exemple 3.156.
Prouvons que $\mathbb{Z}[i\sqrt{2}]$ est un anneau euclidien. Pour cela nous démontrons que

$$\begin{aligned} N\colon \mathbb{Z}[i\sqrt{2}] &\to \mathbb{N} \\ a + bi\sqrt{2} &\mapsto a^2 + 2b^2 \end{aligned} \tag{3.180}$$

est un stathme euclidien.

Soient $z = a + bi\sqrt{2}$, $t = a' + b'i\sqrt{2}$. Nous cherchons q et r tels que la division euclidienne s'écrive $z = qt + r$. Soient $\alpha, \beta \in \mathbb{Q}$ tels que

$$\frac{z}{t} = \alpha + \beta i\sqrt{2}. \tag{3.181}$$

Nous désignons par $\alpha + \epsilon_1$ et $\beta + \epsilon_2$ les entiers les plus proches de α et b. Nous avons $|\epsilon_1|, |\epsilon_2| \leqslant \frac{1}{2}$. Nous posons alors naturellement

$$q = (\alpha + \epsilon_1) + (\beta + \epsilon_2)i\sqrt{2} \tag{3.182}$$

et nous calculons $r = z - qt$:

$$2b'\epsilon_2 - a'\epsilon_1 + i\sqrt{2}\big(\epsilon_1 b' - a'\epsilon_2\big). \tag{3.183}$$

Nous trouvons

$$N(r) = a'^2\epsilon_1^2 + 4b'^2\epsilon_2^2 + 2a'^2\epsilon_1^2 + 2b'^2\epsilon_2^2 \leqslant \frac{3}{4}a'^2 + \frac{3}{2}b'^2. \tag{3.184}$$

D'autre part $N(t) = a'^2 + 2b'^2$, et nous avons donc bien $N(r) < N(t)$.

En ce qui concerne la seconde propriété du stathme, un petit calcul montre que

$$N(zt) = (a^2 + 2b^2)(a'^2 + 2b'^2), \tag{3.185}$$

et tant que $t \neq 0$ nous avons bien $N(zt) > N(z)$. △

Notons en particulier que $\mathbb{Z}[i\sqrt{2}]$ est factoriel et principal.

Exemple 3.157 (Décomposition en facteurs irréductibles dans $\mathbb{Z}[i\sqrt{2}]$)**.**
Les éléments inversibles de $\mathbb{Z}[i\sqrt{2}]$ sont ± 1, donc deux éléments a et b sont associés (définition 3.104) si et seulement si $a = \pm b$.

De plus si p est irréductible[50], alors $-p$ est irréductible. Les éléments irréductibles de $\mathbb{Z}[i\sqrt{2}]$ arrivent donc par pairs d'éléments associés. Soit $\{p_i\}$ une sélection de un élément irréductible parmi chaque paire. Tout élément x de $\mathbb{Z}[i\sqrt{2}]$ peut alors être écrit $x = \pm p_1^{\alpha_1} \dots p_n^{\alpha_n}$. Ce fait va être pratique pour comparer des décompositions en facteurs irréductibles d'éléments. △

50. Définition 3.110

Le lemme suivant fait en pratique partie de l'exemple 3.160, mais nous l'isolons pour plus de clarté [51].

Lemme 3.158.
Si a et b sont deux éléments premiers entre eux de $\mathbb{Z}[i\sqrt{2}]$, et s'il existe $y \in \mathbb{Z}[i\sqrt{2}]$ tel que $ab = y^3$, alors a et b sont des cubes (dans $\mathbb{Z}[i\sqrt{2}]$).

Démonstration. D'après l'exemple 3.157 nous pouvons écrire

$$y = \pm p_1^{\sigma_1} \dots p_n^{\sigma_n} \tag{3.186a}$$
$$a = \pm p_1^{\alpha_1} \dots p_n^{\alpha_n} \tag{3.186b}$$
$$b = \pm p_1^{\beta_1} \dots p_n^{\beta_n} \tag{3.186c}$$

où les p_i sont les irréductibles de $\mathbb{Z}[i\sqrt{2}]$ « modulo ± 1 » au sens où la liste des irréductibles est $\{p_i\} \cup \{-p_i\}$ (union disjointe). Étant donné que a et b sont premiers entre eux, α_i et β_i ne peuvent pas être non nuls en même temps alors que leur somme doit faire $3\sigma_i$. Nous avons donc pour chaque i soit $\alpha_i = 3\sigma_i$ soit $\beta = 3\sigma_i$ (et bien entendu si $\sigma_i = 0$ alors $\alpha_i = \beta_i = 0$).

Étant donné que ± 1 sont également deux cubes, a et b sont bien des cubes.

Notons que nous avons utilisé de façon capitale le fait que $\mathbb{Z}[i\sqrt{2}]$ était factoriel. □

3.13.1 Équations diophantiennes

Exemple 3.159.
L'équation diophantienne

$$x^2 = 3y^2 + 8 \tag{3.187}$$

n'a pas de solutions. En effet si nous prenons l'équation modulo 3 nous obtenons

$$[x^2]_3 = [3y^2 + 8]_3 = [8]_3 = [2]_3. \tag{3.188}$$

Or dans $\mathbb{Z}/3\mathbb{Z}$, aucun carré n'est égal à deux : $0^2 = 0 \neq 2$, $1^2 = 1 \neq 2$ et $2^2 = 4 = 1 \neq 2$. △

Exemple 3.160.
Résolvons l'équation diophantienne

$$x^2 + 2 = y^3. \tag{3.189}$$

Une première remarque est que x doit être impair. En effet si $x = 2k$, nous devons avoir y^3 pair. Mais si un cube pair est divisible par 8, donc $y^3 = 8l$. L'équation devient $4k^2 + 2 = 8l^3$, c'est-à-dire $2k^2 + 1 = 4l^3$. Le membre de gauche est impair tandis que celui de droite est pair. Impossible.

Nous pouvons écrire l'équation sous la forme $x^2 + 2 = (x + i\sqrt{2})(x - i\sqrt{2})$. Et nous considérons $\mathbb{Z}[i\sqrt{2}]$ muni de son stathme N donné par (3.180).

L'élément $i\sqrt{2}$ est irréductible parce que $N(i\sqrt{2}) = 2$, et si nous avions $i\sqrt{2} = pq$, alors nous aurions $N(p)N(q) = 2$, ce qui n'est possible que si $N(p)$ ou $N(q)$ vaut 1.

Nous prouvons maintenant que les éléments $x + i\sqrt{2}$ et $x - i\sqrt{2}$ sont premiers entre eux. Supposons que d soit un diviseur commun ; alors il divise aussi la somme et la différence. Donc d divise à la fois $2x$ et $2i\sqrt{2}$.

Étant donné que $i\sqrt{2}$ est irréductible et que $2i\sqrt{2} = (-i\sqrt{2})^3$, les diviseurs de $2i\sqrt{2}$ sont les puissances de $(-i\sqrt{2})$. Du coup nous devrions avoir $d = (i\sqrt{2})^\beta$ et donc

$$x = (i\sqrt{2})^\beta q \tag{3.190}$$

pour un certain $q \in \mathbb{Z}[i\sqrt{2}]$. Dans ce cas nous avons $N(x) = 2^\beta N(q)$, mais nous avons déjà précisé que x ne pouvait pas être pair, donc $\beta = 0$ et nous avons $d = 1$.

Vu que les nombres $x \pm i\sqrt{2}$ sont premiers entre eux et que leur produit doit être un cube, ils doivent être séparément des cubes (lemme 3.158). Nous devons donc résoudre séparément $x \pm i\sqrt{2} = y^3$.

51. Merci à Marvoir pour m'avoir souligné le manque.

Cherchons les x et y entiers tels que $x + i\sqrt{2} = y^3$. Si nous posons $z = a + bi\sqrt{2}$, il suffit de calculer z^3 :

```
----------------------------------------------------------------------
| Sage Version 4.8, Release Date: 2012-01-20                         |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: var('a,b')
(a, b)
sage: z=a+I*sqrt(2)*b
sage: (z**3).expand()
3*I*sqrt(2)*a^2*b - 2*I*sqrt(2)*b^3 + a^3 - 6*a*b^2
```

En identifiant cela à $x + i\sqrt{2}$ nous trouvons le système

$$\begin{cases} x = a^3 - 6ab^2 & (3.191a) \\ 1 = 3a^2b - 2b^3 & (3.191b) \end{cases}$$

où, nous le rappelons, x, a et b sont des entiers. La seconde équation montre que b doit être inversible : $b(3a^2 - 2b^2) = 1$. Il y a donc les possibilités $b = \pm 1$. Pour $b = 1$ l'équation devient $3a^2 - 2 = 1$, c'est-à-dire $a = \pm 1$. Pour $b = -1$ l'équation devient $3a^2 - 2 = -1$, impossible. En conclusion les possibilités sont

$$(x, z) = (-5, 1 + i\sqrt{2}) \tag{3.192a}$$
$$(x, z) = (5, -1 + i\sqrt{2}) \tag{3.192b}$$
$$\tag{3.192c}$$

Le travail avec $x - i\sqrt{2}$ donne les mêmes résultats.

Les deux solutions de l'équation $x^2 + 2 = y^3$ sont alors $(5, 3)$ et $(-5, 3)$. △

3.13.2 Triplets pythagoriciens et équation de Fermat pour $n = 4$

Définition 3.161.
*Les solutions entières (positives) de l'équation $x^2 + y^2 = z^2$ sont appelés **triplets pythagoriciens**.*

Ils donnent toutes les possibilités de triangles rectangles dont les côtés ont des longueurs entières.

Définition 3.162.
*On dit qu'un triplet pythagoricien est **primitif** si les trois nombres sont premiers dans leur ensemble*[52].

Remarquons que cela est équivalent à montrer que les trois nombres sont premiers deux à deux : en effet, si deux parmi x, y et z sont divisibles par un nombre, alors tous les trois sont divisibles par ce nombre[53], donc les nombres x, y et z sont premiers deux à deux.

Lemme 3.163.
Dans un triplet pythagoricien primitif (x, y, z), on a toujours z impair et :

- *soit x impair et y pair ;*
- *soit x pair et y impair.*

Démonstration. Remarquons que le fait d'imposer que le triplet soit primitif, interdit aux nombres x et y d'être pairs en même temps. En effet, si c'était le cas, alors x^2 et y^2 seraient aussi pairs,

52. Définition 3.11.
53. Parce que k et k^2 ont les mêmes facteurs premiers.

donc leur somme z^2 aussi, d'ou z serait pair et les trois nombres ne seraient pas premiers entre eux.

Nous montrons à présent que les nombres x et y ne sont pas tous les deux impairs. Par l'absurde, si $x = 2a + 1$, nous avons $x^2 = 4a^2 + 4a + 1 \in [1]_4$; de la même manière, $y^2 \in [1]_4$. On en déduit alors que $z^2 = x^2 + y^2 \in [2]_4$. Le nombre z^2 est donc pair, donc z est pair : disons $z = 2c$. Alors, $z^2 = 4c^2 \in [0]_4$. Comme les classes modulo 4 sont disjointes, nous aboutissons à une contradiction. □

Proposition 3.164 (Triplets pythagoriciens[71, 72]).
Un triplet $(x, y, z) \in (\mathbb{N}^*)^3$ est solution de $x^2+y^2 = z^2$ si et seulement s'il existe $d \in \mathbb{N}$ et $u, v \in \mathbb{N}^*$ premiers entre eux tels que

$$\begin{cases} x = d(u^2 - v^2) & (3.193a) \\ y = 2duv & (3.193b) \\ z = d(u^2 + v^2) & (3.193c) \end{cases}$$

ou

$$\begin{cases} x = 2duv & (3.194a) \\ y = d(u^2 - v^2) & (3.194b) \\ z = d(u^2 + v^2) & (3.194c) \end{cases}$$

La différence entre les deux est seulement d'inverser les rôles de x et y.

Démonstration. Montrons d'abord que les formules proposées sont bien des solutions ; nous vérifions (3.193) :

$$x^2 + y^2 = d^2(u^2 - v^2) + 4d^2u^2v^2 = d^2(u^2 + v^2)^2, \tag{3.195}$$

qui correspond bien au z^2 proposé.

Nous allons maintenant prouver la réciproque : toute solution est d'une des deux formes proposées. Déterminer les triplets primitifs suffira parce que si (x, y, z) n'est pas une solution primitive, alors en posant $k = \text{pgcd}(x, y, z)$, le triplet $\left(\frac{x}{k}, \frac{y}{k}, \frac{z}{k}\right)$ est primitif. Connaissant les triplets primitifs, nous obtenons tous les autres par simple multiplication.

Soit donc (x, y, z) un triplet pythagoricien primitif. Sans perte de généralité [54], grâce au lemme 3.163, nous pouvons supposer x est pair tandis que y et z sont impairs. Comme $x^2 = (z+y)(z-y)$, nous avons

$$\left(\frac{x}{2}\right)^2 = \left(\frac{z+y}{2}\right)^2 \left(\frac{z-y}{2}\right)^2. \tag{3.196}$$

Vu que z et y sont premiers entre eux, les nombres $z - y$ et $z + y$ sont également premiers entre eux [55]. Donc les facteurs premiers (qui arrivent au moins au carré) de $(x/2)^2$ sont chacun soit dans $(z + y)/2$ soit dans $(z - y)/2$. Nous en déduisons que ces derniers sont des carrés d'entiers. Nous posons

$$\frac{z-y}{2} = u^2 \quad \frac{z+y}{2} = v^2. \tag{3.197}$$

Bien entendu u et v sont non nuls parce que nous avons exclu la possibilité de triplets dont un élément serait nul. Avec tout cela nous avons $(x/2)^2 = u^2v^2$ et donc $x = 2uv$ puis par somme et différence :

$$\begin{cases} x = 2uv & (3.198a) \\ y = v^2 - u^2 & (3.198b) \\ z = u^2 + v^2, & (3.198c) \end{cases}$$

ce qu'il fallait. □

54. En échangeant les rôles de x et y ici, nous obtenons à la fin la seconde forme des solutions.
55. Si $z - y = kn$ et $z + y = km$, faisant la somme et la différence on voit que y et z sont divisibles par k.

Remarque 3.165.
Les solutions dans lesquelles x, y ou z sont nuls sont faciles à classer. La solution $(1,0,1)$ n'est pas dans les formes proposées. En effet elle ne peut pas être de la première forme : avoir $y = 0$ demanderait qu'un nombre parmi d, u et v soit nul, ce qui est interdit. La solution $(1,0,1)$ ne peut pas non plus être de la seconde forme parce que x y est pair.

Proposition 3.166 ([71])**.**
Les équations $x^4 + y^4 = z^2$ et $x^2 + y^4 = z^4$ n'ont pas de solutions dans $(\mathbb{N}^*)^3$.

Démonstration. Si la première équation n'a pas de solutions, alors la seconde n'en n'a pas non plus parce que z^4 est un carré. Nous nous concentrons donc sur l'équation $x^4 + y^4 = z^2$ et nous supposons qu'il existe au moins une solution dans $(\mathbb{N}^*)^3$. Nous en choisissons une (x, y, z) avec z minimum (les z dans différentes solutions étant dans $\mathbb{N}$, il en existe forcément un minimum [56]). Du coup, les trois nombres x, y et z sont premiers dans leur ensembles parce que une division par leur pgcd donnerait une nouvelle solution qui contredirait la minimalité de z.

Nous posons $x^4 = \bar{x}^2$ et $y^4 = \bar{y}^2$. Ils vérifient l'équation $\bar{x}^2 + \bar{y}^2 = z^2$ et par la proposition 3.164, il existe $u, v \in \mathbb{N}^*$ premiers entre eux tels que, sans perte de généralité [57], on ait

$$\begin{cases} \bar{x} = 2uv & (3.199a) \\ \bar{y} = u^2 - v^2 & (3.199b) \\ z = u^2 + v^2. & (3.199c) \end{cases}$$

Si u est pair, alors v est impair (et inversement) parce que $\operatorname{pgcd}(u, v) = 1$ Si u est pair, alors $u = 2a$ et $v = 2b + 1$, ce qui donne $\bar{y} = 4a^2 - 4b^2 - 4b - 1 \in [-1]_4$. Or nous avons déjà vu qu'un carré est dans $[0]_4$ ou dans $[1]_4$. Il faut donc que u soit impair. Le lemme 3.163 implique alors que v soit pair.

De l'équation 3.199b, nous en déduisons que $v^2 + \bar{y} = u^2$; de plus u^2, v^2 et $\bar{y}$ sont premiers dans leur ensemble : en effet, u et v sont premiers entre eux, et si l'un parmi u^2 et v^2 a un facteur commun avec $\bar{y}$, alors l'autre doit l'avoir aussi (dans une égalité $a + b = c$, si deux des nombres ont un diviseur commun, le troisième l'a aussi). Comme $\bar{y} = y^2$, le triplet (v, y, u) est un triplet pythagoricien primitif. Nous réappliquons la proposition 3.164, en se souvenant que v est pair : il existe donc deux nombres r et s premiers entre eux tels que

$$\begin{cases} v = 2rs & (3.200a) \\ y = r^2 - s^2 & (3.200b) \\ u = r^2 + s^2. & (3.200c) \end{cases}$$

Avec cela, $\bar{x} = 2uv = 4rs(r^2 + s^2)$. Remarquons que les trois nombres r, s et $r^2 + s^2$ sont premiers entre eux dans leur ensemble ; or, comme $\bar{x}$ est un carré ces nombres doivent séparément être des carrés :

$$\begin{cases} r = \alpha^2 & (3.201a) \\ s = \beta^2 & (3.201b) \\ r^2 + s^2 = \gamma^2. & (3.201c) \end{cases}$$

En mettant les deux premiers dans la troisième, nous avons $\alpha^4 + \beta^4 = \gamma^2$. Donc $(\alpha^2, \beta^2, \gamma)$ est une solution. Nous allons prouver que $\gamma < z$, ce qui terminera la preuve, puisque z était supposé minimal. Nous avons :

$$\begin{aligned} z &= u^2 + v^2 && \text{par 3.199c} \\ &= r^2 + s^2 + 4r^2s^2 && \text{par 3.200} \\ &= \gamma^2 + 4r^2s^2 \\ &> \gamma^2, \end{aligned}$$

et a fortiori $\gamma < z$. □

56. Voir quelque chose comme le lemme 1.48.
57. En inversant les rôles de x et y au besoin.

3.14 Polynômes à coefficients dans un anneau commutatif

Lemme 3.167.
Nous considérons un polynôme $P \in A[X]$, et le quotient $A[X]/(P)$. Pour tout polynôme $Q \in A[X]$ nous avons les égalités

$$Q(\bar{X}) = \overline{Q(X)} = \bar{Q}. \tag{3.202}$$

Démonstration. Si $Q = \sum_k a_k X^k$, alors par la linéarité de la prise de classes,

$$\bar{Q} = \sum_k a_k \overline{X^k}. \tag{3.203}$$

Nous insistons sur le fait que cette égalité n'est rien d'autre que l'itération de la définition de la somme dans l'espace quotient : $\bar{x} + \bar{y} = \overline{x+y}$ ainsi que du produit $k\bar{x} = \overline{kx}$ (définition 1.80). Toujours par définition du produit appliqué à l'élément $\bar{X}$ nous avons $(\bar{X})^2 = \overline{X^2}$; par récurrence $\overline{X^k} = \bar{X}^k$, et

$$\bar{Q} = \sum_k a_k \bar{X}^k = Q(\bar{X}). \tag{3.204}$$

Le fait que $\bar{Q} = \overline{Q(X)}$ n'est rien d'autre que le fait que dans $A[X]$ nous avons $Q = Q(X)$, comme expliqué dans le lemme 3.90. □

3.14.1 Monômes

3.168.
Les éléments de la forme λX^k avec $\lambda \in A$ et $k \in \mathbb{N}$ sont des **monômes**.

Nous allons aussi considérer

$$A_n[X] = \{P \in A[X] \text{ tel que } \deg(P) \leqslant n\}. \tag{3.205}$$

Cela est un sous-module libre.

3.14.2 Évaluation

Soit $P \in A[X]$. À priori, P n'est qu'une suite dans A indexée par $\mathbb{N}$.

Nous avons déjà défini son évaluation sur un élément $\alpha \in A$ dans la définition 3.87 :

$$P(\alpha) = \sum_k a_k \alpha^k. \tag{3.206}$$

Cette somme est toujours finie.

3.169.
L'ensemble $A[X]$ est une algèbre et donc un espace vectoriel. Il possède un unique élément nul qui est celui dont tous les coefficients sont nuls ; cela est immédiat par la construction en tant que suites presque nulles.

Il n'y a à priori pas équivalence entre le fait d'être un polynôme nul et le fait de s'évaluer à zéro sur tous les éléments de A. Cela sera discuté dans le théorème 6.101 et l'exemple 18.37.

Définition 3.170.
Soient un anneau A et un anneau B qui contient A (comme sous-anneau). Pour $\alpha \in B$ nous définissons $A[\alpha]_B$ comme étant l'intersection de tous les sous-anneaux de B contenant A.

Comme dit plus haut, nous nous permettons d'écrire $A[\alpha]$ sans préciser B lorsque ce dernier sera clair dans le contexte.

Proposition 3.171.
Soient un anneau A et un anneau B qui contient A (comme sous-anneau). Pour tout $\alpha \in B$ nous avons

$$A[\alpha] = \{P(\alpha) \text{ tel que } P \in A[X]\} \tag{3.207}$$

où encore une fois, $P(\alpha)$ est calculé dans B ; le contexte est clair là-dessus.

Démonstration. Si A' est un sous-anneau de B contenant A et α, alors A' contient tous les $P(\alpha)$ avec $P \in A[X]$. Nous avons donc

$$\{P(\alpha) \text{ tel que } P \in A[X]\} \subset A[\alpha]. \tag{3.208}$$

Par ailleurs, $\{P(\alpha) \text{ tel que } P \in A[X]\}$ est un sous-anneau de B contenant A et α. Donc $A[\alpha]$ y est inclus. □

3.14.3 Polynômes sur un anneau intègre

Théorème 3.172.
L'anneau A est intègre si et seulement si $A[X]$ est intègre.

Démonstration. Soient P et Q des éléments non nuls de $A[X]$. Vu que l'anneau A est intègre, nous avons

$$\deg(PQ) = \deg(P) + \deg(Q) \tag{3.209}$$

et le produit ne peut pas être nul. L'anneau $A[X]$ est donc intègre.

Si $A[X]$ est intègre, A est intègre parce qu'il peut être vu comme sous anneau. □

3.173.
Si A n'est pas intègre, soient $\alpha, \beta \in A$ non nuls tels que $\alpha\beta = 0$. Le produit des polynômes $X \mapsto \alpha X$ et $X \mapsto \beta$ est $(\alpha X)(\beta) = 0$; le degré du produit n'est pas la somme des degrés.

Les personnes qui ont tout compris jusqu'ici remarqueront que la notation « $X \mapsto P(X)$ » n'est pas correcte parce que du point de vue que nous adoptons ici, un polynôme n'est pas une application.

Corolaire 3.174.
Si A est intègre, les inversibles de $A[X]$ sont les éléments de $U(A)$.

Démonstration. Pour que Q soit inversible, il faut un P tel que $PQ = 1$. Mais l'anneau A étant intègre, les degrés s'additionnent. Par conséquent ils doivent être de degrés zéro et il faut que $P, Q \in A$. Enfin pour qu'ils soient inversibles, ils doivent être dans $U(A)$. □

La **valuation** du polynôme $P = \sum_n a_n X^n$, notée $\mathrm{val}(P)$, est

$$\mathrm{val}(P) = \min\{n \text{ tel que } a_n \neq 0\}. \tag{3.210}$$

Nous avons $\mathrm{val}(P) \leqslant \deg(P)$ et $\mathrm{val}(P) = \deg(P)$ si et seulement si P est un monôme. Si $P = 0$, nous convenons que $\mathrm{val}(0) = \infty$ et $\deg(0) = -\infty$.

3.14.4 Division euclidienne

Le théorème suivant établit la **division euclidienne** dans $A[X]$ du polynôme P par un polynôme D.

Théorème 3.175.
Soit $D \neq 0$ dans $A[X]$ de coefficient dominant inversible dans A. Pour tout $P \in A[X]$, il existe $Q, R \in A[X]$ tels que

$$P = QD + R \tag{3.211}$$

avec $\deg(R) < \deg(D)$.

Les polynômes Q et R sont déterminés de façon univoque par cette condition.

Définition 3.176.
Le polynôme Q est le ***quotient*** *et R est le* ***reste*** *de la division euclidienne de P par D. Si le reste de la division de P par D est nul on dit que D* ***divise*** *P et on note $D \mid P$. Autrement dit D divise P s'il existe Q tel que $P = QD$.*[58]

3.177.
Le théorème 3.175 nous incite à utiliser le degré comme stathme euclidien sur $A[X]$ dès que A est un anneau intègre. Or cela ne fonctionne en général pas parce que très peu de polynômes ont à priori un coefficient dominant inversible.

Lemme 3.178 (Thème 46)**.**
Si $\mathbb{K}$ est un corps[59]*, alors l'anneau $\mathbb{K}[X]$ est euclidien et principal.*

Démonstration. Vu que $\mathbb{K}$ est un corps, tous les éléments sont inversibles et le degré donne un stathme par le théorème 3.175. L'anneau $\mathbb{K}[X]$ est donc euclidien et par conséquent principal (proposition 3.153). □

Dans le théorème 6.37 nous donnerons une preuve directe du fait que $\mathbb{K}[X]$ est principal en montrant que tous ses idéaux sont principaux. Nous y démontrerons donc un peu moins pour un peu plus cher, mais avec le plaisir de ne pas devoir passer par un stathme.

Définition 3.179 ([73])**.**
Soit un anneau A. Deux polynômes P et Q dans $A[X]$ sont dits ***étrangers*** *entre eux si 1 est un pgcd*[60] *de P et Q. Un ensemble de polynômes $(P_i)_{i\in I}$ est étranger* ***dans leur ensemble*** *si 1 est un* pgcd *des P_i.*

Les polynômes P et Q sont ***premiers entre eux*** *si les seuls diviseurs communs de P et Q sont les inversibles.*

Les notions de polynômes étrangers entre eux ou de polynômes premiers entre eux ne sont pas identiques, comme le montre l'exemple suivant.

Exemple 3.180 ([1])**.**
Soient dans $\mathbb{Z}[X]$ les polynômes $P(X) = 2X + 2$ et $Q(X) = 2X^2 + 2$. Le nombre 2 est diviseur commun et n'est pas un diviseur de 1. Donc 1 n'est pas un pgcd de P et Q. Ils ne sont pas étrangers.

Mais ils sont premiers entre eux parce qu'ils n'ont pas d'autres diviseurs communs que les inversibles (1 et -1). △

3.14.5 Polynôme primitif

Définition 3.181.
Le ***contenu*** *du polynôme $P = \sum_i a_i X^i \in \mathbb{K}[X]$ est le pgcd de ses coefficients : $c(P) = \mathrm{pgcd}(a_i)$.*

Définition 3.182 (Ordre d'un polynôme)**.**
*Soit P un polynôme irréductible de degré n sur $\mathbb{F}_p[X]$. L'****ordre*** *de P est*

$$\min\{k \text{ tel que } P \mid X^k - 1\}. \tag{3.212}$$

Définition 3.183 (Polynôme primitif)**.**
Soit p, un nombre premier et P un polynôme de degré n dans $\mathbb{F}_p[X]$. Nous disons que P est ***primitif*** *si*

(1) P est unitaire et irréductible,

(2) les racines de P sont d'ordre $p^n - 1$ dans $\mathbb{F}_p[X]/P$.

58. Ceci se rapproche tout naturellement des notions de divisibilité dans un anneau intègre général, vues en sous-section 3.9.2.
59. Définition 1.97.
60. Définition 1.81.

Définition 3.184 (Polynôme primitif au sens du pgcd)**.**
Soit un anneau A. Un polynôme $P \in A[X]$ est ***primitif au sens du pgcd*** *si ses coefficients sont premiers entre eux.*

3.185.
Pour rappel, il y a plusieurs façons de périphraser le fait que les coefficients soient premiers entre eux. Nous pouvons dire ...

(1) Le pgcd de ses coefficients est 1 parce que c'est la définition 3.130 d'avoir des nombres premiers entre eux.
(2) Le contenu de ses coefficients est 1. Parce que le contenu est précisément le pgcd, définition 3.181.

La notion de polynôme primitif au sens du pgcd est particulière aux polynômes à coefficients dans un anneau comme le montre le lemme suivant.

Lemme 3.186.
Si $\mathbb{K}$ est un corps, tout polynôme unitaire dans $\mathbb{K}[X]$ non nul est primitif au sens du pgcd.

Démonstration. Un polynôme unitaire a un 1 parmi ses coefficients, donc le pgcd est forcément 1. □

Lorsque nous utiliserons la notion de polynôme primitif au sens du pgcd, nous le mentionnerons explicitement. C'est pas exemple le cas pour le corolaire 3.196.

3.14.6 Racines des polynômes

Définition 3.187.
Soient A un anneau et $P \in A[X]$. On appelle ***racine*** *un élément $\alpha \in A$ tel que $P(\alpha) = 0$; c'est-à-dire que, en remplaçant toutes les occurrences de X par α dans l'expression de P, on obtient 0.*

Proposition 3.188.
Soient A un anneau et P un polynôme non nul dans $A[X]$. Si $\alpha \in A$ est une racine de P alors $X - \alpha$ divise P, et réciproquement.

Démonstration. Nous notons le polynôme $\mu = X - \alpha$ par analogie avec le polynôme minimal dont il sera question dans la très semblable proposition 6.85. Le sens réciproque est clair : si μ divise P, alors α est racine de P.

Pour le sens direct, remarquons que si α est racine de P, alors P est de degré au moins égal à 1, et nous pouvons donc effectuer la division euclidienne [61] de P par μ : il existe des polynômes Q et R tels que

$$P = Q\mu + R \tag{3.213}$$

avec $\deg(R) < \deg(\mu)$. Donc R est une constante, élément de A : appelons-le a. En évaluant (3.213) en α, il vient

$$0 = P(\alpha) = Q(\alpha)\mu(\alpha) + a, \tag{3.214}$$

et nous en déduisons que $a = 0$, ce qui montre que $P = Q\mu$ et que μ divise P. □

Définition 3.189 (Racine simple et multiple d'un polynôme)**.**
Soit A un anneau ainsi qu'un polynôme $P \in A[X]$ et $\alpha \in A$ racine de P. La ***multiplicité*** *de α par rapport à P est l'entier h tel que P est divisible par $(X - \alpha)^h$ mais pas divisible par $(X - \alpha)^{h+1}$. Nous noterons $\theta_\alpha(P)$ la multiplicité de α par rapport à P.*

3.190.
Pour une définition générale d'une racine simple de l'équation $f(x) = 0$, voir la définition 33.49. La proposition 3.188 nous indique que toute racine est de multiplicité au moins 1.

61. Théorème 3.175.

Proposition 3.191.
Soient un anneau A, un polynôme $P \in \mathbb{A}[X]$ de degré n, ainsi qu'une racine $\alpha \in A$ de P. Alors il existe un polynôme $Q \in \mathbb{A}[X]$ de degré $n-1$ tel que $P = (X-\alpha)Q$.

Proposition 3.192.
L'élément $\alpha \in A$ est une racine de multiplicité h du polynôme P si et seulement s'il existe $Q \in A[X]$ tel que $P = (X-\alpha)^h Q$ avec $Q(\alpha) \neq 0$.

Lemme 3.193.
Soient P et Q des polynômes non nuls de $A[X]$ et $\alpha \in A$. Alors

(1) $\theta_\alpha(P+Q) \leqslant \min\{\theta_\alpha(P), \theta_\alpha(Q)\}$, et l'égalité a lieu si $\theta_\alpha(P) \neq \theta_\alpha(Q)$;

(2) $\theta_\alpha(PQ) \geqslant \theta_{\alpha(P)} + \theta_\alpha(Q)$, et l'égalité a lieu si A est intègre.

Dans le théorème suivant, la partie importante en pratique est la seconde partie parce qu'elle dit que, lorsque nous cherchons les racines d'un polynôme, nous pouvons nous arrêter lorsque nous en avons trouvé autant que le degré, multiplicité comprise.

Théorème 3.194.
Soit A un anneau intègre et $P \in A[X]\backslash\{0\}$, un polynôme de degré n.

(1) Si $\alpha_1, \dots, \alpha_p \in A$ sont des racines deux à deux distinctes de multiplicités $k_1, \dots, k_p$, alors il existe $Q \in A[X]$, de degré $n-p$, tel que $P = Q\prod_{i=1}^p (X-\alpha_i)^{k_i}$ et $Q(\alpha_i) \neq 0$ pour tout i.

(2) La somme des multiplicités des racines de P est au plus $\deg(P)$.

Démonstration. Si $p = 1$, soit α une racine de multiplicité k de P. La définition de la multiplicité d'une racine nous dit que P est divisible par $(X-\alpha)^k$ mais pas par $(X-\alpha)^{k+1}$. Donc il existe $Q \in \mathbb{A}[X]$ tel que $P = Q(X-\alpha)^k$. Il reste à voir que $Q(\alpha) \neq 0$. Cela est une conséquence de la proposition 3.188 : si $Q(\alpha)$ était nul, on pourrait lui factoriser $(X-\alpha)$ et donc avoir $(X-\alpha)^{k+1}$ qui se factorise dans P, ce qui n'est pas possible.

Nous supposons que $p \geqslant 2$ et nous effectuons une récurrence sur p. Nous considérons donc les $p-1$ premières racines $\alpha_1, \dots, \alpha_{p-1}$ et un polynôme $R \in \mathbb{A}[X]$ tel que $R(\alpha_i) \neq 0$ pour $i = 1, \dots, p-1$ et

$$P = \underbrace{(X-\alpha_1)^{k_1} \dots (X-\alpha_{p-1})^{k_{p-1}}}_{S} R. \tag{3.215}$$

Par hypothèse $P(\alpha_p) = S(\alpha_p)R(\alpha_p) = 0$. L'anneau $\mathbb{A}$ étant intègre, $S(\alpha_p) \neq 0$ parce que $\alpha_i \neq \alpha_p$ pour $i \neq p$. Par conséquent, $R(\alpha_p) = 0$.

Nous devons encore vérifier que la multiplicité α_p est k_p par rapport à R. Pour cela nous utilisons le point (2) du lemme 3.193 afin de dire que le degré de α_p pour $P = SR$ est k_p. Par conséquent

$$R = (X-\alpha_p)^{k_p} T \tag{3.216}$$

avec $T(\alpha_p) \neq 0$ et enfin

$$P = \prod_{i=1}^{p} (X-\alpha_i) T. \tag{3.217}$$

De plus $T(\alpha_i) \neq 0$, sinon $R(\alpha_i)$ serait nul. □

Corolaire 3.195.
Un polynôme de degré n possède au maximum n racines distinctes.

Démonstration. Le théorème 3.194(2) dit que la somme des multiplicités des racines de P est au maximum n. Mais la proposition 3.188 dit que toutes les racines ont une multiplicité au moins un. Donc il ne peut pas y en avoir plus de n. □

Corolaire 3.196 (Conséquence du lemme de Gauss[74])**.**
Soient A un anneau factoriel et $\mathrm{Frac}(A)$ *son corps des fractions. Un polynôme non constant $P \in A[X]$ est irréductible (sur A) si et seulement s'il est irréductible et primitif au sens du pgcd*[62] *sur* $\mathrm{Frac}(A)[X]$.

Exemple 3.197.
Il ne faudrait pas croire qu'être irréductible dans un anneau A implique d'être irréductible dans le corps des fractions. En effet soit $A = \mathbb{Z}[\sqrt{5}]$ et $P = X^2 - X - 1$. Nous savons que sa factorisation est

$$P = \left(X - \frac{1+\sqrt{5}}{2}\right)\left(X - \frac{1-\sqrt{5}}{2}\right). \tag{3.218}$$

Si vous ne le saviez pas, faites juste le calcul pour vous en assurer.

Ce polynôme est irréductible sur $\mathbb{Z}[\sqrt{5}]$ mais pas irréductible sur $\mathrm{Frac}\big(\mathbb{Z}[\sqrt{5}]\big)$. △

3.14.7 Quelques identités

Lemme 3.198.
Quelques identités de polynômes.

(1) Si n est impair, alors $1 + X$ divise $1 + X^n$.

(2) Pour tout n nous avons $X^n - 1 = (X - 1)(1 + X + \cdots + X^{n-1})$.

(3) $X^n - a^n = (X - a)\sum_{i=0}^{n-1} a^i X^{n-1-i}$.

Démonstration. Nous démontrons uniquement le point (2), puisque les autres ont été vus en début de chapitre[63]. Le cas $n = 1$ est évident. Procédons alors par récurrence en considérant un nombre entier impair n :

$$\begin{aligned}
1 + X^{n+2} &= 1 + X^n + X^{n+2} - X^n && (3.219a)\\
&= (1+X)P + X^n(X^2 - 1) && (3.219b)\\
&= (1+X)P + X^n(X+1)(X-1) && (3.219c)\\
&= (1+X)\big(P + X^n(X-1)\big). && (3.219d)
\end{aligned}$$

□

62. Définition 3.184.
63. Voir l'égalité (3.1).

Chapitre 4

Espaces vectoriels (début)

4.1 Parties libres, génératrices, bases et dimension

Nous avons déjà défini (dans 3.69) un espace vectoriel comme étant un module sur un corps commutatif. En explicitant un peu, cela donne ceci[75].

Un espace vectoriel sur le corps $\mathbb{K}$ est un ensemble V muni de deux opérations :

— une loi de composition interne $+ \colon V \times V \to V$,

— une loi de composition externe $\cdot \colon \mathbb{K} \times V \to V$

telles que

(1) $(V, +)$ soit un groupe abélien,

(2) pour tout $u, v \in V$ et pour tout $k, k' \in \mathbb{K}$,

$$k(u+v) = (ku) + (kv) \tag{4.1a}$$
$$(kk')u = k(k'u) \tag{4.1b}$$
$$(k+k')u = (ku) + (k'u) \tag{4.1c}$$
$$1u = u \tag{4.1d}$$

où 1 est le neutre de $\mathbb{K}$ et où nous avons directement adopté la notation ku pour $k \cdot u$.

Si $u \in V$, nous notons $-u$ l'inverse de u dans le groupe $(V, +)$.

Définition 4.1 (Partie libre)**.**
Si E est un espace vectoriel, une partie A de E est ***libre*** *si pour tout choix d'un nombre fini d'éléments $\{u_i\}_{i=1,\dots,n}$, l'égalité*

$$a_1 u_1 + \cdots + a_n u_n = 0 \tag{4.2}$$

implique $a_i = 0$ pour tout i (ici les a_i sont dans le corps de base).

Une partie infinie est libre si toutes ses parties finies le sont.

Remarque 4.2.
Notons que le vecteur nul n'est dans aucune partie libre, ne fût-ce que parce que $a0 = 0$ n'implique pas $a = 0$.

Si A est une partie de l'espace vectoriel E nous notons $\mathrm{Span}(A)$ l'ensemble des combinaisons linéaires finies d'éléments de A. Les coefficients de ces combinaisons linéaires sont dans le corps de base $\mathbb{K}$.

Définition 4.3 (Partie génératrice)**.**
Une partie B d'un espace vectoriel E est ***génératrice*** *si* $\mathrm{Span}(B) = E$.

Remarque 4.4.
Ces définitions demandent des commentaires en dimension infinie[1].

1. Nous n'avons pas encore défini le concept de dimension, mais nous nous adressons au lecteur trop pressé.

(1) Tout élément peut être écrit comme combinaison linéaire finie d'une partie génératrice. Cela ne signifie pas que nous pouvons extraire une partie finie qui convient pour tous les éléments à la fois. Lorsque l'espace est de dimension infinie, ceci est particulièrement important.
(2) La définition séparée de liberté dans le cas des parties infinies a son importance lorsqu'on parle d'espaces vectoriels de dimension infinies (en dimension finie, aucune partie infinie n'est libre) parce que cela fera une différence entre une base algébrique et une base hilbertienne par exemple.

Définition 4.5 (Base)**.**
*Une **base** de l'espace vectoriel E est une partie à la fois génératrice et libre.*

Nous prouvons à présent que tout élément non nul d'un espace vectoriel possédant une base [2] se décompose de façon unique en combinaison linéaire finie d'éléments d'une base.

Proposition 4.6 ([1])**.**
Soient un espace vectoriel E sur $\mathbb{K}$ muni d'une base $\{e_i\}_{i\in I}$. Si $v \in E$, alors il existe un unique couple (J, c) où
(1) J est une partie finie de I ;
(2) $c\colon J \to \mathbb{K}$ est une application
(3) $v = \sum_{j\in J} c(j)e_j$.
Les éléments $c(i)$ seront notés v_i et sont les composantes de v dans la base. Sous-entendu, on prolonge c de J vers I par zéro sur $I\backslash J$.

Démonstration. Soit un espace vectoriel E et une base $\{e_i\}_{i\in I}$ où I est un ensemble à priori quelconque. Soit $v \in E$. Vu que $E = \text{Span}\{e_i\}_{i\in I}$, il existe une partie finie J de I et des coefficients $\{v_j\}_{j\in J}$ dans $\mathbb{K}$ tels que

$$v = \sum_{j\in J} v_j e_j. \tag{4.3}$$

Cela donne l'existence.

En ce qui concerne l'unicité, soient J et K des parties finies de I et des coefficients $\{v_j\}_{j\in J}$ et $\{w_k\}_{k\in K}$ tels que

$$v = \sum_{j\in J} v_j e_j = \sum_{k\in K} w_k e_k. \tag{4.4}$$

Nous posons $L = J \cup K$. Remarquez que les unions suivantes sont des unions disjointes :

$$L = (J\backslash K) \cup (J \cap K) \cup (K\backslash J) \tag{4.5a}$$
$$J = (J\backslash K) \cup (J \cap K) \tag{4.5b}$$
$$K = (K \cap J) \cup (K\backslash J). \tag{4.5c}$$

Écrivons $0 = v - v$ en utilisant les expressions de v de (4.4) et en décomposant les sommes :

$$0 = \sum_{j\in J} v_j e_j - \sum_{k\in K} w_k e_k \tag{4.6a}$$
$$= \sum_{j\in J\backslash K} v_j e_j + \sum_{j\in J\cap K} v_j e_j - \sum_{k\in K\cap J} w_k e_k - \sum_{k\in K\backslash J} w_k e_k \tag{4.6b}$$
$$= \sum_{j\in J\backslash K} v_j e_j + \sum_{j\in J\cap K} (v_j - w_j) e_j - \sum_{k\in K\backslash J} w_k e_k. \tag{4.6c}$$

Là-dessus, nous posons

$$\alpha_l = \begin{cases} v_l & \text{si } l \in J\backslash K \\ v_l - w_l & \text{si } l \in J \cap K \\ -w_l & \text{si } l \in K\backslash J. \end{cases} \tag{4.7}$$

2. Nous n'avons pas démontré que tout espace vectoriel possède une base. Donc à notre niveau, il est possible que ce théorème soit sans objet pour beaucoup d'espaces.

Nous avons alors

$$\sum_{l\in L}\alpha_l e_l = 0. \tag{4.8}$$

Vu que $\{e_i\}_{i\in I}$ est libre, la partie $\{e_l\}_{l\in L}$ est également libre. Donc l'équation (4.8) implique que que $\alpha_l = 0$ pour tout $l \in L$.

Nous devons prouver que

$$\{j \in J \text{ tel que } v_j \neq 0\} = \{k \in K \text{ tel que } w_k \neq 0\} \tag{4.9}$$

et que pour l dans cet ensemble, $v_l = w_l$.

Soit $j \in J$; il y a deux possibilités : soit $j \in J\backslash K$, soit $j \in J \cap K$. Dans le premier cas nous avons déjà vu que $\alpha_j = v_j = 0$. Dans le second cas, $\alpha_j = v_j - w_j = 0$, c'est-à-dire $v_j = w_j$.

Donc $j \in J$ vérifiant $v_j \neq 0$ implique $j \in J \cap K$ et l'égalité des coefficients. Idem avec $k \in K$ tel que $w_k \neq 0$ implique $k \in J \cap K$. □

Lemme 4.7 ([1])**.**
Soit un espace vectoriel admettant des bases. Un endomorphisme est une bijection si et seulement si il change toute base en une base.

Démonstration. En deux parties. Soit un espace vectoriel E possédant des bases et un endomorphisme $f\colon E \to E$.

Si f est bijective Soit une base $\{v_i\}_{i\in I}$; nous devons voir que $\{f(v_i)\}_{i\in I}$ est une base.

Libre Si J est une partie finie de I et si les λæ sont des scalaires tels que $\sum_{j\in J}\lambda_j f(v_j) = 0$, alors

$$0 = \sum_{j\in J}\lambda_j f(v_j) = f\big(\sum_{j\in J}\lambda_j v_j\big). \tag{4.10}$$

Mais comme f est bijective, cela implique que $\sum_{j\in J}\lambda_j v_j = 0$. En retour, parce que $\{v_i\}$ est une base, cela implique que $\lambda_j = 0$ pour tout j.

Générateur Soit $x \in E$. Vu que f est bijective, il existe un unique $y \in E$ tel que $x = f(y)$. Comme $\{v_i\}_{i\in I}$ est une base, il existe une partie finie $J \subset I$ et des scalaires $\{\lambda_j\}_{j\in J}$ tels que

$$y = \sum_{j\in J}\lambda_j v_j. \tag{4.11}$$

Nous avons alors

$$x = f(y) = \sum_{j\in J}\lambda_j f(v_j), \tag{4.12}$$

qui montre que $\{f(v_i)\}_{i\in I}$ est bien génératrice de E

Si f change les bases en bases Soit un endomorphisme changeant toute base en une base. Nous devons prouver qu'il est bijectif.

Injective Nous considérons une base $\{v_i\}_{i\in I}$. La partie $\{f(v_i)\}_{i\in I}$ est par hypothèse également une base.

Soient $x, y \in E$ tels que $f(x) = f(y)$. Il existe J et K finis dans I qui permettent de décomposer x et y respectivement dans la base $\{f(v_i)\}_{i\in I}$. Quitte à poser $J' = J \cup K$, nous supposons que J suffit[3]. Il existe donc des scalaires $\{\lambda_j\}_{j\in J}$ et $\{\mu_j\}_{j\in J}$ tels que $x = \sum_{j\in J}\lambda_j f(v_j)$ et $y = \sum_{j\in J}\mu_j f(v_j)$.

La relation $f(x) = f(y)$ donne immédiatement, par la linéarité de f,

$$\sum_{j\in J}(\lambda_j - \mu_j)f(v_j) = 0. \tag{4.13}$$

Du fait que $\{f(v_i)\}_{i\in I}$ soit une base, nous déduisons que $\lambda_j - \mu_j = 0$ pour tout j. Donc $x = y$, et f est injective.

3. Nous utilisons le fait que l'union de deux parties finies d'une ensemble est finie (lemme 1.30).

Surjective Soit $x \in E$. Vu que $\{f(v_i)\}_{i\in I}$ est une base, il existe des scalaires λ_j tels que

$$x = \sum_{j\in J} \lambda_j f(v_j) = f\big(\sum_{j\in J} \lambda_j v_j\big). \tag{4.14}$$

Donc f est surjective.

□

Définition 4.8.
*Un espace vectoriel est **de type fini** s'il contient une partie génératrice finie.*

Nous verrons dans les résultats qui suivent que cette définition est en réalité inutile parce qu'une espace vectoriel sera de type fini si et seulement s'il est de dimension finie.

Lemme 4.9.
Si E a une famille génératrice de cardinal n, alors toute famille de $n+1$ éléments est liée.

Démonstration. Nous procédons par récurrence sur n. Pour $n = 1$, nous avons $E = \mathrm{Span}(e)$ et donc si $v_1, v_2 \in E$ nous avons $v_1 = \lambda_1 e$, $v_2 = \lambda_2 e$ pour certains éléments non nuls λ_1, λ_2 du corps de base. Nous avons donc $\lambda_2 v_1 - \lambda_1 v_1 = 0$. Cela prouve que $\{v_1, v_2\}$ est liée.

Supposons maintenant que le résultat soit vrai pour $k < n$, c'est-à-dire que pour tout espace vectoriel contenant une partie génératrice de cardinal $k < n$, les parties de $k+1$ éléments sont liées. Soit maintenant un espace vectoriel muni d'une partie génératrice $G = \{e_1, \dots, e_n\}$ de n éléments, et montrons que toute partie $V = \{v_1, \dots, v_{n+1}\}$ contenant $n+1$ éléments est liée. Dans nos notations nous supposons que les e_i sont des vecteurs distincts et les v_i également. Nous les supposons également tous non nuls. Étant donné que $\{e_i\}$ est génératrice nous pouvons définir les nombres λ_{ij} par

$$v_i = \sum_{k=1}^{n} \lambda_{ij} e_j \tag{4.15}$$

Vu que

$$v_{n+1} = \sum_{k=1}^{n} \lambda_{n+1,k} e_k \neq 0, \tag{4.16}$$

quitte à changer la numérotation des e_i nous pouvons supposer que $\lambda_{n+1,n} \neq 0$. Nous considérons les vecteurs

$$w_i = \lambda_{n+1,n} v_i - \lambda_{i,n} v_{n+1}. \tag{4.17}$$

En calculant un peu,

$$w_i = \lambda_{n+1,n} \sum_k \lambda_{i,k} e_k - \lambda_{i,n} \sum_k \lambda_{n+1,k} e_k \tag{4.18a}$$

$$= \sum_{k=1}^{n-1} \big(\lambda_{n+1,n}\lambda_{i,k} - \lambda_{i,n}\lambda_{n+1,}\big) e_k \tag{4.18b}$$

parce que les termes en e_n se sont simplifiés. Donc la famille $\{w_1, \dots, w_n\}$ est une famille de n vecteurs dans l'espace vectoriel $\mathrm{Span}\{e_1, \dots, e_{n-1}\}$; elle est donc liée par l'hypothèse de récurrence. Il existe donc des nombres $\alpha_1, \dots, \alpha_n \in \mathbb{K}$ non tous nuls tels que

$$0 = \sum_{i=1}^{n} \alpha_i w_i = \sum_{i=1}^{n} \alpha_i \lambda_{n+1,n} v_i - \left(\sum_{i=1}^{n} \alpha_i \lambda_{i,n}\right) v_{n+1}. \tag{4.19}$$

Vu que $\lambda_{n+1,n} \neq 0$ et que parmi les α_i au moins un est non nul, nous avons au moins un des produits $\alpha_i \lambda_{n+1,n}$ qui est non nul. Par conséquent (4.19) est une combinaison linéaire nulle non triviale des vecteurs de $\{v_1, \dots, v_{n+1}\}$. Cette partie est donc liée. □

Lemme 4.10.
Soit L une partie libre et G une partie génératrice. Si l'ensemble des parties libres L' telles que $L \subset L' \subset G$ possède un élément maximum[4]*, alors cet élément est une base.*

Qu'entend-on par « maximale » ? La partie B doit être libre, contenir L, être contenue dans G et de plus avoir la propriété que $\forall x \in G \backslash B$, la partie $B \cup \{x\}$ est liée.

Démonstration. D'abord si G est une base, alors toutes les parties de G sont libres et le maximum est $B = G$. Dans ce cas le résultat est évident. Nous supposons donc que G est liée.

La partie $B = \{b_1, \ldots, b_l\}$ est libre parce qu'on l'a prise parmi les libres. Montrons que B est génératrice. Soit $x \in G \backslash B$; par hypothèse de maximalité, $B \cup \{x\}$ est liée, c'est-à-dire qu'il existe des nombres λ_i, λ_x non tous nuls tels que

$$\sum_{i=1}^{l} \lambda_i b_i + \lambda_x x = 0. \tag{4.20}$$

Si $\lambda_x = 0$ alors un de λ_i doit être non nul et l'équation (4.20) devient une combinaison linéaire nulle non triviale des b_i, ce qui est impossible parce que B est libre. Donc $\lambda_x \neq 0$ et

$$x = \frac{1}{\lambda_x} \sum_{i=1}^{l} \lambda_i b_i. \tag{4.21}$$

Donc tous les éléments de $G \backslash B$ sont des combinaisons linéaires des éléments de B, et par conséquent, G étant génératrice, tous les éléments de E sont combinaisons linéaires d'éléments de B. □

Théorème 4.11 (Théorème de la base incomplète)**.**
Soit E un espace vectoriel de type fini sur le corps $\mathbb{K}$.

(1) Si L est une partie libre et si G est une partie génératrice contenant L, alors il existe une base B telle que $L \subset B \subset G$.

(2) Toute partie libre peut être étendue en une base.

(3) Toutes les bases sont finies et ont même cardinal.

(4) Si V est un sous-espace vectoriel de E, et si L est une base de V, alors il existe une base E qui contient L.

Démonstration. Point par point.

(1) Vu que E est de type fini, il admet une partie génératrice G de cardinal fini n. Donc une partie libre est de cardinal au plus n par le lemme 4.9. Soit L, une partie libre contenue dans G (ça existe : par exemple $L = \varnothing$). La partie B maximalement libre contenue dans G et contenant L est une base par le lemme 4.10.

(2) Notons que puisque E lui-même est générateur, le point (1) implique que toute partie libre peut être étendue en une base.

(3) Soient B et B', deux bases. En particulier B est génératrice et B' est libre, donc le lemme 4.9 indique que $\mathrm{Card}(B') \leqslant \mathrm{Card}(B)$. Par symétrie on a l'inégalité inverse. Donc $\mathrm{Card}(B) = \mathrm{Card}(B')$.

(4) La partie L étant une base de V, elle est en particulier libre dans E. Par le point (2), L peut être étendue en une base.

□

Remarque 4.12.
Le théorème de la base incomplète 4.11(2) est ce qui permet de construire une base d'une espace vectoriel en « commençant par » une base d'un sous-espace. En effet si H est un sous-espace de E alors une base de H est une partie libre de E et donc peut être étendue en une base de E.

4. Encore une fois, à part quelques cas triviaux, il n'est pas clair à ce point que ce maximum existe.

Définition 4.13.
*La **dimension** d'un espace vectoriel de type fini est le cardinal*[5] *d'une*[6] *de ses bases.*

Il existe une infinité de bases de $\mathbb{R}^m$. On peut démontrer que le cardinal de toute base de $\mathbb{R}^m$ est m, c'est-à-dire que toute base de $\mathbb{R}^m$ possède exactement m éléments.

Exemple 4.14.
La base de **canonique** de $\mathbb{R}^m$ est la partie $\{e_1, \dots, e_m\}$, où le vecteur e_j est

$$e_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{array}{l} \\ \\ \\ \leftarrow \text{j-ème} \\ \\ \\ \\ \end{array} .$$

La composante numéro j de e_i est 1 si $i = j$ et 0 si $i \neq j$. Cela s'écrit $(e_i)_j = \delta_{ij}$ où δ est le **symbole de Kronecker** défini par

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} \tag{4.22}$$

Les éléments de la base canonique de $\mathbb{R}^m$ peuvent donc être écrits $e_i = \sum_{k=1}^m \delta_{ik} e_k$. △

Le théorème suivant est essentiellement une reformulation du théorème 4.11.

Théorème 4.15.
Soit E un espace vectoriel de dimension finie et $\{e_i\}_{i\in I}$ une partie génératrice de E.

(1) Il existe $J \subset I$ tel que $\{e_i\}_{i\in J}$ est une base. Autrement dit : de toute partie génératrice nous pouvons extraire une base.

(2) Soit $\{f_1, \dots, f_l\}$ une partie libre. Alors nous pouvons la compléter en utilisant des éléments e_i. C'est-à-dire qu'il existe $J \subset I$ tel que $\{f_k\} \cup \{e_i\}_{i\in J}$ soit une base.

Proposition 4.16.
Si E est un espace vectoriel de dimension finie n, alors

(1) toute partie contenant $n+1$ éléments est liée.

(2) toute partie libre contenant n éléments est une base,

(3) toute partie génératrice contenant n éléments est une base.

Démonstration. Soit une partie M contenant $n+1$ éléments. L'espace E possède une partie génératrice contenant n éléments (n'importe quelle base). Donc M est liée par le lemme 4.9.

Une partie libre contenant n éléments peut être étendue en une base ; si ladite extension est non triviale (c'est-à-dire qu'on ajoute vraiment au moins un élément) une telle base contiendra une partie de $n+1$ éléments qui serait liée par le lemme 4.9.

Pour l'autre assertion, soit une partie génératrice $\{v_i\}_{i\in I}$ où I contient n éléments. Par le théorème 4.15(2) il existe $J \subset I$ tel que $\{v_j\}_{j\in J}$ soit une base. Si l'inclusion $J \subset I$ est stricte, alors la base $\{v_j\}_{j\in J}$ contiendrait moins de$n$ éléments, ce qui serait en contradiction avec le théorème 4.11(3). □

Définition 4.17.
*Soit F un sous-espace vectoriel de l'espace vectoriel E. La **codimension** de F dans E est*

$$\operatorname{codim}_E(F) = \dim(E/F). \tag{4.23}$$

5. Définition 1.32.
6. Le théorème de la base incomplète 4.11(3) montre que cette définition ne souffre d'aucune ambiguïté.

Problèmes et choses à faire

Voir que E/F a une structure vectorielle, expliciter sa dimension en fonction de celles de E et F.

4.1.1 Et en dimension infinie

Dans ZFC, en dimension infinie, il existe aussi une base pour tout espace vectoriel ainsi qu'un théorème de la base incomplète. Nous ne parlerons pas de ce qu'il se passe lorsque nous ne considérons que ZF [7].

Lemme 4.18 ([76]).
Soient un $\mathbb{K}$-espace vectoriel E et un sous-espace vectoriel V de E. Soient encore deux sous-espaces vectoriels W_1 et W_2 tels que

(1) $V \cap W_1 = \{0\}$;

(2) $V + W_2 = E$.

Alors il existe un supplémentaire W de V tel que $W_1 \subset W \subset W_2$.

Juste une remarque : dans le Frido le symbole « $\subset$ » ne signifie pas une inclusion stricte.

Démonstration. Nous divisons en petits morceaux.

Un gros ensemble Soit $\mathcal{A}$ l'ensemble des sous-espaces vectoriels S de E tels que $W_1 \subset S \subset W_2$ et $S \cap V = \{0\}$. Vu que $W_1 \subset \mathcal{A}$, cet ensemble n'est pas vide. De plus $\mathcal{A}$ est partiellement ordonné pour l'inclusion.

$\mathcal{A}$ est inductif Nous prouvons maintenant que $\mathcal{A}$ est inductif [8]. Pour cela, soit une partie $\mathcal{A}'$ totalement ordonnée et $U = \bigcup_{A \in \mathcal{A}'} A$.

Alors, la partie U est un sous-espace vectoriel de E. En effet si $x, y \in U$, alors il existe $A_1, A_2 \in \mathcal{A}'$ tels que $x \in A_1$ et $y \in A_2$. Vu que $\mathcal{A}'$ est totalement ordonné, l'un des ensembles parmi A_1 et A_2 est inclus dans l'autre. Sans perdre de généralité, disons $A_1 \subset A_2$. Alors les opérations s'effectuent dans A_2 : nous avons $x, y \in A_2$, et donc $\lambda x \in A_2 \subset U$ ainsi que $x + y \in A_2 \subset U$.

De plus, U contient W_1, et est contenu dans W_2. Ainsi, $U \in \mathcal{A}$ et majore $\mathcal{A}'$ pour l'inclusion. En bref, $\mathcal{A}$ est bien inductif.

Utilisation de Zorn Le lemme de Zorn 1.15 nous donne alors un maximum W de $\mathcal{A}$. Ce maximum vérifie

(1) $W \cap V = \{0\}$,

(2) $W_1 \subset W \subset W_2$,

(3) pour tout $W' \in \mathcal{A}$, nous avons $W' \subset W$ parce que W est maximum.

Supplémentaire Montrons que ce W est un supplémentaire de V. Soit $x \in E$. Le but est de trouver une décomposition de x en somme d'un élément de W et un de V. Vu que $V + W_2 = E$ nous avons $v \in V$ et $w_2 \in W_2$ tels que

$$x = v + w_2. \tag{4.24}$$

Si $w_2 \in W$ alors c'est fait. Sinon ...

Soit $X = \mathrm{Span}\{W, w_2\}$. Vu que X contient strictement W et que W est maximum dans $\mathcal{A}$, la partie X n'est pas un élément de $\mathcal{A}$. Vu que X est un sous-espace vectoriel de E tel que $W_1 \subset X \subset W_2$, la seule possibilité pour que X ne soit pas dans $\mathcal{A}$ est que $X \cap V \neq \{0\}$. Soit donc $y \neq 0$ dans $X \cap V$. Par définition de X,

$$y = w' + \lambda w_2 \tag{4.25}$$

7. Si vous ne savez pas ce que signifient les sigles « ZF » et « ZFC » vous ne devriez pas être en train de lire ceci, et encore moins penser à le resservir à un jury d'agrégation.

8. Définition 1.14.

pour $w' \in W$, $w_2 \in W_2$ et $\lambda \in \mathbb{K}$. Nous avons $\lambda \neq 0$, sinon nous aurions $y \in W \cap V$ et donc $y = 0$ puisque W est dans $\mathcal{A}$. La décomposition (4.25) permet alors d'écrire $w_2 = (y - w')/\lambda$ et finalement

$$x = v + \frac{1}{\lambda}(y - w') = \underbrace{v + \frac{1}{\lambda}y}_{\in V} - \underbrace{\frac{1}{\lambda}w'}_{\in W} . \tag{4.26}$$

La somme d'espaces vectoriels $E = V + W$ est donc établie.

□

Corolaire 4.19.
Tout sous-espace vectoriel d'un espace vectoriel possède un supplémentaire.

Démonstration. Soit un espace vectoriel E ainsi qu'un sous-espace vectoriel V. Si $V = E$ nous sommes ok. Sinon nous considérons $v \in E \backslash V$ et nous posons $W_1 = \mathbb{K}v$ et $W_2 = E$.

Vu que V et W_1 sont des espaces vectoriels, nous avons $V \cap W_1 = \{0\}$, et vu que $W_2 = E$ nous avons $V + W_2 = E$. Le lemme 4.18 nous donne alors un supplémentaire de V. □

Proposition 4.20 (Base incomplète)**.**
Tout espace vectoriel (non réduit à $\{0\}$) possède une base.

Démonstration. Soit $\mathcal{A}$ l'ensemble des familles libres de E. Il n'est pas vide parce que $\{v\}$ en est une dès que v est non nul dans E. Rapidement :

— l'ensemble $\mathcal{A}$ est ordonné pour l'inclusion,

— si $\mathcal{A}'$ est une partie totalement ordonnée, l'union est un majorant,

— donc $\mathcal{A}$ est inductif,

— soit un maximum F de $\mathcal{A}$.

La partie F est libre parce qu'elle est dans $\mathcal{A}$. Elle est génératrice parce que si v n'est pas dans $\mathrm{Span}(F)$ alors la partie $F \cup \{v\}$ est encore libre, et majore strictement F pour l'inclusion, ce qui n'est pas possible.

Donc F est une base de E. □

Théorème 4.21 (Base incomplète, dimension quelconque)**.**
Soit une partie $\{e_i\}_{i\in I}$ génératrice de l'espace vectoriel E (ici, I est un ensemble quelconque[9]). Soit $I_0 \in I$ tel que $\{e_i\}_{i\in I_0}$ soit libre.

Alors il existe I_1 tel que $I_0 \subset I_1 \subset I$ tel que $\{e_i\}_{i\in I_1}$ soit une base de E.

Note : une telle partie I_0 existe en prenant un singleton. Mais l'existence n'est pas le sujet ici.

Démonstration. Soit $\mathcal{A}$ l'ensemble des parties J de I telles que $I_0 \subset J \subset I$ et telles que $\{e_i\}_{i\in J}$ soit libre.

Encore une fois, $\mathcal{A}$ est inductif pour l'ordre partiel donné par l'inclusion. Soit J un maximum. Vu que $J \in \mathcal{A}$, la partie $\{e_i\}_{i\in J}$ est libre. Mais elle est également génératrice parce que si e_k n'est pas dedans, J ne serait pas maximum, étant majorée par $J \cup \{k\}$.

Donc $\{e_i\}_{i\in J}$ engendre tous les e_i avec $i \in I$ et donc tous les éléments de E. □

4.1.2 Espace librement engendré

Définition 4.22 ([77])**.**
Soient un ensemble S et un corps $\mathbb{K}$. L'espace vectoriel ***librement engendré*** *sur S, noté $F_{\mathbb{K}}(S)$ est l'ensemble des applications $S \to \mathbb{K}$ qui sont non-nulles en un nombre fini de points de S.*

Autrement dit, $\sigma\colon S \to \mathbb{K}$ est dans $F_{\mathbb{K}}(S)$ si $\{x \in S$ tel que $\delta(x) \neq 0\}$ est fini[10].

9. Un cas d'utilisation intéressant est de poser $I = E$ et $e_i = i$. Pensez-y.
10. Parce que nous l'aimons bien, nous ne résistons pas à faire un renvoi vers la définition 1.29.

Le lemme suivant donne tout son sens à l'expression « librement » engendré. Il dit que $F(S)$ possède une base indexée par S lui-même.

Lemme 4.23.
L'ensemble des applications δ_s données par

$$\begin{aligned} \delta_s \colon S &\to \mathbb{K} \\ t &\mapsto \begin{cases} 1 & si\ t = s \\ 0 & sinon \end{cases} \end{aligned} \tag{4.27}$$

avec $s \in S$ forment une base[11] *de $F(S)$.*

Démonstration. Pour prouver que les δ_s sont générateurs, nous considérons $g \colon S \to \mathbb{K}$ non nul sur la partie finie $\{s_i\}_{i\in I}$ de S. Alors nous avons

$$g = \sum_{i\in I} g(s_i)\delta_{s_i}. \tag{4.28}$$

Pour prouver que les δ_s forment une partie libre, nous supposons avoir $\lambda_i \in \mathbb{K}$ tels que

$$g = \sum_{i\in I} \lambda_i \delta_{s_i} = 0 \tag{4.29}$$

Soit $j \in I$. Nous avons

$$0 = f(s_j) = \sum_{i\in I} \lambda_i \underbrace{\delta_{s_i}(s_j)}_{=\delta_{ij}} = \lambda_j. \tag{4.30}$$

Donc les coefficients λ_i sont tous nuls, et nous avons prouvé que la partie est libre. □

Il est parfois pratique d'écrire les éléments de $F(S)$ comme sommes « formelles » d'éléments de S. Cela va encore lorsque S est un ensemble n'ayant aucune somme bien définie.

Mais attention : si $S = \mathbb{R}$, l'élément $4 + 7$ de $F(\mathbb{R})$ n'est pas 11. L'élément 11 de $F(\mathbb{R})$ est un élément complètement différent. Bref, il n'est pas judicieux d'écrire les éléments de $F(S)$ comme des combinaisons linéaires d'éléments de S. Pour $x \in S$ il vaut mieux écrire explicitement δ_x que x. La somme $\delta_x + \delta_y$ est parfaitement bien définie dans l'esemble des applications de S vers $\mathbb{K}$.

4.2 Applications linéaires

4.2.1 Définition

Définition 4.24.
Soient des espaces vectoriels E et F sur le corps $\mathbb{K}$. Une application $T \colon E \to F$ est dite ***linéaire*** *si*

- $T(x + y) = T(x) + T(y)$ *pour tout x et y dans E,*
- $T(\lambda x) = \lambda T(x)$ *pour tout λ dans $\mathbb{K}$ et x dans E.*

Si vous avez bien suivi, les égalités dans la définition 4.24 sont des égalités dans F.

Lemme-définition 4.25.
L'ensemble de toutes les applications linéaires de E vers F est noté $\mathcal{L}(E, F)$ et devient un espace vectoriel sur $\mathbb{K}$ avec les définitions suivantes :

(1) $(T_1 + T_2)(x) = T_1(x) + T_2(x)$,

(2) $(\lambda T)(x) = \lambda T(x)$.

Exemple 4.26.
Pour tout b dans $\mathbb{R}$ la fonction $T_b(x) = bx$ est une application linéaire de $\mathbb{R}$ dans $\mathbb{R}$. En effet,

11. Définition 4.5.

— $T_b(x+y) = b(x+y) = bx + by = T_b(x) + T_b(y)$,

— $T_b(ax) = b(ax) = abx = aT_b(x)$.

De la même façon on peut montrer que la fonction T_λ définie par $T_\lambda(x) = bx$ est un application linéaire de $\mathbb{R}^m$ dans $\mathbb{R}^m$ pour tout λ dans $\mathbb{R}$ et m dans $\mathbb{N}$. △

Exemple 4.27.
Soit $m = n$. On fixe λ dans $\mathbb{R}$ et v dans $\mathbb{R}^m$. L'application U_λ de $\mathbb{R}^m$ dans $\mathbb{R}^m$ définie par $U_\lambda(x) = \lambda x + v$ n'est pas une application linéaire lorsque $v \neq 0$, parce que si a est un réel différent de 0 et 1, alors $av \neq v$, d'où

$$U_\lambda(ax) = \lambda(ax) + v \neq a(\lambda x + v) = aU_\lambda(x).$$

△

Exemple 4.28.
Soit A une matrice fixée de $\mathcal{M}_{n\times m}$. La fonction $T_A\colon \mathbb{R}^m \to \mathbb{R}^n$ définie par $T_A(x) = Ax$ est une application linéaire. En effet,

— $T_A(x+y) = A(x+y) = Ax + Ay = T_A(x) + T_A(y)$,

— $T_A(ax) = A(ax) = a(Ax) = aT_A(x)$.

△

On peut observer que, si on identifie $\mathcal{M}_{1\times 1}$ et $\mathbb{R}$, on obtient le premier exemple comme cas particulier.

Définition 4.29 (Quelques ensembles d'applications linéaires)**.**
Soient E et F des espaces vectoriels.

— *L'ensemble des applications linéaires de E vers F est noté $\mathcal{L}(E,F)$, comme déjà dit en 4.25.*

— *Une application linéaire $E \to E$ est un **endomorphisme** de E. L'ensemble des endomorphismes de E est noté* $\mathrm{End}(E)$.

— *Un endomorphisme bijectif est un **automorphisme**. L'ensemble des automorphismes de E est noté* $\mathrm{Aut}(E)$.

— *Une application linéaire bijective $E \to F$ est un **isomorphisme** d'espace vectoriel. L'ensemble des isomorphismes $E \to F$ est noté*[12] $\mathrm{GL}(E,F)$.

Remarque 4.30.
Les ensembles définis en 4.29 concernent la structure d'espace vectoriel seulement. Lorsque nous verrons la notion d'espace vectoriel normé, nous demanderons de plus la continuité, laquelle n'est pas automatique en dimension infinie. Voir aussi les définitions 10.19.

Définition 4.31.
*Si E est un espace vectoriel, si X est un espace vectoriel, et si $f\colon X \to E$ est une application, le **noyau** de f est le noyau de f lorsque E est vu comme un groupe pour l'addition*[13]*, c'est-à-dire la partie*

$$\ker(f) = \{x \in X \text{ tel que } f(x) = 0\}. \tag{4.31}$$

Proposition 4.32.
Le noyau d'une application linéaire est un sous-espace vectoriel.

Démonstration. Soit une application linéaire $f\colon E \to F$. Si $x, y \in \ker(f)$ et si $\lambda \in \mathbb{K}$ alors

$$f(x+y) = f(x) + f(y) = 0 + 0 = 0, \tag{4.32}$$

12. Le fait d'utiliser une notation similaire à celle des matrices inversibles n'est pas anodine : le lecteur en est sans doute conscient.
13. Définition 2.27.

donc $x + y \in \ker(f)$ et

$$f(\lambda x) = \lambda f(x) = 0, \tag{4.33}$$

donc $\lambda x \in \ker(f)$. □

Proposition 4.33.
Si E et F sont des espaces vectoriels de dimension n et si $\{e_i\}_{i=1,\dots,n}$ et $\{f_i\}_{i=1,\dots,n}$ sont des bases respectivement de E et F, alors il existe une unique application linéaire $T\colon E \to F$ telle que $T(e_i) = f_i$ pour tout i.

Démonstration. En deux parties.

Existence Soit $v \in E$. Vu que $\{e_i\}$ est une base, il se décompose de façon unique en $v = \sum_i v_i e_i$. Alors définir

$$T(v) = \sum_i v_i f_i \tag{4.34}$$

est une bonne définition et satisfait aux exigences.

Unicité Soient T et U satisfaisant aux exigences. Alors pour tout i nous avons $T(e_i) = U(e_i)$. Si $v \in E$ s'écrit de la forme $v = \sum_i v_i e_i$ alors la linéarité impose $T(v) = \sum_i v_i T(e_i) = \sum_i v_i U(e_i) = U(v)$. Donc $T = U$.

□

Lemme 4.34 ([1]).
Soient des espaces vectoriels V et W de dimension finie. Soient des bases $\{e_i\}$ de V et $\{f_\alpha\}$ de W. Nous posons

$$\begin{aligned} \varphi_{i\alpha}\colon V &\to W \\ v &\mapsto v_i f_\alpha \end{aligned} \tag{4.35}$$

où v_i est défini par la décomposition (unique) $v = \sum_i v_i e_i$.
Alors :

(1) La partie $\{\varphi_{i\alpha}\}$ est une base de $\mathcal{L}(V, W)$.
(2) Au niveau des dimensions : $\dim\big(\mathcal{L}(V, W)\big) = \dim(V)\dim(W)$.

Démonstration. Il faut prouver que $\{\varphi_{i\alpha}\}$ est libre et générateur.

Générateur Soit une application linéaire $b\colon V \to W$. En décomposant $b(v)$ dans la base $\{f_\alpha\}$, nous définissons $b_\alpha\colon V \to \mathbb{K}$ par

$$b(v) = \sum_\alpha b_\alpha(v) f_\alpha. \tag{4.36}$$

Nous posons $b_{\alpha i} = b_\alpha(e_i)$. Ainsi,

$$b(v) = \sum_\alpha v_i b_{\alpha i} f_\alpha = \sum_{\alpha i} b_{\alpha i} \varphi_{i\alpha}(v). \tag{4.37}$$

Donc b peut être écrit comme combinaison linéaire des $\varphi_{i\alpha}$.

Libre Supposons que $\sum_{i\alpha} a_{i\alpha}\varphi_{i\alpha} = 0$ pour certains coefficients $a_{i\alpha} \in \mathbb{K}$. Nous avons, pour tout $v \in V$:

$$0 = \sum_{i\alpha} a_{i\alpha}\varphi_{i\alpha}(v) = \sum_{i\alpha} a_{i\alpha} v_i f_\alpha, \tag{4.38}$$

mais comme les f_α forment une base, chaque terme de la somme sur α est nul :

$$\sum_i a_{i\alpha} v_i = 0. \tag{4.39}$$

Et comme cela est valable pour tout v et donc pour tout choix de v_i, nous avons $a_{i\alpha} = 0$ pour tout i et pour tout α.

La formule de dimension est simplement la cardinalité de la base trouvée ; c'est la définition 4.13.

□

4.2.2 Linéarité et bases

Proposition 4.35 ([78]).
Soient deux espaces vectoriels E et F. Une application linéaire[14] $f\colon E \to F$ *est injective si et seulement si* $\ker\{f\} = \{0\}$.

Démonstration. Nous supposons que f est injective. Si $x \in \ker(f)$, alors $f(x) = 0$. Or f est linéaire, donc $f(0) = 0$. Nous avons donc $f(x) = f(0)$ et donc $x = 0$ parce que f est injective.

Dans l'autre sens, soient x, y tels que $f(x) = f(y)$. Par linéarité de f nous avons $f(x - y) = 0$, et donc $x - y = 0$ parce que $\ker(f) = \{0\}$. Donc $x = y$ et f est injective. □

Proposition 4.36 ([78]).
Soit $f \in \mathcal{L}(E, F)$ où E et F sont deux espaces vectoriels.

(1) Si f est injective et si $\{v_i\}_{i\in I}$ est libre, alors $\{f(v_i)\}_{i\in I}$ est libre.

(2) Si f est surjective et si $\{v_i\}_{i\in I}$ est génératrice, alors $\{f(v_i)\}_{i\in I}$ est génératrice.

(3) Si f est une bijection, alors l'image d'une base par f est une base.

Démonstration. En trois parties.

(1) Nous devons montrer que $\{f(v_j)\}_{j\in J}$ est libre pour tout J fini dans I. Soit donc une partie finie $J \in I$ et des scalaires[15] tels que $\sum_{j\in J} \lambda_j f(v_j) = 0$. La linéarité de f donne[16]

$$f\Big(\sum_{i\in J} \lambda_j v_j\Big) = 0. \tag{4.40}$$

Par injectivité de f nous avons alors $\sum_j \lambda_j v_j = 0$. Vu que les v_j eux-même forment une partie libre, nous avons $\lambda_j = 0$ pour tout $j \in J$.

(2) Soit $y \in F$. Vu que f est surjective, il existe $x \in E$ tel que $f(x) = y$. Étant donné que $\{v ß\}_{i\in I}$ est générateur, il existe une partie finie $J \subset I$ et des scalaires $\lambda_j \in \mathbb{K}$ tels que

$$x = \sum_{j\in J} \lambda_j v_j. \tag{4.41}$$

En appliquant f aux deux côtés, et en tenant compte de la linéarité de f,

$$y = f(x) = \sum_{j\in J} \lambda_j f(v_j), \tag{4.42}$$

ce qui prouve que y est une combinaison linéaire des $f(v_j)$.

(3) Une base est à la fois libre et génératrice et une bijection est à la fois injective et surjective. Les deux premiers points permettent de conclure.

□

Corolaire 4.37 ([1]).
Si E et F sont des espaces vectoriels isomorphes de dimensions finies. Alors leurs dimensions sont égales.

Démonstration. Vu que E et F sont isomorphes, il existe une bijection $f\colon E \to F$. Par la proposition 4.36(3), l'image d'une base de E est une base de F. Donc les espaces E et F ont des bases contenant le même nombre d'éléments. □

14. Définition 4.24.
15. Des éléments du corps de base $\mathbb{K}$.
16. Voir les propriétés de la définition 4.24.

4.2.3 Rang

La proposition 4.38 et le théorème 4.39 sont valables également en dimension infinie ; ce sera une des rares incursions en dimension infinie de ce chapitre.

Proposition-définition 4.38.
L'image d'une application linéaire est un espace vectoriel. La dimension de cet espace est le ***rang*** *de ladite application linéaire.*

Démonstration. Soit une application linéaire $f\colon E \to F$. Nous considérons v, w dans l'image de f ainsi que λ dans le corps de base commun à E et F.

Soient $v_0 \in E$ et $w_0 \in E$ tels que $v = f(v_0)$ et $w = f(w_0)$. Alors $v + w = f(v_0 + w_0)$ et $\lambda v = f(\lambda v_0)$. Donc l'image est bien un espace vectoriel. □

Théorème 4.39 (Théorème du rang)**.**
Soient E et F deux espaces vectoriels (de dimensions finies ou non) et soit $f\colon E \to F$ une application linéaire.

Si $(u_s)_{s\in S}$ est une base de $\ker(f)$ et si $\big(f(v_t)\big)_{t\in T}$ est une base de $\mathrm{Image}(f)$ alors

$$(u_s)_{s\in s} \cup (v_t)_{t\in T} \tag{4.43}$$

est une base de E.

En dimension finie, nous avons en plus la formule suivante :

$$\mathrm{rang}(f) + \dim(\ker f) = \dim E, \tag{4.44}$$

c'est-à-dire que le rang[17] *de f est égal à la codimension*[18] *du noyau.*

Démonstration. Nous devons montrer que

$$(u_s)_{s\in S} \cup (v_t)_{t\in T} \tag{4.45}$$

est libre et générateur.

Soit $x \in E$. Nous définissons les nombres x_t par la décomposition de $f(x)$ dans la base $\big(f(v_t)\big)$:

$$f(x) = \sum_{t\in T} x_t f(v_t). \tag{4.46}$$

Ensuite le vecteur $x = \sum_t x_t v_t$ est dans le noyau de f, par conséquent nous le décomposons dans la base (u_s) :

$$x - \sum_t x_t v_t = \sum_{s\in S} x_s u_s. \tag{4.47}$$

Par conséquent

$$x = \sum_s x_s u_s + \sum_t x_t v_t. \tag{4.48}$$

En ce qui concerne la liberté nous écrivons

$$\sum_t x_t v_t + \sum_s x_s u_s = 0. \tag{4.49}$$

En appliquant f nous trouvons que

$$\sum_t x_t f(v_t) = 0 \tag{4.50}$$

et donc que les x_t doivent être nuls. Nous restons avec $\sum_s x_s u_s = 0$ qui à son tour implique que $x_s = 0$. □

17. Définition 4.38.
18. Définition 4.17.

Un exemple d'utilisation de ce théorème en dimension infinie sera donné dans le cadre du théorème de Fréchet-Riesz, théorème 24.17.

Proposition 4.40 ([79]).
Soit E, un espace vectoriel de dimension finie sur le corps $\mathbb{K}$. Soient V et W des sous-espaces vectoriels de E. Alors

$$\dim(V+W) = \dim(V) + \dim(W) - \dim(V \cap W). \tag{4.51}$$

Démonstration. Nous considérons l'application

$$\begin{aligned} \varphi\colon V \times W &\to E \\ (x,y) &\mapsto x+y. \end{aligned} \tag{4.52}$$

C'est une application linéaire dont l'image est $V+W$. Nous avons donc, pour commencer

$$\dim(V+W) = \dim\big(\mathrm{Image}(\varphi)\big). \tag{4.53}$$

Nous appliquons à présent le théorème du rang 4.39 à l'application φ :

$$\begin{aligned} \dim(V+W) &= \dim\big(\mathrm{Image}(\varphi)\big) && (4.54a)\\ &= \dim(V\times W) - \dim\big(\ker(\varphi)\big) && (4.54b)\\ &= \dim(V) + \dim(W) - \dim\big(\ker(\varphi)\big). && (4.54c) \end{aligned}$$

Nous devons maintenant étudier $\ker(\varphi)$. D'abord, $(v,w) \in V\times W$ appartient à $\ker(\varphi)$ si et seulement si $v+w=0$. Nous avons donc

$$\ker(\varphi) = \{(x,-x) \text{ tel que } x \in V\cap W\}. \tag{4.55}$$

Nous montrons à partir de cela que $\dim\big(\ker(\varphi)\big) = \dim(V\cap W)$ en montrant que l'application

$$\begin{aligned} \psi\colon V\cap W &\to \ker(\varphi) \\ x &\mapsto (x,-x) \end{aligned} \tag{4.56}$$

est un isomorphisme d'espaces vectoriels. D'abord ψ est injective parce que si $\psi(x)=\psi(y)$, alors $(x,-x)=(y,-x)$ et donc $x=y$. Ensuite, ψ est surjective parce qu'un élément générique de $\ker(\varphi)$ est $(x,-x)=\psi(x)$ avec $x\in V\cap W$. L'application ψ étant un isomorphisme d'espaces vectoriels, nous avons bien $\dim\big(\ker(\varphi)\big) = \dim(V\cap W)$. □

Corolaire 4.41.
Soient deux espaces vectoriels E et F de même dimensions finies[19]*. Pour une application linéaire $f\colon E\to F$, les trois conditions suivantes sont équivalentes :*

(1) f est injective ;

(2) f est surjective ;

(3) f est bijective.

Démonstration. Si un endomorphisme $f\colon E\to E$ est surjectif, alors $\mathrm{rang}(f)=\dim(E)$, ce qui donne, par le théorème du rang 4.39, $\dim\big(\ker(f)\big)=0$, c'est-à-dire que f est injectif.

De la même façon, si f est injective, alors $\dim\big(\ker(f)\big)=0$, ce qui donne $\mathrm{rang}(f)=\dim(E)$ ou encore que f est surjective. □

Exemple 4.42.
Le corolaire 4.41 n'est pas correct en dimension infinie. Par exemple en prenant $f(e_1)=f(e_2)=e_1$ et ensuite $f(e_k)=e_{k-1}$ pour tout $k\geqslant 2$. Cette application est surjective mais pas injective. △

19. Les deux mots sont importants : les dimensions doivent être égales et finies.

Une conséquence du théorème du rang est que les endomorphismes ont un inverse à gauche et à droite égaux (lorsqu'ils existent). En résumé, ce que le corolaire 4.43 dit est que si $AB = \mathbb{1}$, alors $BA = \mathbb{1}$.

Corolaire 4.43.
Soit un endomorphisme f d'un espace vectoriel de dimension finie. Si f admet un inverse à gauche, alors

(1) f est bijective,

(2) f admet également un inverse à droite,

(3) les inverses à gauche et à droite sont égaux.

Tout cela tient également en remplaçant « gauche » par « droite ».

Démonstration. Soit g, un inverse à gauche de f : $gf = \mathrm{Id}$. Cela implique que f est injective et que g est surjective, et donc qu'elles sont toutes deux bijectives par le corolaire 4.41. Vu que f est bijective, elle admet également un inverse à droite, soit h. Nous avons : $gf = \mathrm{Id}$ et $fh = \mathrm{Id}$.

Alors $gfh = h$ parce que $gf = \mathrm{Id}$, mais également $gfh = g$ parce que $fh = \mathrm{Id}$. Donc $g = h$.[20] □

C'est ce corolaire qui nous permet d'écrire f^{-1} sans plus de précisions dès que f est une bijection.

Exemple 4.44 (Pas en dimension infinie)**.**
Tout cela ne fonctionne pas en dimension infinie. Par exemple avec une base $\{e_k\}_{k\in\mathbb{N}}$ nous pouvons considérer l'opérateur

$$f(e_k) = e_{k+1}. \tag{4.57}$$

Il est injectif, mais pas surjectif. Si on pose

$$g(e_k) = \begin{cases} e_{k-1} & \text{si } k \geqslant 1 \\ 0 & \text{si } k = 0 \end{cases} \tag{4.58}$$

alors nous avons $gf = \mathrm{Id}$, mais pas $fg = \mathrm{Id}$ parce que ce $(fg)(e_0) = 0$. △

Lemme 4.45.
Si E et F sont des espaces vectoriels et si $f\colon E \to F$ est une application linéaire inversible, alors son inverse est également linéaire.

Démonstration. Nous avons $f^{-1}(x+y) = f^{-1}(x) + f^{-1}(y)$. En effet,

$$f\big(f^{-1}(x) + f^{-1}(y)\big) = f\big(f^{-1}(x)\big) + f\big(f^{-1}(y)\big) = x + y. \tag{4.59}$$

De la même façon,

$$f\big(\lambda f^{-1}(x)\big) = \lambda x, \tag{4.60}$$

donc $f^{-1}(\lambda x) = \lambda f^{-1}(x)$. □

Proposition 4.46.
Soient un espace vectoriel E de dimension finie, un endomorphisme $f\colon E \to E$ et une partie $\{v_i\}_{i\in I}$ tel que $\{f(v_i)\}_{i\in I}$ soit une base.

Alors $\{v_i\}_{i\in I}$ est une base.

Démonstration. Soit $x \in E$. Il existe une partie finie $J \subset I$ et des scalaires λ_j tels que

$$x = \sum_j \lambda_j f(v_j) = f\big(\sum_j \lambda_j v_j\big), \tag{4.61}$$

20. C'est le même argument que celui employé pour la preuve du lemme 1.70 (2), à ceci près que nous devions montrer l'existence de l'inverse à droite.

ce qui prouve que f est surjective. Le corolaire 4.41 nous dit alors que f est une bijection. L'application inverse est également linéaire par le lemme 4.45.

Une application linéaire bijective (comme f^{-1}) transforme une base en une base par la proposition 4.36. Donc

$$f^{-1}\big(\{f(v_i)\}\big) \tag{4.62}$$

est une base. □

Proposition 4.47.
Soit un espace vectoriel E de dimension finie et deux applications linéaires $f, g\colon E \to E$ telles que $g \circ f = \mathrm{Id}$. Alors f et g sont bijectives.

Démonstration. En plusieurs étapes

f est injective Si $f(x) = f(y)$, alors en appliquant g nous avons

$$g\big(f(x)\big) = g\big(f(y)\big), \tag{4.63}$$

ce qui donne $x = y$.

f est surjective C'est maintenant le corolaire 4.41.

g est surjective Pour tout $x \in E$ nous avons $g\big(f(x)\big) = x$. Donc l'image de $f(E)$ par g est E.

g est injective C'est maintenant le corolaire 4.41.

□

4.2.4 Injection, surjection

Définition 4.48.
*Soient deux espaces vectoriels E et V. Une application $f\colon E \to V$ est **affine** si il existe une application linéaire $u\colon E \to V$ et un élément $v \in V$ tel que*

$$f(x) = u(x) + v \tag{4.64}$$

pour tout $x \in E$.

Définition 4.49.
*Soit un espace vectoriel V. Une **droite vectorielle** dans V est un sous-espace vectoriel de dimension 1. Un **plan vectoriel** est un sous-espace vectoriel de dimension 2.*

*Une partie D de V est une **droite** si il existe $v \in V$ tel que $D - v$ soit une droite vectorielle.*

Lemme 4.50.
Si D est une droite et si $a, b \in D$, alors $D - a = D - b$ et $D - a$ est une droite vectorielle.

Démonstration. Vu que D est une droite, il existe $v \in V$ tel que $D - v$ soit une droite vectorielle que nous notons L. Nous allons montrer que $D - a = D - v$. Vu que a est arbitraire, cela suffit.

$D - a \subset D - v$ Un élément de $D - a$ est de la forme $x - a$ avec $x \in D$. Nous écrivons $x - a$ sous la forme $y - v$ et nous espérons que $y \in D$. Allons-y : d'abord nous isolons y dans $x - a = y - v$:

$$y = x - a + v = (x - v) - (a - v) + v. \tag{4.65a}$$

Vu que $x - v$ et $a - v$ sont des éléments de L, la somme est dans L et donc $y = l + v$ pour un certain élément de $l \in L$. Nous avons donc prouvé que $y \in D$ et donc que $x - a = y - v \in D - v$.

$D - v \subset D - a$ Nous notons $x - v$ un élément générique que $D - v$ ($x \in D$). En posant $y - a = x - v$, nous trouvons

$$y = x - v + a = \underbrace{x - v}_{\in L} + \underbrace{(a - v)}_{\in L} + v \tag{4.66}$$

Donc $y \in D$ et $x - v = y - a \in D - a$.

□

Lemme 4.51.
À propos de droites.

(1) *Si L est une droite vectorielle, alors pour tout $a \neq 0$ dans L, nous avons $L = \text{Image}(f)$ où f est l'application linéaire donnée par*

$$\begin{aligned} f\colon \mathbb{K} &\to V \\ \lambda &\mapsto \lambda a. \end{aligned} \tag{4.67}$$

(2) *Si D est une droite affine, alors pour tout $a \neq b$ sur D nous avons $D = \text{Image}(f)$ où f est l'application affine donnée par*

$$\begin{aligned} g\colon \mathbb{K} &\to V \\ \lambda &\mapsto a + \lambda(b - a). \end{aligned} \tag{4.68}$$

Démonstration. En deux parties.

Pour (1) Vu que L est un sous-espace de dimension 1, il possède une baser, disons $\{b\}$. En particulier $a = \mu b$ pour un certain $\mu \in \mathbb{K}$. Si $x \in L$ nous avons $x = \lambda_x b$ pour un certain λ_x, et donc

$$x = \frac{\lambda_x}{\mu} a. \tag{4.69}$$

Donc $x = f(\lambda_x/\mu)$. Cela prouve que $L \subset \text{Image}(f)$.

L'inclusion inverse est simplement le fait que $\lambda a \in L$ dès que $a \in L$ parce que L est vectoriel.

Pour (2) Le lemme 4.50 nous indique qu'il existe une droite vectorielle L telle que $D - x = L$ pour tout $x \in D$.

$D \subset \text{Image}(g)$ Nous nommons $f\colon \mathbb{K} \to V$ l'application linéaire qui donne L. Vu que $b - a \in L$ nous avons

$$f(\lambda) = \lambda(b - a), \tag{4.70}$$

et tout élément de L est de la forme $f(\lambda)$. Nous avons aussi $D = L + a$; donc un élément de D est de la forme $f(\lambda) + a$ et donc de la forme $\lambda(b - a) + a = g(\lambda)$.

$\text{Image}(g) \subset D$ Un élément de $\text{Image}(g)$ est de la forme $a + \lambda(b - a)$ avec $\lambda \in \mathbb{K}$. Mais $b - a \in L$, donc $\lambda(b - a) \in L$ et

$$g(\lambda) = a + \lambda(b - a) \in a + L = D. \tag{4.71}$$

□

Exemple 4.52.
Les exemples les plus courants d'applications affines sont les droites et les plans ne passant pas par l'origine.

Les droites Une droite dans $\mathbb{R}^2$ (ou $\mathbb{R}^3$) qui ne passe pas par l'origine est l'image d'une fonction de la forme $s(t) = ut + v$, avec $t \in \mathbb{R}$, et u et v dans $\mathbb{R}^2$ ou $\mathbb{R}^3$ selon le cas.

En choisissant des coordonnées adéquates, les droites peuvent être aussi vues comme graphes de fonctions affines. Dans le cas de $\mathbb{R}^2$, on retrouve la fonction de l'exemple 4.27, pour $n = m = 1$.

Les plans De la même façon nous savons que tout plan qui ne passe pas par l'origine dans $\mathbb{R}^3$ est le graphe d'une application affine, $P(x, y) = (a, b)^T \cdot (x, y)^T + (c, d)^T$, lorsque les coordonnées sont bien choisies.

△

Lemme 4.53 ([80]).
Soit une application linéaire $f\colon E \to F$.

(1) *L'application f est injective si et seulement s'il existe $g\colon F \to E$ telle que $g \circ f = \text{Id}|_E$.*

(2) L'application f est surjective si et seulement s'il existe $g\colon F \to E$ telle que $f \circ g = \mathrm{Id}\,|_F$.

Démonstration. Nous démontrons séparément les deux affirmations.

(1) Si f est injective, alors $f\colon E \to \mathrm{Image}(f)$ est un isomorphisme. Si V est un supplémentaire de $\mathrm{Image}(f)$ dans F (c'est-à-dire $F = \mathrm{Image}(f) \oplus V$) alors nous pouvons poser $g(x+v) = f^{-1}(x)$ où $x + v$ est la décomposition (unique) d'un élément de F en $x \in \mathrm{Image}(f)$ et $v \in V$. Avec cela nous avons bien $g \circ f = \mathrm{Id}$.

Inversement, s'il existe $g\colon F \to E$ telle que $g \circ f = \mathrm{Id}$ alors $f\colon E \to E$ doit être injective. Parce que si $f(x) = 0$ avec $x \neq 0$ alors $(g \circ f)(x) = 0 \neq x$.

(2) Si f est surjective nous pouvons choisir des éléments $x_1, \dots, x_p$ dans E tels que $\{f(x_i)\}$ soit une base de F. Ensuite nous définissons

$$\begin{aligned} g\colon F &\to E \\ \sum_k a_k f(x_k) &\mapsto \sum_k a_k x_k. \end{aligned} \tag{4.72}$$

Cela donne $f \circ g = \mathrm{Id}\,|_F$ parce que si $v \in F$ alors $v = \sum_k v_k f(x_k)$ avec $v_k \in \mathbb{K}$, et nous avons

$$(f \circ g)(v) = \sum_k v_k (f \circ g)\left(f(x_k)\right) = f\left(\sum_k v_k x_k\right) = \sum_k v_k f(x_k) = v. \tag{4.73}$$

Inversement, s'il existe $g\colon F \to E$ tel que $f \circ g = \mathrm{Id}$ alors f doit être surjective parce que

$$F = \mathrm{Image}(f \circ g) = f\big(\mathrm{Image}(g)\big) \subset \mathrm{Image}(f). \tag{4.74}$$

□

4.3 Matrices

Les matrices et les applications linéaires sont deux choses différentes. Une application linéaire [21] est une application d'un espace vectoriel vers un autre, et une matrice est un simple tableau de nombres sur lesquels nous définissons des opérations, de telle sorte à fournir une structure d'espace vectoriel. Le lien entre ces opérations et les opérations correspondantes sur les applications linéaires sera fait plus tard.

4.3.1 Définitions

Définition 4.54.
Soit un anneau $\mathbb{A}$ ainsi que des entiers m, n strictement positifs. L'ensemble $\mathbb{M}(n \times n, \mathbb{A})$ est l'ensemble des applications

$$\{1, \dots, n\} \times \{1, \dots, m\} \to \mathbb{A}, \tag{4.75}$$

*et est appelé ensemble des **matrices** $n \times m$ sur $\mathbb{A}$.*

Si A est une matrice, nous notons A_{ij} au lieu de $A(i, j)$ l'image de (i, j) par l'application A.

Définition 4.55.
Quelques ensembles de matrices particuliers.

(1) Si $n = m$, alors :
- *nous disons que la matrice est **carrée**,*
- *nous notons $\mathbb{M}(n, \mathbb{A})$ pour $\mathbb{M}(n \times n, \mathbb{A})$,*
- *n est appelée **ordre** de la matrice.*

*(2) Si $n = 1$, alors la matrice est appelée **matrice-ligne**.*

21. Définition 4.24.

(3) Si $m = 1$, alors la matrice est appelée ***matrice-colonne****.*

4.56.
On note les isomorphismes naturels $\mathbb{M}(1 \times m, \mathbb{A}) \simeq \mathbb{A}^m$ et $\mathbb{M}(n \times 1, \mathbb{A}) \simeq \mathbb{A}^n$.

Lemme-définition 4.57.
L'ensemble $\mathbb{M}(n \times m, \mathbb{A})$ muni des opérations

Somme $(A + B)_{ij} = A_{ij} + B_{ij}$,

Produit par un scalaire $(\lambda A)_{ij} = \lambda A_{ij}$,

pour tout $A, B \in \mathbb{M}(n \times m, \mathbb{A})$ et $\lambda \in \mathbb{A}$ est un $\mathbb{A}$-module (définition 3.67).

Lemme-définition 4.58.
Avec la multiplication

$$\begin{aligned} \mathbb{M}(n \times p, \mathbb{A}) \times \mathbb{M}(p \times m, \mathbb{A}) &\to \mathbb{M}(n \times m, \mathbb{A}) \\ (A, B) &\mapsto (AB)_{ij} = \sum_{k=1}^{p} A_{ik} B_{kj}, \end{aligned} \tag{4.76}$$

l'espace $\mathbb{M}(n, \mathbb{K})$ est une $\mathbb{K}$-algèbre[22].

Définition 4.59.
Pour un élément $A \in \mathbb{M}(n \times m, \mathbb{A})$ nous définissons encore

La transposée $A^t_{ij} = A_{ji}$,

La trace $\mathrm{Tr}(A) = \sum_i A_{ii}$.

Remarque 4.60.
Quelques remarques directes sur les définitions.

(1) La motivation de cette définition pour le produit apparaîtra plus loin, mais le Frido n'étant pas un livre d'introduction, j'imagine que le lecteur a déjà une idée.

(2) Nous verrons plus loin en 4.6.2 que la définition de transposée d'une application linéaire n'est pas tout à fait évidente ; elle sera la définition 4.139.

Ici nous avons bien défini la transposée d'une matrice, pas d'une application linéaire.

Remarque 4.61.
Quelques remarques à propos de structures supplémentaires.

(1) Nous utiliserons (presque) tout le temps des matrices à coefficients dans un corps. Il est clair que, si $\mathbb{K}$ est un corps (commutatif), alors $\mathbb{M}(n \times m, \mathbb{K})$ a une structure d'espace vectoriel sur $\mathbb{K}$.

(2) Par ailleurs, sur les matrices carrées d'ordre n fixé, le produit de deux matrices est bien défini. Ainsi, $\mathbb{M}(n, \mathbb{A})$ se voit conférer une structure d'anneau, dont le neutre pour la multiplication est la matrice carrée $\mathbb{1}_n$ (notée aussi $\mathbb{1}$ lorsqu'il n'y a pas d'ambiguïté sur la taille), donnée par

$$\mathbb{1}_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{sinon.} \end{cases} \tag{4.77}$$

Il est vite vu que si A est une matrice carrée d'ordre n, alors $A\mathbb{1} = \mathbb{1}A = A$.

Lemme 4.62 ([1])**.**
Si A, B et C sont des matrices nous avons

(1) $(AB)^t = B^t A^t$,

(2) $\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB)$.

22. Définition 3.83.

Démonstration. La première est un simple calcul :

$$(AB)^t_{ij} = (AB)_{ji} = \sum_k A_{jk}B_{ki} = \sum_k A^t_{kj}B^t_{ik} = (B^tA^t)_{ij}. \tag{4.78}$$

Pour la seconde :

$$\mathrm{Tr}(ABC) = \sum_{ikl} A_{ik}B_{kl}C_{li} = \sum_{ikl} C_{li}A_{ik}B_{kl} = \sum_l (CAB)_{ll} = \mathrm{Tr}(CAB). \tag{4.79}$$

□

4.63.
La seconde égalité est importante et est nommée **invariance cyclique** de la trace. Elle sert entre autres nombreuses choses à prouver que la trace d'une matrice d'une application linéaire ne dépend pas de la base choisie. Ce sera la proposition 9.189.

Lemme 4.64.
Soient des matrices $A, B \in \mathbb{M}(n, \mathbb{K})$. Si pour tout $x, y \in \mathbb{K}^n$ nous avons

$$\sum_{ij} A_{ij}x_iy_j = \sum_{ij} B_{ij}x_iy_j \tag{4.80}$$

alors $A = B$.

Démonstration. Il suffit de choisir $x_i = \delta_{ik}$ et $y_j = \delta_{jl}$, et d'effectuer les sommes ; par exemple

$$\sum_{ij} A_{ij}\delta_{ik}y_j = \sum_j A_{kj}y_j. \tag{4.81}$$

Après avoir effectué toutes les sommes nous nous retrouvons avec $A_{kl} = B_{kl}$, ce qui signifie $A = B$. □

4.3.2 Application linéaire associée

Soient deux espaces vectoriels de dimension finie E, F sur le corps $\mathbb{K}$. Nous considérons les bases [23] $\{e_i\}$ pour E et $\{f_\alpha\}$ pour F.

Définition 4.65.
Nous considérons l'application

$$\begin{aligned} \psi\colon \mathbb{M}(n \times m, \mathbb{K}) &\to \mathcal{L}(E, F) \\ A &\mapsto f_A \end{aligned} \tag{4.82}$$

où f_A est définie par

$$f_A(x) = \sum_{i\alpha} A_{\alpha i}x_if_\alpha \tag{4.83}$$

si x_i sont les coordonnées de $x \in E$ dans la base $\{e_i\}$.

4.66.
Nous allons prouver un certain nombre de résultats montrant que cette application a toutes les propriétés imaginables permettant d'identifier les matrices aux applications linéaires : elle est un isomorphisme pour toutes les structure que vous pouvez raisonnablement imaginer.

À cette application ψ il manque cependant une propriété importante : elle n'est pas canonique. Elle dépend des bases choisies. Autrement dit : nous avons à priori autant d'applications ψ différentes qu'il y a de choix de bases sur E et F [24].

Nous allons prouver maintenant quelques résultats montrant que les matrices et les applications linéaires, dans le cas des espaces vectoriels $\mathbb{K}^n$ sont deux présentations de la même chose.

23. C'est le théorème 4.11 qui nous permet de considérer des bases. Et ce théorème ne fonctionne que parce que nous avons supposé une dimension finie.

24. Bonne question. Est-ce qu'il y a moyen de construire deux choix de bases donnant la même application ψ ? Écrivez-moi si vous savez la réponse.

4.67.
Lorsque $A \in \mathbb{M}(n, \mathbb{K})$ est une matrice et $x \in \mathbb{K}^n$ un vecteur, nous notons Ax l'élément de $\mathbb{K}^n$ donné par

$$(Ax)_i = \sum_j A_{ij} x_j. \tag{4.84}$$

Autrement dit, $Ax = f_A(x)$.

Cette convention et de nombreuses autres à propos de matrice sera rappelée dans -2.1.

Proposition-définition 4.68.
Soient deux espaces vectoriels de dimension finie E, F sur le corps $\mathbb{K}$. Nous considérons les bases $\{e_i\}$ pour E et $\{f_\alpha\}$ pour F.

Nous considérons l'application

$$\begin{aligned} \psi \colon \mathbb{M}(n \times m, \mathbb{K}) &\to \mathcal{L}(E, F) \\ A &\mapsto f_A \end{aligned} \tag{4.85}$$

où f_A est définie par

$$f_A(x) = \sum_{i\alpha} A_{\alpha i} x_i f_\alpha \tag{4.86}$$

si x_i sont les coordonnées de $x \in E$ dans la base $\{e_i\}$.

Alors

(1) Nous avons

$$f_A(e_i)_\alpha = A_{\alpha i}. \tag{4.87}$$

(2) Nous avons

$$f_A(e_i) = \sum_\alpha A_{\alpha i} f_\alpha. \tag{4.88}$$

(3) Nous avons

$$\big(f_A(x)\big)_\alpha = \sum_i A_{\alpha i} x_i. \tag{4.89}$$

(4) L'application ψ est une bijection.

Si f est une application linéaire, alors la matrice $\psi^{-1}(f)$ est la ***matrice associée*** *à f dans les bases choisies.*

Remarque : les bases ne sont supposées être canoniques en aucun sens du terme. Les dimensions de E et F ne sont pas non plus supposées identiques.

Démonstration. En nous rappelant que $(e_j)_i = \delta_{ij}$ nous avons

$$f_A(e_j) = \sum_{i\alpha} A_{\alpha i} (e_j)_i f_\alpha = \sum_\alpha A_{\alpha j} f_\alpha, \tag{4.90}$$

donc $f_A(e_i)_\alpha = A_{\alpha i}$. Cela prouve la formule du point (1).

Le point (2) est une simple somme sur α de (1).

La formule ((3)) est simplement la composante f_α de la définition 4.86.

Prouvons que ψ est injective. Si $f_A = f_B$, nous avons en particulier $f_A(e_i)_\alpha = f_B(e_i)_\alpha$ et donc $A_{\alpha i} = B_{\alpha i}$.

Prouvons que ψ est surjective. Pour cela nous considérons $f \in \mathcal{L}(E, F)$ et nous posons $A_{\alpha i} = f(e_i)_\alpha$. Nous avons alors $f = f_A$ parce que

$$f_A(x) = \sum_{i\alpha} A_{\alpha i} x_i f_\alpha = \sum_{i\alpha} f(e_i)_\alpha x_i f_\alpha = \sum_\alpha f(\sum_i x_i e_i)_\alpha f_\alpha = \sum_\alpha f(x)_\alpha f_\alpha = f(x). \tag{4.91}$$

□

La proposition suivante montre que le produit matriciel correspond à la composition d'applications linéaires, pourvu que l'on travaille avec les bases canoniques sur $\mathbb{K}^n$.

Proposition 4.69 ([1]).
Soit un corps commutatif $\mathbb{K}$. Nous considérons des espaces vectoriels E et F munis de bases $\{e_i\}_{i=1,\dots,n}$ et $\{f_\alpha\}_{\alpha\in 1,\dots,m}$.

L'application déjà définie[25]

$$\psi\colon \mathbb{M}(m\times n,\mathbb{K})\to\mathcal{L}(E,F) \tag{4.92}$$

est un isomorphisme d'espaces vectoriels.

Démonstration. Le fait que ψ soit une bijection est la proposition 4.68. Nous devons montrer que c'est linéaire.

Pour $\lambda\in\mathbb{K}$ nous avons le calcul

$$\psi(\lambda A)(e_k)=f_{\lambda A}(e_k)=\sum_{\alpha i}(\lambda A)_{\alpha i}\underbrace{(e_k)_i}_{=\delta_{ki}}f_\alpha=\lambda\sum_\alpha A_{\alpha k}f_\alpha=\lambda f_A(e_j). \tag{4.93}$$

Donc $\psi(\lambda A)=\lambda\psi(A)$.

Si $A,B\in\mathbb{M}(n,\mathbb{K})$ nous avons de la même façon $f_{A+B}=f_A+f_B$. □

Proposition 4.70.
Soient des espaces vectoriels E, F et G de dimensions n, m et p munis de bases[26] *$\{e_i\}$, $\{f_i\}$ et $\{g_i\}$. Nous considérons les deux applications*

$$\psi\colon \mathbb{M}(m\times n,\mathbb{K})\to\mathcal{L}(E,F) \tag{4.94}$$

et

$$\psi\colon \mathbb{M}(p\times m,\mathbb{K})\to\mathcal{L}(F,G). \tag{4.95}$$

Nous avons

$$f_A\circ f_B=f_{AB} \tag{4.96}$$

pour toutes matrices $A\in\mathbb{M}(p\times m,\mathbb{K})$ et $B\in\mathbb{M}(m\times n,\mathbb{K})$.

Démonstration. Nous considérons les applications linéaires associées à A et B : $f_A\colon F\to G$ et $f_B\colon E\to F$ et la composée $f_A\circ f_B\colon E\to G$. Et puis c'est le calcul :

$$\begin{aligned}(f_A\circ f_B)(e_k)&=f_A\big(\sum_{ij}B_{ij}(e_k)_jf_i\big) &(4.97a)\\ &=\sum_i B_{ik}f_A(f_i) &(4.97b)\\ &=\sum_i B_{ik}\sum_{rs}A_{rs}(f_i)_sg_r &(4.97c)\\ &=\sum_{ir}B_{ik}A_{ri}g_r &(4.97d)\\ &=\sum_r(AB)_{rk}g_r &(4.97e)\\ &=f_{AB}(e_k). &(4.97f)\end{aligned}$$

Donc $f_A\circ f_B=f_{AB}$ comme il se doit. □

Nous pouvons particulariser au cas où $E=F=G$.

25. Notez la position du n et du m. Sachez noter les bornes des sommes écrites dans la démonstration.
26. Avec trois, nous renonçons à utiliser des alphabets différents pour numéroter les éléments des bases.

Proposition 4.71.
Si E est un espace vectoriel muni d'une base $\{e_i\}$, alors l'application

$$\psi\colon \mathbb{M}(n,\mathbb{K}) \to \mathrm{End}(E) \tag{4.98}$$

est un isomorphisme d'algèbre[27] *et d'anneaux*[28].

Démonstration. Le fait que ψ soit un isomorphisme d'algèbre est juste la combinaison entre les propositions 4.69 et 4.70.

En ce qui concerne l'isomorphisme d'anneaux, il faut en plus identifier les neutres. Le neutre pour la composition d'applications linéaires est l'application identité et le neutre pour la multiplication de matrices est la matrice identité. Nous devons donc montrer que $\psi(\delta) = \mathrm{Id}$. Juste un calcul :

$$f_\delta(x) = \sum_{ij} \delta_{ij} x_j e_i = \sum_i x_i e_i = x. \tag{4.99}$$

Donc oui, f_δ est l'identité. □

Voila. Soyez bien conscient que l'application ψ dont nous avons beaucoup parlé est surtout intéressante dans le cas des espaces de la forme $\mathbb{K}^n$. Dans ce cas, nous avons une identification canonique entre $\mathbb{M}(n,\mathbb{K})$ et $\mathrm{End}(\mathbb{K}^n)$ qui est un isomorphisme d'anneaux et d'algèbres.

Nous verrons que ce ψ respecte encore les inverses[29] et les déterminants[30].

4.72.
Il convient de ne pas confondre matrice et application linéaire (bien que nous le ferons sans vergogne). Une matrice est un bête tableau de nombres, tandis qu'une application linéaire est une application entre deux espaces vectoriels vérifiant certaines propriétés.

Cependant si les espaces vectoriels E et F sont munis de bases, alors il y a une application

$$\psi\colon \mathbb{M}(m\times n,\mathbb{K}) \to \mathcal{L}(E,F) \tag{4.100}$$

qui a toutes les propriétés imaginables[31].

Cette application dépend des bases choisies. Il n'y a donc pas de trucs comme « la matrice de telle application linéaire » ou comme « voici une matrice, nous considérons l'application linéaire associée ».

Cependant, sur des espaces comme $\mathbb{R}^n$ ou plus généralement sur $\mathbb{K}^n$, nous avons une base canonique et toute personne raisonnable utilise toujours la base canonique (sauf mention du contraire). Dans ces cas il est sans danger de dire « la matrice associée à telle application linéaire » sans préciser les bases.

Mais si un jour vous utilisez une base autre que la base canonique sur $\mathbb{R}^n$, précisez-le et plutôt deux fois qu'une[32].

4.73.
Tant que nous sommes à parler de matrice et d'applications linéaires, les plus acharnés anti-abus de language[33] remarqueront qu'il n'est pas vrai que « étant donné une base, une application linéaire a une matrice ».

En effet, une base est une partie libre et génératrice (définition 4.5). Or une partie d'un ensemble n'est pas muni d'un ordre. Toutes les permutations de colonnes de la matrice sont encore possible d'après l'ordre que l'on met sur les vecteurs de la base.

27. Définition 3.83.
28. Définition 1.74
29. Proposition 4.90.
30. Proposition 9.11.
31. Et elle en aura encore plus lorsque nous aurons vus les déterminants.
32. Au passage, non, les coordonnées polaires ne sont pas une base de $\mathbb{R}^2$. C'est un système de coordonnées, et ce n'est pas la même chose.
33. Dont l'auteur de ces lignes fait partie.

Encore une fois, la base canonique n'a pas de problèmes parce que les $\{e_i\}$ de $\mathbb{R}^n$ viennent avec un ordre indiscutable. Plus généralement, très souvent, lorsqu'on construit une base, la construction suggère un ordre.

4.3.3 Déterminant

Définition 4.74.
Si $A \in \mathbb{M}(n, \mathbb{K})$ nous définissons le ***déterminant*** *de A par la formule*

$$\det(A) = \sum_{\sigma \in S_n} (-1)^\sigma \prod_{i=1}^{n} A_{i\sigma(i)} \tag{4.101}$$

où la somme est effectuée sur tous les éléments du groupe symétrique[34] *S_n et où $(-1)^\sigma$ représente la parité de la permutation σ.*

En se souvenant que $|S_n| = n!$, nous sommes frappés de stupeur devant le fait que le nombre de termes dans la somme croît de façon factorielle (c'est plus qu'exponentiel, pour info) en la taille de la matrice. Cette formule est donc sans espoir pour une matrice plus grande que 3×3 ou à la rigueur 4×4 à la main. À l'ordinateur, il est possible de monter plus haut, mais pas tellement.

4.3.4 Déterminant en petite dimension

En dimension deux, le déterminant de la matrice $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est le nombre

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - cb. \tag{4.102}$$

Ce nombre détermine entre autres le nombre de solutions que va avoir le système d'équations linéaires associé à la matrice.

Pour une matrice 3×3, nous avons le même concept, mais un peu plus compliqué ; nous avons la formule

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}. \tag{4.103}$$

4.3.5 Manipulations de lignes et de colonnes

Nous voudrions savoir ce qu'il se passe avec le déterminant d'une matrice lorsque nous substituons à une ligne ou une colonne une combinaison des autres lignes et colonnes. Lorsque une matrice est donnée, nous notons C_j sa j^{e}colonne.

Lemme 4.75 ([1])**.**
Si A et B sont des matrices, alors

$$(AB)^t = B^t A^t. \tag{4.104}$$

Démonstration. Il suffit de calculer les éléments de matrice :

$$(AB)^t_{ij} = (AB)_{ji} = \sum_k A_{jk} B_{ki} = \sum_k B^t_{ik} A^t_{kj} = (B^t A^t)_{ij}. \tag{4.105}$$

□

Lemme 4.76 ([1, 81])**.**
Si A est une matrice, alors $\det(A) = \det(A^t)$.

34. Pour le groupe symétrique, c'est la définition 2.63, le fait que ce soit un groupe fini est le lemme 2.65, et pour la somme sur un groupe fini c'est la définition 1.90..

Démonstration. Nous commençons par écrire la définition du déterminant :

$$\det(A^t) = \sum_{\sigma\in S_n} \epsilon(\sigma) \prod_{i=1}^{n} (A^t)_{i,\sigma(i)} = \sum_{\sigma} \epsilon(\epsilon) \prod_i A_{\sigma(i),i}. \tag{4.106}$$

Pour chaque σ séparément, nous utilisant la proposition 1.94 pour ré-indexer le produit :

$$\prod_i A_{\sigma(i),i} = \prod_i A_{i,\sigma^{-1}(i)}. \tag{4.107}$$

Nous profitons du fait que l'application $\varphi\colon S_n \to S_n$ donnée par $\varphi(\sigma) = \sigma^{-1}$ soit une permutation de S_n pour appliquer la définition 1.90 et faire la somme sur σ^{-1} :

$$\det(A^t) = \sum_{\sigma} \epsilon(\sigma) \prod_i A_{i,\sigma^{-1}(i)} = \sum_{\sigma} \epsilon(\sigma^{-1}) \prod_i A_{i,\sigma(i)} = \det(A) \tag{4.108}$$

où nous avons utilisé le fait que $\epsilon(\sigma^{-1}) = \epsilon(\sigma)$ (corolaire 2.77). □

Le fait que $\det(A) = \det(A^t)$ permet, dans toutes les propositions du type « ce qui arrive au déterminant si on change telle ligne ou colonnes » de ne donner qu'une preuve pour la partie « ligne » et déduire automatiquement le cas « colonne ». Le lemme suivant donne un exemple d'utilisation.

Lemme 4.77 ([1])**.**
Soit une matrice A. Nous considérons la matrice B obtenue à partir de A par la permutation de lignes $L_k \leftrightarrow L_l$ ainsi que la matrice C obtenue à partir de A^t par la permutation de colonnes $C_k \leftrightarrow C_l$.

Alors $C^t = B$.

Démonstration. Calculons les éléments de matrice de C :

$$C_{ij} = \begin{cases} (A^t)_{ij} & \text{si } j \neq k, j \neq l \\ (A^t)_{ik} & \text{si } j = l \\ (A^t)_{il} & \text{si } j = k \end{cases} = \begin{cases} A_{ji} & \text{si } j \neq k, j \neq l \\ A_{ki} & \text{si } j = l \\ A_{li} & \text{si } j = k. \end{cases} \tag{4.109}$$

Ensuite nous prouvons que $C^t = B$ en écrivant les éléments de C^t :

$$(C^t)_{ij} = C_{ji} = \begin{cases} A_{ij} & \text{si } i \neq k, i \neq l \\ A_{kj} & \text{si } i = l \\ A_{lj} & \text{si } i = k. \end{cases} \tag{4.110}$$

Cette dernière expression est la matrice A après permutation des lignes $L_k \leftrightarrow L_l$, c'est-à-dire a matrice B. □

Pour la suite nous écrivons δ la matrice « identité », c'est-à-dire celle dont les entrées sont précisément les δ_{ik}. Nous écrivons également E_{ij} la matrice contenant de zéros partout sauf en (i, j) où elle a un 1, c'est-à-dire

$$(E_{ij})_{kl} = \delta_{ik}\delta_{jl}. \tag{4.111}$$

Proposition 4.78 (Permuter des lignes ou des colonnes $L_k \leftrightarrow L_l$[82, 1])**.**
Soient une matrice $A \in \mathbb{M}(n, \mathbb{K})$, deux entiers $k \neq l$ inférieurs ou égaux à n.

(1) Si B est la matrice obtenue à partir de A en permutant deux lignes ou deux colonnes, alors

$$\det(A) = -\det(B). \tag{4.112}$$

(2) Si B est la matrice obtenue à partir de A par la permutation de lignes $L_k \leftrightarrow L_l$. Alors

$$B = SA \tag{4.113}$$

avec $S = \delta + E_{kl} + E_{lk} - E_{kk} - E_{ll}$.

Autrement dit : la matrice S est une matrice de permutations de lignes.

(3) La matrice S vérifie $\det(S) = -1$

(4) Nous avons

$$\det(SA) = \det(S)\det(A). \tag{4.114}$$

Démonstration. Point par point

(1) pour les colonnes Soient k et l fixés, et considérons la permutation des colonnes C_k et C_l. Nous notons α la permutation (kl) dans S_n (groupe symétrique, définition 2.63). Nous avons

$$B_{ij} = A_{i\alpha(j)}, \tag{4.115}$$

ou encore : $A_{ij} = B_{i\alpha(j)}$. Par définition,

$$\det(A) = \sum_{\sigma\in S_n} \epsilon(\sigma) \prod_{i=1}^{n} A_{i\sigma(i)} \tag{4.116}$$

C'est le moment d'utiliser la proposition 1.92 à propos de somme sur des groupes avec $G = S_n$, $h = \alpha$ et

$$f(\sigma) = \epsilon(\sigma) \prod_i A_{i,\sigma(i)}. \tag{4.117}$$

Nous savons que $\epsilon(\alpha) = -1$ et que ϵ est un homomorphisme par la proposition 2.76(1), donc

$$f(\alpha\sigma) = \epsilon(\alpha\sigma) \prod_i A_{i,(\alpha\sigma)(i)} = -\epsilon(\sigma) \prod_i B_{i,\sigma(i)}. \tag{4.118}$$

Avec ça, nous concluons :

$$\det(A) = \sum_{\sigma\in S_n} f(\sigma) = \sum_{\sigma} f(\alpha\sigma) = -\sum_{\sigma\in S_n} \epsilon(\sigma) \prod_{i=1}^{n} B_{i\sigma(i)} = -\det(B). \tag{4.119}$$

(1) pour les lignes Que se passe-t-il si nous permutons les lignes L_k et L_l ?Si nous notons B' la matrice obtenue à partir de A par la permutation de lignes $L_k \leftrightarrow L_l$, et C celle obtenue de A^t après permutation de colonnes $C_k \leftrightarrow C_l$ alors nous avons $C^t = B'$. Le lemme 4.77 nous dit que $C^t = B'$. En utilisant le lemme 4.76 sur le déterminant de la transposée,

$$\det(B') = \det(C^t) = \det(C) = -\det(A^t) = -\det(A). \tag{4.120}$$

Voila qui prouve le résultat pour les permutation de lignes.

(2) Si $k = l$, il n'y a pas de permutations, et il est vite vu que la matrice S est l'identité parce qu'il y a quatre fois le terme E_{kk}. Nous supposons donc que $k \neq l$; en particulier $\delta_{kl} = 0$.
Il s'agit surtout d'un beau calcul :

$$(SA)_{ij} = \sum_m S_{im}A_{mj} = A_{ij} + \sum_m (\delta_{ki}\delta_{lm} + \delta_{li}\delta_{lm} - \delta_{ki}\delta_{km} - \delta_{li}\delta_{lm})A_{mj} \tag{4.121a}$$

$$= A_{ij} + \delta_{ki}A_{lj} + \delta_{li}A_{kj} - \delta_{ki}A_{kj} - \delta_{li}A_{lj}. \tag{4.121b}$$

Si $i \neq j$ et $i \neq l$, alors $(SA)_{ij} = A_{ij}$. Si $i = k$, alors

$$(SA)_{kj} = A_{kj} + A_{lj} - A_{kj} = A_{lj}, \tag{4.122}$$

c'est-à-dire que la k^{e}ligne de SA est la l^{e}ligne de A.
Avec $i = l$ nous obtenons la k^{e}ligne de A.
Tout cela montre que SA est la matrice A dans laquelle les lignes k et l ont été inversées, c'est-à-dire $SA = B$.

(3) En utilisant la définition du déterminant,

$$\det(S) = \sum_{\sigma\in S_n} \epsilon(\sigma) \prod_{i=1}^{n} S_{i\sigma(i)} \tag{4.123a}$$

$$= \sum_{\sigma} \epsilon(\sigma) \prod_{i} \big(\delta_{i\sigma(i)} + \delta_{ki}\delta_{l\sigma(i)} + \delta_{li}\delta_{k\sigma(i)} - \delta_{ki}\delta_{k\sigma(i)} - \delta_{li}\delta_{l\sigma(i)}\big). \tag{4.123b}$$

Nous utilisons l'associativité et la commutativité du produit pour séparer les facteurs $i = k$ et $i = l$ des autres :

$$\det(S) = \sum_{\sigma} \epsilon(\sigma) \prod_{\substack{i\neq k\\ i\neq l}} \delta_{i\sigma(i)}(\delta_{k\sigma(k)} + \delta_{l\sigma(k)} - \delta_{k\sigma(k)})(\delta_{l\sigma(l)} + \delta_{k\sigma(l)} - \delta_{l\sigma(l)}). \tag{4.124}$$

À cause des facteurs $i \neq k$ et $i \neq l$, les σ pour lesquels le tout n'est pas nuls doivent vérifier $\delta_{i\sigma(i)} = 1$ pour tout i différent de k et l. Les deux seuls sont donc $\sigma = \mathrm{Id}$ et la permutation $\sigma = (k, l)$. Pour $\sigma = \mathrm{Id}$, nous avons

$$\prod_{\substack{i\neq k\\ i\neq l}} \delta_{ii}(\delta_{kk} + \delta_{lk} - \delta_{kk})(\delta_{ll} + \delta_{kl} - \delta_{ll}) = 0. \tag{4.125}$$

Dernier espoir : $\sigma = (k, l)$. Pour ce terme nous avons $\epsilon(\sigma) = -1$ et

$$\prod_{\substack{i\neq k\\ i\neq l}} \delta_{ii}(\delta_{kl} + \delta_{ll} - \delta_{kl})(\delta_{lk} + \delta_{kk} - \delta_{lk}) = 1. \tag{4.126}$$

Au final dans $\det(S)$ il n'y a de non nul que le terms $\sigma = (k, l)$ et il vaut -1. Donc

$$\det(S) = -1. \tag{4.127}$$

(4) Il s'agit de mettre bout à bout les points déjà prouvés :

$$\det(SA) = -\det(A) = \det(S)\det(A). \tag{4.128}$$

□

Corolaire 4.79 ([82]).
Soit une matrice $A \in \mathbb{M}(n, \mathbb{K})$. Si deux lignes ou deux colonnes de A sont égales, alors $\det(A) = 0$.

Démonstration. Si deux colonnes sont égales, la matrice ne change pas lorsqu'on les permute, alors que le déterminant change de signes. La seule possibilité est que $\det(A) = -\det(A)$, ce qui signifie que $\det(A) = 0$. □

Notons que si pour $k \neq l$ nous avons $C_k = \lambda C_l$, alors nous avons aussi $\det(A) = 0$.

La réciproque n'est pas vraie : il existe des matrices dont le déterminant est nul et dont aucune entrée n'est nulle. Par exempe

$$\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}. \tag{4.129}$$

Proposition 4.80 ([82]).
Soient $A \in \mathbb{M}(n, \mathbb{K})$, et $v \in \mathbb{K}^n$. Si B est la matrice A avec la substitution $L_j \to L_j + v$ et C est la matrice A avec la substitution $L_j \to v$, alors

$$\det(B) = \det(A) + \det(C). \tag{4.130}$$

Démonstration. En utilisant l'associativité de la multiplication,

$$\det(B) = \sum_{\sigma\in S_n} \epsilon(\sigma)\prod_{i=1}^{n} B_{i\sigma(i)} \tag{4.131a}$$
$$= \sum_{\sigma} \epsilon(\sigma)\big(\prod_{i\neq j} B_{i\sigma(i)}\big)B_{j\sigma(j)} \tag{4.131b}$$
$$= \sum_{\sigma} \epsilon(\sigma)\big(\prod_{i\neq j} A_{i\sigma(i)}\big)(A_{j\sigma(j)} + v_{\sigma(j)}) \tag{4.131c}$$
$$= \sum_{\sigma} \epsilon(\sigma)\prod_{i} A_{i\sigma(i)} + \sum_{\sigma} \epsilon(\sigma)\prod_{i\neq j} C_{i\sigma(i)}v_{\sigma(j)} \tag{4.131d}$$
$$= \det(A) + \sum_{\sigma} \epsilon(\sigma)\prod_{i\neq j} C_{i\sigma(i)}C_{j\sigma(j)} \tag{4.131e}$$
$$= \det(A) + \det(C). \tag{4.131f}$$

Justifications :

— 4.131d parce que pour $i \neq j$ nous avons $A_{i\sigma(i)} = C_{i\sigma(i)}$

— 4.131e parce que $v_{\sigma(j)} = C_{j\sigma(j)}$.

□

Proposition 4.81 (Combinaison de lignes ou colonnes $L_k \to L_k + \lambda L_l$[82])**.**
Soient une matrice $A \in \mathbb{M}(n, \mathbb{K})$, deux entiers $k \neq l$ inférieurs ou égaux à n.

(1) Si B est la matrice obtenue à partir de A par la substitution $L_k \to L_k+\lambda L_l$ ou $C_k \to C_k+\lambda C_l$, alors

$$\det(A) = \det(B). \tag{4.132}$$

(2) Si B est la matrice A dans laquelle nous avons fait la substitution $L_k \to L_k + \lambda L_l$, alors

$$B = UA \tag{4.133}$$

avec $U = \delta + \lambda E_{kl}$, c'est-à-dire que U est une matrice de combinaison de lignes.

(3) La matrice U vérifie $\det(U) = 1$.

(4) Nous avons

$$\det(UA) = \det(U)\det(A). \tag{4.134}$$

Démonstration. Point par point.

(1) Soit la matrice C obtenue à partir de A par $L_k \to \lambda L_l$. En considérant le vecteur $v = \lambda L_l$, nous sommes dans la situation de la proposition 4.80. Donc

$$\det(B) = \det(A) + \det(C). \tag{4.135}$$

Mais dans la matrice C, nous avons $L_k = \lambda L_l$, ce qui implique $\det(C) = 0$ par le corolaire 4.79. Donc $\det(A) = \det(B)$ comme il se devait.

(2) Encore un calcul :

$$(UA)_{ij} = \sum_m \big(\delta_{im} + \lambda(E_{kl})_{im}\big)A_{mj} = A_{ij} + \lambda\sum_m \delta_{ki}\delta_{lm}A_{mj} = A_{ij} + \lambda\delta_{li}A_{kj}. \tag{4.136}$$

Cela donne, pour $i = k$ la ligne

$$(UA)_{kj} = A_{kj} + \lambda A_{lj}, \tag{4.137}$$

ce qui correspond bien à $L_k \to L_k + \lambda L_l$.

(3) Nous calculons le déterminant de $U = \delta + \lambda E_{kl}$ avec $k \neq l$. Nous avons dans un premier temps :

$$\det(U) = \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^{n} (\delta_{i\sigma(i)} + \lambda \delta_{ki} \delta_{l\sigma(i)}). \tag{4.138}$$

Vu que nous avons toujours $\delta_{ki}\delta_{li} = 0$, le terme $\sigma = \mathrm{Id}$ donne 1.

Pour les $\sigma \neq \mathrm{Id}$, le facteur $\lambda \delta_{ki} \delta_{l\sigma(i)}$ ne s'annulle pas uniquement si $i = k$ et $\sigma(i) = k$. Donc le seul terme non nul autre que $\sigma = \mathrm{Id}$ peut provenir de $\sigma = (k, l)$. Pour ce terme, nous isolons les termes $i = l$ et $i = k$:

$$(\delta_{k\sigma(k)} + \lambda \delta_{kk} \delta_{k\sigma(k)})(\delta_{l\sigma(l)} + \lambda \delta_{kl} \delta_{k\sigma(l)}). \tag{4.139}$$

Le dernier facteur est nul.

(4) En mettant bout à bout les résultats prouvés,

$$\det(UA) = \det(A) = \det(U)\det(A). \tag{4.140}$$

□

Proposition 4.82 (Multiplication par un scalaire d'une ligne ou colonne $L_k \to \lambda L_k$[82])**.**
Soient une matrice $A \in \mathbb{M}(n, \mathbb{K})$, un entier $k \neq l$ inférieurs ou égal à n. Soit la matrice B obtenue à partir de A en multipliant la ligne L_k par $\lambda \in \mathbb{K}$.

(1) $\det(B) = \lambda \det(A)$

(2) En considérant la matrice $T = \delta + (\lambda - 1)E_{kk}$, nous avons

$$B = TA, \tag{4.141}$$

c'est-à-dire que a matrice T est une matrice de multiplication de ligne par un scalaire.

(3) Nous avons $\det(T) = \lambda$.

(4) Et aussi : $\det(TA) = \det(T)\det(A)$

Démonstration. Point par point.

(1) La matrice B est donnée par les éléments

$$B_{ij} = \begin{cases} A_{ij} & \text{si } j \neq k \\ \alpha A_{ij} & \text{si } j = kn \end{cases} \tag{4.142}$$

c'est-à-dire $B_{ij} = \big(1 + (\alpha - 1)\delta_{jk}\big)A_{ij}$. Nous mettons cela dans la définition du déterminant de B :

$$\det(B) = \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^{n} B_{i\sigma(i)} = \sum_{\sigma} \prod_{i} \big(1 + (\alpha - 1)\delta_{\sigma(i)k} A_{i\sigma(i)}\big). \tag{4.143}$$

L'associativité du produit dans $\mathbb{K}$ nous permet de séparer le produit de la façon suivante :

$$\prod_{i=1}^{n} \big(1 + (\alpha - 1)\delta_{\sigma(i)k}\big) A_{i\sigma(i)} = \prod_{i} \big(1 + (\lambda - 1)\delta_{\sigma(i)k}\big) \prod_{i} A_{i\sigma(i)} = \lambda \prod_{i} A_{i\sigma(i)}. \tag{4.144}$$

En remettant dans (4.143), nous trouvons $\det(B) = \det(A)$.

(2) C'est un cas particulier de la proposition 4.81(2) en prenant $k = l$ et en adaptant le λ.

(3) Nous calculons le déterminant de la matrice $T = \delta + (\lambda - 1)E_{kk}$. La formule du déterminant donne

$$\det(T) = \sum_{\sigma} \epsilon(\sigma) \prod_{i=1}^{n} \big(\delta_{i\sigma(i)} + (\lambda - 1)\delta_{ki}\delta_{k\sigma(i)}\big). \tag{4.145}$$

Si $i \neq \sigma(i)$, alors non seulement $\delta_{i\sigma(i)} = 0$, mais en plus $\delta_{ki}\delta_{k\sigma(i)} = 0$. Donc suel $\sigma = \mathrm{Id}$ reste dans la somme sur $\sigma \in S_n$. Il reste donc

$$\det(T) = \prod_{i=1}^{n} \big(1 + (\lambda - 1)\delta_{ki}\big) = \left(\prod_{i\neq k} 1\right)(1 + (\lambda - 1)) = \lambda \tag{4.146}$$

où nous avons utilisé encore l'associativité pour isoler le facteur $i = k$.

(4) Il faut mettre bout à bout les résultats déjà faits :

$$\det(TA) = \lambda \det(A) = \det(T)\det(A). \tag{4.147}$$

□

4.3.6 Réduction de Gauss

Nous avons vu les matrices d'opérations élémentaire sur les lignes et colonnes :

— Permutation de lignes $L_k \leftrightarrow L_l$: $S(n; k, l) = \delta + E_{kl} + E_{lk} - E_{kk} - E_{ll}$, proposition 4.78.

— Combinaisons de lignes $L_k \to L_k + \lambda L_l$: $U(n; k, l, \lambda) = \delta + \lambda E_{kl}$, proposition 4.81.

— Multiplication d'une ligne par un scalaire $L_k \to \lambda L_k$: $T = \delta + (\lambda - 1)E_{kk}$, proposition 4.82.

Ces matrices seront dans la suite notées G. Et elles vérifient la grosse propriété

$$\det(GA) = \det(G)\det(A) \tag{4.148}$$

pour toute matrice A.

Proposition 4.83 (Réduction de Gauss[1])**.**
Soit une matrice $A \in \mathbb{M}(n, \mathbb{K})$ de déterminant non nul : $\det(A) \neq 0$. Alors il existe des matrices $G_1, \dots, G_N$ toutes de type S, U ou T telles que

$$G_1 \dots G_N A = \delta. \tag{4.149}$$

Démonstration. Nous faisons une récurrence sur n. D'abord pour $n = 1$, la matrice A contient un seul élément A_{11} qui est non nul par hypothèse. Nous pouvons multiplier sa ligne par $1/A_{11}$ pour obtenir le résultat. Plus précisément, nous avons l'égalité

$$T(1; 1, \frac{1}{A_{11}})A = \delta \tag{4.150}$$

dans $\mathbb{M}(1, \mathbb{K})$. Notons que $\mathbb{K}$ est un corps (donc A_{11} est inversible) commutatif, ce qui permet d'écrire $1/A_{11}$ sans ambiguïtés.

Supposons le résultat prouvé pour n, et voyons ce qu'il se passe pour $n+1$. Vu que $\det(A) \neq 0$, aucune de ses colonnes n'est nulle (corolaire 4.79). Il existe donc un k tel que $A_{k1} \neq 0$.

Par la proposition 4.78, la matrice

$$B^{(1)} = S(n + 1; k, 1)A \tag{4.151}$$

est une matrice telle que $B^{(1)}_{11} = A_{k1} \neq 0$. Ensuite, par la proposition 4.82 la matrice

$$B^{(2)} = T(n + 1; 1, \frac{1}{A_{k1}})B^{(1)} \tag{4.152}$$

vérifie $B^{(2)}_{11} = 1$.

Vu que la multiplication par la matrice $U(n+1; k; l; \lambda)$ fait par la proposition 4.81 la substitution $L_k \to L_k + \lambda L_l$, la matrice

$$B^{(3)} = \prod_{k=2}^{n+1} U(n + 1; k, 1, -B^{(1)}_{k1})B^{(1)} \tag{4.153}$$

a toute sa première colonne nulle à l'exception de $B^{(3)}_{11} = 1$.

Nous n'avons pas donné de nom ni démontré de théorèmes à propos de la substitution $C_k \to C_k + \lambda C_l$. En passant éventuellement par les transposées et en utilisant les lemmes 4.75 et 4.76 nous obtenons une matrice $U'(n+1; k, l, \lambda)$ ayant la propriété que la matrice

$$B^{(4)} = \prod_{k=2}^{n+1} U'(n+1; k, 1, -B^{(3)}_{1k}) B^{(3)} \tag{4.154}$$

vérifie $B^{(4)}_{1j} = B^{(4)}_{j1} = 0$ pour tout j sauf $j = 1$. En d'autres termes, la matrice $B^{(4)}$ est de la forme

$$B^{(4)} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{pmatrix} \tag{4.155}$$

où A' est une matrice de taille n.

Voyons quelques propriétés de A'. Nous savons que

$$B^{(4)} = \prod_i G_i A \tag{4.156}$$

où les G_i sont de type S, T ou U. Vu que $\det(SA) = \det(S)\det(A)$ (et idem pour T et U), nous avons

$$\det(B^{(4)}) = \prod_i \det(G_i)\det(A), \tag{4.157}$$

et comme aucun des $\det(G_i)$ n'est nul, nous avons encore $\det(B^{(4)}) \neq 0$, ce qui implique $\det(A') \neq 0$.

La récurrence peut avoir lieu. Il existe des matrices G'_i telles que

$$G'_1 \dots G'_M A' = \delta \tag{4.158}$$

où les G'_i sont de taille n, ainsi que le δ. En remarquant que

$$S(n+1; k, l) = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & S(n; k-1, l-1) & \\ 0 & & & \end{pmatrix}, \tag{4.159}$$

et pareillement pour les matrices T et U, nous voyons qu'en prenant

$$G_i = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & G'_i & \\ 0 & & & \end{pmatrix}, \tag{4.160}$$

nous avons

$$\prod_{i=1}^{M} G_i B^{(3)} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \prod_{i=1}^{M} G'_i A' & \\ 0 & & & \end{pmatrix} = \delta_{n+1} \tag{4.161}$$

où nous avons mis un indice sur le dernier δ pour être plus explicite. □

Proposition 4.84.
Si $A \in \mathbb{M}(n, \mathbb{K})$ est telle que $\det(A) = 0$, alors il existe des matrices de manipulation de lignes et de colonnes $G_1, \dots, G_N$ telles que $G_1 \dots G_N A$ ait une colonne de zéros.

Démonstration. Si la matrice A elle-même n'a pas de colonnes de zéros, alors nous pouvons faire un pas de réduction de Gauss[35] et obtenir des matrices $G_1, \dots, G_{N_1}$ telles que

$$G_1 \dots G_{N_1} A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A^{(1)} & \\ 0 & & & \end{pmatrix}. \tag{4.162}$$

Si $A^{(1)}$ ne possède pas de colonnes de zéros, nous pouvons continuer.

Si nous parvenons à faire n pas de la sorte, alors nous aurions

$$G_1 \dots G_N A = \delta, \tag{4.163}$$

et donc $\det(G_1 \dots G_N)\det(A) = 1$, ce qui est impossible lorsque $\det(A) = 0$. Nous en concluons que le processus doit s'arrêter et qu'une des matrices $A^{(k)}$ doit avoir une colonne de zéros[36]. □

4.3.7 Matrices inversibles

Proposition-définition 4.85.
Soit une matrice $A \in \mathbb{M}(n, \mathbb{K})$. Si les matrices B_1 et B_2 de $\mathbb{M}(n, \mathbb{K})$ vérifient

$$AB_1 = B_1 A = \delta \tag{4.164}$$

et

$$AB_2 = B_2 A = \delta, \tag{4.165}$$

alors $B_1 = B_2$. Dans ce cas, nous disons que A est inversible et nous notons A^{-1} l'unique matrice telle que $AA^{-1} = A^{-1}A = \delta$.

Démonstration. La preuve est réalisée dans le cas général par le lemme 1.70. Mais si vous en voulez une preuve avec les notations d'ici, en voici une.

Nous avons $AB_1 = AB_2$. En multipliant à gauche par B_1, nous trouvons $B_1AB_1 = B_1AB_2$. En remplaçant B_1A par δ des deux côtés, il reste $B_1 = B_2$. □

Lemme 4.86 ([82])**.**
Si $A \in \mathbb{M}(n, \mathbb{K})$, alors il existe au plus une matrice $B \in \mathbb{M}(n, \mathbb{K})$ telle que $AB = \delta$.

Démonstration. Soient des matrices $B, C \in \mathbb{M}(n, \mathbb{K})$ telles que $AB = AC = \delta$. Nous allons montrer que $B = C$.

Pour cela nous considérons les applications linéaires $f_A, f_B, f_C \in \mathrm{End}(\mathbb{K}^)$ associées par la proposition 4.68. Vu que $AB = \delta$, par la proposition 4.70, nous avons $f_A \circ f_B = f_{AB} = \mathrm{Id}$. La proposition 4.47 nous dit alors que f_A et f_B sont bijectives.

En particulier, vu que $\{e_i\}$ est une base, son image par f_B est une base par la proposition 4.36. La proposition 4.46 dit alors que $\{f_B(e_i)\}$ est une base. Nous décomposons $f_B(e_k) - f_C(e_k)$ dans cette base :

$$f_B(e_k) - f_C(e_k) = \sum_j \alpha_j f_B(e_j) \tag{4.166}$$

où les α_j dépendent à priori de k. Vu que $f_A \circ (f_B - f_C) = 0$, nous avons

$$0 = f_A\big(f_B(e_k) - f_C(e_k)\big) = \sum_j (f_A \circ f_B)(e_j) = \sum_j \alpha_j e_j. \tag{4.167}$$

Donc les α_j sont tous nuls.

Nous en déduisons que $f_B(e_k) = f_C(e_k)$, et donc $f_B = f_C$. Cela implique que $B = C$ par la proposition 4.68(4). □

35. La réduction de Gauss est le théorème 9.171, mais le lien avec ce que nous disons ici n'est peut-être pas directement clair.

36. En réalité, le processus tel que nous l'avons décrit ne s'arrête que lorsque la première colonne est remplie de zéros.

Proposition 4.87 ([82]).
Si $A, B \in \mathbb{M}(n, \mathbb{K})$ vérifient $AB = \delta$, alors $BA = \delta$.

Démonstration. L'astuce est de poser $C = BA - \delta + B$ et de montrer que $C = B$. Pour cela, un rapide calcul commence par montrer que

$$AC = ABA - A + AB = AB = \delta. \tag{4.168}$$

Donc C est également un inverse à droite de A. Le lemme 4.86 donne alors $C = B$. □

Corolaire 4.88.
Soit $A \in \mathbb{M}(n, \mathbb{K})$. Si il existe $B \in \mathbb{M}(n, \mathbb{K})$ tel que $AB = \delta$, alors A est inversible et son inverse est B.

Démonstration. Il s'agit d'une paraphrase de la proposition 4.87 et de la définition 4.85. □

Lemme 4.89.
Si une matrice A n'est pas inversible, alors le produit AB n'est inversible pour aucune matrice B.

Démonstration. Supposons que AB soit inversible. Alors

$$AB(AB^{-1}) = \delta, \tag{4.169}$$

ce qui dirait que $B(AB^{-1})$ serait un inverse de A. □

Proposition 4.90.
Une matrice est inversible si et seulement si son application linéaire associée est inversible. Dans ce cas, nous avons

$$f_A^{-1} = f_{A^{-1}}. \tag{4.170}$$

Démonstration. Dans le sens direct, si A est inversible nous avons $AA^{-1} = \delta$. Donc

$$f_A \circ f_{A^{-1}} = f_{AA^{-1}} = f_\delta = \mathrm{Id} \tag{4.171}$$

où nous avons utilisé la proposition 4.70 pour la composition et la proposition 4.71 pour l'identité. L'égalité (4.171) indique que f_A est inversible et que son inverse est $f_{A^{-1}}$.

Dans l'autre sens, l'application f_A^{-1} existe. Soit $B \in \mathbb{M}(n, \mathbb{K})$ sa matrice. Alors nous avons

$$f_\delta = \mathrm{Id} = f_A \circ f_B = f_{AB}. \tag{4.172}$$

Le fait que l'application $\psi\colon A \to f_A$ soit une bijection[37] implique que $AB = \delta$, c'est-à-dire que A est inversible et que $B = A^{-1}$. □

4.3.8 Inversibilité et déterminant

Proposition 4.91.
Une matrice au déterminant non nul est inversible.

Démonstration. Si A est une matrice telle que $\det(A) \neq 0$, alors la proposition 4.83 nous donne des matrices $G_1, \dots, G_N$ telles que

$$G_1 \dots G_N A = \delta. \tag{4.173}$$

Donc la matrice $G_1 \dots G_N$ est un inverse de A par le corolaire 4.88. □

Proposition 4.92.
Si une matrice A a une ligne ou une colonne de zéros, alors

(1) $\det(A) = 0$,

(2) A n'est pas inversible.

37. Proposition 4.68(4).

Démonstration. Par définition nous avons

$$\det(A) = \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^{n} A_{i\sigma(i)}. \tag{4.174}$$

Si la k^{e} ligne est nulle, alors $A_{k\sigma(k)} = 0$ pour tout σ. Donc tous les produits contiennent un facteur nul. Donc $\det(A) = 0$.

Pour toute matrice B nous avons

$$(AB)_{kk} = \sum_{l} A_{kl} B_{lk}. \tag{4.175}$$

Si la k^{e}ligne de A est nulle nous avons $(AB)_{kk} = 0$ et donc pas $AB = \delta$. Donc A n'est pas inversible. □

Proposition 4.93.
Une matrice dont le déterminant est nul n'est pas inversible.

Démonstration. Par la proposition 4.84, il existe des matrices de manipulation de lignes et de colonnes $G_1, \dots, G_N$ telles que la matrice $G_1 \dots G_N A$ ait une colonne de zéros. De là, la proposition 4.92 implique que la matrice

$$G_1 \dots G_N A \tag{4.176}$$

n'est pas inversible. Vu les déterminants des matrices G_i, la proposition 4.91 implique que $G_1 \dots G_N$ est inversible. Si A était inversible, nous aurions

$$G_1 \dots G_N A A^{-1} (G_1 \dots G_N)^{-1} = \delta, \tag{4.177}$$

c'est-à-dire que $A^{-1}(G_1 \dots G_N)^{-1}$ serait un inverse de la matrice (4.176). Cette dernière n'ayant pas d'inverse, nous concluons que A n'en a pas non plus. □

Théorème 4.94.
Une matrice sur un corps commutatif est inversible si et seulement si son déterminant est non nul.

Démonstration. Dans un sens c'est la proposition 4.91 et dans l'autre sens c'est la proposition 4.93. □

4.3.9 Quelques ensembles de matrices particuliers

Certains ensembles de matrices ont une importance particulière, que nous développerons plus tard.

Définition 4.95 (Groupe linéaire de matrices)**.**
On note $\mathrm{GL}(n, \mathbb{A})$ *l'ensemble des matrices carrées d'ordre* n *à coefficients dans* $\mathbb{A}$*, qui sont inversibles. En d'autres termes,* $\mathrm{GL}(n, \mathbb{A}) = U(\mathbb{M}(n, \mathbb{A}))$.

Définition 4.96 (Groupe orthogonal de matrices)**.**
On dit qu'une matrice A *est* ***orthogonale*** *si son inverse est sa transposée, c'est-à-dire si* $A^{-1} = A^t$*. On note* $\mathrm{O}(n, \mathbb{A})$ *l'ensemble des matrices carrées d'ordre* n *à coefficients dans* $\mathbb{A}$*, qui sont orthogonales.*

4.3.10 Déterminant et combinaisons de lignes et colonnes

Proposition 4.97.
Soient des matrices $A, B \in \mathbb{M}(n, \mathbb{K})$ *telles que* $\det(A) \neq 0 \neq \det(B)$*. Alors*

$$\det(AB) = \det(A)\det(B). \tag{4.178}$$

Démonstration. La proposition 4.83 nous donne des matrices de permutations de lignes et de colonnes $G_1, \dots, G_N$ et $G'_1, \dots, G'_N$ telles que[38]

$$G_1 \dots G_N A = \delta \tag{4.179a}$$
$$G'_1 \dots G'_N B = \delta. \tag{4.179b}$$

Nous avons

$$(G'_1 \dots G'_N) \underbrace{(G_1 \dots G_N) A}_{=\delta} B = \delta. \tag{4.180}$$

En prenant le déterminant des deux côtés et en tenant compte de (4.148),

$$1 = \det(\delta) = \det\big(G'_1 \dots G'_N G_1 \dots G_N AB\big) = \det(G'_1 \dots G'_N) \det(G_1 \dots G_N) \det(AB). \tag{4.181}$$

Mais en même temps, les équations 4.179 donnent

$$\det(G_1 \dots G_N) = \det(A)^{-1} \tag{4.182a}$$
$$\det(G'_1 \dots G'_N) = \det(B)^{-1}. \tag{4.182b}$$

Cela pour dire que

$$1 = \det(A)^{-1} \det(B)^{-1} \det(AB), \tag{4.183}$$

et donc ce qu'il nous fallait. □

Proposition 4.98.
Soient des matrices $A, B \in \mathbb{M}(n, \mathbb{K})$ telles que $\det(A) = 0$ et $\det(B) \neq 0$. Alors

$$\det(AB) = \det(BA) = \det(A) \det(B). \tag{4.184}$$

Démonstration. Il existe des matrices de manipulations de lignes et de colonnes $G_1, \dots, G_N$ telles que $G_1 \dots G_N B = \delta$. Donc

$$0 = \det(A) = \det(G_1 \dots G_N BA) = \det(G_1 \dots G_N) \det(BA). \tag{4.185}$$

Donc $\det(BA) = 0$. □

Proposition 4.99.
Soient des matrices A et B sur un corps commutatif. Alors

$$\det(AB) = \det(A) \det(B). \tag{4.186}$$

Démonstration. Les propositions 4.97 et 4.98 ont déjà fait une grosse partie du travail. Il ne reste que le cas où $\det(A) = \det(B) = 0$.

Dans ce cas, les matrices A et B ne sont pas inversibles (proposition 4.94). Le produit AB n'est alors pas inversible non plus[39]. La proposition 4.94, utilisée dans le sens inverse, nous dit alors que $\det(AB) = 0$.

Au final dans le cas $\det(A) = \det(B) = 0$ nous avons $0 = \det(AB) = \det(A) \det(B) = 0$. □

Faisons maintenant le cas général des manipulations de lignes et colonnes.

Proposition 4.100.
Soit une matrice carré $A \in \mathbb{M}(n, \mathbb{K})$. La matrice B obtenue par la substitution simultanée

$$C_j \to \sum_k a_{kj} C_k \tag{4.187}$$

a pour déterminant

$$\det(B) = \det(a) \det(A). \tag{4.188}$$

38. Les plus acharnés préciseront que pour avoir le même N des deux côtés, il a fallu compléter avec des matrices δ là où il y en avait le moins.

39. Citez le lemme 4.89 si vous voulez justifier ça.

Démonstration. L'élément B_{ij} de la matrice B est une combinaison linéaire de tous les éléments de sa ligne :

$$B_{ij} = \sum_k a_{kj} A_{ik} = (Aa)_{ij}. \tag{4.189}$$

Donc $B = Aa$. La proposition 4.99 nous dit alors que $\det(B) = \det(a)\det(A)$. □

4.3.11 Transvections

Nous nommons E_{ij} la matrice remplie de zéros sauf à la case ij qui vaut 1. Autrement dit

$$(E_{ij})_{kl} = \delta_{ik}\delta_{jl}. \tag{4.190}$$

Définition 4.101.
*Une **matrice de transvection** est une matrice de la forme*

$$T_{ij}(\lambda) = \mathrm{Id} + \lambda E_{ij} \tag{4.191}$$

avec $i \neq j$.

*Une **matrice de dilatation** est une matrice de la forme*

$$D_i(\lambda) = \mathrm{Id} + (\lambda - 1)E_{ii}. \tag{4.192}$$

Ici le $(\lambda - 1)$ sert à avoir λ et non $1 + \lambda$. C'est donc une matrice qui dilate d'un facteur λ la direction i tout en laissant le reste inchangé.

*Si σ est une permutation (un élément du groupe symétrique S_n) alors la **matrice de permutation** associée est la matrice d'entrées*

$$(P_\sigma)_{ij} = \delta_{i\sigma(j)}. \tag{4.193}$$

Lemme 4.102.
La matrice $T_{ij}(\lambda)A = (\mathbb{1} + \lambda E_{ij})A$ est la matrice A à qui on a effectué la substitution

$$L_i \to L_i + \lambda L_j. \tag{4.194}$$

La matrice $AT_{ij}(\lambda)$ est la substitution

$$C_j \to C_j + \lambda C_i. \tag{4.195}$$

La matrice AP_σ est la matrice A dans laquelle nous avons permuté les colonnes avec σ.

La matrice $P_\sigma A$ est la matrice A dans laquelle nous avons permuté les lignes avec σ^{-1}.

Démonstration. Calculons la composante kl de la matrice $E_{ij}A$:

$$(E_{ij}A)_{kl} = \sum_m (E_{ij})_{km} A_{ml} \tag{4.196a}$$

$$= \sum_m \delta_{ik}\delta_{jm} A_{ml} \tag{4.196b}$$

$$= \delta_{ik} A_{jl}. \tag{4.196c}$$

C'est donc la matrice pleine de zéros, sauf la ligne i qui est donnée par la ligne j de A. Donc effectivement la matrice

$$A + \lambda E_{ij} A \tag{4.197}$$

est la matrice A à laquelle on a substitué la ligne i par la ligne i plus λ fois la ligne j.

En ce qui concerne l'autre assertion sur les transvections, le calcul est le même et nous obtenons

$$(AE_{ij}) = A_{ki}\delta_{jl}. \tag{4.198}$$

Pour les matrices de permutation, nous avons

$$(AP_\sigma)_{kl} = A_{k\sigma(l)} \tag{4.199}$$

et

$$(P_\sigma A)_{kl} = \sum_m \delta_{k\sigma(m)} A_{ml} = \sum_m \delta_{\sigma^{-1}(k)m} A_{ml} = A_{\sigma^{-1}(k)l}. \tag{4.200}$$

□

4.3.12 Mineur, rang

Pour la définition du rang d'une matrice, nous en donnons une qui est clairement inspirée de l'application linéaire associée.

Définition 4.103 ([82]).
*Le **rang** d'une matrice de $\mathbb{M}(n,\mathbb{K})$ est la dimension de la partie de $\mathbb{K}^n$ engendrée par ses colonnes.*

Il est possible d'exprimer le rang d'une matrice de façon plus « intrinsèque » via le concept de mineur.

Définition 4.104 ([83]).
Les mineurs d'une matrice sont les déterminants de ses sous-matrices carrées.

Dans la suite nous désignerons souvent par le mot « mineur » la sous-matrice carrée elle-même au lieu de son déterminant.

Proposition 4.105.
Le rang d'une matrice est la taille de son plus grand mineur non nul.

Lorsque A est une matrice, nous notons f_A l'application linéaire associée à la matrice A par l'application (4.82).

Le rang d'une matrice est également le rang de son application linéaire associée.

Proposition 4.106 (Rang).
Si A est une matrice carré, alors

$$\operatorname{rang}\big(f_A\big) = \operatorname{rang}(A). \tag{4.201}$$

À gauche, c'est le rang d'une application linéaire 4.38 ; à droite c'est le rang d'une matrice 4.103.

Lemme 4.107.
Soit $\mathbb{K}$ un corps commutatif[40]*. Si A est une matrice carrée d'ordre n et de rang r à coefficients dans $\mathbb{K}$, alors il existe des vecteurs $(x_i)_{i=1,\dots,n}$ formant une base de $\mathbb{K}^n$ tels que*

$$f_A(x_i) \neq 0 \tag{4.202}$$

pour $x \leqslant r$ et

$$f_A(x_i) = 0 \tag{4.203}$$

pour $i > r$.

Démonstration. Soit V le sous-espace de $\mathbb{K}^n$ engendré par les colonnes de A. Nous considérons la base canonique $\{e_i\}$ de $\mathbb{K}^n$, ainsi que v_i le vecteur créé par la i^{e}colonne de A. Nous avons

$$v_i = f_A(e_i). \tag{4.204}$$

Les vecteurs v_i engendrent V, donc nous pouvons en extraire une base par le théorème 4.15(1). Soit donc $\{v_j\}_{i\in J}$ une base de V avec $J \subset \{1,\dots,n\}$.

La base de $\mathbb{K}^n$ que nous cherchons commence par les vecteurs $\{e_j\}_{j\in J}$. Ces vecteurs vérifient $f_A(e_j) = v_j \neq 0$ parce que des vecteurs d'une base ne sont jamais nuls.

Pour la suite de la base, nous pourrions penser au théorème de la base incomplète[41], mais les vecteurs ainsi complétant la base ne sont pas garantis de s'annuler par f_A. Voir l'exemple 4.108.

L'idée est d'utiliser le noyau de f_A qui est un sous-espace vectoriel par la proposition 4.32. Soit une base[42] $\{z_k\}$ de $\ker(f)$. Les vecteurs $\{e_j\}_{j\in J}$ forment une base de $\operatorname{Image}(f_A)$. Vu que les z_i forment une base de $\ker(f_A)$, le théorème du rang 4.39 dit alors que $\{e_j\}_{j\in J} \cup \{z_k\}$ est une base de $\mathbb{K}^n$.

40. Comme toujours.
41. Théorème 4.11(2).
42. Cette base contient $n-r$ éléments, mais ce n'est pas très important pour la suite.

Il y a r éléments dans J parce que l'espace engendré par les colonnes de A est de dimension r par hypothèse. Donc il y a $n-r$ éléments dans les z_k pour que le tout ait le bon nombre d'éléments. □

Exemple 4.108.
Soit la matrice

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}. \tag{4.205}$$

Elle est de rang 1. En suivant l'idée de la démonstration, nous commençons la base de $\mathbb{R}^2$ par le vecteur e_1 qui vérifie

$$f_A(e_1) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}. \tag{4.206}$$

L'utilisation du théorème de la base incomplète ne permet pas de trouver un second vecteur de base v tel que $f_A(v) = 0$. En effet ce théorème donne juste l'existence d'une completion de la base, mais pas de propriétés particulières de la base obtenue. Elle pourrait donner $v = e_2$ comme second vecteur de base. Mais alors

$$f_A(v) = f_A(e_2) = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \neq 0. \tag{4.207}$$

Au contraire, le noyau de f_A est donné par le sous-espace engendré par $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Une base convenable est donc $\{e_1, e_1 - e_2\}$. △

Proposition 4.109.
Le rang d'une application linéaire est égal au rang de sa matrice dans n'importe quelle base.

4.3.13 Matrices équivalentes et semblables

Définition 4.110.
Deux relations d'équivalence entre les matrices.

*(1) Deux matrices A et B sont **équivalentes** dans $\mathbb{M}(n, \mathbb{K})$ s'il existe $P, Q \in \mathrm{GL}(n, \mathbb{K})$ telles que $A = PBQ^{-1}$.*

*(2) Deux matrices sont **semblables** s'il existe une matrice $P \in \mathrm{GL}(n, \mathbb{K})$ telle que $A = PBP^{-1}$.*

Lemme 4.111.
Une matrice de rang[43] *r dans $\mathbb{M}(n, \mathbb{K})$ est équivalente à la matrice par blocs*

$$J_r = \begin{pmatrix} \mathbb{1}_r & 0 \\ 0 & 0 \end{pmatrix}. \tag{4.208}$$

Démonstration. Nous devons prouver que pour toute matrice $A \in \mathbb{M}(n, \mathbb{K})$ de rang r, il existe $P, Q \in \mathrm{GL}(n, \mathbb{K})$ telles que $QAP = J_r$. Soit $\{e_i\}$ la base canonique de $\mathbb{K}^n$, puis $\{f_i\}$ une base telle que $Af_i = 0$ dès que $i > r$, qui existe par le lemme 4.107.

Nous considérons la matrice inversible P telle que $Pe_i = f_i$; ses colonnes sont donc précisément les f_i, si bien que

$$APe_i = Af_i = \begin{cases} 0 & \text{si } i > r \\ \neq 0 & \text{sinon.} \end{cases} \tag{4.209}$$

La matrice AP se présente donc sous la forme

$$AP = \begin{pmatrix} M & 0 \\ * & 0 \end{pmatrix} \tag{4.210}$$

43. Définition 4.105.

où M est une matrice $r \times r$. Nous considérons maintenant une base $\{g_i\}_{i=1,\dots,n}$ dont les r premiers éléments sont les r premières colonnes de AP et une matrice inversible Q telle que $Qg_i = e_i$. Alors

$$QAPe_i = \begin{cases} e_i & \text{si } i < r \\ 0 & \text{sinon.} \end{cases}. \tag{4.211}$$

Cela signifie que QAP est la matrice J_r. □

Corolaire 4.112 (Équivalence et rang)**.**
Deux matrices sont équivalentes[44] *si et seulement si elles sont de même rang.*

Démonstration. D'abord il y a des implicites dans l'énoncé. Vu que nous voulons soit par hypothèse soit par conclusion que les matrices A et B soient équivalentes, nous supposons qu'elles ont même dimension. Soient donc A et B deux matrices carrées d'ordre n.

Par le lemme 4.111, deux matrices de même rang r sont équivalentes à J_r. Elles sont donc équivalentes entre elles.

Inversement, supposons que A et B soient deux matrices équivalentes : $A = PBQ^{-1}$ avec P et Q inversibles. Alors

$$\begin{aligned} \text{Image}(PBQ^{-1}) &= \{PBQ^{-1}v \text{ tel que } v \in \mathbb{K}^n\} && (4.212a)\\ &= PB\underbrace{\{Q^{-1}v \text{ tel que } v \in \mathbb{K}^n\}}_{=\mathbb{K}^n} && (4.212b)\\ &= P\big(B(\mathbb{K}^n)\big). && (4.212c) \end{aligned}$$

L'ensemble $B(\mathbb{K}^n)$ est un sous-espace vectoriel de $\mathbb{K}^n$. Vu que le rang de P est maximum, la dimension de $P\big(B(\mathbb{K}^n)\big)$ est la même que celle de $B(\mathbb{K}^n)$. Par conséquent

$$\dim\Big(\text{Image}(PBQ^{-1})\Big) = \dim\big(B(\mathbb{K}^n)\big) = \text{rang}(B). \tag{4.213}$$

Le membre de gauche de cela n'est autre que $\text{rang}(A) = \dim\big(\text{Image}(PBQ^{-1})\big)$. □

4.3.14 Algorithme des facteurs invariants

Proposition 4.113 (Algorithme des facteurs invariants[54])**.**
Soit $(\mathbb{A}, \delta)$ *un anneau euclidien muni de son stathme et* $U \in \mathbb{M}(n \times m, \mathbb{A})$*. Alors il existe* $d_1, \dots, d_s \in \mathbb{A}^*$ *et des matrices* $P \in \text{GL}(m, \mathbb{A})$*,* $Q \in \text{GL}(n, \mathbb{A})$ *tels que nous ayons*

$$U = P\begin{pmatrix} d_1 & & & \\ & \ddots & & 0 \\ & & d_s & \\ & 0 & & 0 \end{pmatrix}Q \tag{4.214}$$

avec $d_i \mid d_{i+1}$ *pour tout* i*.*

Démonstration. Nous allons donner la preuve plus ou moins sous forme d'algorithme.

D'abord si $U = 0$ c'est bon, on a la réponse. Sinon, nous prenons l'élément (i_0, j_0) dont le stathme est le plus petit et nous l'amenons en $(1, 1)$ par les permutations

$$\begin{aligned} C_1 &\leftrightarrow C_{j_0} \\ L_1 &\leftrightarrow L_{i_0} \end{aligned} \tag{4.215}$$

Ensuite nous traitons la première colonne jusqu'à amener des zéros partout en dessous de u_{11} de la façon suivante : pour chaque ligne successivement nous calculons la division euclidienne

$$u_{i1} = qu_{11} + r_i, \tag{4.216}$$

44. Définition 4.110(1).

et nous faisons

$$L_i \to L_i - qL_1, \tag{4.217}$$

c'est-à-dire que nous enlevons le maximum possible et il reste seulement r_i en u_{i1}. Vu que le but est de ne laisser que des zéros dans la première colonne, si le reste n'est pas zéro nous ne sommes pas content [45]. Dans ce cas nous permutons $L_1 \leftrightarrow L_i$, ce qui aura pour effet de strictement diminuer le stathme de u_{11} parce qu'on va mettre en u_{11} le nombre r_i dont le stathme est strictement plus petit que celui de u_{11}.

En faisant ce jeu de division euclidienne puis échange, on diminue toujours le stathme de u_{11}, donc ça finit par s'arrêter, c'est-à-dire qu'à un certain moment la division euclidienne de u_{i1} par u_{11} va donner un reste zéro et nous serons content.

Une fois la première colonne ramenée à la forme

$$C_1 = \begin{pmatrix} u_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{4.218}$$

nous faisons tout le même jeu avec la première ligne en faisant maintenant des sommes divisions et permutations de colonnes. Notons que ce faisant nous ne changeons plus la première colonne.

En fin de compte nous trouvons une matrice [46]

$$U = \begin{pmatrix} u_{11} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A & \\ 0 & & & \end{pmatrix} \tag{4.219}$$

Si l'élément u_{11} ne divise pas un des éléments de A, disons a_{ij}, alors nous faisons

$$C_1 \to C_1 - C_j. \tag{4.220}$$

Cela nous détruit un peu la première colonne, mais ne change pas u_{11}. Nous avons maintnant

$$U = \begin{pmatrix} u_{11} & 0 & \dots & 0 \\ 0 & & & \\ * & & & \\ u_{ij} & & A & \\ * & & & \\ 0 & & & \end{pmatrix} \tag{4.221}$$

Et nous refaisons tout le jeu depuis le début. Cependant lorsque nous allons nous attaquer à la ligne i, u_{11} ne divisera pas u_{ij}, ce qui donnera lieu à une division euclidienne et un échange $L_1 \leftrightarrow L_i$. L'échange consistant à mettre r_i à la place de u_{11} et inversement diminuera encore strictement le stathme. Encore une fois nous allons travailler jusqu'à avoir la matrice sous la forme

$$U = \begin{pmatrix} u_{11} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A & \\ 0 & & & \end{pmatrix}, \tag{4.222}$$

sauf que cette fois le stathme de u_{11} est strictement plus petit que la fois précédente. Si u_{11} ne divise toujours pas tous les éléments de A, nous recommençons encore et encore. En fin de compte nous finissons par avoir une matrice de la forme (4.222) avec u_{11} qui divise tous les éléments de A.

45. S'il est zéro, nous passons à la ligne suivante

46. Nous nommons toujours par la même lettre U la matrice originale et la modifiée, comme il est d'usage en informatique.

Une fois que cela est fait, il faut continuer en recommençant tout sur la matrice A. Nous avons maintenant

$$U = \begin{pmatrix} u_{11} & & 0 \\ & u_{22} & \\ 0 & & B \end{pmatrix}. \tag{4.223}$$

Sous cette forme nous avons $u_{11} \mid u_{22}$ et u_{11} divise tous les éléments de B. En effet u_{11} divisant tous les éléments de A, il divise toutes les combinaisons de ces éléments. Or tout l'algorithme ne consiste qu'à prendre des combinaisons d'éléments.

Nous finissons donc bien sûr une matrice comme annoncée. De plus n'ayant effectué que des combinaisons de lignes, nous avons seulement multiplié par des matrices inversibles (lemme 4.102).

□

4.4 Changement de base

Soit un espace vectoriel V muni de deux bases $(e_i)_{i=1,\dots,n}$ et $(f_\alpha)_{\alpha=1,\dots,n}$. Les deux bases sont liées entre elles par

$$f_\alpha = \sum_i Q_{i\alpha} e_i. \tag{4.224}$$

Ici Q n'est pas une application linéaire $V \to V$: Q est seulement un tableau de nombres, donnant les coordonnées des vecteurs f_α dans la base de e_i. Éventuellement Q peut être vu comme une application linéaire $\mathbb{K}^n \to \mathbb{K}^n$.

Dans la suite nous nommerons Q^{-1} la matrice inverse de Q. Inverse au sens des bêtes tableaux de nombres, sans interprétations en tant qu'application linéaire. De même pour Q^t qui est la transposée de Q.

4.4.1 Changement de base : vecteurs de base

Lemme 4.114.
Soit un espace vectoriel V sur $\mathbb{K}$ ainsi que deux bases $(e_i)_{i=1,\dots,n}$, $(f_\alpha)_{\alpha=1,\dots,n}$ de V liées par $f_\alpha = \sum_i Q_{i\alpha} e_i$. Alors

$$e_i = \sum_\alpha Q^{-1}_{\alpha i} f_\alpha. \tag{4.225}$$

Démonstration. Nous multiplions l'égalité $f_\alpha = \sum_i Q_{i\alpha} e_i$ par le nombre[47] $Q^{-1}_{\alpha j} \in \mathbb{K}$ et nous sommons sur α :

$$\sum_\alpha Q^{-1}_{\alpha j} f_\alpha = \sum_{i\alpha} (A_{i\alpha} Q^{-1}_{\alpha j}) e_i = e_j. \tag{4.226}$$

□

4.4.2 Changement de base : coordonnées

Proposition 4.115.
Soit un espace vectoriel V sur $\mathbb{K}$. Soient deux bases $(e_i)_{i=1,\dots,n}$ et $(f_\alpha)_{\alpha=1,\dots,n}$ liées par $f_\alpha = \sum_i Q_{i\alpha} e_i$. Nous considérons un même vecteur dans les deux bases : $\sum_i x_i e_i = \sum_\alpha y_\alpha f_\alpha$. Alors

(1) $y_\alpha = \sum_i Q^{-1}_{\alpha i} x_i$

(2) $x_i = \sum_\alpha Q_{i\alpha} y_\alpha$.

Démonstration. Soit un vecteur $x \in V$. Il peut être écrit dans les deux bases :

$$x = \sum_i x_i e_i = \sum_\alpha y_\alpha f_\alpha. \tag{4.227}$$

47. Attention à la bonne interprétation de ce nombre : on fait bien référence à l'élément situé en (α, j) de la matrice Q^{-1}, et pas autre chose.

En remplaçant e_i par sa valeur (4.225) nous avons l'égalité

$$\sum_{i\alpha} x_i Q^{-1}_{\alpha i} f_\alpha = \sum_\alpha y_\alpha f_\alpha. \tag{4.228}$$

Vu que les f_α sont linéairement indépendants, l'égalité des sommes donne l'égalité de chacun de termes :

$$y_\alpha = \sum_i x_i Q^{-1}_{\alpha i}. \tag{4.229}$$

En identifiant $x \in V$ au vecteur dans $\mathbb{K}^n$ de ses coordonnées dans la base $\{e_i\}$ nous pouvons écrire

$$y_\alpha = (Q^{-1}x)_\alpha, \tag{4.230}$$

Le point (1) est prouvé.

En ce qui concerne le point (2), nous repartons encore de (4.227), mais nous y substituons la définition des f_α :

$$\sum_i x_i e_i = \sum_{\alpha i} y_\alpha Q_{i\alpha} e_i. \tag{4.231}$$

Vous voulez des détails ? Allez, une étape de plus que le strict nécessaire : nous écrivons

$$\sum_i \big(x_i - \sum_\alpha y_\alpha Q_{i\alpha}\big) e_i = 0. \tag{4.232}$$

Par linéaire indépendance des e_i, nous avons annulation de tous les coefficients, c'est à dire

$$x_i = \sum_\alpha Q_{i\alpha} y_\alpha, \tag{4.233}$$

comme annoncé. □

4.116.
Attention à l'ordre des indices dans la dernière égalité : la matrice Q vient avec les indices dans l'ordre $i\alpha$, tandis que la matrice Q^{-1} vient avec les indices dans l'ordre opposé : αi. C'est pour cela qu'il est intéressant de noter avec des lettres latines les indices se rapportant à la première base et avec des lettres grecques ceux se rapportant à la seconde base.

4.117.
Les formules de changement de coordonnées de la proposition 4.115 s'écrivent souvent de la façon suivante :

(1) $y_\alpha = (Q^{-1}x)_\alpha$
(2) $y = Q^{-1}x$.
(3) $x_i = (Qy)_i$
(4) $x = Qy$

Ces égalités reposent sur un petit paquet d'abus de notations qu'il convient de bien comprendre. Ici, x et y sont les éléments de $\mathbb{K}^n$ donnés par les composantes de x dans les bases $\{e_i\}$ et $\{f_\alpha\}$, et Q est vu comme une matrice, un opérateur linéaire sur $\mathbb{K}^n$. Autrement dit, le choix des bases permet d'identifier V avec $\mathbb{K}^n$ et la matrice Q avec l'application linéaire f_Q de la proposition 4.68.

4.4.3 Changement de base : matrice d'une application linéaire

Proposition 4.118.
Soit une application linéaire $t\colon V \to V$ de matrices A et B dans les bases $\{e_i\}$ et $\{f_\alpha\}$. Si les bases sont liées par

$$f_\alpha = \sum_i Q_{i\alpha} e_i, \tag{4.234}$$

alors les matrices A et B sont liées par

$$B = Q^{-1}AQ. \tag{4.235}$$

Démonstration. L'hypothèse sur le fait que A et B sont les matrices de t signifie que pour tout $x \in V$,

$$t(x) = \sum_{ij} A_{ji} x_i e_j = \sum_{\alpha\beta} B_{\alpha\beta} y_\beta f_\alpha. \tag{4.236}$$

En remplaçant e_j par son expression (4.225) en termes des f_α et x_i par son expression $x_i = (Qy)_i$ (proposition 4.115), nous avons

$$(By)_\alpha = \sum_{ij\alpha} A_{ji} (Qy)_i Q^{-1}_{\alpha j} f_\alpha \tag{4.237a}$$

$$= \sum_{i\alpha} (Q^{-1}A)_{\alpha i} (Qy)_i f_\alpha \tag{4.237b}$$

$$= \sum_{\alpha} (Q^{-1}AQy)_\alpha f_\alpha. \tag{4.237c}$$

Vu que les f_α forment une base nous en déduisons $Q^{-1}AQy = By$. Et vu que y est un élément quelconque de $\mathbb{K}^n$, nous en déduisons l'égalité de matrices

$$B = Q^{-1}AQ. \tag{4.238}$$

□

Il s'agit bien d'une égalité de matrices, ou à la limite d'applications linéaires sur $\mathbb{K}^n$, et non d'une égalité d'application linéaire sur V.

4.5 Espaces de polynômes

Attention : les polynômes en soi sont définis par la définition 3.86.

Pour chaque $k > 0$ donné nous définissons

$$\mathcal{P}^k_{\mathbb{R}} = \{p : \mathbb{R} \to \mathbb{R} \,|\, p : x \mapsto a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k,\ a_i \in \mathbb{R},\ \forall i = 0, \ldots, k\}. \tag{4.239}$$

Il est facile de se convaincre que la somme de deux polynômes de degré inférieur ou égal à k est encore un polynôme de degré inférieur ou égal à k. En outre il est clair que la multiplication par un scalaire ne peut pas augmenter le degré d'un polynôme. L'ensemble $\mathcal{P}^k_{\mathbb{R}}$ est donc un espace vectoriel muni des opérations héritées de $\mathcal{P}_{\mathbb{R}}$.

La base canonique de l'espace $\mathcal{P}^k_{\mathbb{R}}$ est donné par les monômes $\mathcal{B} = \{x \mapsto x^j \,|\, j = 0, \ldots, k\}$. Le fait que cela soit une base est vraiment facile à démontrer et est un exercice très utile si vous ne l'avez pas encore vu dans un cours précédent.

Nous allons maintenant étudier trois applications linéaires de $\mathcal{P}^k_{\mathbb{R}}$ vers des autres espaces vectoriels

L'isomorphisme canonique $\phi : \mathcal{P}^k_{\mathbb{R}} \to \mathbb{R}^{k+1}$ Nous définissons ϕ par les relations suivantes

$$\phi(x^j) = e_{j+1}, \qquad \forall j \in \{0, \ldots, k\}.$$

Cela veut dire que pour tout p dans $\mathcal{P}^k_{\mathbb{R}}$, avec $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k$, l'image de p par ϕ est

$$\phi(p) = \phi\left(\sum_{j=0}^{k} a_j x^j\right) = \sum_{j=0}^{k} a_j e_{j+1}.$$

Exemple 4.119.
Soit $k = 5$ on a

$$\phi(-8 - 7x - 4x^2 + 4x^3 + 2x^5) = \begin{pmatrix} -8 \\ -7 \\ -4 \\ 4 \\ 0 \\ 2 \end{pmatrix}. \tag{4.240}$$

△

Cette application est clairement bijective et respecte les opérations d'espace vectoriel, donc elle est un isomorphisme d'espaces vectoriels. L'existence d'un isomorphisme entre $\mathcal{P}_{\mathbb{R}}^k$ et $\mathbb{R}^{k+1}$ est un cas particulier du théorème qui dit que pour chaque m dans $\mathbb{N}_0$ fixée, tous les espaces vectoriels sur $\mathbb{R}$ de dimension m sont isomorphes à $\mathbb{R}^m$. Vous connaissez peut être déjà ce théorème depuis votre cours d'algèbre linéaire.

La dérivation $d : \mathcal{P}_{\mathbb{R}}^k \to \mathcal{P}_{\mathbb{R}}^{k-1}$ L'application de dérivation d fait exactement ce qu'on s'attend d'elle

$$d(x^0) = d(1) = 0, \qquad d(x^j) = jx^{j-1}, \quad \forall j \in \{1, \dots, k\}.$$

Cette application n'est pas injective, parce que l'image de p ne dépend pas de la valeur de a_0, donc si deux polynômes sont les mêmes à une constante près ils auront la même image par d.

Exemple 4.120.
Soit $k = 3$ on a

$$d(-8 - 12x + 4x^3) = -12(1) + 4(3x^2) = -12 + 12x^2. \tag{4.241}$$

Noter que $d(-30 - 12x + 4x^3) = d(-8 - 12x + 4x^3)$. Cela confirme, comme mentionné plus haut, que la dérivée n'est pas injective. △

L'intégration $I : \mathcal{P}_{\mathbb{R}}^k \to \mathcal{P}_{\mathbb{R}}^{k+1}$ Nous pouvons définir une application que est «à une constante prés» l'application inverse de la dérivation. Cette application est définie sur les éléments de base par

$$I(x^j) = \frac{x^{j+1}}{j+1}. \tag{4.242}$$

Bien entendu la raison d'être et la motivation de cette définition apparaîtra lorsque nous développerons une théorie générale de l'intégration.

Exemple 4.121.
Soit $k = 4$ on a

$$I(6 + 2x + x^2 + x^4) = 6x + x^2 + \frac{x^3}{3} + \frac{x^5}{5}. \tag{4.243}$$

△

Remarquez que, étant donné que dans la définition de I nous avons décidé d'intégrer entre zéro et x, tous les polynômes dans $\mathcal{P}_{\mathbb{R}}^{k+1}$ qui sont l'image par I d'un polynôme de $\mathcal{P}_{\mathbb{R}}^k$ ont $a_0 = 0$. Cela veut dire que nous pouvons générer toute l'image de I en utilisant un sous-ensemble de la base canonique de $\mathcal{P}_{\mathbb{R}}^{k+1}$, en particulier $\mathcal{B}_1 = \{x \mapsto x^j \,|\, j = 1, \dots, k\} \subset \mathcal{B}$ nous suffira. Cela n'est guère surprenant, parce que l'image par une application linéaire d'un espace vectoriel de dimension finie ne peut pas être un espace de dimension supérieure.

Les applications de dérivation et intégration correspondent évidemment à des applications linéaires de $\mathcal{P}_{\mathbb{R}}$ dans lui-même.

L'espace de tous les polynômes étant de dimension infinie, il peut servir de contre-exemple assez simple. Dans la sous-section 7.10.2, nous verrons que toutes les normes ne sont pas équivalentes sur l'espace des polynômes.

4.5.1 Projection et orthogonalité

Proposition 4.122 (Produit scalaire dans $\mathbb{R}^n$)**.**
Si $x, y \in \mathbb{R}^n$, nous définissons

$$x \cdot y = \sum_{i=1}^{n} x_i y_i. \tag{4.244}$$

Cela est un produit scalaire qui vérifie

Symétrie $X \cdot Y = Y \cdot X$;

Linéarité $(\lambda X + \mu X') \cdot Y = \lambda(X \cdot Y) + \mu(X' \cdot Y)$ *pour tout* λ *et* μ *dans* $\mathbb{R}$ *;*
Défini positif $X \cdot X \geqslant 0$ *et* $X \cdot X = 0$ *si et seulement si* $X = 0$.
Ce produit scalaire est le ***produit scalaire*** *qui sera toujours considéré. C'est de lui qui découle toujours la norme, et la topologie de* $\mathbb{R}^n$.

Proposition 4.123.
Soit une application linéaire $A\colon \mathbb{R}^n \to \mathbb{R}^n$. *Nous considérons le produit scalaire usuel sur* $\mathbb{R}^n$. *Alors :*

(1) Les éléments de matrice de A *sont donnés par* $A_{ij} = e_i \cdot Ae_j$.
(2) Nous avons la formule $x \cdot Ay = \sum_{kl} A_{kl}x_k y_l$.

Démonstration. Pour Ae_j nous utilisons la formule 4.88 avec des notations plus décontractées : $Ae_j = \sum_k A_{kj}e_k$. Ensuite nous faisons un calcul avec la formule (4.244) :

$$e_i \cdot Ae_j = e_i \cdot \sum_k A_{kj}e_k = \sum_k A_{kj}\delta_{ik} = A_{ij}. \tag{4.245}$$

La seconde formule à prouver est du même tonneau, en utilisant cette fois la formule (4.89) :

$$x \cdot Ay = \sum_k x_k(Ay)_k = \sum_{kl} x_k A_{kl} y_l = \sum_{kl} A_{kl}x_k y_l. \tag{4.246}$$

□

Définition 4.124.
La ***norme*** *du vecteur* X, *notée* $\|X\|$, *est définie par*

$$\|X\| = \sqrt{X \cdot X} = \sqrt{x^2 + y^2 + z^2} \tag{4.247}$$

si $X = (x, y, z)$. *Cette norme sera parfois nommée « norme euclidienne ».*

Cette définition est motivée par le théorème de Pythagore. Le nombre $X \cdot X$ est bien la longueur de la « flèche » X. Plus intrigante est la définition suivante :

Définition 4.125.
Deux vecteurs X *et* Y *sont* ***orthogonaux*** *si* $X \cdot Y = 0$.

Cette définition de l'orthogonalité est motivée par la proposition suivante.

Proposition 4.126.
Si nous écrivons $\mathtt{proj}_Y$ *l'opération de projection sur la droite qui sous-tend* Y, *alors nous avons*

$$\|\mathtt{proj}_Y X\| = \frac{X \cdot Y}{\|Y\|}. \tag{4.248}$$

Démonstration. Les vecteurs X et Y sont des flèches dans l'espace. Nous pouvons choisir un système d'axe orthogonal tel que les coordonnées de X et Y soient

$$X = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}, \quad Y = \begin{pmatrix} l \\ 0 \\ 0 \end{pmatrix} \tag{4.249}$$

où l est la longueur du vecteur Y. Pour ce faire, il suffit de mettre le premier axe le long de Y, le second dans le plan qui contient X et Y, et enfin le troisième axe dans le plan perpendiculaire aux deux premiers.

Un simple calcul montre que $X \cdot Y = xl + y \cdot 0 + 0 \cdot 0 = xl$. Par ailleurs, nous avons $\|\mathtt{proj}_Y X\| = x$. Par conséquent,

$$\|\mathtt{proj}_Y X\| = \frac{X \cdot Y}{l} = \frac{X \cdot Y}{\|Y\|}. \tag{4.250}$$

□

Corolaire 4.127.
Si la norme de Y est 1, alors le nombre $X \cdot Y$ est la longueur de la projection de X sur Y.

Démonstration. Poser $\|Y\| = 1$ dans la proposition 4.126. □

4.6 Dualité

Proposition 4.128.
Si A est la matrice d'une application linéaire, alors le rang de cette application linéaire est égal au rang de A, c'est-à-dire à la taille de la plus grande matrice carrée de déterminant non nul contenue dans A.

Définition 4.129.
Soit E un espace vectoriel sur $\mathbb{K}$.

*Une **forme linéaire** sur E est une application linéaire de E sur son corps de base $\mathbb{K}$.*

*Le **dual algébrique** de E, noté E^*, l'ensemble des formes linéaires sur E. Ainsi, $E^* = \mathrm{GL}(E, \mathbb{K})$.*

Nous verrons plus tard qu'en dimension infinie, les applications linéaires ne sont pas toujours continues. Nous définirons donc aussi un concept de dual topologique. Voir la proposition 10.18, la remarque 10.30 et la définition 10.32.

Définition 4.130.
*Si E est un espace vectoriel et si $\{e_i\}$ est une base de E, alors nous définissons la **base duale** de E^* par*

$$e_i^*(e_j) = \delta_{ij} \tag{4.251}$$

est sa prolongation par linéarité.

Notons que si $v \in E$ est un vecteur, ça n'a aucun sens à priori de parler de v^*. Il s'agit bien de définir *toute* la base $\{e_i^*\}$ à partir de toute la base $\{e_i\}$.

Lemme 4.131 ([84])**.**
Soit un espace vectoriel E de dimension finie.

(1) Si α est une forme linéaire non nulle sur E, alors il existe $x \in E$ tel que $\alpha(x) = 1$.

(2) Si $x \neq 0$ dans E, il existe une forme linéaire α sur E telle que $\alpha(x) = 1$.

Démonstration. En deux parties.

Pour (1) Vu que α est non nulle nous pouvons considérer $v \in E$ tel que $\alpha(v) \neq 0$. Alors en posant $x = \alpha(v)^{-1}v$, nous avons le résultat.

Pour (2) Soit un vecteur non nul que nous écrivons sous la forme $x = \sum_{i=1}^n x_i e_i$ (pour une certaine base $(e_i)_{i=1,\dots,n}$ de E). Supposons que $x_k \neq 0$. Alors la forme

$$\begin{aligned} \alpha\colon E &\to \mathbb{K} \\ y &\mapsto y_k/x_k \end{aligned} \tag{4.252}$$

fait l'affaire.

□

Lemme 4.132.
Soit un espace vectoriel de dimension finie E sur le corps $\mathbb{K}$. Si $(\alpha_1, \dots, \alpha_n)$ est une base de E^*, alors l'application

$$\begin{aligned} \Phi\colon E &\to \mathbb{K}^n \\ x &\mapsto \big(\alpha_1(x), \dots, \alpha_n(x)\big) \end{aligned} \tag{4.253}$$

est un isomorphisme d'espaces vectoriels.

Démonstration. En deux parties.

Φ est injective Soit $z \in \ker(\Phi)$. Nous avons $\alpha_i(z) = 0$ pour tout i. Si $z \neq 0$, alors le lemme 4.131 dit qu'il existe $\beta \in E^*$ tel que $\beta(z) \neq 0$.

Décomposons ce β dans la base de (α_i) :

$$\beta = \sum_{i=1}^{n} \beta_i \alpha_i. \tag{4.254}$$

Alors nous avons

$$0 = \beta(z) = \sum_{i=1}^{n} \beta_i \underbrace{\alpha_i(z)}_{=0} = 0. \tag{4.255}$$

Contradiction. Donc $\ker(\Phi) = \{0\}$ et Φ est injective.

Φ est surjective Les espaces vectoriels E, E^* et $\mathbb{K}^n$ ont tout trois dimension n. Donc Φ est une application linéaire injective entre deux espaces de même dimension. Elle est donc surjective par le corollaire 4.41.

□

Proposition-définition 4.133 ([84]).
*Soit un espace vectoriel E de dimension finie sur le corps $\mathbb{K}$. Toute base du dual E^* est duale d'une unique base de E. Cette base est dite **préduale**.*

Démonstration. Nous considérons une base $\mathcal{F} = (\alpha_1, \dots, \alpha_n)$ de E^*. Nous devons prouver qu'il existe une unique base de E dont la base duale est $\mathcal{F}$.

Existence Le lemme 4.132 nous indique que l'application

$$\begin{aligned} \Phi \colon E &\to \mathbb{K}^n \\ x &\mapsto \big(\alpha_1(x), \dots, \alpha_n(x)\big) \end{aligned} \tag{4.256}$$

est un isomorphisme d'espaces vectoriels.

Soit la base canonique de $\mathbb{K}^n$: $(\epsilon_1, \dots, \epsilon_n)$. Vu que Φ est une isomorphisme, $\big(\Phi^{-1}(\epsilon_i)\big)_{i=1,\dots,n}$ est une base de E. Nous allons montrer qu'elle est préduale de (α_i). Nous posons $e_i = \Phi^{-1}(\epsilon_i)$ et nous calculons :

$$\begin{aligned} \alpha_i(e_j) &= \alpha_i\big(\Phi^{-1}(\epsilon_j)\big) && (4.257a) \\ &= \Phi\big(\Phi^{-1}(\epsilon_i)\big)_j && (4.257b) \\ &= (\epsilon_j)_i && (4.257c) \\ &= \delta_{ij} && (4.257d) \end{aligned}$$

Justifications :

— Pour 4.257b, nous remarquons que $\alpha_i(x) = \Phi(x)_i$.

— Pour 4.257d, nous utilisons le fait que les ϵ_j forment la base canonique de $\mathbb{K}^n$.

Unicité Soit une base préduale (e_i) de (α_i). Nous avons, par définition, que $\alpha_i(e_j) = \delta_{ij}$. Donc

$$\big(\alpha_1(e_j), \dots, \alpha_n(e_j)\big) = \epsilon_j. \tag{4.258}$$

Nous appliquons Φ^{-1} à cette dernière équation pour obtenir $e_j = \Phi^{-1}(\epsilon_j)$. Donc les e_j sont déterminés de façon unique à partir des α_i.

□

4.6.1 Orthogonal

Définition 4.134.
*Soit E, un espace vectoriel, et F une sous-espace de E. L'**orthogonal** de F est la partie $F^\perp \subset E^*$ donnée par*

$$F^\perp = \{\alpha \in E^* \text{ tel que } \forall x \in F, \alpha(x) = 0\}. \tag{4.259}$$

Cette définition d'orthogonal via le dual n'est pas du pur snobisme. En effet, la définition « usuelle » qui ne parle pas de dual,

$$F^\perp = \{y \in E \text{ tel que } \forall x \in F, y \cdot x = 0\}, \tag{4.260}$$

demande la donnée d'un produit scalaire. Évidemment dans le cas de $\mathbb{R}^n$ munie du produit scalaire usuel et de l'identification usuelle entre $\mathbb{R}^n$ et $(\mathbb{R}^n)^*$ via une base, les deux notions d'orthogonal coïncident.

La définition 4.134, au contraire, est intrinsèque : elle ne dépend que de la structure d'espace vectoriel.

Si $B \subset E^*$, on note B^o son orthogonal :

$$B^o = \{x \in E \text{ tel que } \omega(x) = 0 \, \forall \omega \in B\}. \tag{4.261}$$

Notons qu'on le note B^o et non $B^\perp$ parce qu'on veut un peu s'abstraire du fait que $(E^*)^* = E$. Du coup on impose que B soit dans un dual et on prend une notation précise pour dire qu'on remonte au pré-dual et non qu'on va au dual du dual.

Proposition 4.135.
Soient un espace vectoriel E et un sous-espace vectoriel F. Nous avons

$$\dim F + \dim F^\perp = \dim E. \tag{4.262}$$

Démonstration. Soit $\{e_1, \dots, e_p\}$ une base de F que nous complétons en une base $\{e_1, \dots, e_n\}$ de E par le théorème 4.11. Soit $\{e_1^*, \dots, e_n^*\}$ la base duale. Alors nous prouvons que $\{e_{p+1}^*, \dots, e_n^*\}$ est une base de $F^\perp$.

Déjà c'est une partie libre en tant que partie d'une base.

Ensuite ce sont des éléments de $F^\perp$ parce que si $i \leqslant p$ et si $k \geqslant 1$, nous avons $e_{p+k}^*(e_i) = 0$; donc oui, $e_{p+k}^* \in F^\perp$.

Enfin $F^\perp \subset \operatorname{Span}\{e_k^*, k \in \{p+1, \dots, n\}\}$ parce que si $\omega = \sum_{k=1}^n \omega_k e_k^*$, alors $\omega(e_i) = \omega_i$, mais nous savons que si $\omega \in F^\perp$, alors $\omega(e_i) = 0$ pour $i \leqslant p$. Donc $\omega = \sum_{k=p+1}^n \omega_k e_k^*$. □

La proposition 10.108 donnera une version plus terre à terre de la proposition 4.135 en disant que si nous avons un produit scalaire, alors $V = F \oplus F^\perp$ où $F^\perp$ est cette fois défini comme l'orthogonal pour le produit scalaire.

4.6.2 Transposée : pas d'approche naïve

Il est légitime, si $t \colon E \to E$ est une application linéaire, de dire que sa transposée soit l'application linéaire $t^t \colon E \to E$ dont la matrice est la matrice transposée de celle de t. Lorsque nous travaillons sur $\mathbb{R}^n$ muni de la base canonique, cela ne pose pas de problèmes et nous pouvons écrire des égalités du type $\langle x, Ay \rangle = \langle A^t x, y \rangle$.

Proposition 4.136 (Matrice transposée et produit scalaire)**.**
Soit une matrice réelle A. En utilisant l'application linéaire associée[48] *$f_A \colon \mathbb{R}^n \to \mathbb{R}^n$, nous avons*

$$x \cdot f_A(y) = f_{A^t}(x) \cdot y. \tag{4.263}$$

Cette formule est souvent écrite $x \cdot Ay = A^t x \cdot y$ ou $\langle x, Ay \rangle = \langle A^t x, y \rangle$.

48. Définition 4.65. Ici nous considérons la base canonique sur $\mathbb{R}^n$

Démonstration. Il s'agit d'une calcul utilisant la formule (4.83) et le produit scalaire (4.244) :

$$x \cdot f_A(y) = \sum_i x_i \big(f_A(y)\big)_i = \sum_i x_i \sum_j A_{ij} y_j = \sum_{ij} A^t_{ji} x_i y_j = \sum_j f_{A^t}(x)_j y_j = f_{A^t}(x) \cdot y. \tag{4.264}$$

□

Hélas nous allons voir que cette façon de définir une transposée est mauvaise.

Soit une application linéaire $t\colon E \to E$ de matrice A dans la base $\{e_i\}_{i=1,\dots,n}$ et de matrice B dans la base $\{f_\alpha\}_{\alpha=1,\dots,n}$.

Nous nommons t_1 l'application linéaire associée à A^t dans la base $\{e_i\}$ et t_2 l'application linéaire associée à la matrice B^t dans la base $\{f_\alpha\}$. Définir la transposée d'une application linéaire comme étant l'application linéaire associée à la transposée de sa matrice ne sera une bonne définition que si $t_1 = t_2$.

La première chose facile à voir est

$$t_1(e_i)_j = \sum_k (A^t)_{jk}(e_i)_k = A^t_{ji} = A_{ij}. \tag{4.265}$$

Pour calculer $t_2(e_i)_j$, c'est un peu plus laborieux :

$$t_2(e_i) = \sum_\alpha Q^{-1}_{\alpha i} t_2(f_\alpha) = \sum_{\beta\gamma\alpha} Q^{-1}_{\alpha i} B^t_{\gamma\beta} \underbrace{(t_\alpha)_\beta}_{\delta_{\alpha\beta}} f_\gamma = \sum_{\beta\gamma} Q^{-1}_{\beta i} B^t_{\gamma\beta} f_\gamma \tag{4.266a}$$

$$= (B^t Q^{-1})_{\gamma i} Q_{j\gamma} e_j \tag{4.266b}$$

$$= \sum_j (QB^tQ^{-1})_{ji} e_j. \tag{4.266c}$$

Donc $t_2(e_i)_j = (QB^tQ^{-1})_{ji}$. En tenant compte du fait que $B = Q^{-1}AQ$ nous avons

$$t_2(e_i)_j = (QQ^tA^t(Q^{-1})^tQ^{-1})_{ji}. \tag{4.267}$$

Cela est égal à l'expression (4.265) lorsque $Q^t = Q^{-1}$. Nous voyons que confondre transposée d'une application linéaire avec transposée de la matrice associée n'est valable que si nous sommes certain de ne considérer que des changements de base par des matrices orthogonales.

Cela est la situation typique dans laquelle nous nous trouvons lorsque nous considérons des applications linéaires sur $\mathbb{R}^n$ muni de la base canonique et que nous n'avons aucune intention de changer de base, et encore moins de chercher une base non orthonormale. Cette situation est clairement la situation la plus courante.

Exemple 4.137 ([41]).
Soit la base canonique $\{e_1, e_2\}$ de $\mathbb{R}^2$. Nous considérons l'application linéaire $t\colon \mathbb{R}^2 \to \mathbb{R}^2$ définie par

$$t(e_1) = e_1 \tag{4.268a}$$

$$t(e_2) = 0. \tag{4.268b}$$

La matrice de t dans cette base est

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{4.269}$$

Elle est symétrique : elle vérifie $A^t = A$. Si nous comptions sur la transposée de matrice pour définir la transposée de t, nous aurions $t^t = t$.

Soit maintenant la base $f_1 = e_1$, $f_2 = e_1 + e_2$. Nous avons $t(f_1) = f_1$ et

$$t(f_2) = t(e_1) + t(e_2) = e_1 = f_1. \tag{4.270}$$

Donc la matrice de t dans cette base est

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}. \tag{4.271}$$

Et là, nous avons $B^t \neq B$. Donc en comptant sur cette base pour définir la transposée de t nous aurions $t^t \neq t$. △

4.138.
Autrement dit, la façon « usuelle » de voir la transposée d'une application linéaire ne fonctionne dans les livres pour enfant uniquement parce qu'on n'y considère toujours $\mathbb{R}^n$ muni de la base canonique ou de bases orthonormées.

Notons que nous avons tout de même les notions d'opérateur adjoint et autoadjoint pour parler d'application orthogonale sans passer par la transposée, voir 9.39.

4.6.3 Transposée : la bonne approche

Définition 4.139.
Si $f\colon E \to F$ est une application linéaire entre deux espaces vectoriels, la ***transposée*** *est l'application $f^t\colon F^* \to E^*$ donnée par*

$$f^t(\omega)(x) = \omega\big(f(x)\big). \tag{4.272}$$

pour tout $\omega \in F^*$ et $x \in E$.

Lemme 4.140.
Soit E muni de la base $\{e_i\}$ et F muni de la base $\{g_i\}$ et une application $f\colon E \to F$. Si A est la matrice de f dans ces bases, alors A^t est la matrice de f^t dans les bases $\{e_i^*\}$ et $\{g_i^*\}$ de E^* et F^*.

Démonstration. Nous allons montrer que les formes $f^t(g_i^*)$ et $\sum_k (A^t)_{ik} g_k^*$ sont égales en les appliquant à un vecteur.

Par définition de la matrice d'une application linéaire dans une base,

$$f^t(g_i^*) = \sum_j (f^t)_{ij} e_j^*, \tag{4.273}$$

et

$$f(e_k) = \sum_l A_{kl} g_l. \tag{4.274}$$

Du coup, si $x = \sum_k x_k e_k$, nous avons

$$f^t(g_i^*)x = \sum_{kl} x_k g_i^* A_{kl} g_l = \sum_{kl} x_k A_{kl} \delta_{il} = \sum_k x_k A_{ki} = \sum_k (A^t)_{ik} x_k. \tag{4.275}$$

D'autre part,

$$\sum_k (A^t)_{ik} g_k^* x = \sum_{kl} (A^t)_{ik} g_k^* x_l e_l = \sum_k (A^t)_{ik} x_k. \tag{4.276}$$

Le fait que (4.275) et (4.276) donnent le même résultat prouve le lemme. □

En corolaire, les rangs de f et de f^t sont égaux parce que le rang est donné par la plus grande matrice carrée de déterminant non nul. Nous prouvons cependant ce résultat de façon plus intrinsèque.

Lemme 4.141 (Gilles Dubois)**.**
Si $f\colon E \to F$ est une application linéaire, alors

$$\operatorname{rang}(f) = \operatorname{rang}(f^t). \tag{4.277}$$

Démonstration. Nous posons $\dim\ker(f) = p$ et donc $\operatorname{rang}(f) = n - p$. Soit $\{e_1, \dots, e_p\}$ une base de $\ker(f)$ que l'on complète en une base $\{e_1, \dots, e_n\}$ de E. Nous considérons maintenant les vecteurs

$$g_i = f(e_{p+i}) \tag{4.278}$$

pour $i = 1, \dots, n-p$. C'est-à-dire que les g_i sont les images des vecteurs qui ne sont pas dans le noyau de f. Prouvons qu'ils forment une famille libre. Si

$$\sum_{k=1}^{n-p} a_k f(e_{p+k}) = 0, \tag{4.279}$$

alors $f\big(\sum_k a_k e_{p+k}\big) = 0$, ce qui signifierait que $\sum_k a_k e_{p+k}$ se trouve dans le noyau de f, ce qui est impossible par construction de la base $\{e_i\}_{i=1,\dots,n}$. Étant donné que les vecteurs $g_1, \dots, g_{n-p}$ sont libres, nous les complétons en une base

$$\{\underbrace{g_1, \dots, g_{n-p}}_{\text{images}}, \underbrace{g_{n-p+1}, \dots, g_r}_{\text{complétion}}\} \tag{4.280}$$

de F.

Nous prouvons maintenant que $\operatorname{rang}(f^t) \geqslant n-p$ en montrant que les formes $\{g_i^*\}_{i=1,\dots,n-p}$ forment une partie libre (et donc l'espace image de f^t est au moins de dimension $n-p$). Pour cela nous prouvons que $f^t(g_i^*) = e_{i+p}^*$. En effet

$$f^t(g_i^*)e_k = g_i^*(fe_k), \tag{4.281}$$

Si $k = 1, \dots, p$, alors $fe_k = 0$ et donc $g_i^*(fe_k) = 0$; si $k = p + l$ alors

$$f^t(g_i^*)e_k = g_i^*(fe_{k+l}) = g_i^*(g_l) = \delta_{i,l} = \delta_{i,k-p} = \delta_{k,i+p}. \tag{4.282}$$

Donc $f^t(g_i^*) = e_{i+p}^*$. Cela prouve que les formes $f^t(g_i^*)$ sont libres et donc que

$$\operatorname{rang}(f^t) \geqslant n-p = \operatorname{rang}(f). \tag{4.283}$$

En appliquant le même raisonnement à f^t au lieu de f, nous trouvons

$$\operatorname{rang}\big((f^t)^t\big) \geqslant \operatorname{rang}(f^t) \tag{4.284}$$

et donc, vu que $(f^t)^t = f$, nous obtenons $\operatorname{rang}(f) = \operatorname{rang}(f^t)$.

□

Proposition 4.142 ([85]).
Si f est une application linéaire entre les espaces vectoriels E et F, alors nous avons

$$\operatorname{Image}(f^t) = \ker(f)^\perp. \tag{4.285}$$

Démonstration. Soient donc l'application $f\colon E \to F$ et sa transposée $f^t\colon F^* \to E^*$. Nous commençons par prouver que $\operatorname{Image}(f^t) \subset (\ker f)^\perp$. Pour cela nous prenons $\omega \in \operatorname{Image}(f^t)$, c'est-à-dire $\omega = \alpha \circ f$ pour un certain élément $\alpha \in F^*$. Si $z \in \ker(f)$, alors $\omega(z) = (\alpha \circ f)(z) = 0$, c'est-à-dire que $\omega \in (\ker f)^\perp$.

Pour prouver qu'il y a égalité, nous n'allons pas démontrer l'inclusion inverse, mais plutôt prouver que les dimensions sont égales. Après, on sait que si $A \subset B$ et si $\dim A = \dim B$, alors $A = B$. Nous avons

$$\begin{aligned}
\dim\big(\operatorname{Image}(f^t)\big) &= \operatorname{rang}(f^t) && & (4.286a)\\
&= \operatorname{rang}(f) && \text{lemme 4.141} & (4.286b)\\
&= \dim(E) - \dim\ker(f) && \text{théorème 4.39} & (4.286c)\\
&= \dim\big((\ker f)^\perp\big) && \text{proposition 4.135.} & (4.286d)
\end{aligned}$$

□

Lemme 4.143 ([80]).
Soit $\mathbb{K}$ un corps, E et F deux $\mathbb{K}$-espaces vectoriels de dimension finie et une application linéaire $f\colon E \to F$. L'application f est injective si et seulement si sa transposée[49] *f^t est surjective.*

Démonstration. Supposons que f soit injective. Alors par le lemme 4.53, il existe $g\colon F \to E$ tel que $g \circ f = \mathrm{Id}|_E$. Nous avons alors aussi $(g \circ f)^t = \mathrm{Id}|_{E^*}$, mais $(g \circ f)^t = f^t \circ g^t$, donc f^t est surjective.

Inversement, nous supposons que $f^t\colon F^* \to E^*$ est surjective. Alors en nous souvenant que E et F sont de dimension finie et en faisant jouer les identifications $(f^t)^t = f$ et $(E^*)^* = E$ nous savons qu'il existe $s\colon E^* \to F^*$ tel que $f^t \circ s = \mathrm{Id}|_{E^*}$. En passant à la transposée,

$$s^t \circ f = \mathrm{Id}|_E, \tag{4.287}$$

qui implique que f est injective. □

4.6.4 Polynômes de Lagrange

Soit $E = \mathbb{R}_n[X]$ l'ensemble des polynômes à coefficients réels de degré au plus n. Soient les $n+1$ réels distincts $a_0, \dots, a_n$. Nous considérons les formes linéaires associées $f_i \in E^*$,

$$f_i(P) = P(a_i). \tag{4.288}$$

Lemme 4.144.
Ces formes forment une base de E^*.

Démonstration. Nous prouvons que l'orthogonal est réduit au nul :

$$\mathrm{Span}\{f_0, \dots, f_n\}^\perp = \{0\} \tag{4.289}$$

pour que la proposition 4.135 conclue. Si $P \in \mathrm{Span}\{f_i\}^\perp$, alors $f_i(P) = 0$ pour tout i, ce qui fait que $P(a_i) = 0$ pour tout $i = 0, \dots, n$. Un polynôme de degré au plus n qui s'annule en $n+1$ points est automatiquement le polynôme nul. □

Les **polynômes de Lagrange** sont les polynômes de la base (pré)duale de la base $\{f_i\}$.

Proposition 4.145.
Les polynômes de Lagrange sont donnés par

$$P_i = \prod_{k \neq i} \frac{X - a_k}{a_i - a_k}. \tag{4.290}$$

Démonstration. Il suffit de vérifier que $f_j(P_i) = \delta_{ij}$. Nous avons

$$f_j(P_i) = P_i(a_j) = \prod_{k \neq i} \frac{a_j - a_k}{a_i - a_k}. \tag{4.291}$$

Si $j \neq i$ alors un des termes est nul. Si au contraire $i = j$, tous les termes valent 1. □

4.6.5 Dual de $\mathbb{M}(n, \mathbb{K})$

Proposition 4.146 ([54]).
Soit $\mathbb{K}$, un corps. Les formes linéaires sur $\mathbb{M}(n, \mathbb{K})$ sont les applications de la forme

$$\begin{aligned} f_A\colon \mathbb{M}(n, \mathbb{K}) &\to \mathbb{K} \\ M &\mapsto \mathrm{Tr}(AM). \end{aligned} \tag{4.292}$$

49. Définition 4.139.

Démonstration. Nous considérons l'application

$$\begin{aligned} f\colon \mathbb{M}(n,\mathbb{K}) &\to \mathbb{M}(n,\mathbb{K})^* \\ A &\mapsto f_A \end{aligned} \tag{4.293}$$

et nous voulons prouver que c'est une bijection. Étant donné que nous sommes en dimension finie, nous avons égalité des dimensions de $\mathbb{M}(n,\mathbb{K})$ et $(\mathbb{M}(n,\mathbb{K}))^*$, et il suffit de prouver que f est injective. Soit donc A telle que $f_A = 0$. Nous l'appliquons à la matrice $(E_{ij})_{kl} = \delta_{ik}\delta_{jl}$:

$$0 = f_A(E_{ij}) = \sum_k (AE_{ij})_{kk} = \sum_{kl} A_{kl}(E_{ij})_{lk} = \sum_{kl} A_{kl}\delta_{il}\delta_{jk} = A_{ij}. \tag{4.294}$$

Donc $A = 0$. □

Corolaire 4.147 ([54]).
Soit $\mathbb{K}$ un corps et $\phi \in \mathbb{M}(n,\mathbb{K})^*$ telle que pour tout $M, N \in \mathbb{M}(n,\mathbb{K})$ on ait

$$\phi(MN) = \phi(NM). \tag{4.295}$$

Alors il existe $\lambda \in \mathbb{K}$ tel que $\phi = \lambda \operatorname{Tr}$.

Démonstration. La proposition 4.146 nous donne une matrice $A \in \mathbb{M}(n,\mathbb{K})$ telle que $\phi = f_A$. L'hypothèse nous dit que $f_A(MN) = f_A(NM)$, c'est-à-dire

$$\operatorname{Tr}(AMN) = \operatorname{Tr}(ANM) \tag{4.296}$$

pour toutes matrices $M, N \in \mathbb{M}(n,\mathbb{K})$. L'invariance cyclique de la trace [50] appliqué au membre de droite nous donne $\operatorname{Tr}(AMN) = \operatorname{Tr}(MAN)$, ce qui signifie que

$$\operatorname{Tr}\big((AM - MA)N\big) = 0 \tag{4.297}$$

ou encore que $f_{AM-MA} = 0$, et ce, pour toute matrice M. La fonction f étant injective nous en déduisons que la matrice A doit satisfaire

$$AM = MA \tag{4.298}$$

pour tout $M \in \mathbb{M}(n,\mathbb{K})$. En particulier, en prenant pour M les fameuses matrices E_{ij} et en calculant un peu,

$$A_{li}\delta_{jm} = \delta_{il}A_{jm} \tag{4.299}$$

pour tout i, j, l, m. Cela implique que $A_{ll} = A_{mm}$ pour tout l et m et que $A_{jm} = 0$ dès que $j \neq m$. Il existe donc $\lambda \in \mathbb{K}$ tel que $A = \lambda\mathbb{1}$. En fin de compte,

$$\phi(X) = f_{\lambda\mathbb{1}}(X) = \lambda \operatorname{Tr}(X). \tag{4.300}$$

□

Corolaire 4.148 ([54]).
Soit $\mathbb{K}$ un corps. Tout hyperplan de $\mathbb{M}(n,\mathbb{K})$ coupe $\mathrm{GL}(n,\mathbb{K})$.

Démonstration. Soit $\mathcal{H}$ un hyperplan de $\mathbb{M}$. Il existe une forme linéaire ϕ sur $\mathbb{M}(n,\mathbb{K})$ telle que $\mathcal{H} = \ker(\phi)$. Encore une fois la proposition 4.146 nous donne $A \in \mathbb{M}$ telle que $\phi = f_A$; nous notons r le rang de A. Par le lemme 4.111 nous avons $A = PJ_rQ$ avec $P, Q \in \mathrm{GL}(n,\mathbb{K})$ et

$$J_r = \begin{pmatrix} \mathbb{1}_r & 0 \\ 0 & 0 \end{pmatrix}. \tag{4.301}$$

50. Lemme 4.62.

Pour tout $M \in \mathbb{M}$ nous avons

$$\phi(M) = \mathrm{Tr}(AM) = \mathrm{Tr}(PJ_rQM) = \mathrm{Tr}(J_rQMP), \tag{4.302}$$

la dernière égalité découlant de l'invariance cyclique de la trace[51]. Ce que nous cherchons est $M \in \mathrm{GL}(n, \mathbb{K})$ telle que $\phi(M) = 0$. Nous commençons par trouver $N \in \mathrm{GL}(n, \mathbb{K})$ telle que $\mathrm{Tr}(J_rN) = 0$. Celle-là est facile : c'est

$$N = \begin{pmatrix} 0 & 1 \\ \mathbb{1}_{n-1} & 0 \end{pmatrix}. \tag{4.303}$$

Les éléments diagonaux de J_rN sont tous nuls. Par conséquent en posant $M = Q^{-1}NP^{-1}$ nous avons notre matrice inversible dans le noyau de ϕ. □

4.7 Représentation de groupe

Définition 4.149 (Représentation)**.**
Soit un groupe G et un espace vectoriel V. Nous disons qu'une application $\rho\colon G \to \mathrm{GL}(V)$ est une ***représentation*** *de G sur V si pour tout $g, h \in G$ nous avons*

$$\rho(g) \circ \rho(g) = \rho(gh). \tag{4.304}$$

Très souvent, nous disons que la représentation est le couple (V, ρ).

Définition 4.150.
Une représentation[52] *est* ***fidèle*** *si elle est injective en tant que application $G \to \mathrm{GL}(V)$. Ce ne sont pas chacun des $\rho(g)$ qui doivent être injectifs. La dimension de V est le* ***degré*** *de la représentation (V, ρ).*

Proposition 4.151.
Soit un corps $\mathbb{K}$. Si G est un groupe dans $\mathbb{M}(n, \mathbb{K})$ (c'est-à-dire un groupe de matrices à coefficients dans $\mathbb{K}$), alors l'application

$$\begin{aligned} \rho\colon G &\to \mathrm{GL}(\mathbb{K}^n) \\ A &\mapsto f_A \end{aligned} \tag{4.305}$$

où f_A est l'application linéaire associée à A est une représentation de G.

51. Lemme 4.62.
52. Définition 4.149.

Chapitre 5

Classification de certains groupes

5.1 Théorèmes de Sylow

Lemme 5.1.
Soient H et K des sous-groupes finis de G. Alors

$$\mathrm{Card}(HK) = \frac{|H| \cdot |K|}{|H \cap K|}. \tag{5.1}$$

Attention : dans ce lemme, l'ensemble HK n'est pas spécialement un groupe. Ce serait le cas si H normaliserait K, c'est-à-dire si nous avions $hkh^{-1} \in K$, $\forall h, k \in H \times K$.

Théorème 5.2 (Théorème de Cauchy[86])**.**
Soit G un groupe fini et p un nombre premier divisant $|G|$. Alors

(1) G contient un élément d'ordre p.

(2) Si G est un p-groupe, il existe un élément central d'ordre p dans G.

Lemme 5.3 (Théorème de Cayley)**.**
Si G est un groupe d'ordre n alors il est isomorphe à un sous-groupe du groupe symétrique S_n.

Démonstration. L'action à gauche de G sur lui-même

$$\begin{aligned} \varphi\colon G &\to S_n \\ \varphi(x)g &\mapsto xg \end{aligned} \tag{5.2}$$

est une permutation des éléments de G. Cela donne un morphisme injectif parce que si $\varphi(x) = \varphi(y)$ nous avons $xg = yg$ pour tout g et en particulier pour $g = e$ nous trouvons $x = y$. □

Lemme 5.4.
Soit p un diviseur premier de n. Alors le groupe symétrique S_n se plonge dans $\mathrm{GL}_n(\mathbb{F}_p)$.

Démonstration. Soit $\{e_i\}$ la base canonique de $\mathbb{F}_p$. Nous avons le morphisme injectif $\varphi\colon S_n \to \mathrm{GL}(n, \mathbb{F})$ donné par $\varphi(\sigma)e_i = e_{\sigma(i)}$. □

Remarque 5.5.
En mettant bout à bout les lemmes 5.3 et 5.4, nous trouvons que si p est un diviseur premier de $|G|$, alors G peut être vu comme un sous-groupe de $\mathrm{GL}(n, \mathbb{F}_p)$.

Définition 5.6.
*Soit p un nombre premier. Un **p-groupe** est un groupe dont tous les éléments sont d'ordre p^m pour un certain m (dépendant de l'élément).*

*Soit G un groupe fini et p, un diviseur premier de $|G|$. Un **p-Sylow** dans G est un p-sous-groupe d'ordre p^n où p^n est la plus grande puissance de p divisant $|G|$.*

Notons que si p est un nombre premier, alors tout groupe d'ordre p^m est un p-groupe.

Lemme 5.7.
Soit G un groupe fini et P, Q des p-sous-groupes. Nous supposons que Q normalise P. Alors PQ est un p-sous-groupe de G.

Si S est un p-Sylow, alors p ne divise pas le nombre $|G : S| = |G|/|S|$.

Proposition 5.8.
Soit le corps fini $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ (p premier). Soit T le sous-ensemble de $\mathrm{GL}_n(\mathbb{F}_p)$ formé des matrices triangulaires supérieures de rang[1] *n et dont les éléments diagonaux sont 1. Alors T est un p-Sylow de $\mathrm{GL}_n(\mathbb{F}_p)$.*

Démonstration. Nous commençons par étudier le cardinal de $\mathrm{GL}_n(\mathbb{F}_p)$. Pour la première colonne, la seule contrainte à vérifier est qu'elle ne soit pas nulle. Il y a donc $p^n - 1$ possibilités. Pour la seconde, il faut ne pas être multiple de la première. Il y a donc $p^n - p$ possibilités (parce qu'il y a p multiples possibles de la premières colonne). Pour la k-ième colonne, il faut éviter toutes les combinaisons linéaires des $(k-1)$ premières colonnes. Il y a p^{k-1} telles combinaisons et donc $p^n - p^{k-1}$ possibilités pour la k-ième colonne. Nous avons donc

$$\begin{aligned}
\mathrm{Card}\big(\mathrm{GL}(n, \mathbb{F}_p)\big) &= (p^n - 1)(p^n - p)\dots(p^n - p^{n-1}) && (5.3a)\\
&= p \cdot p^2 \cdots p^{n-1}(p^n - 1)(p^{n-1} - 1)\dots(p - 1) && (5.3b)\\
&= p^{\frac{n(n-1)}{2}} m && (5.3c)
\end{aligned}$$

où m est un entier qui ne divise pas p.

En ce qui concerne le cardinal de T, le calcul est plus simple : pour la première ligne nous avons p^{n-1} choix (parce qu'il y a un 1 qui est imposé sur la diagonale), pour la seconde p^{n-2}, etc. En tout nous avons alors

$$|T| = p^{\frac{n(n-1)}{2}}, \tag{5.4}$$

et T est un p-Sylow de $\mathrm{GL}_n(\mathbb{F}_p)$. □

Proposition 5.9.
Soit p un nombre premier. Un groupe fini G est un p-groupe si et seulement l'ordre de G est p^n pour un certain n.

Démonstration. Supposons que G est un p-groupe. Soit q un nombre premier divisant $|G|$. Par le théorème de Cauchy (5.2), le groupe G contient un élément d'ordre q, soit g un tel élément. Étant donné que G est un p-groupe, $g^{p^n} = g^q = e$ pour un certain n. Donc $q = p^n$ et $q = p$ parce que q est premier. Nous venons de prouver que p est le seul nombre premier qui divise $|G|$. L'ordre de G est par conséquent une puissance de p.

Nous nous intéressons maintenant à l'implication inverse. Nous supposons que $|G| = p^n$ pour un certain entier $n \geqslant 0$. Soit $g \in G$; nous notons r l'ordre de G. Le sous-groupe $\mathrm{gr}(g)$ est d'ordre r, donc r divise $|G|$ (par le théorème 2.34 de Lagrange). Le nombre r est alors une puissance de p. □

Lemme 5.10.
Soit G, un groupe fini de cardinal $|G| = n$ et p, un diviseur premier de n. Nous notons $n = p^m \cdot r$ où p ne divise pas r. Soit H un sous-groupe de G et S, un p-Sylow de G. Alors il existe $g \in G$ tel que

$$gSg^{-1} \cap H \tag{5.5}$$

soit un p-Sylow de H.

1. Définition 4.38.

Démonstration. Nous considérons l'ensemble G/S sur lequel H agit. Si $a \in G$, le stabilisateur de $[a]$ dans G/S est

$$\begin{aligned}\operatorname{Fix}\big([a]\big) &= \{h \in H \text{ tel que } [ha] = [a]\} && (5.6a)\\ &= \{h \in H \text{ tel que } a^{-1}ha \in S\} && (5.6b)\\ &= aSa^{-1} \cap H. && (5.6c)\end{aligned}$$

Nous cherchons $a \in G$ tel que l'entier

$$\frac{\operatorname{Card}(H)}{\operatorname{Card}\big(aSa^{-1} \cap H\big)} \tag{5.7}$$

soit premier avec p. En effet, dans ce cas le groupe $\operatorname{Fix}([a])$ est un p-Sylow de H parce que $|H : aSa^{-1} \cap H|$ ne divise pas p. La formule des orbites (équation (2.79)) nous dit que

$$\frac{|H|}{|aSa^{-1} \cap H|} = \operatorname{Card}\big(\mathcal{O}_{[a]}\big). \tag{5.8}$$

Supposons que toutes les orbites aient un cardinal divisible par p. Étant donné que G/S est une réunion disjointe de ses orbites, nous aurions

$$p \mid \operatorname{Card}(G/S) = \frac{|G|}{|S|} \tag{5.9}$$

alors que S étant un p-Sylow, p ne peut pas diviser $|G|/|S|$. Toutes les orbites n'ont donc pas un cardinal divisible par p, et il existe un $a \in G$ tel que (5.7) soit vérifiée. □

Théorème 5.11 (Théorème de Sylow)**.**
Soit G un groupe fini et p, un diviseur premier de $|G|$. Alors

(1) G possède au moins un p-Sylow[2]*.*

(2) Tout p-sous-groupe de G est contenu dans un p-Sylow.

(3) Les p-Sylow de G sont conjugués.

(4) Si n_p est le nombre de p-Sylow de G, alors n_p divise $|G|$ et $n_p \in [1]_p$.

Démonstration. En plusieurs points.

(1) Nous savons de la remarque 5.5 que G est un sous-groupe de $\mathrm{GL}_n(\mathbb{F}_p)$ et que ce dernier a un p-Sylow par la proposition 5.8. Par conséquent G possède un p-Sylow par le lemme 5.10.

(2) Soit H un p-sous-groupe de G et S, un p-Sylow de G (qui existe par le point précédent). Par le lemme 5.10 il existe $a \in G$ tel que $aSa^{-1} \cap H$ soit un p-Sylow de H. Mais H est un p-groupe et un p-Sylow dans un p-groupe est automatiquement le groupe entier. Par conséquent,

$$H = aSa^{-1} \cap H \tag{5.10}$$

et $H \subset aSa^{-1}$, ce qui signifie que H est inclus dans un p-Sylow.

(3) Soit H un p-Sylow. Nous venons de voir que si S est un p-Sylow quelconque, alors H est inclus au p-Sylow aSa^{-1} pour un certain $a \in G$. Donc H est un p-Sylow inclus dans le p-Sylow aSa^{-1}, donc $H = aSa^{-1}$.

(4) Le fait que n_p divise n est parce que tous les p-Sylow ont le même nombre d'éléments (ils sont conjugués) et sont deux à deux disjoints. Donc ils forment une partition de G et $|G| = n_p|S|$ si S est un p-Sylow quelconque.

Montrons maintenant que n_p est congru à un modulo p. Soit E l'ensemble des p-Sylow de G. Le groupe G agit sur E par conjugaison. Soit S un p-Sylow et considérons l'ensemble

$$E_S = \{T \in E \text{ tel que } s \cdot T = T \forall s \in S\}. \tag{5.11}$$

2. Définition 5.6.

où l'action est celle par conjugaison. C'est l'ensemble des points fixes de E sous l'action de S. L'ensemble E est la réunion des orbites sous S et chacune de ces orbites a un cardinal qui divise $|S| = p^m$. Par conséquent $|\mathcal{O}_T|$ vaut 1 lorsque $T \in E_S$ et est un multiple de p sinon. Nous avons donc

$$|E| \equiv |E_S| \mod p. \tag{5.12}$$

Nous voulons obtenir $|E_S| = 1$. Évidemment $S \in E_S$ parce que si $s \in S$ alors $sSs^{-1} = S$. Nous voudrions montrer que S est le seul élément de E_S. Soit $T \in E_S$, c'est-à-dire que T est un p-Sylow de G tel que

$$sTs^{-1} = T \tag{5.13}$$

pour tout $s \in S$. Soit N le groupe engendré par S et T. Montrons que T est normal dans N. Un élément g dans N s'écrit

$$g = s_1 t_1 \cdots s_r t_r \tag{5.14}$$

avec $s_i \in S$ et $t_i \in T$. Si $t \in T$, en utilisant le fait que T est un groupe et le fait que S le normalise, nous avons

$$gtg^{-1} = s_1 t_1 \ldots s_r t_r t t_r^{-1} s_r^{-1} \ldots t_1^{-1} s_1^{-1} \in T. \tag{5.15}$$

Donc T est un sous-groupe normal de N. Mais S et T sont conjugués dans N (parce que ils sont des p-Sylow de N), donc il existe un élément $a \in N$ tel que $aTa^{-1} = S$. Mais étant donné que T est normal,

$$S = aTa^{-1} = T. \tag{5.16}$$

Ceci achève la démonstration des théorèmes de Sylow.

□

Proposition 5.12.
Si S est un p-Sylow dans le groupe G alors pour tout $g \in G$, l'ensemble gSg^{-1} est encore un p-groupe.

Démonstration. Si les éléments de S sont d'ordre p^n, alors nous avons

$$(gsg^{-1})^q = gs^q g^{-1} = e. \tag{5.17}$$

Pour avoir $gs^q g^{-1} = e$, il faut et suffit que $gs^q = g$, alors $s^q = e$, c'est-à-dire $q = p^n$. Donc gSg^{-1} est encore un p-Sylow. □

Lemme 5.13 ([87])**.**
Soit G, un groupe fini et p, un nombre premier. Si H et K sont des groupes distincts d'ordre p, alors $H \cap K = \{e\}$.

Démonstration. L'ensemble $H \cap K$ est un sous-groupe de H. Par conséquent son ordre divise celui de H qui est un nombre premier. Par conséquent soit $|H \cap K| = 1$, soit $|H \cap K| = |H|$. Dans le second cas nous aurions $H = K$, alors que nous avons supposé que H et K étaient distincts. □

Proposition 5.14 ([87])**.**
Soit G un groupe fini et n le nombre de sous-groupes d'ordre p dans G. Alors le nombre d'éléments d'ordre p dans G vaut $n(p-1)$.

Démonstration. Si g est un élément d'ordre p dans G, le groupe H engendré par g est d'ordre p. Réciproquement si H est un groupe d'ordre p, tous les éléments de $H \backslash \{e\}$ sont d'ordre p (parce que l'ordre d'un élément divise l'ordre du groupe). Donc l'ensemble des éléments d'ordre p dans G est la réunion des ensembles $H \backslash \{e\}$ où H parcourt les sous-groupes d'ordre p dans G. Chacun de ces ensembles possède $p-1$ éléments et le lemme 5.13 nous assure qu'ils sont disjoints. Par conséquent nous avons $n(p-1)$ éléments d'ordre p dans G. □

Corolaire 5.15.
Un groupe d'ordre premier est cyclique.

Démonstration. Soit p l'ordre de G. Le nombre de sous-groupes d'ordre p est $n = 1$ (et c'est G lui-même). La proposition 5.14 nous dit alors que le nombre d'éléments d'ordre p dans G est $p-1$. Donc tout élément est générateur. □

5.2 Groupe monogène

Le théorème suivant donne quelques informations à propos de groupes monogènes. Il impliquera dans le corolaire 18.13 qu'un groupe monogène d'ordre n possède $\varphi(n)$ générateur où φ est la fonction indicatrice d'Euler définie en 18.9.

Théorème 5.16.
Un groupe monogène est abélien. Plus précisément,

(1) un groupe monogène infini est isomorphe à $\mathbb{Z}$,

(2) un groupe monogène fini est isomorphe à $\mathbb{Z}/n\mathbb{Z}$ pour un certain n.

Démonstration. Le groupe est abélien parce que $g = a^n$, $g' = a^{n'}$ implique $gg' = q^{n+n'} = g'g$. Nous considérons un générateur a de G (qui existe parce que G est monogène) et le morphisme surjectif

$$\begin{aligned} f\colon \mathbb{Z} &\to G \\ p &\mapsto a^p. \end{aligned} \tag{5.18}$$

Si G est infini, alors f est injective parce que si $a^n = a^{n'}$, alors $a^{n-n'} = e$, ce qui rendrait G cyclique et par conséquent non infini. Nous concluons que si G est infini, alors f est une bijection et donc un isomorphisme $\mathbb{Z} \simeq G$.

Si G est fini, alors f n'est pas injective et a un noyau $\ker f$. Étant donné que $\ker f$ est un sous-groupe de G, il existe un (unique) n tel que $\ker f = n\mathbb{Z}$ et le premier théorème d'isomorphisme (théorème 2.28) nous indique que

$$\mathbb{Z}/\ker f = \mathbb{Z}/n\mathbb{Z} = \operatorname{Image} f = G. \tag{5.19}$$

□

Le lemme suivant donne une démonstration alternative, avec une construction plus explicite de l'isomorphisme.

Lemme 5.17 ([1])**.**
À propos de groupes monogènes[3]

(1) Soit un groupe monogène G d'ordre fini n dont g est un générateur. Alors il existe un isomorphisme

$$\phi\colon G \to (\mathbb{Z}/n\mathbb{Z}, +) \tag{5.20}$$

tel que $\phi(g) = 1$.

(2) Si G est un groupe monogène d'ordre infini et si g est un générateur, alors il existe un isomorphisme

$$\phi\colon G \to (\mathbb{Z}, +) \tag{5.21}$$

tel que $\phi(g) = 1$.

(3) Soient G et H deux groupes monogènes de même ordre. Soient g un générateur de G et h, un générateur de H. Il existe un isomorphisme de G sur H qui envoie g sur h.

3. Définition 2.11.

Démonstration. Commençons par enfoncer une porte ouverte : vu que le groupe est monogène, l'ordre du groupe est égal à l'ordre de son générateur. Nous séparons les cas selon quel l'ordre soit fini ou non.

L'ordre de G est fini et vaut n Si $k \in \mathbb{Z}$, nous notons $[k]_n$ la classe de k modulo n, c'est-à-dire l'ensemble $\{k + pn$ tel que $p \in \mathbb{Z}\}$.

Nous construisons l'isomorphisme $\phi\colon G \to \mathbb{Z}/n\mathbb{Z}$ de la façon suivante :

$$\phi(g^m) = [m]_n. \tag{5.22}$$

Cela est une bonne définition parce que une égalité du type $g^m = g^{m'}$ implique que m et m' soient dans la même classe modulo n. Nous vérifions que cela est une isomorphisme entre G et $\mathbb{Z}/n\mathbb{Z}$.

Morphisme Pour l'identité, si $x = e$ alors $m = 0$ et $\phi(e) = [0]_n$. Et si $x = g^k$, $y = g^l$ alors $\phi(xy) = \phi(g^{k+l}) = [k + l]_n = [k]_n + [l]_n = \phi(x) + \phi(y)$.

Injectif Supposons $\phi(g^k) = \phi(g^l)$ avec $k \geqslant l$. Nous avons $h^k = h^l$, dont $h^{k-l} = e$, ce qui donne $k - l \in [0]_n$ ou encore $[k]_n = [l]_n$. En particulier $g^k = g^l$.

Surjectif La classe $[k]_n$ est l'image de g^k.

L'ordre de G est infini Si l'ordre de G est infini alors un élément $x \in G$ s'écrit de façon unique sous la forme $x = g^m$ avec $m \in \mathbb{Z}$. Dans ce cas nous définissons directement $\phi(g^m) = m$.

Le reste de la preuve est alors identique au cas d'ordre fini, mais sans les complications liées au modulo.

La dernière assertion s'obtient des précédentes par composition d'isomorphismes.

□

5.3 Automorphismes du groupe $\mathbb{Z}/n\mathbb{Z}$

Notons que $\mathbb{Z}/n\mathbb{Z} = \mathbb{F}_n$ est un groupe pour l'addition tandis que $(\mathbb{Z}/n\mathbb{Z})^*$ est un groupe pour la multiplication. Il ne peut donc pas y avoir d'équivoque.

Théorème 5.18 ([88]).
Pour chaque $x \in (\mathbb{Z}/n\mathbb{Z})^*$ nous considérons l'application

$$\begin{aligned} \sigma_x\colon \mathbb{Z}/n\mathbb{Z} &\to \mathbb{Z}/n\mathbb{Z} \\ y &\mapsto xy. \end{aligned} \tag{5.23}$$

L'application

$$\sigma\colon \big((\mathbb{Z}/n\mathbb{Z})^*,\ \cdot\ \big) \to \mathrm{Aut}\,\big(\mathbb{Z}/n\mathbb{Z}, +\big) \tag{5.24}$$

ainsi définie est un isomorphisme de groupes.

L'énoncé de ce théorème s'écrit souvent rapidement par

$$\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z}) = (\mathbb{Z}/n\mathbb{Z})^*, \tag{5.25}$$

mais il faut bien garder à l'esprit qu'à gauche on considère le groupe additif et à droite celui multiplicatif.

Démonstration. Nous notons $[x]$ la classe de x dans $\mathbb{Z}/n\mathbb{Z}$. Nous avons $\mathbb{Z}/n\mathbb{Z} = [1]$. Soit f un automorphisme de $(\mathbb{Z}/n\mathbb{Z}, +)$; pour tout $r \in \mathbb{Z}$ nous avons

$$f([r]) = f(r[1]) = rf([1]) = [r]f([1]). \tag{5.26}$$

En particulier, vu que f est surjective, il existe un r tel que $f([r]) = [1]$. Pour un tel r nous avons $[1] = [r]f([1])$, c'est-à-dire que nous avons montré que $f([1])$ est inversible dans $\big((\mathbb{Z}/n\mathbb{Z})^*, \ \cdot\ \big)$. Nous montrons à présent que[4]

$$\begin{aligned} \sigma\colon \operatorname{Aut}((\mathbb{Z}/n\mathbb{Z},+)) &\to \big((\mathbb{Z}/n\mathbb{Z})^*, \ \cdot\ \big) \\ f &\mapsto f([1]) \end{aligned} \tag{5.27}$$

est un isomorphisme.

Nous commençons par la surjectivité. Soit $[a] \in (\mathbb{Z}/n\mathbb{Z})^*$. Les élément $[a]$ et $[1]$ étant tous deux des générateurs de $(\mathbb{Z}/n\mathbb{Z},+)$, il existe un automorphisme de $\mathbb{Z}/n\mathbb{Z}$ qui envoie $[1]$ sur $[a]$ par le lemme 5.17. Cela prouve la surjectivité de σ.

En ce qui concerne l'injectivité, considérons des automorphismes f_1 et f_2 de $(\mathbb{Z}/n\mathbb{Z},+)$ tels que $f_1([1]) = f_2([1])$. Les automorphismes f_1 et f_2 prennent la même valeur sur un générateur et donc sur tout le groupe. Donc $f_1 = f_2$.

Enfin nous prouvons que σ est un morphisme, c'est-à-dire que $\sigma(f \circ g) = \sigma(f)\sigma(g)$. Nous avons

$$f\big(g([1])\big) = f\big(g([1])[1]\big) = g([1])f([1]) = \sigma(f)\sigma(g). \tag{5.28a}$$

□

Ce dernier résultat s'étend aux groupes cycliques.

Proposition 5.19.
Si G est un groupe cyclique[5] *d'ordre n, alors*

$$\operatorname{Aut}(G) = (\mathbb{Z}/n\mathbb{Z})^*. \tag{5.29}$$

Corolaire 5.20.
Si p divise $q-1$ alors $\operatorname{Aut}(\mathbb{F}_q)$ possède un unique sous-groupe d'ordre p.

Démonstration. Si a est un générateur de $\mathbb{F}_q^*$ alors le groupe

$$\operatorname{gr}\left(a^{\frac{q-1}{p}}\right) \tag{5.30}$$

est un sous-groupe d'ordre p. En ce qui concerne l'unicité, soit S un sous-groupe d'ordre p. Il est donc d'indice $(q-1)/p$ dans $\mathbb{F}_q^*$ et le lemme 3.37 nous enseigne que le groupe donné en (5.30) est contenu dans S. Il est donc égal à S parce qu'il a l'ordre de S. Le fait que S soit normal est dû au fait que $\mathbb{F}_q^*$ est abélien. □

5.4 Groupes abéliens finis

Source : [36].

Nous rappelons que l'exposant d'un groupe fini est le ppcm des ordres de ses éléments. Dans le cas des groupes abéliens finis, l'exposant joue un rôle important du fait qu'il existe un élément dont l'ordre est l'exposant. Cela est le théorème suivant.

Théorème 5.21 (Exposant dans un groupe abélien fini)**.**
Un groupe abélien fini contient un élément dont l'ordre est l'exposant du groupe.

Démonstration. Soit G un groupe abélien fini et $x \in G$, un élément d'ordre maximum m. Nous montrons par l'absurde que l'ordre de tous les éléments de G divise m. Soit donc $y \in G$, un élément dont l'ordre ne divise pas m ; nous notons q son ordre. Vu que q ne divise pas m, le

4. Le σ donné ici est l'inverse de celui donné dans l'énoncé. Cela ne change évidemment rien à la validité de l'énoncé et de la preuve.

5. Définition 2.12.

nombre q possède au moins un facteur premier plus de fois que m : soit p premier tel que la décomposition de q contienne p^β et celle de m contienne p^α avec $\beta > \alpha$. Autrement dit,

$$m = p^\alpha m' \tag{5.31a}$$
$$q = p^\beta q' \tag{5.31b}$$

où m' et q' ne contiennent plus le facteur p. L'élément x étant d'ordre m, l'élément x^{p^α} est d'ordre m'. De la même manière, l'élément $y^{q'}$ est d'ordre p^β. Étant donné que p^β et m' sont premiers entre eux, l'élément $x^{p^\alpha}y^{q'}$ est d'ordre $p^\alpha m' > m$. D'où une contradiction avec le fait que x était d'ordre maximal.

Par conséquent l'ordre de tous les éléments de G divise celui de x qui est alors le ppcm des ordres de tous les éléments de G, c'est-à-dire l'exposant de G. □

Proposition 5.22.
Soit G un groupe abélien fini et $x \in G$, un élément d'ordre maximum. Alors

(1) Il existe un morphisme $\varphi\colon G \to \mathrm{gr}(x)$ tel que $\varphi(x) = x$.

(2) Il existe un sous-groupe K de G tel que $G = \mathrm{gr}(x) \oplus K$.

Démonstration. Nous notons a l'ordre de x qui est également l'exposant du groupe G.

Nous allons prouver la première partie par récurrence sur l'ordre du groupe. Si $G = \mathrm{gr}(x)$, alors c'est évident. Soit H un sous-groupe propre de G contenant x et tel que le problème soit déjà résolu pour H : il existe un morphisme $\varphi\colon H \to \mathrm{gr}(x)$ tel que $\varphi(x) = x$. Soit $y \in G\backslash H$, d'ordre b. Nous allons trouver un morphisme $\hat{\varphi}\colon \mathrm{gr}(H, y) \to \mathrm{gr}(x)$ telle que $\hat{\varphi}(x) = x$.

Pour cela nous commençons par construire les applications suivantes :

$$\begin{aligned} \tilde{\varphi}\colon \mathbb{Z}/b\mathbb{Z} \times H &\to \mathrm{gr}(x) \\ (\bar{k}, h) &\mapsto x^{kl}\varphi(h) \end{aligned} \tag{5.32}$$

où l est encore à déterminer, et

$$\begin{aligned} p\colon \mathbb{Z}/b\mathbb{Z} \times H &\to \mathrm{gr}(y, H) \\ (\bar{k}, h) &\mapsto y^k h. \end{aligned} \tag{5.33}$$

Pour que $\tilde{\varphi}$ soit bien définie, il faut que a divise bl. L'application p est bien définie parce que $\bar{k}$ est pris dans $\mathbb{Z}/b\mathbb{Z}$ et que b est l'ordre de y.

Nous allons construire le morphisme $\hat{\varphi}$ en considérant le diagramme

$$\begin{tikzcd} \ker(p) \arrow[r, hook] & \mathbb{Z}/b\mathbb{Z} \times H \arrow[r, "p"] \arrow[d, "\tilde{\varphi}"'] & \mathrm{gr}(y, H) \arrow[dl, "\hat{\varphi}"] \\ & \mathrm{gr}(x) & \end{tikzcd} \tag{5.34}$$

que l'on voudra être commutatif. Vu que p est surjective, les théorèmes d'isomorphismes nous disent que

$$\mathrm{gr}(y, H) \simeq \frac{\mathbb{Z}/b\mathbb{Z} \times H}{\ker p}. \tag{5.35}$$

Si $[\bar{k}, h]$ est la classe de $(\bar{k}, h)$ modulo $\ker(p)$ alors nous voudrions définir $\hat{\varphi}$ par

$$\hat{\varphi}\big([\bar{k}, h]\big) = \tilde{\varphi}(\bar{k}, h). \tag{5.36}$$

Pour que cela soit bien définit, il faut que si $(\bar{r}, z) \in \ker p$,

$$\hat{\varphi}\big([\bar{k}\bar{r}, hz]\big) = \hat{\varphi}\big([\bar{k}, h]\big), \tag{5.37}$$

c'est-à-dire que $\tilde{\varphi}(\bar{r}, z) = e$. Du coup la définition (5.36) n'est bonne que si et seulement si

$$\ker(p) \subset \ker(\tilde{\varphi}). \tag{5.38}$$

Nous pouvons obtenir cela en choisissant bien l.

Déterminons d'abord le noyau de p. Pour cela nous considérons un nombre β divisant b tel que $\mathrm{gr}(y) \cap H = \mathrm{gr}(y^\beta)$. Nous aurons $p(\bar{k}, h) = e$ si et seulement si $y^h = e$. En particulier $h = y^{-k} \in \mathrm{gr}(y) \cap H = \mathrm{gr}(y^\beta)$. Si $h = (y^\beta)^m = y^{m\beta}$, alors $k = -m\beta$ et nous avons

$$\ker(p) = \{(-m\beta, y^{m\beta}) \text{ tel que } m \in \mathbb{Z}\}. \tag{5.39}$$

En plus court : $\ker(p) = \mathrm{gr}(\beta, y^{-\beta})$. Nous devons donc fixer l de telle sorte que $\tilde{\varphi}(\beta, y^{-\beta}) = e$. Étant donné que φ prend ses valeurs dans $\mathrm{gr}(x)$, il existe un entier α tel que $\varphi(y^{-\beta}) = x^\alpha$; en utilisant cet α, nous écrivons

$$\tilde{\varphi}(\beta, y^{-\beta}) = x^{\beta l}\varphi(y^{-\beta}) = x^{\beta l + \alpha}. \tag{5.40}$$

Par conséquent nous choisissons $l = -\alpha/\beta$. Nous devons maintenant vérifier que ce choix est légitime, c'est-à-dire que a divise bl et que α/β est un entier.

Étant donné que y est d'ordre b,

$$e = \varphi(y^b) = \varphi(y^{-\beta b/\beta}) = \varphi(y^{-\beta})^{b/\beta} = x^{b\beta/\alpha}. \tag{5.41}$$

Par conséquent a divise $\frac{b\alpha}{\beta} = -bl$.

Pour voir que l est entier, nous nous rappelons que a est l'exposant de G (parce que x est d'ordre maximum) et que par conséquent b divise a. Mais a divise $\alpha\frac{b}{\beta}$. Donc α/β est entier.

Nous passons maintenant à la seconde partie de la preuve. Nous considérons un morphisme $\varphi\colon G \to \mathrm{gr}(x)$ tel que $\varphi(x) = x$. La première partie nous en assure l'existence. Nous montrons que

$$\begin{aligned} \psi\colon G &\to \mathrm{gr}(x) \oplus \ker(\varphi) \\ g &\mapsto \big(\varphi(g), g\varphi(g)^{-1}\big) \end{aligned} \tag{5.42}$$

est un isomorphisme. D'abord $g\varphi(g)^{-1}$ est dans le noyau de φ parce que $\varphi(g)^{-1}$ étant dans $\mathrm{gr}(x)$, et φ étant un morphisme,

$$\varphi\big(g\varphi(g)^{-1}\big) = \varphi(g)\varphi(g)^{-1} = e. \tag{5.43}$$

L'application ψ est un morphisme parce que, en utilisant le fait que G est abélien,

$$\begin{aligned} \psi(g_1g_2) &= \big(\varphi(g_1g_2), g_1g_2\varphi(g_1g_2)^{-1}\big) && (5.44a)\\ &= \big(\varphi(g_1)\varphi(g_2), g_1\varphi(g_1)^{-1}g_2\varphi(g_2)^{-1}\big) && (5.44b)\\ &= \psi(g_1)\psi(g_2). && (5.44c) \end{aligned}$$

L'application ψ est injective parce que si $\psi(g) = (e, e)$ alors $\varphi(g) = e$ et $g\varphi(g)^{-1} = e$, ce qui implique $g = e$.

Enfin ψ est surjective parce qu'elle est injective et que les ensembles de départ et d'arrivée ont même cardinal. En effet par le premier théorème d'isomorphisme (théorème 2.28) appliqué à φ nous avons

$$|G| = |\,\mathrm{gr}(x)| \cdot |\ker(\varphi)|. \tag{5.45}$$

□

Théorème 5.23.
Tout groupe abélien fini (non trivial) se décompose en

$$G \simeq \mathbb{Z}/d_1\mathbb{Z} \oplus \ldots \oplus \mathbb{Z}/d_r\mathbb{Z} \tag{5.46}$$

avec $d_1 \geqslant 1$ et d_i divise d_{i+1} pour tout $i = 1, \ldots, r-1$.

De plus la liste $(d_1, \ldots, d_r)$ vérifiant ces propriétés est unique.

Démonstration. Soit x_1 un élément d'ordre maximal dans G. Soit n_1 son ordre et

$$H_1 = \mathrm{gr}(x_1) = \mathbb{F}_{n_1}. \tag{5.47}$$

D'après la proposition 5.22(2), il existe un supplémentaire K_1 tel que $G = \mathbb{F}_{n_1} \oplus K_1$. Si $K_1 = \{1\}$ on s'arrête et on garde $G = \mathbb{F}_{n_1}$. Sinon on continue de la sorte en prenant x_2 d'ordre maximal dans K_1 etc.

Nous devons maintenant prouver l'unicité de cette décomposition. Soit

$$G = \mathbb{F}_{d_1} \oplus \ldots \oplus \mathbb{F}_{d_r} = \mathbb{F}_{s_1} \oplus \ldots \oplus \mathbb{F}_{s_q}. \tag{5.48}$$

L'exposant de G est d_r et s_q. Donc $d_r = s_q$. Les complémentaires étant égaux nous avons

$$\mathbb{F}_{d_1} \oplus \ldots \oplus \mathbb{F}_{d_{r-1}} = \mathbb{F}_{s_1} \oplus \ldots \oplus \mathbb{F}_{s_{q-1}}. \tag{5.49}$$

En continuant nous trouvons $r = q$ et $d_i = s_i$. □

5.5 Groupes d'ordre pq

Lemme 5.24.
Soit G un groupe d'ordre pq où p et q sont des nombres premiers distincts. Nous supposons que $p < q$.

(1) Le groupe G possède un unique q-Sylow.

(2) Cet unique q-Sylow est normal dans G.

(3) Il n'est ni $\{e\}$ ni G.

(4) Le groupe G n'est pas un groupe simple[6].

Démonstration. Soit n_q le nombre de q-Sylow ; par le théorème de Sylow 5.11(1) le groupe G possède des q-Sylow et par 5.11(4),

$$n_q \in [1]_q. \tag{5.50}$$

De plus le nombre n_q divise $|G| = pq$. Donc n_q vaut p, q ou 1. Avoir $n_q = p$ n'est pas possible parce que $n_q \in [1]_q$ et $p < q$. Avoir $n_q = q$ n'est pas possible non plus, pour la même raison. Donc $n_q = 1$. Notons H l'unique q-Sylow de G.

Le fait que H soit normal est une conséquence de 5.11(3) parce que le conjugué de H est encore un q-Sylow alors que H est l'unique q-Sylow.

Vu que

$$1 < p = |H| < pq = |G|, \tag{5.51}$$

le sous-groupe H n'est ni réduit à l'identité ni le groupe entier.

Par conséquent G n'est pas simple parce qu'il contient un sous-groupe normal non trivial. □

Avant le lire le théorème suivant, n'oubliez pas de lire la définition d'un produit semi-direct 2.79.

Théorème 5.25 ([89]).
Soient deux nombres premiers distincts[7] *p et q avec $q > p$.*

(1) Si p ne divise pas $q - 1$ alors tout groupe d'ordre pq est cyclique et plus précisément le seul groupe (à isomorphisme près) d'ordre pq est $\mathbb{Z}/pq\mathbb{Z}$.

(2) Si $p \mid q - 1$, alors il n'existe que deux groupes d'ordre pq :

- *Le groupe abélien et cyclique $\mathbb{Z}/pq\mathbb{Z}$.*
- *Le produit semi-direct non abélien*

$$G = \mathbb{Z}/q\mathbb{Z} \times_\varphi \mathbb{Z}/p\mathbb{Z} \tag{5.52}$$

où $\varphi(\bar{1})$ est d'ordre p dans $\mathrm{Aut}(\mathbb{Z}/q\mathbb{Z})$.

6. Pas de sous-groupes normaux non triviaux, 2.5.
7. Le cas $p = q$ sera traité par la proposition 5.28.

(3) Si p et q sont premiers entre eux, le produit est direct[8].

Démonstration. Division de la preuve en plusieurs parties.

Préliminaires avec Sylow Soit un groupe G d'ordre pq. Soient H, un q-Sylow et K, un p-Sylow de G. Ils existent parce que p et q sont des diviseurs premiers de $|G|$ (théorème de Sylow 5.11). Si n_q est le nombre de q-Sylow dans G alors n_q divise $|G|$ et $n_q = 1 \mod q$. Donc d'abord n_q vaut 1, p ou q. Ensuite $n_q = q$ est exclu par la condition $n_q = 1 \mod q$; la possibilité $n_q = p$ est également impossible parce que $p = 1 \mod q$ est impossible avec $p < q$. Donc $n_q = 1$ et H est normal dans G.

L'ensemble $H \cap K$ est un sous-groupe à la fois de H et de K, ce qui entraine que (théorème de Lagrange 2.34) $|H \cap K|$ divise à la fois p et q. Nous en déduisons que $|H \cap K| = 1$ et donc que $H \cap K = \{e\}$.

Étant donné que H est normal, l'ensemble HK est un sous-groupe de G. De plus l'application

$$\begin{aligned} \psi\colon H \times K &\to HK \\ (h,k) &\mapsto hk \end{aligned} \tag{5.53}$$

est un bijection. Nous ne devons vérifier seulement l'injectivité. Supposons que $hk = h'k'$. Alors $e = h^{-1}h'k'k^{-1}$, et donc

$$h^{-1}h' = (k'k^{-1})^{-1} \in H \cap K = \{e\}. \tag{5.54}$$

Par conséquent $|pq| = |H \times K| = |HK|$, et $HK = G$. Le corolaire 2.81 nous indique que

$$G = H \times_\varphi K \tag{5.55}$$

où φ est l'action adjointe. Nous devons maintenant identifier cette action. En d'autres termes, nous savons que $H = \mathbb{Z}/q\mathbb{Z}$ et $K = \mathbb{Z}/p\mathbb{Z}$ et que $\varphi\colon \mathbb{Z}/p\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/q\mathbb{Z})$ est un morphisme. Nous devons déterminer les possibilités pour φ.

Soit n_p le nombre de p-Sylow de G. Comme précédemment, n_p vaut 1, p ou q et la possibilité $n_p = p$ est exclue. Donc n_p est 1 ou q.

Si p ne divise pas $q-1$ Si p ne divise pas $q-1$ alors il n'est pas possible d'avoir $n_p = q$ parce que $n_p \in [1]_p$. Or dire $n_p = q$ demanderait $q \in [1]_p$, c'est-à-dire $q = kp + 1$, qui impliquerait que p divise $q - 1$.

La seule possibilité est que $n_p = 1$. Dans ce cas, K est également normal dans G. Du coup le produit semi-direct (5.55) est en réalité un produit direct (φ est triviale) et nous avons

$$G = \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} = \mathbb{Z}/pq\mathbb{Z}. \tag{5.56}$$

Si p divise $q-1$ Cette fois $n_p = 1$ et $n_p = q$ sont tous deux possibles. Ce que nous savons est que $\varphi(\mathbb{Z}/p\mathbb{Z})$ est un sous-groupe de $\mathrm{Aut}(\mathbb{Z}/q\mathbb{Z})$. Par le premier théorème d'isomorphisme 2.28, nous avons

$$|\varphi(\mathbb{Z}/p\mathbb{Z})| = \frac{|\mathbb{Z}/p\mathbb{Z}|}{|\ker\varphi|}, \tag{5.57}$$

ce qui signifie que $|\varphi(\mathbb{Z}/p\mathbb{Z})|$ divise $|\mathbb{Z}/p\mathbb{Z}| = p$. Par conséquent, $|\varphi(\mathbb{Z}/p\mathbb{Z})|$ est égal à 1 ou p. Si c'est 1, alors l'action est triviale et le produit est direct.

Nous supposons que $|\varphi(\mathbb{Z}/p\mathbb{Z})| = p$. Le corolaire 5.20 nous indique que $\mathrm{Aut}(\mathbb{Z}/q\mathbb{Z})$ possède un unique sous-groupe d'ordre p que nous notons Γ ; c'est-à-dire que $\Gamma = \mathrm{Image}(\varphi)$. Vu que $\varphi\colon \mathbb{Z}/p\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/q\mathbb{Z})$ est un morphisme, Γ est généré par $\varphi(\bar{1})$ qui est alors un élément d'ordre p, comme annoncé.

8. Cette affirmation me semble très bizarre. Comment deux nombres premiers distincts pourraient ne pas être premiers entre eux ? ? ?

Unicité Nous nous attaquons maintenant à l'unicité. Soient φ et φ' deux morphismes non triviaux $\mathbb{Z}/p\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/q\mathbb{Z})$. Étant donné que $\mathrm{Aut}(\mathbb{Z}/q\mathbb{Z})$ ne possède qu'un seul sous-groupe d'ordre p, nous savons que $\mathrm{Image}(\varphi) = \mathrm{Image}(\varphi') = \Gamma$. Nous pouvons donc parler de φ'^{-1} en tant qu'application de $\mathbb{Z}/p\mathbb{Z}$ dans Γ. Nous montrons que

$$\begin{aligned} f\colon \mathbb{Z}/q\mathbb{Z} \times_\varphi \mathbb{Z}/p\mathbb{Z} &\to \mathbb{Z}/q\mathbb{Z} \times_{\varphi'} \mathbb{Z}/p\mathbb{Z} \\ (h,k) &\mapsto (h, \alpha(k)) \end{aligned} \tag{5.58}$$

où $\alpha = \varphi'^{-1} \circ \varphi$ est un isomorphisme de groupes. Le calcul est immédiat :

$$f(h_1,k_1)f(h_2mk_2) = \big(h_1, \alpha(k_1)\big)(h_2, \alpha(k_2)) \tag{5.59a}$$
$$= \big(h_1\varphi'(\alpha(k_1))h_2m\alpha(k_1k_2)\big) \tag{5.59b}$$
$$= f\big(h_1\varphi(k_1)h_2, k_1k_2\big) \tag{5.59c}$$
$$= f\big((h_1,k_1),(h_2,k_2)\big). \tag{5.59d}$$

Par conséquent $\mathbb{Z}/q\mathbb{Z} \times_\varphi \mathbb{Z}/p\mathbb{Z} \simeq \mathbb{Z}/q\mathbb{Z} \times_{\varphi'} \mathbb{Z}/p\mathbb{Z}$.

□

Note : il existe des nombres premiers p et q tels que $q = 1 \mod p$. Par exemple $7 = 1 \mod 3$.

Proposition 5.26 ([31]).
Soit G un groupe fini d'ordre pq où p et q sont deux nombres premiers distincts vérifiant

$$\begin{cases} p \neq 1 \mod q & (5.60a) \\ q \neq 1 \mod p. & (5.60b) \end{cases}$$

Alors G est cyclique, abélien et

$$G \simeq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}. \tag{5.61}$$

Démonstration. Soient n_p et n_q les nombres de p-Sylow et q-Sylow. Par le théorème de Sylow 5.11, n_p divise pq et $n_p = 1 \mod p$. Le second point empêche n_p de diviser p. Par conséquent n_p divise q et donc n_p vaut 1 ou q. La possibilité $n_p = q$ est exclue par l'hypothèse $q \neq 1 \mod p$. Donc $n_p = 1$, et de la même façon nous obtenons $n_q = 1$.

Soient S l'unique p-Sylow et T, l'unique q-Sylow. Pour les mêmes raisons que celles exposée plus haut, ce sont deux sous-groupes normaux dans G. Étant donné que S est d'ordre p^n pour un certain n et que l'ordre de S doit diviser celui de G, nous avons $|S| = p$. De la même façon, $|T| = q$. Par conséquent S est un groupe cyclique d'ordre p et nous considérons x, un de ses générateurs. De la même façon soit y, un générateur de T.

Nous montrons maintenant que x et y commutent, puis que xy engendre G. Nous savons que $S \cap T$ est un sous-groupe à la fois de S et de T, de telle façon que $|S \cap T|$ divise à la fois $|S| = p$ et $|T| = q$. Nous avons donc $|S \cap T| = 1$ et donc $S \cap T$ se réduit au neutre. Par ailleurs, S et T sont normaux, donc

$$(xyx^{-1})y^{-1} \in T \tag{5.62a}$$
$$x(yx^{-1})y^{-1}) \in S, \tag{5.62b}$$

donc $xyx^{-1}y^{-1} = e$, ce qui montre que $xy = yx$.

Montrons que xy engendre G. Soit $m > 0$ tel que $(xy)^m = e$. Pour ce m nous avons $x^m = y^{-m}$ et $y^{-m} = x^m$, ce qui signifie que x^m et y^m appartiennent à $S \operatorname{cat} T$ et donc $x^m = y^m = e$. Les nombres p et q divisent donc tous deux m ; par conséquent $\mathrm{ppcm}(p,q) = pq$ divise m. Nous en concluons que xy est d'ordre pq (il ne peut pas être plus) et qu'il est alors générateur.

Pour la suite nous allons d'abord prouver que $G = ST$ puis que $G \simeq S \times T$. Nous savons déjà que $|S \cap T| = 1$, ce qui nous amène à dire que $|ST| = |S||T|$. En effet si $s, s' \in S$ et $t, t' \in t$ et si $st = s't'$, alors $t = s^{-1}s't'$, ce qui voudrait dire que $s^{-1}s' \in T$ et donc que $s^{-1}s' = e$. Au final nous avons

$$|ST| = |S||T| = pq = |G|. \tag{5.63}$$

Par conséquent $G = ST$. En nous rappelant du fait que $S \cap T = \{e\}$ et que S et T sont normaux, le lemme 2.17 nous dit que $G \simeq S \times T$. Le groupe S étant cyclique d'ordre p nous avons $S = \mathbb{Z}/p\mathbb{Z}$ et pour T, nous avons la même chose : $T = \mathbb{Z}/q\mathbb{Z}$. Nous concluons que

$$G \simeq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}. \tag{5.64}$$

□

Théorème 5.27 (Théorème de Burnside[36])**.**
Le centre d'un p-groupe non trivial est non trivial.

Démonstration. Soit G un p-groupe non trivial. Nous considérons l'action adjointe G sur lui-même. Les points fixes de cette action sont les éléments du centre :

$$\mathcal{Z}_G = \{z \in G \text{ tel que } \sigma_x(z) = z \forall x \in G\} = \mathrm{Fix}_G(G). \tag{5.65}$$

Nous utilisons l'équation aux classes (2.57) pour dire que $|G| = |\mathcal{Z}_G| \mod p$. Mais $|\mathcal{Z}_G|$ n'est pas vide parce qu'il contient l'identité. Donc $|\mathcal{Z}_G|$ est au moins d'ordre p. □

Proposition 5.28.
Si p est un nombre premier, tout groupe d'ordre p ou p^2 est abélien.

Rappel : un groupe d'ordre p ou p^2 est automatiquement un p-groupe.

Démonstration. Si $|G| = p$, alors le théorème de Cauchy 5.2 nous donne l'existence d'un élément d'ordre p. Cet élément est alors automatiquement générateur, G est cyclique et donc abélien.

Si par contre G est d'ordre p^2, alors les choses se compliquent (un peu). D'après le théorème de Burnside 5.27, le centre $\mathcal{Z}$ n'est pas trivial ; il est alors d'ordre p ou p^2. Supposons qu'il soit d'ordre p et prenons $x \in G \backslash \mathcal{Z}$. Alors le stabilisateur de x pour l'action adjointe contient au moins $\mathcal{Z}$ et x, c'est-à-dire que $|\mathrm{Fix}_G(x)| \geqslant p + 1$. Étant donné que $\mathrm{Fix}_G(x)$ est un sous-groupe, son ordre est automatiquement 1, p ou p^2. En l'occurrence, il doit être p^2 (parce que plus grand que p), et donc x doit être central, ce qui est une contradiction. □

5.6 Groupe symétrique, groupe alterné

La définition des permutations et du groupe symétrique sont 2.63. Voir aussi le thème 61.

5.6.1 Le groupe alterné

Définition 5.29.
Le groupe A_n des permutations paires[9] *dans S_n est la* ***groupe alterné****.*

Proposition 5.30.
À propos du groupe alterné dans le groupe symétrique.

(1) Le groupe alterné A_n est un sous-groupe caractéristique[10] *de S_n*

(2) Le sous-groupe A_n est d'indice 2 dans S_n.

(3) Le sous-groupe A_n est l'unique sous-groupe d'indice[11] *2 de S_n.*

Démonstration. Soit $\alpha \in \mathrm{Aut}(S_n)$. Étant donné que $\epsilon \circ \alpha$ est un homomorphisme surjectif sur $\{-1, 1\}$, par unicité de cet homomorphisme, nous avons $\epsilon \circ \alpha = \epsilon$, et donc $\alpha(A_n) = A_n$. Par le premier théorème d'isomorphisme 2.28, il existe un isomorphisme

$$f \colon S_n/\ker(\epsilon) \to \mathrm{Image}(\epsilon). \tag{5.66}$$

9. Définition 2.74.
10. Définition 2.4.
11. Définition 2.33.

En égalant le nombre d'éléments nous avons $|S_n : \ker\epsilon| = |S_n : A_n| = 2$.

Nous prouvons maintenant l'unicité. Soit H un sous-groupe d'indice 2 dans S_n. Par le lemme 3.36, H est distingué et nous pouvons considérer le groupe S_n/H. Ce dernier ayant 2 éléments, il est isomorphe à $\{-1,1\}$. Soit θ l'isomorphisme. On note φ le morphisme canonique $\varphi\colon S_n \to S_n/H$:

$$S_n \xrightarrow{\varphi} S_n/H \xrightarrow{\theta} \{-1,1\}. \tag{5.67}$$

La composition $\varphi\circ\theta$ est alors un homomorphisme surjectif de S_n sur $\{-1,1\}$ et nous avons $\varphi\circ\theta = \epsilon$ par la proposition 2.76. L'enchainement (5.67) nous montre que $H = \ker(\theta\circ\varphi) = \ker(\epsilon) = A_n$. □

Proposition 5.31 ([90]).
Le groupe symétrique S_n peut être écrit comme un produit semi-direct[12] *du groupe alterné :*

$$S_n = A_n \times_\varphi \mathbb{Z}/2\mathbb{Z} \tag{5.68}$$

où l'action de $\mathbb{Z}/2\mathbb{Z}$ sur A_n est la conjugaison par $\sigma = (12)$, c'est-à-dire $\rho(-1)\tau = \sigma\tau\sigma^{-1}$.

Démonstration. Nous avons la suite exacte

$$1 \xrightarrow{i} A_n \xrightarrow{i} S_n \xrightarrow{\epsilon} \{\pm 1\} \longrightarrow 1 \tag{5.69}$$

où les i représentent des inclusions et ϵ est la signature définie en 2.68. Grâce à cette suite et au fait que la signature soit un isomorphisme à partir de la partie $\{\mathrm{Id},\sigma\}$ (pour σ d'ordre 2, par exemple $\sigma = (12)$), le théorème 2.80 nous dit que

$$S_n \simeq A_n \times_\varphi \{\mathrm{Id},\sigma\} \tag{5.70}$$

où φ est l'action adjointe de $\{\mathrm{Id},\sigma\}$ sur A_n. □

Proposition 5.32.
Si $\beta \in S_n$ est une transposition, nous avons les égalités suivante d'ensembles :

$$S_n = A_n \cup A_n\beta = A_n \cup \beta A_n. \tag{5.71}$$

Démonstration. Les parties A_n et βA_n ont le même nombre d'éléments. En effet, l'application

$$\begin{aligned}\varphi\colon A_n &\to A_n\beta\\ \sigma &\mapsto \sigma\beta\end{aligned} \tag{5.72}$$

est une bijection.

De plus ces deux ensembles sont disjoints à cause de la proposition 2.76. En effet si $\sigma \in A_n$, alors $\epsilon(\sigma) = 1$. Mais un élément de $A_n\beta$ est de la forme $\sigma\beta$ avec $\sigma \in A_n$. Or ϵ est une homomorphisme, donc $\epsilon(\sigma\beta) = \epsilon(\sigma)\epsilon(\beta) = -1$.

Enfin, la proposition 5.30(2) dit que A_n est d'indice deux dans S_n. Donc la partie

$$A_n \cup A_n\beta \tag{5.73}$$

contient $|S_n|/2 + |S_n|/2 = |S_n|$ éléments. C'est donc S_n. □

Lemme 5.33.
Le groupe dérivé du groupe symétrique est le groupe alterné : $D(S_n) = A_n$.

Démonstration. Tout élément de $D(S_n)$ s'écrit sous la forme $ghg^{-1}h^{-1}$. Quel que soit le nombre de transpositions dans g et h, le nombre de transpositions dans $[g,h]$ est pair. □

Proposition 5.34 ([91]).
Soit $n \geqslant 3$. Les 3-cycles $c_i = (1,2,i)$ avec $i = 3,\dots,n$ engendrent le groupe alterné A_n.

12. Définition 2.79.

Démonstration. Soit H, le groupe engendré par les c_i. D'abord nous avons

$$c_i = (1, 2, i) = (1, 2)(2, i), \tag{5.74}$$

de telle sorte que $\epsilon(c_i) = 1$. Par conséquent nous avons $H \subset A_n$. Nos montrons par récurrence que $A_n \subset H$.

Pour $n = 3$ il suffit de vérifier que $A_3 = \{\mathrm{Id}, c_3, c_3^2\}$. Supposons avoir obtenu le résultat pour A_{n-1}, et prouvons le pour A_n. Soit $s \in A_n$.

Si $s(n) = n$, alors s se décompose de la même manière que sa restriction s' à $\{1, \dots, n-1\}$. Par l'hypothèse de récurrence, cette restriction, appartenant à A_{n-1}, se décompose en produit des $c_3, \dots, c_{n-1}$ et de leurs inverses.

Si $s(n) = k$ alors nous considérons l'élément $c_n^2 c_k s$. Cet élément envoie n sur n et peut donc être décomposé avec les c_i $(i = 1, \dots, n-1)$ en vertu du point précédent. □

Proposition 5.35.
Lorsque $n \geqslant 5$, tous les 3-cycles de A_n sont conjugués. Autrement dit, la classe de conjugaison d'un 3-cycle est l'ensemble des 3-cycles.

Démonstration. Soient les 3-cycles $\sigma = (i_1, i_2, i_3)$ et $\varphi = (j_1, j_2, j_3)$. Nous considérons une bijection α de $\{1, \dots, n\}$ telle que $\alpha(i_s) = j_s$. Nous avons immédiatement que $\alpha \in S_n$ et que $\alpha\sigma\alpha^{-1} = \varphi$. Donc les 3-cycles sont conjugués dans S_n. Il reste à prouver qu'ils le sont dans A_n.

Si α est une permutation paire, la preuve est terminée. Si α est impaire, alors nous devons un peu la modifier. Vu que $n \geqslant 5$, nous pouvons prendre s et t, des éléments distincts dans $\{1, \dots, n\} \backslash \{j_1, j_2, j_3\}$ et poser $\tau = (st)$. Vu que la signature est un homomorphisme et que τ et α sont impairs, l'élément $\tau\alpha$ est pair (lemme et proposition 2.75 et 2.73) et est donc dans A_n. Les supports de τ et φ étant disjoints, ces derniers commutent et nous avons

$$(\tau\alpha)\sigma(\tau\alpha)^{-1} = \tau(\alpha\sigma\alpha^{-1})\tau^{-1} = \tau\varphi\tau^{-1} = \varphi. \tag{5.75}$$

Donc σ et φ sont conjugués par $\tau\alpha$ qui est dans A_n. □

Théorème 5.36 ([31]).
Le groupe alterné A_n est simple[13] *pour $n \geqslant 5$.*

Démonstration. Soit N, un sous-groupe normal de A_n non réduit à l'identité. Étant donné que les 3-cycles engendrent A_n (proposition 5.34) et que tous les 3-cycles sont conjugués dans A_n (proposition 5.35), il suffit de montrer que N contient un 3-cycle. En effet si N contient un 3-cycle, le fait qu'il soit normal implique (par conjugaison) qu'il les contienne tous et donc qu'il contient une partie génératrice de A_n.

Soit donc $\sigma \in N$ différent de l'identité. Nous prenons i dans le support de σ et $j = \sigma(i)$. Nous choisissons ensuite $k \in \{1, \dots, n\} \backslash \{i, j, \sigma^{-1}(i)\}$ et $m = \sigma(k)$. Nous considérons la permutation $\alpha = (ijk)$. Étant donné que N est normal l'élément

$$\theta = (\alpha^{-1}\sigma\alpha)\sigma^{-1} \tag{5.76}$$

est dans N. De plus en utilisant le lemme 2.69 et le fait que $\alpha^{-1} = (ikj)$ nous avons

$$\theta = (ikj)(j\sigma(j)m). \tag{5.77}$$

Cela n'est pas spécialement un 3-cycle, mais nous allons en construire un. Nous allons déterminer que θ est soit un 5-cycle, soit un 3-cycle , soit un 2×2-cycle suivant les valeurs de $\sigma(j)$ et m.

Souvenons-nous que nous avons :

— $i \neq j = \sigma(i)$, puisque i est dans le support de σ ;

— $k \neq i$ et $k \neq j$, par définition de k (rappelons aussi que $k \neq \sigma^{-1}(i)$) ;

13. Pas de sous-groupes normaux non triviaux, définition 2.5.

— $m \neq i$, $m \neq j$ et $m \neq \sigma(j)$ puisque $m = \sigma(k)$.

Il ne nous reste alors seulement les deux possibilités suivantes :

(1) soit $m = k$, soit $m \neq k$, d'une part ;

(2) soit $\sigma(j) = i$, soit $\sigma(j) = k$, soit $\sigma(j)$ n'est ni i, ni k, ni m, d'autre part.

Supposons dans un premier temps que $m = k$; alors

$$\theta = (ik)(j\sigma(j)). \tag{5.78}$$

C'est à priori un 2×2-cycle. Mais si de plus $\sigma(j) = i$, alors

$$\theta = (ijk) \tag{5.79}$$

qui est un 3-cycle ; et si $\sigma(j) = k$, alors

$$\theta = (ikj) \tag{5.80}$$

qui est un autre 3-cycle.

Supposons à présent que $m \neq k$. Si $\sigma(j)$ n'est ni i, ni k, ni m, alors i, j, k, $\sigma(j)$ et m sont cinq nombres différents, et

$$\theta = (i, j, \sigma(j), m, k) \tag{5.81}$$

est un 5-cycle. Si $\sigma(j) = i$, alors

$$\theta = (ikj)(jim) = (imk) \tag{5.82}$$

qui est un 3-cycle. Si $\sigma(j) = k$, alors

$$\theta = (ikj)(jkm) = (ikm) \tag{5.83}$$

qui est encore un 3-cycle.

Bref nous avons montré que θ est soit un 3-cycle, soit un 5-cycle, soit un 2×2-cycle. Si θ est un 3-cycle, la preuve est terminée.

Si $\theta = (ab)(cd)$, alors on considère $e \in \{1, \dots, n\} \backslash \{a, b, c, d\}$ et nous avons

$$\underbrace{(abe)^{-1}\theta(abe)}_{\in N}\theta^{-1} = (aeb)(ab)(cd)(abe)(an)(cd) = (abe) \in N. \tag{5.84}$$

Si θ est le 5-cycle $(abcde)$, alors l'élément suivant est dans N :

$$(abc)^{-1}\theta(abc)\theta^{-1} = (acb)(abcde)(abc)(aedcb) = (acd). \tag{5.85}$$

Dans tous les cas nous avons trouvé un 3-cycle dans N et nous avons par conséquent $N = A_n$, ce qui fait que A_n ne contient pas de sous-groupes normaux non triviaux. Le groupe alterné A_n est donc simple. □

Nous en déduisons immédiatement que si $n \geqslant 5$, le groupe dérivé de A_n est A_n parce que A_n ne contient pas d'autres sous-groupes non triviaux.

Lemme 5.37.
Le groupe alterné[14] A_6 *n'accepte pas de sous-groupes normaux d'ordre* 60.

Démonstration. Soit G normal dans A_6, et a, un élément d'ordre 5 dans G (qui existe parce que 5 divise 60). Soit aussi un élément b d'ordre 5 dans A_6. Les groupes $\mathrm{gr}(a)$ et $\mathrm{gr}(b)$ sont deux 5-Sylow dans A_6. En effet, 5 un nombre premier et est la plus grande puissance de 5 dans la décomposition de 60 ; donc $\mathrm{gr}(a)$ est un 5-Sylow dans G. D'autre part, l'ordre de A_6 (qui est $\frac{1}{2}6!$) ne possède également que 5 à la puissance 1 dans sa décomposition.

En vertu du théorème de Sylow 5.11(3), les 5-Sylow $\mathrm{gr}(a)$ et $\mathrm{gr}(b)$ sont conjugués et il existe $\tau \in A_6$ tel que $b = \tau a \tau^{-1}$. Mais G étant normal dans A_6, l'élément $\tau a \tau^{-1}$ est encore dans G, de telle sorte que $b \in G$. Du coup G doit contenir tous les éléments d'ordre 5 de A_6.

14. Définition 5.29.

Les éléments d'ordre 5 de A_6 doivent fixer un des points de $\{1,2,3,4,5,6\}$ puis permuter les autres de façon à n'avoir qu'un seul cycle. Un cycle correspond à écrire les nombres $1,2,3,4,5$ dans un certain ordre. Ce faisant, le premier n'a pas d'importance parce qu'on considère la permutation cyclique, par exemple $(3,5,2,1,4)$ est la même chose que $(5,2,1,4,3)$. Le nombre de cycles sur $\{1,2,3,4,5\}$ est donc de $4!$, et par conséquent le nombre d'éléments d'ordre 5 dans A_6 est $6 \cdot 4! = 144$.

Le groupe G doit contenir au moins 144 éléments alors que par hypothèse il en contient 60 ; contradiction. □

Le théorème suivant montre que tout groupe peut être vu, en agissant sur lui-même, comme une partie du groupe symétrique.

Théorème 5.38.
Un groupe G est isomorphe à un sous-groupe de son groupe symétrique $S(G)$.

Démonstration. Nous considérons φ, la translation à gauche :

$$\begin{aligned} \varphi\colon G &\to S(G) \\ g &\mapsto t_g \end{aligned} \tag{5.86}$$

où $f_g(h) = gh$. Étant donné que

$$\varphi(gh) = ghx = g(t_h x) = t_g \circ t_h(x), \tag{5.87}$$

l'application φ est un morphisme de groupes. Il est injectif parce que si $gx = hx$ pour tout x, en particulier pour $x = e$ nous trouvons $g = h$.

De la même manière, $\varphi(g)x = \varphi(g)y$ implique $x = y$. Cela montre que l'image est bien dans le groupe symétrique.

L'ensemble $\text{Image}(\varphi)$ est donc un sous-groupe de $S(G)$, et φ est un isomorphisme vers ce groupe. □

Lemme 5.39.
Si $n \geqslant 3$, alors

(1) Le centre de S_n est trivial.

(2) Le groupe S_n est non abélien.

Démonstration. Soit $s \in Z(S_n)$ et trois éléments distincts a, b et c de $\{1,\dots,n\}$. Nous posons $\tau = (ab)$ et nous avons $s\tau = \tau s$. En notant $a' = s(a)$ et $b' = s(b)$ nous avons

$$a' = s(a) = (\tau s \tau^{-1})(a) = (\tau s)(b) = \tau(b') \tag{5.88a}$$

$$b' = s(b) = (\tau s \tau^{-1})(b) = (\tau s)(a) = \tau(a'). \tag{5.88b}$$

Donc τ permute a' et b'. Mais comme τ ne permute que a et b, en tant qu'ensembles, $\{a,b\} = \{s(a), s(b)\}$. Le même raisonnement sur $\{b,c\}$ donne $\{b,c\} = \{s(b), s(c)\}$. Et vu que a, b et c sont distincts,

$$\{b\} = \{b,c\} \cap \{a,b\} = \{s(b)\}. \tag{5.89}$$

Cela montre que $s(b) = b$, et donc que le centre de S_n est réduit à la permutation identité.

En ce qui concerne le fait que S_n est non abélien, si nous avions $st = ts$ pour tout $s,t \in S_n$ alors $s = tst^{-1}$ pour tout t. Alors s serait dans le centre de S_n. En bref, si S_n était abélien, son centre serait S_n et non $\{\text{Id}\}$.

□

Proposition 5.40 ([92, 87])**.**
Tout groupe simple[15] *d'ordre 60 est isomorphe au groupe alterné A_5.*

15. Définition 2.5.

Démonstration. Nous avons la décomposition en nombres premiers $60 = 2^2 \cdot 3 \cdot 5$. Déterminons pour commencer le nombre n_5 de 5-Sylow dans G. Le théorème de Sylow 5.11(4) nous renseigne que n_5 doit diviser 60 et doit être égal à $1 \mod 5$. Les deux seules possibilités sont $n_5 = 1$ et $n_5 = 6$. Étant donné que tous les p-Sylow sont conjugués, si $n_5 = 1$ alors le 5-Sylow serait un sous-groupe invariant à l'intérieur de G, ce qui est impossible vu que G est simple. Donc $n_5 = 6$.

Par le point (3) du théorème de Sylow, le groupe G agit transitivement sur l'ensemble des 5-Sylow par l'action adjointe :

$$g \cdot S = gSg^{-1}. \tag{5.90}$$

Cela donne donc un morphisme $\theta\colon G \to S_6$. Le noyau de θ est un sous-groupe normal. En effet si $k \in \ker\theta$ et si $g \in G$ nous avons

$$\begin{aligned}
(gkg^{-1}) \cdot S &= gkg^{-1}Ggk^{-1}g^{-1} && (5.91a)\\
&= gkTk^{-1}g^{-1} && (5.91b)\\
&= gTg^{-1} && (5.91c)\\
&= S && (5.91d)
\end{aligned}$$

où T est le Sylow $T = g^{-1}Sg$. Étant donné que $k \in \ker\theta$ nous avons utilisé $kTk^{-1} = aT$. Au final $gkg^{-1} \cdot S = S$, ce qui prouve que $gkg^{-1} \in \ker\theta$.

Étant donné que $\ker\theta$ est normal dans G, soit est soit réduit à $\{e\}$ soit il vaut G. La seconde possibilité est exclue parce qu'elle reviendrait à dire que G agit trivialement, ce qui n'est pas correct étant donné qu'il agit transitivement. Nous en déduisons que $\ker\theta = \{e\}$, que θ est injective et que G est isomorphe à un sous-groupe de S_6.

Par ailleurs le groupe dérivé de G est un sous-groupe normal (et non réduit à l'identité parce que G est non commutatif). Donc $D(G) = G$. Étant donné que $G \subset S_6$, nous avons

$$G = D(G) \subset D(S_6) = A_6 \tag{5.92}$$

parce que le groupe dérivé du groupe symétrique est le groupe alterné (lemme 5.33).

L'ensemble $\theta^{-1}(A_6)$ est distingué dans G. En effet si $\sigma \in A_6$ et si $g \in G$ nous avons

$$\theta\big(g\theta^{-1}(\sigma)g^{-1}\big) = \theta(g)\sigma\theta(g)^{-1} \in A_6. \tag{5.93}$$

Nous en déduisons que $\theta^{-1}(A_6)$ est soit G entier soit réduit à $\{e\}$. Si $\theta^{-1}(A_6) = \{e\}$, alors pour tout $g \in G$ nous aurions $g^2 = e$ parce que $\theta(g^2) \in A_6$. L'ordre de G étant 60, il n'est pas possible que tous ses éléments soient d'ordre 2. Nous en déduisons que $\theta(G) \subset A_6$.

Nous nommons $H = \theta(G)$ et nous considérons l'ensemble $X = A_6/H$ où les classes sont prises à gauche, c'est-à-dire

$$[\sigma] = \{h\sigma \text{ tel que } h \in H\}. \tag{5.94}$$

Évidemment A_6 agit sur X de façon naturelle. Au niveau de la cardinalité,

$$\mathrm{Card}(X) = \frac{|A_6|}{|G|} = \frac{360}{60} = 6. \tag{5.95}$$

Le groupe A_6 agit sur X qui a 6 éléments. Nous avons donc une application $\varphi\colon A_6 \to A_6$. Encore une fois, la simplicité de A_6 montre que $\varphi(A_6) = A_6$.

Nous étudions maintenant $\varphi(H)$ agissant sur X. Un élément $x \in A_6$ fixe la classe de l'unité $[e]$ si et seulement si $x \in H$ et par conséquent $\varphi(H)$ est la fixateur de $[e]$ dans X. À la renumérotation près, nous pouvons identifier $\varphi(H)$ au sous-groupe de A_6 agissant sur $\{1, \dots, 6\}$ et fixant 6. Nous avons alors $\varphi(H) = S_5 \cap A_6 = A_5$. Nous venons de prouver que φ fournit un isomorphisme entre A_5 et H. Étant donné que H était isomorphe à G, nous concluons que G est isomorphe à A_6. $\square$

5.6.2 Sous-groupes normaux

5.41 ([93]).
Soit le groupe V_4 engendré par les bitranspositions de S_4. Nous savons de l'exemple 2.72(5) que ce groupe contient exactement 3 éléments non triviaux et l'identité. De plus, comme c'est une classe de conjugaison, V_4 est normal dans S_4.

Lemme 5.42.
Les sous-groupes $\mathrm{Fix}_{S_n}(a)$ *(avec* $a \in \{1, \dots, n\}$*) sont conjugués entre eux.*

Démonstration. Soit $\sigma \in \mathrm{Fix}(a)$ et $s \in S_n$ nous devons prouver que $s\sigma s^{-1}$ est le fixateur d'un élément de $\{1, \dots, n\}$. Nous notons $s(a) = b$. Alors

$$(s\sigma s^{-1})(b) = (s\sigma)(a) = s(a) = b. \tag{5.96}$$

Donc $s\,\mathrm{Fix}(a)s^{-1} \subset \mathrm{Fix}(b)$.

Dans l'autre sens, si $\sigma \in \mathrm{Fix}(b)$ alors $s^{-1}\sigma s \in \mathrm{Fix}(a)$. Mais $\sigma = s(s^{-1}\sigma s)s^{-1}$, donc $\sigma \in s\,\mathrm{Fix}(a)s^{-1}$. □

Proposition 5.43 (Sous-groupes normaux de S_n [93]).
Les sous-groupes normaux de S_n *ne sont pas légions.*

(1) Pour $n = 4$*, le sous-groupes normaux de* S_4 *sont* $\{\mathrm{Id}\}$*,* V_4*,* A_4 *et* S_4*.*

(2) Pour $n \neq 4$*, les sous-groupes normaux de* S_n *sont* $\{\mathrm{Id}\}$*,* A_n *et* S_n*.*

Démonstration. Les cas $n \leqslant 2$ sont un peu triviaux, donc nous faisons $n \geqslant 3$. Soit H normal dans S_n et $s \neq \mathrm{Id}$ dans H ; par le lemme 5.39, s n'est pas dans le centre de S_n et il existe $u \in S_n$ tel que $us \neq su$. Vu que u est un produit de transpositions (proposition 2.73), il existe une transposition t telle que $st \neq ts$. Le sous-groupe H est normal et que $s \in H$ nous avons aussi $ts^{-1}t^{-1} \in H$. Mais en même temps, la combinaison sts^{-1} est le conjugué d'une transposition et est donc également une transposition (classe de conjugaison de S_4 dans 2.72). Nous en concluons que $sts^{-1}t^{-1}$ est un produit de deux transpositions appartenant H.

Nous venons de prouver que H contient au moins un produit de deux transpositions. Et ce produit est différent de Id parce que $sts^{-1}t^{-1} = \mathrm{Id}$ impliquerait $st = ts$.

Soient donc deux transpositions $t_1, t_2 \in H$ telles que $t_1t_2 \neq \mathrm{Id}$. Les supports de t_1 et t_2 ont soit 1 soit aucun élément communs.

Premier cas Supposons $t_1 = (a, b)$, $t_2 = (b, c)$ avec a, b, c distincts dans $\{1, \dots, n\}$. Dans ce cas $t_1t_2 = (a, b, c)$ et H contient un cycle de longueur 3. Vu que H est normal et que les cycles de longueur trois sont une classe de conjugaison (exemple 2.72) et que A_n est engendré par ceux-ci (proposition 5.34), $A_n \subset H$. Mais A_n est d'indice deux dans S_n (proposition (2)(2)). Quel nombre plus grand que $n!/2$ divise $n!$? Seulement n lui-même. Donc H est soit A_n soit S_n.

Second cas Le groupe H contient un élément de la forme $(ab)(cd)$ avec a, b, c, d distincts dans $\{1, \dots, n\}$.

Si $n = 3$ Impossible parce que avec $n = 3$ nous n'avons pas quatre éléments distincts.

Si $n = 4$ Le sous-groupe H de S_4 contient un élément de V_4 qui n'est pas l'identité. Par normalité et classes de conjugaison, H contient V_4. Nous devons maintenant prouver que si H n'est pas V_4 alors H est A_4 ou S_4. Nous avons les inclusions $V_4 \subset H \subset S_4$ et donc les inégalités

$$4 \leqslant |H| \leqslant 24. \tag{5.97}$$

Donc le nombre $|H|$ est un multiple de 4 qui divise 24. Les possibilités sont $|H| = 4, 8, 12, 24$. La possibilité $|H| = 4$ donne $H = V_4$; si $|H| = 24$ alors $H = S_4$; si $|H| = 12$ alors H est d'indice 2 dans S_4 et $H = A_n$ (proposition 5.30(3)). Quid de $|H| = 8$?

D'après le corolaire 2.35 au théorème de Lagrange, l'ordre d'un élément divise l'ordre du groupe. Soit x dans H mais pas dans V_4. L'ordre de x peut être 1, 2, 4 ou 8. Ordre 1 serait $x = \mathrm{Id}$. Ordre 8, pas possible parce que S_4 n'a pas d'éléments d'ordre 8.

x d'ordre 2 Prenons la décomposition de x en cycles disjoints. Vu qu'on est dans S_4, ces cycles ne peuvent être que des transpositions. Soit il y en a un (alors H contient une transposition et donc $H = S_4$), soit il y en a deux et alors x est dans V_4.

x d'ordre 4 L'élément x serait alors un cycle de longueur 4. Et alors H contient tous les cycles de longueur 4. Par exemple il contient le produit $(abcd)(bacd) = (adc)$. Le sous-groupe H contient alors A_4 (parce qu'il contient tous les 3-cycles).

Si $n \geqslant 5$ Soit un élément e distinct de a, b, c et d. Par notre liste préférée des classes de conjugaisons (exemple 2.72(5)), le 2-cycle $(c, e)(a, b)$ est conjugué à $(a, b)(c, d)$ et appartient donc à H. Mais alors le produit suivant est également dans H :

$$(ce)(ab)(ab)(cd) = (ce)(cd) = (ecd). \tag{5.98}$$

Donc H contient un 3-cycle, et par conséquent tous les 3-cycles. Encore une fois, cela prouve que H est soit A_n soit S_n.

Pourquoi $n = 4$ est spécial ? Dans le premier cas, nous montrons tout de suite que $H = V_4$ n'est pas possible. Dans le deuxième cas, nous montrons que, grâce à un élément différent de a, b, c et d, la possibilité $H = V_4$ est exclue. La possibilité $H = V_4$ n'existe que pour $n = 4$.

□

5.6.3 Indice

Théorème 5.44.
Tout sous-groupe d'indice n dans S_n est isomorphe à S_n.

Démonstration. Pour $n = 1$, il n'y a rien. Pour $n = 2$, un sous-groupe d'indice 2 ne peut contenir que 1 élément qui est donc l'identité. Ok pour que $\{\mathrm{Id}\}$ soit égal à S_1 ?

Pour les autres, il y a un peu plus de travail.

Pour $n = 3$ Nous avons $|S_3| = 6$. Donc un sous-groupe d'indice 3 dans S_3 contient exactement 2 éléments. Il contient Id et un autre élément $\sigma \in S_3$ qui doit vérifier $\sigma^2 = \mathrm{Id}$ ou $\sigma^2 = \sigma$. Aucun élément de S_3 ne vérifie $\sigma^2 = \sigma$ (à part l'identité). Donc $\sigma^2 = \mathrm{Id}$, ce qui fait que σ est une transposition. Donc

$$H = \{\mathrm{Id}, (12)\} \tag{5.99}$$

ou l'identité avec (23) ou avec (13). Dans tous les cas c'est isomorphe à S_2.

Pour $n = 4$ Nous avons $|S_4 : H| = 4$, donc $|H| = 6$. Mais $6 = 2 \times 3$ et $2 \mid 3 - 1$, donc le théorème 5.25 nous dit que H est soit cyclique[16] (et donc abélien), soit un produit semi-direct. Vu que S_4 n'a pas d'éléments d'ordre 6, aucun sous-groupe d'ordre 6 ne peut être cyclique. Nous sommes donc dans le cas du produit semi-direct

$$H = \mathbb{Z}_3 \times_\varphi \mathbb{Z}_2 \tag{5.100}$$

où $\varphi\colon \mathbb{Z}_2 \to \mathrm{Aut}(\mathbb{Z}_3)$ et $\varphi(1)$ est d'ordre 2 dans $\mathrm{Aut}(\mathbb{Z}_3)$. Il convient de nous attarder un peu pour être sûr d'avoir bien compris tout ce qui se trouve dans l'identification (5.100). D'abord un point de notations : ici nous considérons les groupes $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ munis de l'addition. Donc 1 n'est pas le neutre. Ensuite nous savons du théorème 5.18 que $\mathrm{Aut}(\mathbb{Z}/3\mathbb{Z}) = (\mathbb{Z}/3\mathbb{Z})^*$, et que via cette identification, $\varphi(1) = 2 \in (\mathbb{Z}/3\mathbb{Z})^*$ au sens où $\varphi(1)x = 2x$. Nous avons alors $\varphi(1)^2 x = 4x = x$ dans $\mathbb{Z}/3\mathbb{Z}$. Cela montre bien que $\varphi(1)$ est d'ordre 2.

Par rapport à la proposition 5.31, ici nous écrivons $\mathbb{Z}_2 = \big(\{0, 1\}, +\big)$ alors que là nous écrivons $\mathbb{Z}_2 = \big(\{-1, 1\}, \cdot\big)$. Ce sont les mêmes groupes, mais il convient de remarquer que le 1 ici est le -1 là.

Nous savons par la proposition 5.31 que $S_n = A_n \times_\varphi \mathbb{Z}_2$; en comparant avec (5.100) nous voyons qu'il suffit de prouver que $A_3 = \mathbb{Z}/3\mathbb{Z}$ pour avoir $H = S_3$.

16. Définition 2.12.

Le groupe A_3 possède $|S_3|/2 = 3$ éléments. Il est vite vu que $A_3 = \{\mathrm{Id}, (12)(31), (12)(32)\}$: ce sont trois éléments de signature paire dans S_3 ; donc c'est S_3. La correspondance $\mathrm{Id} \mapsto 0$, $(12)(13) \mapsto 1$, $(13)(12) \mapsto 2$ donne un isomorphisme avec $(\mathbb{Z}_3, +)$.

Pour $n \geqslant 5$ Soit un sous-groupe H d'indice n dans S_n et l'action à gauche de S_n sur $E = S_n/H$ (qui n'est à priori pas un groupe) donnée par $g \cdot [s] = [gs]$.

Morphisme $\varphi\colon S_n \to S_E$ Le φ définit par l'action est un morphisme parce que

$$\varphi(g_1g_2)[s] = [g_1g_2s] = \varphi(g_1)[g_2s] = \varphi(g_1)\varphi(g_2)[s]. \tag{5.101}$$

Mais il faut également vérifier que pour chaque $g \in G$, l'application $\varphi(g)\colon E \to E$ est bien une permutation. Pour l'injectivité, si $\varphi(g)[s_1] = \varphi(g)[s_2]$ alors $[gs_1] = [gs_2]$, donc il existe $h \in H$ tel que $gs_1 = gs_2h$, ce qui prouve que $s_1 = s_2h$ et donc que $[s_1] = [s_2]$. Pour la surjectivité, soit $[t] \in S_n/H$ et résolvons $\varphi(g)[s] = [t]$ par rapport à s. L'élément $s = g^{-1}t$ fonctionne.

$\ker(\varphi)$ est normal Soit $z \in \ker(\varphi)$, c'est-à-dire que $\varphi(z) = \mathrm{Id}_E$. Alors pour $\sigma \in S_n$ nous avons $\varphi(\sigma z \sigma^{-1}) = \varphi(\sigma)\varphi(z)\varphi(\sigma^{-1}) = \mathrm{Id}_E$.

$\ker(\varphi) = \bigcap_{g\in S_n} gHg^{-1}$ Supposons que $z \in gHg^{-1}$ pour tout g, et calculons $\varphi(z)[s]$. D'abord par hypothèse il existe $h \in H$ tel que $z = shs^{-1}$, donc

$$\varphi(z)[s] = [zs] = [zhs^{-1}s] = [s], \tag{5.102}$$

ce qui prouve que $\varphi(z) = \mathrm{Id}$.

Dans l'autre sens, soit $z \in \ker(\varphi)$. Donc $\varphi(z)[s] = [s]$. Il existe donc $h \in H$ tel que $zs = sh$, c'est-à-dire tel que $z = shs^{-1}$. La formule demandée est donc prouvée.

Questions d'ordre Nous savons que $|H| = (n-1)!$ alors que $|A_n| = \frac{n!}{2}$. Donc $|H| < |A_n|$ avec une inégalité stricte. En même temps nous avons $|\ker(\varphi)| \leqslant |H|$ parce que $\ker(\varphi)$ est une intersection dont un des termes est H lui-même. Nous avons alors les inégalités

$$|\ker(\varphi)| \leqslant |H| = (n-1)! < |A_n|. \tag{5.103}$$

Mais le seul sous-groupes normaux de S_n sont A_n et S_n et $\{\mathrm{Id}\}$ (proposition 5.43). Donc $\ker(\varphi) = \mathrm{Id}$ et φ est une injection entre deux ensembles finis de même cardinalité. Cela fait de φ une bijection et donc un isomorphisme de groupes

$$\varphi\colon S_n \to S_E. \tag{5.104}$$

Soit une fonction de numérotation $\psi\colon E \to \{1, \dots, n\}$. Avec cela nous définissons un isomorphisme de groupes

$$\begin{aligned} \tilde{\psi}\colon S_E &\to S_n \\ \sigma &\mapsto \psi\sigma\psi^{-1}. \end{aligned} \tag{5.105}$$

Fixateur Nous montrons à présent que $(\tilde{\psi} \circ \varphi)(H) = \mathrm{Fix}\big(\psi[\mathrm{Id}]\big)$ où le stabilisateur est pris dans S_n. Pour la première inclusion, soit $h \in H$. Nous avons $(\tilde{\psi} \circ \varphi)(h) = \psi \circ \varphi(h)\psi^{-1}$, qui nous appliquons à $\psi[\mathrm{Id}]$:

$$(\tilde{\psi} \circ \varphi)(h)\psi[\mathrm{Id}] = \psi \circ \varphi(h)[\mathrm{Id}] = \psi[h] = \psi[\mathrm{Id}]. \tag{5.106}$$

Donc $(\tilde{\psi} \circ \varphi)(H) \subset \mathrm{Fix}\big(\psi[\mathrm{Id}]\big)$.

Pour l'autre inclusion, soit $\sigma \in S_n$ tel que $\sigma\psi[\mathrm{Id}] = \psi[\mathrm{Id}]$. Vu que $\sigma \in S_n$ nous avons $s \in S_E$ tel que $\sigma = \tilde{\psi}(s)$. Pour ce s nous avons donc

$$\big(\tilde{\psi}(s) \circ \psi\big)[\mathrm{Id}] = \psi[\mathrm{Id}], \tag{5.107}$$

d'où nous déduisons $s[\mathrm{Id}] = [\mathrm{Id}]$. Cela prouve que s stabilise $[\mathrm{Id}]$ dans S_E. Donc $s = \varphi(h)$ pour un certain $h \in H$, et au final $\sigma = \tilde{\psi}\big(\varphi(h)\big)$.

Conclusion L'application $\tilde{\psi} \circ \varphi \colon H \to S_n$ est une application dont l'image est le fixateur d'un point. Plus précisément,

$$\tilde{\psi} \circ \varphi \colon H \to \operatorname{Fix}\big(\psi[\mathrm{Id}]\big) \tag{5.108}$$

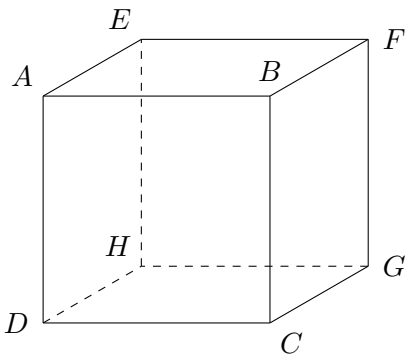
est un isomorphisme de groupe. Mais le stabilisateur d'un point dans S_n est S_{n-1}.

□

5.7 Isométriques du cube

Les isométries du cube proviennent de [54].

Nous considérons le cube centré en l'origine de $\mathbb{R}^3$ et G, le groupe des isométries de $\mathbb{R}^3$ préservant ce cube. Nous notons aussi G^+ le sous-groupes de G constitué des éléments de déterminant positif. Nous notons



$$\mathcal{D} = \{D_1, \dots, D_4\} \tag{5.109}$$

l'ensemble des grandes diagonales, c'est-à-dire les segments $[AG]$, $[EC]$, $[FD]$, et $[BH]$. Nous savons que G préserve les longueurs et que ces segments sont les plus longs possibles à l'intérieur du cube. Donc G agit sur $\mathcal{D}$ parce qu'il ne peut transformer une grande diagonale qu'en une autre grande diagonale. Nous avons donc un morphisme de groupes

$$\rho \colon G \to S_4. \tag{5.110}$$

Nous montrons ce que morphisme est surjectif en montrant qu'il contient les transpositions. Le groupe G contient la symétrie axiale passant par le milieu M de $[AE]$ et le milieu N de CG. Il est facile de voir que cette symétrie permute $[AG]$ avec $[EC]$. De plus elle laisse $[FD]$ inchangée. En effet, aussi incroyable que cela paraisse en regardant le dessin, nous avons $FD \perp MN$, parce qu'en termes de vecteurs directeurs,

$$\overrightarrow{ON} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \overrightarrow{OF} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}. \tag{5.111}$$

Étudions à présent le noyau $\ker(\rho)$. Si $f \in \ker(\rho)$ n'est pas l'identité, alors $f(D_i) = D_i$ pour tout i, mais au moins pour une des diagonales les sommets sont inversés. Quitte à renommer les sommets du cube nous supposons que la diagonale $[AG]$ est retournée : $f(A) = G$ et $f(G) = A$. Regardons où peut partir B sous l'effet de f. Étant donné que f préserve les diagonales, $f(B) \in \{B, C\}$, mais étant donné que f est une isométrie, $d\big(f(B), f(G)\big) = d(B, G)$, et nous concluons que $f(B) = H$. Donc la diagonale $[BH]$ est retournée sous l'effet de f. En raisonnant de même, nous voyons que f retourne toutes les diagonales. Donc les éléments non triviaux de $\ker(\rho)$ retournent toutes les diagonales ; il n'y en a donc qu'un seul et c'est la symétrie centrale :

$$\ker(\rho) = \{\mathrm{Id}, s_0\}. \tag{5.112}$$

Le premier théorème d'isomorphisme 2.28 nous permet d'écrire le quotient de groupes :

$$\frac{G}{\{\mathrm{Id}, s_0\}} \simeq S_4. \tag{5.113}$$

Une classe d'équivalence modulo $\ker(\rho)$ dans G est donc toujours de la forme $\{f, f \circ s_0\}$. Et vu que s_0 est de déterminant -1, une classe contient toujours exactement un élément de déterminant 1 et un de déterminant -1.

D'autre part $\ker(\rho)$ est normal dans G parce que en tant que matrice, $s_0 = -\mathbb{1}$, donc les problèmes de commutativité ne se posent pas. L'application

$$\begin{aligned} \varphi\colon \frac{G}{\{\mathrm{Id}, s_0\}} &\to G^+ \\ [g] &\mapsto \begin{cases} g & \text{si } \det(g) > 0 \\ g \circ s_0 & \text{sinon} \end{cases} \end{aligned} \tag{5.114}$$

est un isomorphisme de groupes. Et enfin nous pouvons écrire

$$G^+ \simeq S_4. \tag{5.115}$$

Nous allons maintenant utiliser le corolaire 2.81 pour montrer que $G = G^+ \times_\sigma \ker(\rho)$. Les conditions sont remplies :

— $\ker(\rho)$ normalise G^+ parce que $\ker(\rho)$ ne contient que $\pm\mathbb{1}$.

— $\ker(\rho) \cap G^+ = \{\mathrm{Id}\}$.

— $\ker(\rho)G^+ = G$ parce que les classes d'équivalence de G modulo $\ker(\rho)$ sont composées de $\{f, f \circ s_0\}$.

Vu que $G^+ \simeq S_4$ et $\ker(\rho) \simeq \mathbb{Z}/2\mathbb{Z}$ nous pouvons écrire de façon plus brillante que

$$G \simeq S_4 \times_\sigma \mathbb{Z}/2\mathbb{Z}. \tag{5.116}$$

Mais étant donné que la conjugaison par s_0 est triviale, le produit semi-direct est un produit direct :

$$G \simeq S_4 \times \mathbb{Z}/2\mathbb{Z}. \tag{5.117}$$

Il est maintenant du meilleur gout de pouvoir identifier géométriquement ces éléments. Les éléments de $\mathbb{Z}/2\mathbb{Z} = \{\mathrm{Id}, s_0\}$ ne font pas de mystères. Dans S_4 nous avons les classes de conjugaison des éléments Id, (12), (123), (1234) et (12)(34) déterminées durant l'exemple 2.72.

(1) L'élément (12) consiste à permuter deux diagonales et laisser les autres en place. Nous avons déjà vu que c'était une symétrie axiale passant par les milieux de deux côtés opposés. Cela fait 6 axes d'ordre 2.

(2) L'élément (123) fixe une des diagonales. C'est donc la symétrie axiale le long de la diagonale fixée. Par exemple la symétrie d'axe (AG) fait bouger le point B de la façon suivante :

$$B \to D \to E \to B. \tag{5.118}$$

C'est une rotation est d'angle $\frac{2\pi}{3}$. Cela sont 8 rotations d'ordre 3.

Notons à ce propos que la différence entre (234) et (243) est que la première fait une rotation d'angle $2\pi/3$ tandis que la seconde fait une rotation d'angle $-2\pi/3$.

(3) L'élément (1234) ne maintient aucune des diagonales et est d'ordre 4. C'est donc la rotation d'angle $\pi/2$ ou $-\pi/2$ autour de l'axe passant par les milieux de deux faces opposées. Il y en a 6 comme ça (3 paires de faces puis pour chaque il y a $\pi/2$ et $-\pi/2$), et ça tombe bien 6 est justement la taille de la classe de conjugaison de (1234) dans S_4.

(4) L'élément (12)(34) est le carré de la précédente [17], c'est-à-dire les rotations d'angle π autour des mêmes axes. Cela fait 3 éléments d'ordre 2.

17. En fait c'est (13)(24), le carré de la précédente, mais c'est la même classe de conjugaison.

Chapitre 6

Corps

6.1 Généralités

6.1.
Nous trouvons parfois le terme **anneau à division**. Cela provient du fait que dans beaucoup de cas on considère uniquement des corps commutatifs ; donc on voudrait une façon de parler d'un anneau dont tous les éléments non nuls sont inversibles. Dans ce cadre on dit :

— Un anneau à division est un anneau dont tous les éléments non nuls sont inversibles,

— Un corps est un anneau à division commutatif.

Pour prendre un exemple de cette différence, le théorème de Wedderburn 18.31 est énoncé ici sous les termes « Tout corps fini est commutatif ». Sous-entendu : la commutativité ne fait pas partie de la définition d'un corps. Par contre dans [54] il est énoncé sous les termes « Tout anneau à division fini est un corps ». Chez lui, un corps est toujours commutatif et un anneau à division est ce que nous appelons ici un corps.

6.1.1 Corps ordonnés

Nous avons vu la définition de corps totalement ordonné en 1.109.

Définition 6.2 ([94])**.**
Un corps est ***formellement réel*** *si* -1 *n'est pas une somme de carrés.*

Proposition 6.3.
Un corps totalement ordonné est formellement réel.

Démonstration. Soit un corps totalement ordonné $(\mathbb{K}, \leqslant)$ et $a \in \mathbb{K}$ alors $a^2 \geqslant 0$. En effet si $a \geqslant 0$ alors $a^2 = a \times a \geqslant 0$ directement par la définition 1.109(1b). Si $a \leqslant 0$ alors $-a \geqslant 0$ et

$$a^2 = (-a)^2 \geqslant 0. \tag{6.1}$$

Vu que $-1 < 0$, il ne peut donc pas être écrit comme un carré. A fortiori comme somme de carrés. □

6.1.2 Automorphismes de $\mathbb{R}$ et $\mathbb{C}$

Proposition 6.4 ([95, 1])**.**
L'identité est l'unique automorphisme du corps $\mathbb{R}$.

Démonstration. Soit un automorphisme $\sigma\colon \mathbb{R} \to \mathbb{R}$. Comme pour tout automorphisme,

$$\sigma(a) = \sigma(1a) = \sigma(1)\sigma(a). \tag{6.2}$$

Donc $\sigma(1) = 1$.

Identité sur les rationnels De plus

$$\sigma(n) = \sigma(1 + \ldots + 1) = \sigma(1) + \ldots + \sigma(1) = n, \tag{6.3}$$

et

$$\sigma\left(\frac{1}{n}\right) + \ldots + \sigma\left(\frac{1}{n}\right) = \sigma\left(\frac{1}{n} + \ldots + \frac{1}{n}\right) = \sigma(1) = 1. \tag{6.4}$$

Donc $\sigma(1/n) = 1/n$.

Nous en déduisons que pour tout $q \in \mathbb{Q}$, $\sigma(q) = q$. Cela ne suffit pas pour déduire $\sigma(x) = x$ pour tout $x \in \mathbb{R}$ parce que rien n'indique que σ soit continue.

Positive sur les positifs Si $x > 0$ alors $\sigma(x) = \sigma(\sqrt{x})^2 > 0$.

Croissance Si $x > y$ alors $x - y > 0$ et $\sigma(x - y) > 0$. Cela donne $\sigma(x) > \sigma(y)$.

Identité sur les réels Soit un irrationnel $x \in \mathbb{R}$ et une suite (q_i) dans $\mathbb{Q}$ qui converge de façon croissante vers x. Soit $\epsilon > 0$ dans $\mathbb{Q}$. Il existe N tel que si $i > N$ alors $q_i + \epsilon > x$; en appliquant σ à cette inégalité et en se souvenant que σ est l'identité sur $\mathbb{Q}$,

$$q_i + \epsilon > \sigma(x). \tag{6.5}$$

Mais de plus, $q_i < x$ donne $\sigma(q_i) < \sigma(x)$, c'est-à-dire $q_i < \sigma(x)$. En regroupant ces deux inégalités,

$$q_i < \sigma(x) < q_i + \epsilon \tag{6.6}$$

pour tout $\epsilon > 0$ dans $\mathbb{Q}$ et $i > N$. Ce ϵ étant fixé nous pouvons prendre la limite des inégalités (6.6) :

$$x \leqslant \sigma(x) \leqslant x + \epsilon. \tag{6.7}$$

Cela étant valable pour tout $\epsilon > 0$ dans $\mathbb{Q}$, nous avons bien $x = \sigma(x)$.

□

Remarque 6.5.
Certains[95] pensent que l'énoncé de cette proposition, ne parlant que de *corps* $\mathbb{R}$ n'autorise pas l'utilisation d'autre structure réelle que celle de corps. Du coup il faut reconstruire la notion d'ordre à partir seulement du langage des corps. Par exemple en disant que $a > b$ si et seulement si il existe k tel que $a = b + k^2$.

On peut s'en sortir en donnant l'énoncé suivant : « Si $\mathbb{K}$ est un corps isomorphe (en tant que corps) à $\mathbb{R}$ alors son unique automorphisme est l'identité ». Cela se démontre immédiatement en disant que si f est un automorphisme de $\mathbb{K}$ et si ϕ est un isomorphisme $\mathbb{K} \to \mathbb{R}$ alors $\phi \circ f \circ \phi^{-1}$ est un automorphisme de $\mathbb{R}$. Donc il est l'identité et f l'est également.

Attention cependant à prouver que ϕ^{-1} est un morphisme. En effet en posant $\phi^{-1}(x) = a$ et $\phi^{-1}(y) = b$ nous avons

$$\phi\big(\phi^{-1}(x) + \phi^{-1}(y)\big) = x + y \tag{6.8}$$

parce que ϕ est un morphisme. D'autre part,

$$\phi\big(\phi^{-1}(x) + \phi^{-1}(y)\big) = \phi(a + b). \tag{6.9}$$

Donc

$$\phi^{-1}(x + y) = \phi^{-1}\big(\phi(a) + \phi(b)\big) = \phi^{-1}\big(\phi(a + b)\big) = a + b = \phi^{-1}(x) + \phi^{-1}(y). \tag{6.10}$$

Proposition 6.6.
Un automorphisme du corps $\mathbb{C}$ qui fixe $\mathbb{R}$ est soit l'identité soit la conjugaison complexe[1].

Démonstration. Soit un automorphisme σ vérifiant la condition de fixer $\mathbb{R}$. Alors la restriction de σ à $\mathbb{R}$ est un automorphisme de $\mathbb{R}$ et y est donc l'identité par la proposition 6.4.

En ce qui concerne les imaginaires purs,

$$-1 = \sigma(-1) = \sigma(ii) = \sigma(i)^2. \tag{6.11}$$

Donc $\sigma(i)$ est un élément de $\mathbb{C}$ vérifiant $\sigma(i)^2 = -1$. C'est-à-dire $\sigma(i) = \pm i$.

Si $\sigma(i) = i$ alors $\sigma = \mathrm{Id}$. Si $\sigma(i) = -i$ alors σ est la conjugaison complexe. □

1. Par « fixer $\mathbb{R}$ » nous entendons que $\sigma(\mathbb{R}) = \mathbb{R}$, pas spécialement que $\sigma(x) = x$ pour tout $x \in \mathbb{R}$

6.1.3 Corps premier

Définition 6.7.
Un corps est ***premier*** *s'il est son seul sous corps. Le* ***sous corps premier*** *d'un corps est l'intersection de tous ses sous corps.*

Lemme 6.8.
Un corps premier est commutatif.

Démonstration. Le centre d'un corps est certainement un sous corps. Par conséquent un corps premier doit être contenu dans son propre centre, c'est-à-dire être commutatif. □

Définition 6.9.
Soit p un nombre premier. Nous notons $\mathbb{F}_p = \mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$.

Nous verrons plus loin (section 18.5) comment nous pouvons définir $\mathbb{F}_{p^n}$ lorsque p est premier, ainsi que l'unicité d'un tel corps.

Nous avons par exemple

$$\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z} = \{0, 1\} \tag{6.12}$$

avec la loi $2 = 0$.

Notons que $\mathbb{F}_p$ est un corps possédant p éléments. L'ensemble $\mathbb{F}_p^*$ est un groupe d'ordre $p - 1$.

Lemme 6.10.
Les corps $\mathbb{Q}$ et $\mathbb{F}_p$ (avec p premier) sont premiers.

Démonstration. Tout sous corps de $\mathbb{Q}$ doit contenir 1, et par conséquent $\mathbb{Z}$. Devant également contenir tous les inverses, il contient $\mathbb{Q}$.

Tout sous corps de $\mathbb{F}_p$ doit contenir 1 et donc $\mathbb{F}_p$ en entier. Par ailleurs nous savons de la proposition 3.109 que $\mathbb{F}_p$ est un corps lorsque p est premier. □

Proposition 6.11.
Soit $\mathbb{K}$ un corps de caractéristique p et $\mathbb{P}$ son sous corps premier. Si $p = 0$ alors $\mathbb{P} = \mathbb{Q}$. Si $p > 0$, alors $\mathbb{P} = \mathbb{F}_p$.

Démonstration. Notons d'abord que la caractéristique d'un corps est toujours soit 0 soit un nombre premier, parce qu'un corps est en particulier un anneau intègre (proposition 3.100).

Étant donné que 1 est dans tout sous corps, nous devons avoir $\mathbb{Z}1 \subseteq \mathbb{P}$. Si $p = 0$, alors $\mathbb{Z}1 \simeq \mathbb{Z}$, et nous avons

$$\mathbb{Z}1_{\mathbb{A}} \subset \mathbb{P} \subset \mathbb{K}. \tag{6.13}$$

Pour chaque $(n, m) \in \mathbb{Z}1_{\mathbb{A}} \times (\mathbb{Z}1_{\mathbb{A}})^*$ l'élément $nm^{-1} \in \mathbb{K}$ est dans $\mathbb{P}$ parce que $\mathbb{P}$ est un corps. Nous en déduisons que le corps des fractions de $\mathbb{Z}$ est contenu dans $\mathbb{P}$ par conséquent $\mathbb{P} = \mathbb{Q}$ (théorème 1.106).

Si par contre la caractéristique de $\mathbb{K}$ est $p \neq 0$, nous avons $\mathbb{Z}1_{\mathbb{A}} \simeq \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ par le lemme 3.61. L'ensemble $\mathbb{F}_p$ étant un corps, c'est le corps premier de $\mathbb{K}$. □

Proposition 6.12.
Soit $\mathbb{K}$ un corps et $\mathbb{P}$ son sous-corps premier. Si $\sigma \in \mathrm{Aut}(\mathbb{K})$ alors $\sigma|_{\mathbb{P}} = \mathrm{Id}$, c'est-à-dire que $\sigma(x) = x$ pour tout $x \in \mathbb{P}$.

6.1.4 Petit théorème de Fermat

Théorème 6.13 (Petit théorème de Fermat)**.**
Soit p un nombre premier. Si $x \in \mathbb{F}_p$ alors $x^p = x$. Si $x \in (\mathbb{F}_p)^*$, alors $x^{p-1} = 1$.

En particulier si $x \in \mathbb{F}_p^*$ alors $x^{-1} = x^{p-2}$.

Démonstration. Étant donné que $\mathbb{F}_p$ est un corps commutatif et que p est premier, la proposition 3.64 nous indique que $\sigma(x) = x^p$ est un automorphisme. La proposition 6.12 nous indique alors que

$$a^p = a. \tag{6.14}$$

Si a est inversible alors $a^{p-1} = a^p a^{-1} = 1$. □

Remarque 6.14.
Une autre façon d'énoncer le petit théorème de Fermat 6.13 est que si p est premier et si a est premier avec p, alors $a^{p-1} \in [1]_p$. Le nombre a n'est pas premier avec p uniquement lorsque a est multiple de p. Dans ce cas c'est $a = 0$ dans $\mathbb{F}_p$ et donc $a^{p-1} = 0$.

Exemple 6.15.
Soit $\mathbb{K} = \mathbb{F}_{29}$. Le nombre 29 étant premier, $\mathbb{K}$ est un corps premier. C'est le corps des entiers modulo 29. Nous avons donc

$$-142 = -113 = -84 = -55 = -26 = 3 = 32 = 61 = 90 = 119. \tag{6.15}$$

Le petit théorème de Fermat nous permet aussi de calculer des exposants et des inverses. En effet, puisque $1 = x^{28}$ pour tout $x \in \mathbb{F}_{29}^*$, nous avons $x^{-1} = x^{27}$, et par suite, pour tout entier a,

$$x^{-a} = (x^a)^{27} = x^{27a}. \tag{6.16}$$

Le nombre $27a$ peut être grand par rapport à 29. Mais en réutilisant le fait que $1 = x^{28}$, on obtient

$$x^{-a} = x^{[27a]_{28}}. \tag{6.17}$$

Cette expression doit être comprise comme disant que pour tout $k \in [27a]_{28}$ nous avons $x^{-a} = x^k$.

Chose à retenir : dans les exposants nous calculons modulo 28. △

6.2 Théorème des deux carrés

Proposition 6.16.
Soit p un nombre premier et P un élément de $\mathbb{F}_p[X]$. L'anneau $\mathbb{F}_p[X]/(P)$ est intègre si et seulement si P est irréductible dans $\mathbb{F}_p[X]$.

Démonstration. Supposons que P soit réductible dans $\mathbb{F}_p[X]$, c'est-à-dire qu'il existe $Q, R \in \mathbb{F}_p[X]$ tels que $P = QR$. Dans ce cas, $\bar{Q}$ est diviseur de zéro dans $\mathbb{F}_p[X]/(P)$ parce que $\bar{Q}\bar{R} = 0$.

Nous supposons maintenant que $\mathbb{F}_p[X]/(P)$ ne soit pas intègre : il existe des polynômes $R, Q \in \mathbb{F}_p[X]$ tels que $\bar{Q}\bar{R} = 0$. Dans ce cas le polynôme QR est le produit de P par un polynôme : $QR = PA$. Tous les facteurs irréductibles de A étant soit dans Q soit dans R, il est possible de modifier un peu Q et R pour obtenir $QR = P$, ce qui signifie que P n'est pas irréductible. □

6.2.1 Un peu de structure dans $\mathbb{Z}[i]$

Lemme 6.17.
L'application

$$\begin{aligned} N\colon \mathbb{Z}[i] &\to \mathbb{N} \\ a + bi &\mapsto a^2 + b^2 \end{aligned} \tag{6.18}$$

est un stathme euclidien pour $\mathbb{Z}[i]$.

Démonstration. Soient $t, z \in \mathbb{Z}[i]\backslash\{0\}$ dont le quotient s'écrit

$$\frac{z}{t} = x + iy \tag{6.19}$$

dans $\mathbb{C}$. Nous considérons $q = a + bi$ où a et b sont les entiers les plus proches de x et y. S'il y a *ex aequo*, on prend au hasard[2]. Alors nous avons

$$|\frac{z}{t} - q| \leqslant \frac{|1+i|}{2} = \frac{\sqrt{2}}{2} < 1. \tag{6.20}$$

On pose $r = z - qt$ qui est bien un élément de $\mathbb{Z}[i]$. De plus

$$|r| = |z - qt| = |t||\frac{z}{t} - q| < |t|, \tag{6.21}$$

c'est-à-dire que $|r|^2 < |t|^2$ et donc $N(r) < N(t)$. □

Étant donné que $\mathbb{Z}[i]$ est euclidien, il est principal (proposition 3.153).

Lemme 6.18.
Les éléments inversibles de $\mathbb{Z}[i]$ sont $\{\pm 1, \pm i\}$.

Démonstration. Déterminons les éléments inversibles de $\mathbb{Z}[i]$. Si $z \in \mathbb{Z}[i]^*$, alors il existe $z' \in \mathbb{Z}[i]^*$ tel que $zz' = 1$. Dans ce cas nous aurions

$$1 = N(zz') = N(z)N(z'), \tag{6.22}$$

ce qui est uniquement possible avec $N(z) = N(z') = 1$, c'est-à-dire $z = \pm 1$ ou $z = \pm i$. Nous avons donc

$$\mathbb{Z}[i]^* = \{\pm 1, \pm i\}. \tag{6.23}$$

□

Définition 6.19 ([96])**.**
*Un **monoïde** est un triple $(E, *, e)$ où E est un ensemble, e est un élément de E et $*\colon E \times E \to E$ est une loi de composition telle que pour tout $x, y \in E$,*

(1) $x * (y * z) = (x * y) * z$ (associativité)

(2) $e * x = x * e = x$ (e est un neutre).

Nous notons $\Sigma = \{a^2 + b^2 \text{ tel que } a, b \in \mathbb{N}\}$.

Lemme 6.20.
L'ensemble $\Sigma = \{a^2 + b^2 \text{ tel que } a, b \in \mathbb{N}\}$ est un sous-monoïde[3] de $\mathbb{N}$.

Démonstration. Il suffit de prouver que si $m, n \in \Sigma$, alors le produit mn est également dans Σ. Si N est le stathme euclidien sur $\mathbb{Z}[i]$, alors $n \in \Sigma$ si et seulement s'il existe $z \in \mathbb{Z}[i]$ tel que $N(z) = n$. Si $z, z' \in \mathbb{Z}[i]$, alors $zz' \in \mathbb{Z}[i]$ et de plus

$$N(zz') = N(z)N(z') = nm. \tag{6.24}$$

Donc nm est l'image de zz' par N, ce qui prouve que $nm \in \Sigma$. □

Théorème 6.21 (Théorème des deux carrés, version faible)**.**
Un nombre premier est somme de deux carrés si et seulement si $p = 2$ ou $p \in [1]_4$.

Remarque 6.22.
Il n'est pas dit que les nombres dans $[1]_4$ sont premiers ($9 = 8 + 1$ ne l'est pas par exemple). Le théorème signifie que (à part 2), si un nombre premier est dans $[1]_4$ alors il est somme de deux carrés, et inversement, si un nombre premier est somme de deux carrés, il est dans $[1]_4$.

2. Dans l'exemple 3.152, nous prenions toujours l'inférieur parce que le stathme tenait compte de la positivité.
3. Définition 6.19.

Démonstration. Soit p un nombre premier dans Σ. Si $a = 2k$, alors $a^2 = 4k^2$ et $a^2 = 0 \mod 4$. Si au contraire a est impair, $a = 2k+1$ et $a^2 = 4k^2 + 1 + 4k = 1 \mod 4$. La même chose est valable pour b. Par conséquent, $a^2 + b^2$ est automatiquement $[0]_4$, $[1]_4$ ou $[2]_4$. Évidemment les nombres de la forme $0 \mod 4$ ne sont pas premiers ; parmi les $2 \mod 4$, seul $p = 2$ est premier (et vaut $1^2 + 1^2$).

Nous avons démontré que les seuls premiers de la forme $a^2 + b^2$ sont $p = 2$ et les $p = 1 \mod 4$. Il reste à faire le contraire : démontrer que si un nombre premier p vaut $1 \mod 4$, alors il est premier. Nous considérons l'anneau

$$\mathbb{Z}[i] = \{a + bi \text{ tel que } a, b \in \mathbb{Z}\}. \tag{6.25}$$

puis l'application

$$\begin{aligned} N\colon \mathbb{Z}[i] &\to \mathbb{N} \\ a + bi &\mapsto a^2 + b^2. \end{aligned} \tag{6.26}$$

Un peu de calcul dans $\mathbb{C}$ montre que pour tout $z, z' \in \mathbb{Z}[i]$, $N(zz') = N(z)N(z')$.

Nous savons que les éléments inversibles de $\mathbb{Z}[i]$ sont ± 1 et $\pm i$ (lemme 6.18).

Le lemme 6.17 montre que $\mathbb{Z}[i]$ est un anneau euclidien parce que N est un stathme. L'anneau $\mathbb{Z}[i]$ étant euclidien, il est principal (proposition 3.153).

Pour la suite, nous allons d'abord montrer que $p \in \Sigma$ si et seulement si p n'est pas irréductible dans $\mathbb{Z}[i]$, puis nous allons voir quels sont les irréductibles de $\mathbb{Z}[i]$.

Soit p, un nombre premier dans Σ. Si $p = a^2 + b^2$, alors nous avons $p = (a + ib)(a - bi)$, mais étant donné que p est premier, nous avons $a \neq 0$ et $b \neq 0$. Du coup p n'est pas inversible dans $\mathbb{Z}[i]$, mais il peut être écrit comme le produit de deux non inversibles. Le nombre p est donc non irréductible dans $\mathbb{Z}[i]$.

Dans l'autre sens, nous supposons que p est un nombre premier non irréductible dans $\mathbb{Z}[i]$. Nous avons alors $p = zz'$ avec ni z ni z' dans $\{\pm 1, \pm i\}$. En appliquant N nous avons

$$p^2 = N(p) = N(z)N(z'). \tag{6.27}$$

Vu que p est premier, cela est uniquement possible avec $N(z) = N(z') = p$ (avoir $N(z) = 1$ est impossible parce que cela dirait que z est inversible). Si $z = a + ib$, alors $p = N(z) = a^2 + b^2$, et donc $p \in \Sigma$.

Nous savons déjà que $\mathbb{Z}[i]$ est un anneau principal et n'est pas un corps ; la proposition 3.124 s'applique donc et p sera non irréductible si et seulement si l'idéal (p) sera non premier. Le fait que (p) soit un idéal non premier implique que le quotient $\mathbb{Z}[i]/(p)$ est non intègre (c'est la définition d'un idéal premier). Nous cherchons donc les nombres premiers pour lesquels le quotient $\mathbb{Z}[i]/(p)$ n'est pas intègre.

Nous commençons par écrire le quotient $\mathbb{Z}[i]/(p)$ sous d'autres formes. D'abord en remarquant que si I et J sont deux idéaux, on a $(\mathbb{A}/I)/J \simeq (\mathbb{A}/J)/I$, du coup, en tenant compte du fait que $\mathbb{Z}[i] = \mathbb{Z}[X]/(X^2 + 1)$, nous avons

$$\mathbb{Z}[i]/(p) = (\mathbb{Z}[X]/(p))/(X^2 + 1) = \mathbb{F}_p[X]/(X^2 + 1). \tag{6.28}$$

Nous avons donc équivalence des propositions suivantes :

$$p \in \Sigma \tag{6.29a}$$

$$\mathbb{F}_p[X]/(X^2 + 1) \text{ n'est pas intègre} \tag{6.29b}$$

$$X^2 + 1 \text{ n'est pas irréductible dans } \mathbb{F}_p \tag{6.29c}$$

$$X^2 + 1 \text{ admet une racine dans } \mathbb{F}_p \tag{6.29d}$$

$$-1 \in (\mathbb{F}_p^*)^2 \tag{6.29e}$$

$$\exists y \in \mathbb{F}_p^* \text{ tel que } y^2 = -1. \tag{6.29f}$$

Le point (6.29c) vient de la proposition 6.16. Maintenant nous utilisons le fait que p soit un premier impair (parce que le cas de $p = 2$ est déjà complètement traité), donc $(p-1)/2 \in \mathbb{N}$ et nous avons, pour le y de la dernière ligne,

$$(-1)^{(p-1)/2} = (y^2)^{(p-1)/2} = y^{p-1} = 1 \tag{6.30}$$

parce que dans $\mathbb{F}_p$ nous avons $y^{(p-1)} = 1$ par le petit théorème de Fermat (théorème 6.13). Du coup p doit vérifier

$$1 = (-1)^{(p-1)/2}, \tag{6.31}$$

c'est-à-dire $\frac{p-1}{2} = 0 \mod 2$ ou encore $p = 1 \mod 4$. □

Théorème 6.23 (Théorème des deux carrés[54])**.**
Soit $n \geqslant 2$ un nombre dont nous notons

$$n = \prod_{p \in \mathcal{P}} p^{v_p(n)} \tag{6.32}$$

où $\mathcal{P}$ est l'ensemble des nombres premiers. Alors $n \in \Sigma$ si et seulement si pour tout $p \in \mathcal{P} \cap [3]_4$, nous avons $v_p(n) \in [0]_2$ (c'est-à-dire $v_p(n)$ est pair).

Démonstration. **Condition suffisante.** Le produit (6.32) est évidemment un produit fini que nous pouvons alors regrouper en quatre parties : $\mathcal{P} \cap [0]_4$, $\mathcal{P} \cap [1]_4$, $\mathcal{P} \cap [2]_4$ et $\mathcal{P} \cap [3]_4$.

— Il n'y a pas de nombres premiers dans $[0]_4$.

— Les nombres premiers de $[1]_4$ sont dans Σ. Le produit d'éléments de Σ étant dans Σ, nous avons

$$\prod_{p \in \mathcal{P} \cap [1]_4} p^{v_p(n)} \in \Sigma. \tag{6.33}$$

— Le seul nombre premier dans $[2]_4$ est 2. C'est un élément de Σ.

— Le produit

$$\prod_{p \in \mathcal{P} \cap [3]_4} p^{v_p(n)} \tag{6.34}$$

est par hypothèse un produit de carrés ($v_p(n)$ est pair), et est donc un carré.

Au final le produit $\prod_{p \in \mathcal{P}} p^{v_p(n)}$ est un produit d'un carré par un élément de Σ, ce qui est encore un élément de Σ.

Pour cette partie, nous avons utilisé et réutilisé le lemme 6.20.

Condition nécessaire. Soit p, un nombre premier. Nous voulons montrer que

$$\{v_p(n) \text{ tel que } n \in \Sigma\} \subset [2]_2. \tag{6.35}$$

Pour montrer cela nous allons procéder par récurrence sur les ensembles

$$E_k = \{v_p(n) \text{ tel que } n \in \Sigma\} \cap \{0, \dots, k\}. \tag{6.36}$$

Il est évident que les éléments de E_0 sont pairs, vu qu'il n'y a que zéro, qui est pair.

Supposons que $E_k \subset [0]_2$, et montrons que $E_{k+1} \subset [0]_2$. Soit un élément de E_{k+1}, c'est-à-dire $v_p(n) \leqslant k+1$ avec $n = a^2 + b^2$. Si $v_p(n) = 0$ alors l'affaire est réglée ; sinon c'est que p divise n. Mais dans $\mathbb{Z}[i]$ nous avons

$$n = a^2 + b^2 = (a + bi)(a - bi) \tag{6.37}$$

Vu que $\mathbb{Z}[i]$ est principal, le lemme de Gauss 3.134 nous dit que si p divise n, alors il doit diviser soit $a + bi$, soit $a - bi$ (et du coup en fait les deux). Nous avons alors $p \mid a$ et $p \mid b$ en même temps. Du coup

$$p^2 \mid a^2 + b^2 = n. \tag{6.38}$$

Posons $a = pa'$ et $b = pb'$ avec $a', b' \in \mathbb{N}$. Nous avons

$$\frac{n}{p^2} = \frac{p^2a'^2 + p^2b'^2}{p^2} = a'^2 + b'^2 \in \Sigma. \tag{6.39}$$

Mais par construction,

$$v_p\left(\frac{n}{p^2}\right) = v_p(n) - 2 < k. \tag{6.40}$$

Donc $v_p(\frac{n}{p^2})$ est pair et du coup $v_p(n)$ doit également être pair.

□

6.2.2 Résultats chinois

Nous allons maintenant parler du système d'équations

$$\begin{cases} x = a_1 \mod p & (6.41a) \\ x = a_2 \mod q & (6.41b) \end{cases}$$

avec a_1, a_2 donnés dans $\mathbb{Z}$ et p, q des entiers premiers entre eux. Le lemme chinois nous donne la liste des solutions ainsi qu'une manière de les construire. Le théorème chinois en sera une espèce de corolaire qui établira l'isomorphisme d'anneaux $\mathbb{Z}/pq\mathbb{Z} \simeq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}$. Voir les-mathematiques.net.

Lemme 6.24 (Lemme chinois [97])**.**
Soient n_1, n_2 deux entiers premiers entre eux. Soient $a_1, a_2 \in \mathbb{Z}$. Les solutions du système

$$\begin{cases} x = a_1 \mod n_1 & (6.42a) \\ x = a_2 \mod n_2 & (6.42b) \end{cases}$$

pour $x \in \mathbb{Z}/n_1n_2\mathbb{Z}$ sont données de la façon suivante. Soient u_1, u_2 deux entiers qui satisfont la relation de Bézout[4]

$$u_1n_1 + u_2n_2 = 1, \tag{6.43}$$

et

$$a = \big(a_1u_2n_2 + a_2u_1n_1\big) \mod (n_1). \tag{6.44}$$

Alors $x = a \mod (n_1n_2)$.

Démonstration. Vérifions que le x donné par $x = a \mod (n_1n_2)$ est bien une solution. D'abord

$$\begin{aligned} a \mod n_2 &= a_1u_2n_2 \mod n_1 && (6.45a) \\ &= a_1(1 - u_1n_1) \mod n_1 && (6.45b) \\ &= a_1 \mod n_1 && (6.45c) \end{aligned}$$

où nous avons utilisé l'identité de Bézout (6.43). La vérification de $a \mod n_2 = a_2 \mod n_2$ est la même.

Soit maintenant $x \in \mathbb{Z}/n_1n_2\mathbb{Z}$ une solution du système (6.42) et a donné par la formule (6.44). Alors

$$\begin{aligned} (x - a) \mod n_1 &= \Big(a_1 - (a_1n_2u_2 + a_2u_1n_1)\Big) \mod n_1 && (6.46a) \\ &= a_1 - a_1u_2n_2 \mod n_1 && (6.46b) \\ &= 0, && (6.46c) \end{aligned}$$

donc $(x - a) \mod n_1 = 0$, ce qui signifie que $x - a$ est divisible par n_1. De la même façon, $(x - a) \mod n_2 = 0$ et $x - a$ est divisible par n_2. Nous savons maintenant que $x - a$ est divisible par n_1 et n_2. Étant donné que n_1 et n_2 sont premiers entre eux, nous en déduisons que $x - a$ est divisible par n_1n_2, ou encore que $x = a \mod n_1n_2$. □

4. voir le théorème 3.13

Théorème 6.25 (Théorème chinois)**.**
Soient p, q deux naturels premiers entre eux. Si $p, q \geqslant 2$ alors l'application

$$\begin{aligned} \phi\colon \mathbb{Z}/pq\mathbb{Z} &\to \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z} \\ [x]_{pq} &\mapsto \big([x]_p, [x]_q\big) \end{aligned} \tag{6.47}$$

est un isomorphisme d'anneaux.

Démonstration. Nous devons prouver que l'application ϕ respecte la somme, le produit et qu'elle est bijective. En ce qui concerne la somme,

$$\begin{aligned} \phi([q]_{pq} + [y]_{pq}) &= \phi\big((x+y) \mod pq\big) && (6.48a) \\ &= \big([x+y]_p, [x+y]_q\big) && (6.48b) \\ &= \big([x]_p + [y]_p, [x]_q + [y]_q\big) && (6.48c) \\ &= \big([x]_p, [x]_q\big) + \big([y]_p, [y]_q\big) && (6.48d) \\ &= \phi(x) + \phi(y). && (6.48e) \end{aligned}$$

En ce qui concerne le produit, c'est le même jeu : nous obtenons

$$\phi\big([xy]_{pq}\big) = \phi([x]_{pq}])\phi([y]_{pq}) \tag{6.49}$$

en utilisant le fait que $[xy]_p = [x]_p[y]_p$.

Montrons maintenant que ϕ est surjective. Soient $y_1, y_2 \in \mathbb{Z}$ et $x \in \mathbb{Z}$. Demander

$$\phi([x]_{pq}) = \big([y_1]_p, [y_2]_q\big) \tag{6.50}$$

revient à demander que $[x]_p = [y_1]_p$ et $[x]_q = [y_2]_q$, c'est-à-dire que x résolve le système

$$\begin{cases} x = y_1 \mod p & (6.51a) \\ x = y_2 \mod q. & (6.51b) \end{cases}$$

Le lemme chinois 6.24 nous assure qu'une solution existe.

En ce qui concerne l'injectivité, nous supposons que x et y soient deux entiers tels que

$$\phi([x]_{pq}) = \phi([y]_{pq}). \tag{6.52}$$

Nous en déduisons le système

$$\begin{cases} x \mod p = y \mod p & (6.53a) \\ x \mod q = y \mod q & (6.53b) \end{cases}$$

c'est-à-dire qu'il existe des entiers k et l tels que $x = y + kp$ et $x = y + lq$ ou encore tels que

$$kp + lq = 0. \tag{6.54}$$

Étant donné que p et q sont premiers entre eux, la seule possibilité est $k = l = 0$, c'est-à-dire $x = y$. □

Théorème 6.26 (Théorème chinois[1])**.**
Soit A un anneau commutatif. Soient $n \geqslant 2$, des éléments $x_1, \dots, x_n$ dans A et des idéaux $I_1, \dots, I_n$ tels que $I_i + I_j = A$ pour tout $i \neq j$.

Alors il existe un $x \in A$ tel que $x - x_i \in I_i$ pour tout $1 \leqslant i \leqslant n$.

Démonstration. En plusieurs points.

Définition de J_k Pour $k \in \{1, \dots, n\}$ nous notons J_k le produit $J_k = \prod_{l \neq k} I_l$. Étant donné que chaque I_k est un idéal, nous avons

$$I_k \subset J_l \tag{6.55}$$

dès que $k \neq l$.

$I_i + J_i = A$ En effet, soit $k \neq i$. Nous avons $J_i \subset I_k$ et donc

$$I_i + J_i \subset I_i + I_k = A. \tag{6.56}$$

Des α et des β Soit $i \in \{1, \ldots, n\}$. Vu que $I_i + J_i = A$, nous donc prendre $\alpha_i \in I_i$ et $\beta_i \in J_i$ tels que

$$\alpha_i + \beta_i = 1. \tag{6.57}$$

Et enfin Nous posons $x = \sum_{i=1}^{n} \beta_i x_i$. Pour chaque k nous avons

$$x - x_k = \sum_{i=1}^{n} \beta_i x_i - x_k \tag{6.58a}$$
$$= \sum_{i \neq k} \beta_i x_i + (\beta_k - 1) x_k \tag{6.58b}$$
$$= \sum_{i \neq k} \beta_i x_i - \alpha_k x_k. \tag{6.58c}$$

Les deux termes sont dans I_k. En effet, d'une part $\beta_i \in J_i \subset I_k$ et I_k est un idéal, donc $\beta_i x_i \in I_k$, et d'autre part $\alpha_k \in I_k$, donc $\alpha_k x_k \in I_k$.

Un idéal étant stable par somme, la somme du tout est dans I_k.

□

Remarque 6.27.
Ce théorème chinois est bien une généralisation du lemme chinois 6.24. En effet, l'élément x dont il est question est solution du problème $x = x_i \mod I_i$. L'hypothèse $I_i + I_j = A$ n'est pas nouvelle non plus étant donné que si p et q sont des entiers premiers entre eux nous avons $p\mathbb{Z} + q\mathbb{Z} = \mathbb{Z}$ par le corolaire 3.14.

6.3 Polynômes à coefficients dans un corps

Nous supposons que $\mathbb{K}$ est un corps commutatif, et nous étudions l'anneau $\mathbb{K}[X]$, défini en 3.86.

Proposition 6.28.
L'anneau $\mathbb{K}[X]$ des polynômes sur un corps commutatif $\mathbb{K}$ est factoriel.

Le théorème suivant est un cas particulier pour $\mathbb{K}[X]$ du théorème chinois 3.131.

Théorème 6.29 (Théorème chinois)**.**
Si P et Q sont deux polynômes premiers entre eux, alors nous avons l'isomorphisme

$$\mathbb{K}[X]/(P, Q) \simeq \mathbb{K}[X]/(P) \times \mathbb{K}[X]/(Q). \tag{6.59}$$

6.3.1 Irréductibilité

Définition 6.30 ([98])**.**
Un polynôme à coefficients dans un anneau commutatif est irréductible si il

(1) n'est pas inversible,

(2) n'est pas le produit de deux non inversibles.

Un polynôme est irréductible dans $\mathbb{K}[X]$ au sens de la définition 3.110 si et seulement s'il est irréductible au sens de la définition 6.30 parce que seules les constantes (non nulles) sont inversibles dans $\mathbb{K}[X]$.

Exemple 6.31.
Si un polynôme $P \in \mathbb{Z}[X]$ n'a que des racines complexes, ça ne l'empêche pas d'être réductible sur $\mathbb{Z}$. La réductibilité ne signifie pas qu'on peut mettre des racines en évidence. Par exemple le polynôme $P = X^4 + 3X^2 + 2$ est réductible sur $\mathbb{Z}$ en

$$P = (X^2 + 1)(X^2 + 2), \tag{6.60}$$

mais n'a pas de racines dans $\mathbb{Z}$. Par contre, il est vrai que si on veut réduire plus, il faut sortir de $\mathbb{Z}$.

△

Définition 6.32.
*Nous disons que $P \in \mathbb{K}[X]$ est **scindé** sur $\mathbb{K}$ s'il est produit dans $\mathbb{K}[X]$ de polynômes de degré 1.*

Définition 6.33 (Polynôme scindé)**.**
*Soit $P \in \mathbb{K}[X]$ un polynôme irréductible, et $\mathbb{L}$ un corps, extension du corps $\mathbb{K}$. On dit que P est **scindé** dans $\mathbb{L}$ si P se décompose en un produit de polynômes de degré 1 dans $\mathbb{L}[X]$.*

Note : les constantes ne sont donc pas des polynômes scindés.

Proposition 6.34 (Critère d'Eisenstein)**.**
Soit le polynôme $P = \sum_{k=0}^{n} a_n X^n$ dans $\mathbb{Z}[X]$. Nous supposons avoir un nombre premier p tel que

(1) p divise tous les $a_0, \dots, a_{n-1}$,

(2) p ne divise pas a_n,

(3) p^2 ne divise pas a_0.

Alors P est irréductible dans $\mathbb{Q}[X]$.

Si de plus P est primitif au sens du pgcd *(définition 3.184) alors P est irréductible dans $\mathbb{Z}[X]$.*

Démonstration. Nous considérons $\bar{P}$ le polynôme réduit modulo p, c'est-à-dire $\bar{P} \in \mathbb{F}_p[X]$. Étant donné que par hypothèse tous les coefficients sont multiples de p sauf a_n, nous avons $\bar{P} = cX^n$. Supposons par l'absurde que $P = QR$ avec $Q, R \in \mathbb{Q}[X]$. Alors le lemme de Gauss (3.134) impose $P, Q \in \mathbb{Z}[X]$.

Nous avons aussi, au niveau des réductions modulo p que $\bar{Q}\bar{R} = \bar{P}$. Or $\bar{P}$ est un monôme, donc $\bar{Q}$ et $\bar{R}$ doivent également l'être. Donc $\bar{Q} = dX^k$ et $\bar{R} = eX^{n-k}$ et en particulier $\bar{Q}(0) = \bar{R}(0) = 0$, c'est-à-dire que $Q(0)$ et $R(0)$ sont divisibles par p. Cela impliquerait que $a_0 = Q(0)R(0)$ soit divisible par p^2, ce qui est exclu par les hypothèses. Donc P est irréductible.

Supposons de surcroît que P est primitif au sens du pgcd. Il est donc irréductible et primitif sur $\mathbb{Q}[X]$ et le corolaire 3.196 nous dit alors que P est irréductible sur $\mathbb{Z}[X]$. □

Exemple 6.35.
Soit le polynôme $P = 3X^4 + 15X^2 + 10$. Pour faire fonctionner le critère d'Eisenstein il nous faut un nombre premier p divisant 15 et 10, mais pas 3 et dont le carré ne divise pas 10. C'est vite vu que $p = 5$ fait l'affaire. Le polynôme P est donc irréductible sur $\mathbb{Q}[X]$. △

6.3.2 Idéaux

Soit $P \in \mathbb{K}[X]$ un polynôme. Nous notons (P) l'idéal engendré par P :

$$(P) = \{PR \text{ tel que } R \in \mathbb{K}[X]\}. \tag{6.61}$$

Lemme 6.36.
Nous avons

(1) $(P) \subset (Q)$ si et seulement si Q divise P,

(2) $(P) = (Q)$ si et seulement si P et Q sont multiples (non nuls) l'un de l'autre.

Démonstration. Si $(P) \subset (Q)$, en particulier $P \in (Q)$ et il existe $R \in \mathbb{K}[X]$ tel que $P = QR$, ce qui signifie que Q divise P.

Si les idéaux de P et de Q sont identiques, l'un divise l'autre et l'autre divise l'un. Ils sont donc multiples l'un de l'autre. □

Théorème 6.37.
Soit $\mathbb{K}$ un corps commutatif.

(1) L'anneau $\mathbb{K}[X]$ est euclidien et principal.

(2) Si I est un idéal dans $\mathbb{K}[X]$ et si $P \in I$ est de degré minimal, alors $(P) = I$.

(3) De plus si $I \neq \{0\}$, il existe un unique polynôme unitaire μ tel que $I = (\mu)$.

Démonstration. Le point (1) a déjà été démontré dans le lemme 3.178 via le fait que $\mathbb{K}[X]$ est euclidien. Nous allons cependant donner ici une preuve directe que tous les idéaux de $\mathbb{K}[X]$ sont principaux. Si $I = \{0\}$, le résultat est évident. Nous supposons donc I non nul. Soit P de degré minimum parmi les éléments de I. Évidemment $(P) \subset I$. Nous allons démontrer qu'en réalité $(P) = I$.

Soit $P' \in I$. Par le théorème 3.175 de la division euclidienne[5], il existe Q et R dans $\mathbb{K}[X]$ tels que $P' = PQ + R$ avec $\deg(R) < \deg(P)$. Étant donné que $R = P' - PQ$ nous avons $R \in I$ et par conséquent $R = 0$ parce que P a été choisi de degré minimum dans I. Nous avons donc $P' = PQ$ et $I \subset (P)$.

L'existence d'un polynôme unitaire qui génère I est obtenu en choisissant $\mu = P/a_n$ où a_n est le coefficient du terme de plus haut degré. L'unicité d'un tel polynôme est obtenu par le fait que si μ et μ' génèrent le même idéal, alors ils sont multiples l'un de l'autre, or puisqu'ils sont unitaires, ils sont égaux. □

Nous voyons que n'importe quel polynôme de degré minimum dans un idéal génère l'idéal. Une importante conséquence du théorème 6.37 que nous verrons plus bas est que tout polynôme annulateur d'un endomorphisme est divisé par le polynôme minimal (proposition 9.96).

Corolaire 6.38.
Si $\mathbb{K}$ est un corps et si P est un polynôme irréductible, alors l'ensemble $\mathbb{L} = \mathbb{K}[X]/(P)$ est un corps. De plus $\mathbb{L}$ est un espace vectoriel de dimension $\deg(P)$.

Démonstration. En effet $\mathbb{K}[X]$ est un anneau principal par le théorème 6.37, par conséquent la proposition 3.126(2) déduit que $\mathbb{K}[X]/(P)$ est un corps.

Une base de $\mathbb{L}$ est donnée par les projections de $1, X, X^2, \dots, X^{n-1}$. En effet ces éléments forment une famille libre parce que si $\sum_{k=0}^{n-1} a_k \bar{X}^n = 0$ alors un représentant de cette classe doit être de la forme SP dans $\mathbb{K}[X]$, c'est-à-dire

$$\sum_{k=0}^{n-1} a_k X^k = SP, \tag{6.62}$$

ce qui n'est possible que si $S = 0$ et $a_k = 0$. D'autre part c'est un système générateur. En effet si $P = X^n + Q$ avec $\deg(Q) = n - 1$ alors

$$\bar{X}^{n+l} = \bar{X}^n \bar{X}^l = (\bar{P} - \bar{Q})\bar{X}^l = \bar{Q}\bar{X}^l. \tag{6.63}$$

Nous avons donc exprimé $\bar{X}^{n+l}$ comme une somme de termes de degré $n + l - 1$. Par récurrence nous pouvons exprimer tout $\bar{X}^{n+l}$ comme combinaison d'éléments de degré plus petit que n. □

6.39.
Ce corolaire prendra une nouvelle jeunesse lorsque nous parlerons de polynômes d'endomorphismes, en particulier la proposition 9.106 va donner des précisions.

5. Ici $\mathbb{K}$ est un corps et donc l'hypothèse d'inversibilité est automatiquement vérifiée.

Lemme 6.40 ([99]).
Soit un isomorphisme de corps $\tau\colon \mathbb{K} \to \mathbb{K}'$. Alors

(1) L'application étendue

$$\begin{aligned} \tau\colon \mathbb{K}[X] &\to \mathbb{K}'[X] \\ \sum_i a_i X^i &\mapsto \sum_i \tau(a_i) X^i \end{aligned} \tag{6.64}$$

est un isomorphisme d'anneaux ;

(2) pour tout $P \in \mathbb{K}[X]$, le passage au quotient

$$\begin{aligned} \phi_\tau\colon \mathbb{K}[X]/(P) &\to \mathbb{K}'[X]/\big(\tau(P)\big) \\ \bar{Q} &\mapsto \overline{\tau(Q)} \end{aligned} \tag{6.65}$$

est un isomorphisme d'anneaux (et d'abord est bien définie).

Démonstration. Nous n'allons pas faire explicitement toutes les vérifications, mais tout de même les principales. Montrons que τ respecte le produit entre $\mathbb{K}[X]$ et $\mathbb{K}'[X]$. Nous rappelons que ce produit est défini par a formule (3.118). En notant P_i les coefficients de P et Q_i ceux de Q et en remarquant que la définition de τ est essentiellement que $\tau(P)_i = \tau(P_i)$, nous avons :

$$\tau(PQ) = \tau\Big(\sum_k \big(\sum_{l=0}^{k} P_l Q_{k-l}\big) X^k\Big) \tag{6.66a}$$

$$= \sum_k X^k \sum_{l=0}^{k} \tau(P_l Q_{k-l}) \tag{6.66b}$$

$$= \sum_k X^k \sum_{l=0}^{k} \tau(P_l)\tau(Q_{k-l}) \tag{6.66c}$$

$$= \sum_k X^k \sum_{l=0}^{k} \tau(P)_l \tau(Q)_{k-l} \tag{6.66d}$$

$$= \sum_i \big(\tau(P)_i X^i\big) \sum_j \big(\tau(Q)_j X^j\big) \tag{6.66e}$$

$$= \tau(P)\tau(Q). \tag{6.66f}$$

Passons à l'isomorphisme d'anneaux donné par ϕ_τ.

Bien définie Si $\bar{Q}_1 = \bar{Q}_2$ alors $Q_2 = Q_1 + RP$ pour un certain $R \in \mathbb{K}[X]$. Dans ce cas,

$$\phi_\tau(Q_2) = \overline{\tau(Q_2)} = \overline{\tau(Q_1) + \tau(RP)} \tag{6.67a}$$

$$= \overline{\tau(Q_1) + \tau(R)\tau(P)} \tag{6.67b}$$

$$= \overline{\tau(Q_1)}. \tag{6.67c}$$

Ok pour bien définie.

Injection Si $\phi_\tau(\bar{Q}_1) = \phi_\tau(\bar{Q}_2)$ alors $\overline{\tau(Q_1)} = \overline{\tau(Q_2)}$, ce qui signifie que

$$\tau(Q_1) = \tau(Q_2) + R\tau(P) \tag{6.68}$$

pour un certain $R \in \mathbb{K}'[X]$. Vu que $\tau\colon \mathbb{K}[X] \to \mathbb{K}'[X]$ est un isomorphisme, nous pouvons y appliquer τ^{-1} pour trouver :

$$Q_1 = Q_2 + \tau^{-1}(R)P, \tag{6.69}$$

ce qui signifie que $\bar{Q}_1 = \bar{Q}_2$.

Surjection Un élément de $\mathbb{K}'[X]/\big(\tau(P)\big)$ est de la forme $\bar{Q}$ avec $Q \in \mathbb{K}'[X]$. Cela est l'image par ϕ_τ de l'élément $\overline{\tau^{-1}(Q)} \in \mathbb{K}[X]/(P)$.

Morphisme Nous vous laissons vérifier que l'application ϕ_τ est un morphisme d'anneaux.

□

6.3.3 Bézout

Théorème 6.41 (Bézout)**.**
Les polynômes $P_1, \dots, P_n$ dans $\mathbb{K}[X]$ sont étrangers entre eux si et seulement s'il existe des polynômes $Q_1, \dots, Q_n \in \mathbb{K}[X]$ tels que

$$P_1Q_1 + \cdots + P_nQ_n = 1. \tag{6.70}$$

Deux polynômes P et Q ne sont donc pas premiers entre eux s'il existe des polynômes x et y tels que l'identité de Bézout soit vérifiée :

$$xP + yQ = 0; \tag{6.71}$$

cette dernière pourra être écrite en termes de la matrice de Sylvester, voir sous-section 9.1.7.

Lemme 6.42.
Soient $(P_i)_{i=1,\dots,n} \in \mathbb{K}[X]$ des polynômes étrangers deux à deux. Alors les polynômes

$$Q_i = \prod_{j \neq i} P_j \tag{6.72}$$

sont étrangers entre eux[6].

Lemme 6.43 ([100])**.**
Soit $\mathbb{K}$ un corps commutatif et $\mathbb{A} \subset \mathbb{K}$ un sous anneau de $\mathbb{K}$. Alors $\mathbb{A}[X]$, vu comme idéal de $\mathbb{K}[X]$, est un idéal premier.

En d'autres termes, si $\phi \in \mathbb{K}[X]$, et s'il existe $Q \in \mathbb{K}[X]$ unitaire tel que $\phi Q \in \mathbb{A}[X]$, alors $\phi \in \mathbb{A}[X]$.

6.3.4 Lemme et théorème de Gauss

Théorème 6.44 (Théorème de Gauss)**.**
Soient $P, Q, R \in \mathbb{K}[X]$ tels que P soit premier avec Q et divise QR. Alors P divise R.

Démonstration. Étant donné que P est premier avec Q, le théorème de Bézout[7] nous donne $U, V \in \mathbb{K}[X]$ tels que $PU + QV = 1$. De plus il existe un polynôme S tel que $PS = QR$. En multipliant l'identité de Bézout par R, nous obtenons

$$R = PUR + QVR = PUR + VPS = P(UR + VS), \tag{6.73}$$

ce qui signifie que P divise R. □

Le lemme suivant est une généralisation du lemme de Gauss dans $\mathbb{Z}$ (lemme 3.134).

Lemme 6.45 (Lemme de Gauss[71])**.**
Soient les polynômes unitaires $P, Q \in \mathbb{Q}[X]$. Si $PQ \in \mathbb{Z}[X]$, alors P et Q sont tous deux dans $\mathbb{Z}[X]$.

Démonstration. Soit $a > 0$ le plus petit entier tel que $aP \in \mathbb{Z}[X]$ (c'est le PPCM des dénominateurs) et de la même façon $b > 0$ le plus petit entier tel que $bQ \in \mathbb{Z}[X]$. On pose $P_1 = aP$ et $Q_1 = bQ$.

Si $ab = 1$, alors $a = b = 1$ et nous avons tout de suite $P, Q \in \mathbb{Z}[X]$. Nous supposons donc $ab > 1$ et nous considérons p, un diviseur premier de ab. Ensuite nous considérons la projection

$$\pi_p \colon \mathbb{Z}[X] \to (\mathbb{Z}/p\mathbb{Z})[X]. \tag{6.74}$$

6. Et non seulement deux à deux.
7. théorème 6.41.

Par définition $abPQ = P_1Q_1 \in \mathbb{Z}[X]$; en prenant la projection,

$$\pi_p(P_1)\pi_p(Q_1) = \pi_p(P_1Q_1) = \pi_P(ab)\pi_p(PQ) = 0 \tag{6.75}$$

parce que $\pi_p(ab) = 0$. Étant donné que $(\mathbb{Z}/p\mathbb{Z})[X]$ est intègre (théorème 3.172), nous avons soit $\pi_p(P_1) = 0$ soit $\pi_p(Q_1) = 0$. Supposons pour fixer les idées que $\pi_p(P_1) = 0$. Alors $P_1 = pP_2$ pour un certain $P_2 \in \mathbb{Z}[X]$. Par ailleurs P est unitaire et $P_1 = aP$, donc le coefficient de plus haut degré de P_1 est a, et nous concluons que p divise a.

Mettons $a = pa'$. Dans ce cas, $pa'P = P_1 = pP_2$, et donc $a'P = P_2 \in \mathbb{Z}[X]$. Cela contredit la minimalité de a. □

6.3.5 Polynômes sur un corps et pgcd

Nous savons qu'un corps est un anneau intègre (lemme 1.100). De plus l'ensemble des polynômes sur un anneau intègre est lui-même un anneau intègre (théorème 3.172). Donc la notion de pgcd à utiliser dans le cas de $\mathbb{K}[X]$ est celle de la définition 1.81.

Lemme 6.46 (Unicité du pgcd à inversibles près)**.**
Soit un corps commutatif $\mathbb{K}$ et $S \subset \mathbb{K}[X]$. Si δ_1 et δ_2 sont des pgcd[8] *de S, alors $\delta_1 = k\delta_2$ avec $k \in \mathbb{K}$.*

Démonstration. Nous savons que δ_1 est un pgcd de S, mais que δ_2 divise S. Donc $\delta_2 \mid \delta_1$. De la même manière, $\delta_1 \mid \delta_2$. Il existe donc $A, B \in \mathbb{K}[X]$ tels que $\delta_1 = A\delta_2$ et $\delta_2 = B\delta_1$. En substituant,

$$\delta_1 = AB\delta_1. \tag{6.76}$$

Mais $\mathbb{K}[X]$ possède la propriété de simplification par la proposition 1.88(3). Donc $AB = 1$. Cela signifie entre autres que A et B sont des inversibles de $\mathbb{K}[X]$.

Or les seuls inversibles dans $\mathbb{K}[X]$ sont les éléments de $\mathbb{K}$; si vous en doutez, pensez que le degré de AB est supérieur ou égal à celui de A. □

6.47.
En général, lorsque nous dirons « le » pgcd d'un ensemble, nous parlerons du pgcd unitaire, qui existe et est bien défini par le lemme 6.46.

Lemme 6.48 ([101])**.**
Soit un corps commutatif $\mathbb{K}$, deux polynômes quelconques $A, B \in \mathbb{K}[X]$ et un polynôme unitaire G.

Nous avons $G = \operatorname{pgcd}(A, B)$ si et seulement si les deux conditions suivantes sont satisfaites :

(1) Il existe $U, V \in \mathbb{K}[X]$ tels que $AU + BV = G$,

(2) G divise A et B.

Démonstration. Une implication dans chaque sens.

$\Rightarrow$ Si G est le pgcd de A et B, il est clair que $G|A$ et $G|B$. Il reste donc à montrer l'existence des polynômes U et V vérifiant $AU + BV = G$. Vu que G divise A et B, il existe des polynômes A_1, B_1 tels que $A = GA_1$ et $B = GB_1$.

Nous montrons que les polynômes A_1 et B_1 sont premiers entre eux. S'ils ont un diviseur commun D, alors GD est un diviseur commun à A et B. Or, G est le pgcd de A et B donc $GD|G$; D ne peut être qu'un polynôme constant (c'est-à-dire un élément de $\mathbb{K}$). Mais comme G est unitaire, le coefficient du terme de plus haut degré de GD doit être 1. Donc $D = 1$. L'élément 1 est l'unique diviseur commun de A_1 et B_1 ; donc A_1 et B_1 sont donc bien premiers entre eux.

D'après le théorème de Bézout 6.41, il existe donc U et V tels que $A_1U + B_1V = 1$. En multipliant par G, nous obtenons l'égalité voulue : $AU + BV = G$.

8. Définition 1.81.

$\Leftarrow$ Si G vérifie les deux conditions, montrons que G est le pgcd de A et B. Nous savons déjà (par hypothèse) que G divise A et B, il reste à montrer que tous les diviseurs commun à A et B divisent aussi G. Soit donc D un diviseur commun à A et B : il existe A_1 et B_1 tels que $A = DA_1$ et $B = DB_1$. Nous savons que $G = AU + BV$ donc $G = D(A_1U + B_1V)$, et $D|G$. Par définition, G est bien le pgcd de A et B.

□

Notons qu'en supprimant la condition d'unitarité de G, le résultat tient presque : il suffit de remplacer partout « le pgcd » par « un pgcd ».

Lemme 6.49 ([101]).
Soient deux polynômes A, B premiers entre eux. Si le polynôme P est divisible par A et par B alors P est divisible par AB.

Démonstration. Vu que $A \mid P$, il existe $Q_1 \in \mathbb{K}[X]$ tel que $P = AQ_1$. Mais B divise $P = AQ_1$ alors que B est premier avec A ; donc d'après le théorème de Gauss 6.44 : $B|Q_1$.

Il existe donc $Q_2 \in \mathbb{K}[X]$ tel que $Q_1 = BQ_2$. On a donc $P = ABQ_2$: P est bien divisible par AB. □

Lemme 6.50 ([101]).
Quelques propriétés du PGCD[9] *dans les polynômes. Soient des polynômes $P, Q, R \in \mathbb{K}[X]$.*

(1) Nous avons l'égalité[10]

$$\operatorname{pgcd}(P, PQ + R) = \operatorname{pgcd}(P, R). \tag{6.77}$$

(2) Si Q et R sont premiers entre eux,

$$\operatorname{pgcd}(P, QR) = \operatorname{pgcd}(P, Q)\operatorname{pgcd}(P, R) \tag{6.78}$$

(3) Si P et Q sont premiers entre eux,

$$\operatorname{pgcd}(P, QR) = \operatorname{pgcd}(P, R) \tag{6.79}$$

Démonstration. Dans la suite si A et B sont des polynômes, nous dirons « les diviseurs de $\{A, B\}$ » pour parler des diviseurs communs de A et B.

(1) Nous montrons que $\{P, PQ + R\}$ a les mêmes diviseurs que $\{P, R\}$.
D'une part, si $A \mid \{P, PQ + R\}$, alors il existe des polynômes B_1 et B_2 tels que $P = AB_1$ et $PQ + R = AB_2$. Donc

$$R = AB_2 - PQ = AB_2 - AB_1Q = A(B_2 - B_1Q), \tag{6.80}$$

et nous concluons que A divise R.
D'autre part, si $A \mid \{P, R\}$ alors il existe des polynômes B_1 et B_2 tels que $P = AB_1$ et $R = AB_2$. Donc

$$PQ + R = AB_1Q + AB_2 = A(B_1Q + B_2), \tag{6.81}$$

et A divise $PQ + R$.
Conclusion : les paires $\{P, PQ + R\}$ et $\{P, R\}$ ont même ensemble de diviseurs, et donc même pgcd.

(2) Nous avons trois polynômes P, Q, R et nous savons que Q et R sont premiers entre eux. Nous notons : $G_1 = \operatorname{pgcd}(P, Q)$ et $G_2 = \operatorname{pgcd}(P, R)$. Il faut montrer que G_1G_2 est le pgcd de P et QR ; pour cela nous allons utiliser le lemme 6.48.

9. Définition 1.81.
10. Notez l'analogie avec le lemme 3.18.

$\exists U, V$ tels que $G_1G_2 = PU + QRV$ Vu que $G_1 = \text{pgcd}(P, Q)$, il existe U_1 et V_1 tels que $G_1 = PU_1 + QV_1$ (lemme 6.48). On a de même : $G_2 = PU_2 + RV_2$. En prenant le produit :

$$G_1G_2 = (PU_1 + QV_1)(PU_2 + RV_2) = P(PU_1U_2 + RU_1V_2 + QV_1V_2) + QR(V_1V_2). \quad (6.82)$$

Donc c'est bon pour ce point.

G_1 et G_2 sont premiers entre eux Si D est un diviseur commun à G_1 et G_2, alors D divise Q et R qui sont premiers entre eux ; D ne peut être qu'un polynôme constant. Tous les diviseurs communs de G_1 et G_2 sont dans $\mathbb{K}$. Mais le pgcd est par définition un diviseur commun unitaire, donc $\text{pgcd}(G_1, G_2) = 1$. Cela signifie que G_1 et G_2 sont premiers entre eux (définition 3.11).

$G_1G_2 \mid QR$ En effet : $G_1|Q$ et $G_2|R$ donc $G_1G_2|QR$.

$G_1G_2 \mid P$ Le polynôme P est divisible par G_1 et par G_2, et de plus G_1 et G_2 sont premiers entre eux. Donc le lemme 6.49 conclu que P est divisible par G_1G_2.

(3) Supposons d'abord que $A \in \mathbb{K}[X]$ divise P et QR. Le théorème de Bézout 6.41 assure l'existence de polynômes U et V tels que $PU + QV = 1$. Ensuite l'hypothèse de division nous donne des polynômes B_1 et B_2 tels que $P = AB_1$ et $QR = AB_2$. Nous avons :

$$1 = PU + QV = AB_1U + QV. \quad (6.83)$$

Cela prouve que A est premier avec Q grâce encore à Bézout, mais dans l'autre sens. Donc A est premier avec Q et $A \mid QR$. Donc $A|R$ par le théorème de Gauss 6.44.

Dans l'autre sens, si $A|R$ alors on a évidemment : $A|QR$.

Les diviseurs de $\{P, QR\}$ sont exactement les diviseurs de $\{P, R\}$. En conséquence, nous concluons que les paires $\{P, QR\}$ et $\{P, R\}$ ont le même pgcd.

□

6.4 Extension de corps

Lemme 6.51.
Soit $\mathbb{L}$ un corps[11] *fini et $\mathbb{K}$ un sous corps de $\mathbb{L}$. Alors il existe $s \in \mathbb{N}$ tel que*

$$\text{Card}(\mathbb{L}) = \text{Card}(\mathbb{K})^s. \quad (6.84)$$

Démonstration. Le corps $\mathbb{L}$ est un $\mathbb{K}$-espace vectoriel de dimension finie. Si s est la dimension alors nous avons la formule (6.84) parce que chaque élément de $\mathbb{L}$ est un s-uple d'éléments de $\mathbb{K}$. □

Définition 6.52 ([102]).
Soit $\mathbb{K}$ un corps commutatif. Une ***extension*** *de $\mathbb{K}$ est un couple $(\mathbb{L}, j)$ où $\mathbb{L}$ est un corps et $j : \mathbb{K} \to \mathbb{L}$ est un morphisme de corps.*

Nous identifions le plus souvent $\mathbb{K}$ avec $i(\mathbb{K}) \subset \mathbb{L}$, mais il faut savoir que le corps $\mathbb{L}$ étendant $\mathbb{K}$ n'est pas toujours un sur-corps de $\mathbb{K}$.

Lemme 6.53.
Si $(\mathbb{L}, i)$ est une extension de $\mathbb{K}$, alors $\mathbb{L}$ est un espace vectoriel sur $\mathbb{K}$.

Démonstration. Il faut définir le produit d'un élément de $\mathbb{L}$ par un élément de $\mathbb{K}$; si $\lambda \in \mathbb{K}$ et $x \in \mathbb{L}$ nous la définissons par

$$\lambda \cdot x = i(\lambda)x \quad (6.85)$$

où la multiplication du membre de droite est celle du corps $\mathbb{L}$. □

11. Définition 1.97.

Définition 6.54.
*Le **degré** de $\mathbb{L}$ est la dimension de cet espace vectoriel. Il est noté $[\mathbb{L} : \mathbb{K}]$; notons qu'il peut être infini.*

Exemple 6.55.
L'ensemble $\mathbb{C}$ est une extension de $\mathbb{R}$ et son degré est $[\mathbb{C} : \mathbb{R}] = 2$. △

Proposition 6.56 (Composition des degrés[103])**.**
Si $\mathbb{L}_2$ est une extension de $\mathbb{L}_1$ qui est elle-même une extension de $\mathbb{K}$, alors $\mathbb{L}_2$ est une extension de $\mathbb{K}$ et on a :

$$[\mathbb{L}_2 : \mathbb{K}] = [\mathbb{L}_2 : \mathbb{L}_1][\mathbb{L}_1 : \mathbb{K}]. \tag{6.86}$$

Dans ce cas, si $\{v_i\}_{i\in I}$ est une $\mathbb{K}$-base de $\mathbb{L}_1$ et si $\{w_\alpha\}_{\alpha\in A}$ est une $\mathbb{L}_1$-base de $\mathbb{L}_2$ alors $\{v_i w_\alpha\}_{\substack{i\in I\\ \alpha\in A}}$ est une $\mathbb{K}$-base de $\mathbb{L}_2$.

Notons que la formule (6.86) n'est pas très instructive dans le cas des extensions non finies. La seconde partie, sur les bases, est en réalité nettement plus intéressante.

Démonstration. Soit $a \in \mathbb{L}_2$. Vu que les w_α forment une $\mathbb{L}_2$-base nous avons une décomposition

$$a = \sum_\alpha a_\alpha w_\alpha \tag{6.87}$$

pour des éléments $a_\alpha \in \mathbb{L}_1$. Mais les v_i forment une $\mathbb{K}$-base de $\mathbb{L}_1$, donc chacun des a_α peut être décomposé comme $a_\alpha = \sum_i a_{\alpha i} v_i w_\alpha$. Donc :

$$a = \sum_{\alpha i} a_{\alpha i} v_i w_\alpha, \tag{6.88}$$

qui donne une décomposition de a en éléments de $\{v_i w_\alpha\}$ à coefficients dans $\mathbb{K}$. La partie proposée est donc génératrice.

Pour prouver qu'elle est également libre, nous supposons avoir des éléments $a_{\alpha i} \in \mathbb{K}$ tels que

$$\sum_{\alpha i} a_{\alpha i} v_i w_\alpha = 0. \tag{6.89}$$

En récrivant sous la forme

$$\sum_\alpha \Big(\sum_i a_{\alpha i} v_i \Big) w_\alpha = 0, \tag{6.90}$$

nous reconnaissons une combinaison linéaire nulle des w_α à coefficients dans $\mathbb{L}_1$. Les coefficients sont donc nuls : $\sum_i a_{\alpha i} v_i = 0$. Cela est une combinaison linéaire nulle des v_i à coefficients dans $\mathbb{K}$. Vu que les v_i forment une base, les coefficients sont nuls : $a_{\alpha i} = 0$. □

6.4.1 Extension et polynôme minimal

Lemme-définition 6.57 (Polynôme minimal)**.**
Soit $\mathbb{L}$ une extension de $\mathbb{K}$ et $a \in \mathbb{L}$. Nous considérons la partie

$$I_a = \{P \in \mathbb{K}[X] \text{ tel que } P(a) = 0\} \tag{6.91}$$

que nous supposons non réduite à $\{0\}$ [12]

(1) La partie I_a est un idéal dans $\mathbb{K}[X]$,

(2) la partie I_a est un idéal principal dans $\mathbb{K}[X]$,

(3) l'idéal I_a possède un unique générateur unitaire.

*Cet unique générateur unitaire est le **polynôme minimal** de a sur $\mathbb{K}$.*

12. La non trivialité de I_a est une vraie hypothèse. En effet si nous prenons $\mathbb{K} = \mathbb{Q}$ et l'extension $\mathbb{L} = \mathbb{R}$, alors il suffit de prendre un réel a non algébrique sur $\mathbb{Q}$ pour que I_a soit réduit au seul polynôme identiquement nul.

Démonstration. En plusieurs parties.

Pour (1) Soit $P \in I_a$: $P(a) = 0$. Si $Q \in \mathbb{K}[X]$ alors la proposition 3.88 nous indique que

$$(PQ)(a) = P(a)Q(a) = 0. \tag{6.92}$$

Donc $PQ \in I_a$. Comme de plus I_a est clairement vectoriel, I_a est un idéal.

Notez que nous avons utilisé la règle du produit nul justifiée par le fait que $\mathbb{K}$ soit un corps[13] et donc soumis au point (3) de la proposition 1.88.

Pour (2) Nous savons par le théorème 6.37 que $\mathbb{K}[X]$ est un anneau principal. En particulier, tous ses idéaux sont principaux, c'est dans la définition 3.118 d'un anneau principal.

Pour (3) Le théorème 6.37(3) nous informe alors que I_a possède un unique générateur unitaire.

□

Si nous avons un corps et un élément dans une extension du corps, il n'est pas autorisé de dire « soit le polynôme minimal de cet élément dans le premier corps » parce qu'il n'existe peut-être pas de polynôme annulateur.

Exemple 6.58.
Le polynôme minimal dépend du corps sur lequel on le considère. Par exemple le nombre imaginaire pur i accepte $X - i$ comme polynôme minimal sur $\mathbb{C}$ et $X^2 + 1$ sur $\mathbb{Q}[X]$. △

Proposition 6.59 ([1])**.**
Soit $\mathbb{L}$ une extension de $\mathbb{K}$ et $a \in \mathbb{L}$ dont le polynôme minimal sur $\mathbb{K}$ est $\mu_a \in \mathbb{K}[X]$. Alors

(1) le polynôme μ_a est irréductible[14] *sur $\mathbb{K}$;*

(2) Le polynôme μ_a est premier[15] *avec tout polynôme de $\mathbb{K}[X]$ non annulateur de a.*

Démonstration. Une chose à la fois.

(1) D'abord le polynôme μ_a n'est pas inversible parce que seuls les éléments de $\mathbb{K}$ (ceux de degré zéro) peuvent être inversibles[16]. Mais ces polynômes sont constants et ne peuvent donc pas être des polynômes annulateurs de quoi que ce soit.

Ensuite, supposons la décomposition $\mu_a = PQ$ avec $P, Q \in \mathbb{K}[X]$. En évaluant cette égalité en a nous avons

$$0 = P(a)Q(a). \tag{6.93}$$

Vu que nous sommes sur un corps, nous avons la règle du produit nul[17] et nous déduisons que soit $P(a)$ soit $Q(a)$ est nul, ou les deux. Pour fixer les idées, nous supposons $P(a) = 0$.

Dans ce cas, P fait partie de l'idéal annulateur de a, lequel idéal est engendré par μ_a. Donc il existe $S \in \mathbb{K}[X]$ tel que $P = S\mu_a$. En récrivant $\mu_a = PQ$ avec cela nous avons :

$$\mu_a = S\mu_a Q \tag{6.94}$$

ou encore : $SQ = 1$, ce qui signifie que S et Q sont dans $\mathbb{K}$ et inversibles.

Nous concluons que μ_a ne peut pas être écrit sous forme de produit de deux non inversibles.

(2) Soit Q un polynôme non annulateur de a. Soit aussi un diviseur commun P de Q et μ_a dans $\mathbb{K}[X]$. Nous devons prouver que P est un inversible, c'est-à-dire un élément de $\mathbb{K}$ (le fait que P ne soit pas le polynôme nul est évident). Nous avons $\mu_a = PR_1$ et $Q = PR_2$ pour certains polynômes $R_1, R_2 \in \mathbb{K}[X]$. Vu que μ_a est irréductible par (1), il n'est pas produit de deux

13. Si vous connaissez un contre-exemple à cette proposition dans le cas où $\mathbb{K}$ serait remplacé par un anneau, écrivez-moi.
14. Définition 6.30.
15. Définition 3.179.
16. Et d'ailleurs, le sont, mais ce n'est pas important ici.
17. Parce que un corps est un anneau intègre par le lemme 1.100 et qu'un anneau intègre est justement un anneau sur lequel nous avons la règle du produit nul, voir la définition 1.88.

non inversibles. En d'autres termes, soit P soit R_1 est inversible. Si P n'est pas inversible, alors R_1 est inversible ; disons $R_1 = k \in \mathbb{K}$. Alors

$$0 = \mu_a(a) = P(a)k, \tag{6.95}$$

donc $P(a) = 0$. Mais alors

$$Q(a) = P(a)R_2(a) = 0, \tag{6.96}$$

ce qui est contraire à l'hypothèse selon laquelle Q n'était pas annulateur de a.

Nous retenons donc que P est inversible, ce qu'il fallait montrer.

□

Définition 6.60.
*Deux éléments α et β dans $\mathbb{L}$ sont dit **conjugués** s'ils ont même polynôme minimal. Par exemple i et $-i$ sont conjugués dans $\mathbb{C}$ vu comme extension de $\mathbb{Q}$.*

6.4.2 Extensions algébriques et éléments transcendants

6.4.2.1 Éléments algébriques et transcendants

Définition 6.61.
L'ensemble $A[X]$ devient un $\mathbb{K}$-espace vectoriel avec la définition

$$(\lambda P)_k = \lambda P_k. \tag{6.97}$$

Voici une définition d'un élément algébrique sur un corps. Une caractérisation plus « pratique » sera donnée dans le lemme 6.64.

Lemme-définition 6.62 (Élément algébrique et transcendant[104])**.**
Soit une extension $\mathbb{L}$ de $\mathbb{K}$ et $\alpha \in \mathbb{L}$. Nous considérons l'application

$$\begin{aligned} \varphi\colon \mathbb{K}[X] &\to \mathbb{L} \\ P &\mapsto P(\alpha). \end{aligned} \tag{6.98}$$

Alors

(1) L'application φ est un morphisme d'anneaux[18]*.*

(2) L'application φ est un morphisme de $\mathbb{K}$-espace vectoriel.

*Si φ est injective, nous disons que α est **transcendant**. Sinon, nous disons qu'il est **algébrique**.*

Démonstration. Le fait que φ soit un morphisme d'anneaux est le lemme 3.88 déjà prouvé.

Pour le morphisme de $\mathbb{K}$-espace vectoriel, il faut seulement ajouter le calcul

$$\varphi(\lambda P) = (\lambda P)(\alpha) = \lambda P(\alpha) = \lambda\varphi(P). \tag{6.99}$$

Notons la justification suivante qui n'est pas tout à fait triviale :

$$(\lambda P)(\alpha) = \sum_k (\lambda P)_k \alpha^k = \sum_k \lambda P_k \alpha^k = \lambda P(\alpha) \tag{6.100}$$

qui utilise la définition 6.61. □

Exemple 6.63.
L'injectivité de φ n'est pas automatique. Prenons par exemple $\mathbb{L} = \mathbb{Q}[\sqrt{2}]$ dans $\mathbb{R}$. Les polynômes dans $\mathbb{Q}[X]$ ont des degrés arbitrairement élevés en X, tandis que les éléments de $\mathbb{L}$ n'ont pas de degré très élevés en $\sqrt{2}$ parce que $\sqrt{2}\sqrt{2} = 2$. L'ensemble $\mathbb{Q}[\sqrt{2}]$ ne contient donc que des éléments de la forme $a + b\sqrt{2}$ avec $a, b \in \mathbb{Q}$.

Si par contre $x_0 \in \mathbb{R}$ n'est racine d'aucun polynôme (cela existe parce que $\mathbb{R}$ n'est pas dénombrable), alors $\mathbb{Q}[x_0]$ contient tous les $\sum_{k=0}^{N} a_k x_0^k$ avec N arbitrairement grand. Et tous ces nombres sont différents. △

18. Définition 1.74.

Le lemme suivant donne une caractérisation d'élément algébrique moins abstraire que la définition 6.62.

Lemme 6.64.
Soit $\mathbb{K}$, un corps et $\mathbb{L}$, une extension de $\mathbb{K}$. Un élément $\alpha \in \mathbb{L}$ est algébrique sur $\mathbb{K}$ si et seulement si existe un polynôme non nul $P \in \mathbb{K}[X]$ tel que $P(a) = 0$.

Démonstration. Nous considérons l'application φ de la définition 6.62. Si φ n'est pas injective, c'est qu'il existe un polynôme P dans $\mathbb{K}[X]$ tel que $\varphi(P) = 0$. Dans ce cas, $P(\alpha) = 0$.

À l'inverse si il existe P non nul dans $\mathbb{K}[X]$ tel que $P(\alpha) = 0$, alors $\varphi(P) = 0$ et φ n'est pas injective. □

Définition 6.65.
*Un corps $\mathbb{K}$ est **algébriquement clos** si tout polynôme non constant à coefficients dans $\mathbb{K}$ contient au moins une racine dans $\mathbb{K}$.*

Définition 6.66 (Extension algébrique, clôture algébrique)**.**
Soient un corps $\mathbb{K}$ et une extension $\alpha\colon \mathbb{K} \to \mathbb{L}$.

*(1) L'extension $\mathbb{L}$ est une extension **algébrique** de $\mathbb{K}$ si tous ses éléments sont algébriques*[19] *sur $\mathbb{K}$, c'est à dire sont racines de polynômes à coefficients dans $\alpha(\mathbb{K})$, voir le lemme 6.64.*

*(2) L'extension $\mathbb{L}$ est **algébriquement close** si tout $\mathbb{L}$ est algébriquement clos (définition 6.65).*

*(3) L'extension $\mathbb{L}$ est une **clôture algébrique** du corps $\mathbb{K}$ si elle est une extension algébrique qui est algébriquement close.*

6.67.
Donc une extension est algébrique si elle contient seulement des racines de polynômes ; elle est close si elle contient au moins une racine de chaque polynôme. Elle est une clôture algébrique si elle est les deux en même temps.

Exemple 6.68.
Le corps $\mathbb{R}$ n'est pas une extension algébrique de $\mathbb{Q}$. En effet il existe seulement une infinité *dénombrable* de polynômes dans $\mathbb{Q}[X]$ et donc une infinité dénombrable de racines de tels polynômes. Toute extension algébrique de $\mathbb{Q}$ est donc dénombrable. Voir aussi la proposition 6.121. △

Lemme 6.69.
Un corps est algébriquement clos si et seulement si tous ses polynômes sont scindés[20].

Démonstration. Si tout polynôme est scindé, tout polynôme possède des racines ; c'est l'autre sens qui est plus consistant.

Soit un corps algébriquement clos $\mathbb{K}$. Nous allons par récurrence sur le degré des polynômes. Si P est un polynôme de degré 1, alors il est scindé.

Nous supposons que tous les polynômes de degré $n-1$ sont scindés. Soit un polynôme P de degré n. Le corps étant algébriquement clos, le polynôme P a une racine que nous notons $a_n \in \mathbb{K}$. La proposition 3.191 nous explique qu'il existe un polynôme Q de degré $n-1$ tel que $P = (X - a_n)Q$.

Par hypothèse de récurrence, le polynôme Q est scindé : il existe $\{a_i\}_{i=1,\dots,n-1}$ dans $\mathbb{K}$ tels que $Q = \prod_{i=1}^{n-1}(X - a_i)$. Au final,

$$P = (X - a_n)Q = \prod_{k=1}^{n}(X - a_n) \tag{6.101}$$

et P est scindé. □

19. Définition 6.62.
20. Définition 6.32

Lemme 6.70.
Soient un corps $\mathbb{K}$ et un polynôme $P \in \mathbb{K}[X]$. Nous supposons que P est scindé :

$$P = \prod_{k=1}^{n} (X - a_k). \tag{6.102}$$

Si α est une racine de P, alors α est l'un des a_k.

Démonstration. Dire que α est une racine de P revient à dire que

$$\prod_{k=1}^{n} (\alpha - a_k) = 0 \tag{6.103}$$

Un corps est toujours un anneau intègre (lemme 1.100), c'est à dire que la règle du produit nul est utilisable. Dans notre cas, le produit nul (6.103) nous indique que $\alpha - a_i = 0$ pour (au moins) un des i. Donc effectivement α est l'un des a_i. □

Lemme 6.71.
Soient un corps algébriquement clos $\mathbb{K}$ ainsi qu'une extension algébrique $\alpha\colon \mathbb{K} \to \mathbb{L}$. Alors $\alpha(\mathbb{K}) = \mathbb{L}$.

Démonstration. Nous allons montrer que tous les éléments de $\mathbb{L}$ sont dans l'image de α. Soit donc $l \in \mathbb{L}$. Vu que l'extension $\alpha\colon \mathbb{K} \to \mathbb{L}$ est une extension algébrique, il existe un polynôme $P \in \alpha(\mathbb{K})[X]$ tel que $P(l) = 0$.

Étant donné que α est injective, il est possible de considérer le polynôme $Q = \alpha^{-1}(P)$, c'est à dire que, si $P = \sum_k a_k X^k$, nous posons $Q = \sum_k \alpha^{-1}(a_k) X^k$.

Le corps $\mathbb{K}$ étant algébriquement clos, le polynôme Q est scindé (proposition 6.69) :

$$Q = \prod_{k=1}^{n} (X - b_k) \tag{6.104}$$

avec $b_k \in \mathbb{K}$. Nous avons alors aussi la factorisation

$$P = \prod_{k=1}^{n} (X - \alpha(b_k)) \tag{6.105}$$

dans $\mathbb{L}[X]$. Nous avons vu que l était une racine de P. Donc l est un des $\alpha(b_k)$ (lemme 6.70). Cela prouve que $l \in \alpha(\mathbb{K})$. □

6.4.3 Extension algébrique et polynôme minimal

Proposition 6.72 ([41]).
Soit une extension algébrique[21] *$\mathbb{L}$ du corps $\mathbb{K}$.*

(1) Pour tout $a \in \mathbb{L}$, il existe un polynôme $P \in \mathbb{K}[X]$ tel que $P(a) = 0$.

(2) Le polynôme minimal de a dans $\mathbb{K}[X]$ est l'unique polynôme unitaire irréductible annulant a.

Démonstration. Le premier point est seulement la définition 6.66 d'une extension algébrique.

L'idéal annulateur $I_a = \{P \in \mathbb{K}[X] \text{ tel que } P(a) = 0\}$ n'est pas réduit à $\{0\}$ parce que $\mathbb{L}$ est une extension algébrique. L'existence du polynôme minimal est le lemme 6.57 et le fait qu'il soit irréductible est la proposition 6.59(1).

Ce qui nous intéresse ici est l'unicité. Soit $\mu_1 \in \mathbb{K}[X]$, un polynôme annulateur de a irréductible et unitaire. Vu que $\mu_1 \in I_a$ et que par définition, $I_a = (\mu)$, il existe $P \in \mathbb{K}[X]$ tel que $\mu_1 = P\mu$. Vu que μ n'est pas inversible et que μ_1 est irréductible, P doit être inversible : $\mu_1 = k\mu$ pour un certain $k \in \mathbb{K}$.

Vu que μ et μ_1 sont unitaires, $k = 1$. Donc $\mu_1 = \mu$. □

21. Définition 6.66.

Lemme 6.73.
Soient un corps $\mathbb{K}$, une extension $\mathbb{L}$ de $\mathbb{K}$ et $\alpha \in \mathbb{L}$, un élément algébrique[22] *sur $\mathbb{K}$. Si μ est le polynôme minimal de α sur $\mathbb{K}$ alors*

$$\begin{aligned} \varphi\colon \mathbb{K}[\alpha] &\to \mathbb{K}[X]/(\mu) \\ Q(\alpha) &\mapsto \bar{Q} \end{aligned} \tag{6.106}$$

avec $Q \in \mathbb{K}[X]$ est un isomorphisme de corps et de $\mathbb{K}$-espaces vectoriels.

Démonstration. D'abord, α est algébrique, donc l'idéal annulateur I_α n'est pas réduit à $\{0\}$, et l'existence d'un polynôme minimal est assurée par le lemme 6.57.

Ensuite, le fait que $\mathbb{K}[X]/(\mu)$ soit un corps est le corolaire 6.38. Nous montrons à présent que φ est un isomorphisme (d'anneaux) ; cela suffit pour en déduire que $\mathbb{K}[\alpha]$ est également un corps.

Ces préliminaires étant dits, nous commençons.

Bien définie Nous devons prouver que φ est bien définie, c'est-à-dire que tout élément de $\mathbb{K}[\alpha]$ peut être écrit sous la forme $Q(\alpha)$ pour un $Q \in \mathbb{K}[X]$, et que si $Q_1(\alpha) = Q_2(\alpha)$ alors $\bar{Q}_1 = \bar{Q}_2$. Le fait que tous les éléments de $\mathbb{K}[\alpha]$ peuvent être écrits sous la forme $Q(\alpha)$ est le proposition 3.171. Supposons que $Q_1(\alpha) = Q_2(\alpha)$. Alors nous définissons $R \in \mathbb{K}[X]$ par $Q_1 = Q_2 + R$, et en évaluant cette égalité en α nous avons

$$Q_1(\alpha) = Q_2(\alpha) + R(\alpha), \tag{6.107}$$

autrement dit $R(\alpha) = 0$. Donc R est dans l'idéal annulateur de α et est donc dans (μ), c'est-à-dire que dans le quotient $\mathbb{K}[X]/(\mu)$ nous avons $\bar{R} = 0$ et donc $\bar{Q}_1 = \bar{Q}_2$.

Surjective Tout élément de $\mathbb{K}[X]/(\mu)$ est de la forme $\bar{Q}$ pour un $Q \in \mathbb{K}[X]$. Or ces éléments sont ceux de l'ensemble d'arrivée de φ.

Injective Si $\bar{Q}_1 = \bar{Q}_2$, alors $Q_1 = Q_2 + R$ avec R dans l'idéal engendré par μ, c'est-à-dire entre autres $R(\alpha) = 0$. Donc $Q_1(\alpha) = Q_2(\alpha)$.

Nous devons encore montrer que nous avons là un morphisme de $\mathbb{K}$-espaces vectoriels.

(1) Si $k \in \mathbb{K}$ alors $\varphi\big(kQ(\alpha)\big) = \overline{kQ}$. Mais par définition de la structure d'espace vectoriel sur $\mathbb{K}[X]/(\mu)$, $\overline{kQ} = k\bar{Q}$ (vérifier que cette définition de la multiplication par un scalaire sur $\mathbb{K}[X]/(\mu)$ est correcte).

(2) Nous avons aussi $\varphi\big(Q_1(\alpha) + Q_2(\alpha)\big) = \varphi\big((Q_1 + Q_2)(\alpha)\big) = \overline{Q_1 + Q_2} = \bar{Q}_1 + \bar{Q}_2$.

□

6.4.4 Extensions et polynômes

Nous savons déjà depuis la définition 3.86 ce qu'est $A[X]$ pour tout anneau A et donc a fortiori pour un corps.

Définition 6.74.
Soit un corps commutatif[23]*. Nous notons $\mathbb{K}(X)$ le corps des fractions*[24] *de $\mathbb{K}[X]$.*

Lemme-définition 6.75.
Si $R \in \mathbb{K}(X)$, avec $R = P/Q$ et si $\mathbb{L}$ est une extension[25] *de $\mathbb{K}$ contenant l'élément α, alors nous définissons*

$$R(\alpha) = P(\alpha)Q(\alpha)^{-1}. \tag{6.108}$$

Cela est une bonne définition au sens où elle ne dépend pas du choix du représentant (P, Q) pris dans la classe P/Q.

22. Définition 6.62.
23. Sauf mention du contraire, tous les corps du Frido sont commutatifs.
24. Définition 1.103.
25. Définition 6.52.

Démonstration. Supposons $R = P_1/Q_1 = P_2/Q_2$. Par définition des classes (définition 1.103) nous avons

$$P_1Q_2 = Q_1P_2. \tag{6.109}$$

Vu que l'évaluation est un morphisme $\mathbb{K}[X] \to \mathbb{K}$ [26] nous pouvons évaluer l'équation (6.109) en α :

$$P_1(\alpha)Q_2(\alpha) = Q_1(\alpha)P_2(\alpha). \tag{6.110}$$

Cette dernière est une égalité dans le corps $\mathbb{K}$. Nous pouvons donc la multiplier par $Q_2(\alpha)^{-1}P_2(\alpha)^{-1}$ (et utiliser toutes les hypothèses de commutativité des anneaux et corps) pour obtenir

$$P_1(\alpha)Q_1(\alpha)^{-1} = P_2(\alpha)Q_2(\alpha)^{-1}, \tag{6.111}$$

c'est-à-dire

$$(P_1/Q_1)(\alpha) = (P_2/Q_2)(\alpha). \tag{6.112}$$

□

Proposition-définition 6.76 ([1]).
Soient un corps $\mathbb{K}$, une extension $(\mathbb{L}, j_{\mathbb{L}})$ de $\mathbb{K}$ et un élément $\alpha \in \mathbb{L}$. Nous définissons $\mathbb{K}(\alpha)_{\mathbb{L}}$ comme étant l'intersection de tous les sous-corps de $\mathbb{L}$ contenant $j_{\mathbb{L}}(\mathbb{K})$ et α.

Alors

(1) $\mathbb{K}(\alpha)_{\mathbb{L}}$ est un sous-corps de $\mathbb{L}$,

(2) $\mathbb{K}(\alpha)_{\mathbb{L}}$ est une extension [27] de $\mathbb{K}$.

Démonstration. Nous commençons par prouver que $\mathbb{K}(\alpha)_{\mathbb{L}}$ est bien un corps. Si $a, b \in \mathbb{K}(\alpha)_{\mathbb{L}}$ alors il suffit de calculer ab, $a + b$ et a^{-1} dans n'importe quel sous-corps de $\mathbb{L}$ contenant $\mathbb{K}$ et α ; nous avons une garantie que a, b, ab, $a + b$ et a^{-1} sont dans tous les tels sous-corps.

Pour prouver que $\mathbb{K}(\alpha)_{\mathbb{L}}$ est bien une extension, nous devons trouver une homomorphisme de corps $j\colon \mathbb{K} \to \mathbb{K}(\alpha)_{\mathbb{L}}$. Il se fait que prendre $j = j_{\mathbb{L}}$ fonctionne parce que par définition, $\mathbb{K}(\alpha)_{\mathbb{L}}$ est une partie de $\mathbb{L}$ contenant l'image de $j_{\mathbb{L}}$. □

Lemme 6.77.
Soit n tel que $\sqrt{n}$ ne soit pas un rationnel. Si $\alpha \in \{a + b\sqrt{n}\}_{a,b\in\mathbb{Q}}$, alors il existe un unique choix $(x, y) \in \mathbb{Q}^2$ tel que

$$\alpha = x + y\sqrt{n}. \tag{6.113}$$

Exemple 6.78.
Nous avons

$$\mathbb{Q}(\sqrt{2})_{\mathbb{R}} = \{a + b\sqrt{2}\}_{a,b\in\mathbb{Q}} \tag{6.114}$$

où à droite nous calculons les sommes et et les produits dans $\mathbb{R}$. Le tout est un sous-ensemble de $\mathbb{R}$ qui se révèle être un corps contenant $\mathbb{Q}$ et $\sqrt{2}$.

En particulier, dans $\mathbb{Q}(\sqrt{2})_{\mathbb{R}}$ nous avons $\sqrt{2}\sqrt{2} = 2$. △

Lemme 6.79.
Les corps $\mathbb{Q}(\sqrt{2})_{\mathbb{R}}$ et $\mathbb{Q}(\sqrt{3})_{\mathbb{R}}$ ne sont pas isomorphes.

Démonstration. Supposons l'existence d'un morphisme de corps [28]

$$\psi\colon \mathbb{Q}(\sqrt{2})_{\mathbb{R}} \to \mathbb{Q}(\sqrt{3})_{\mathbb{R}}. \tag{6.115}$$

Nous notons « 1 » à la fois le neutre de la multiplication dans $\mathbb{Q}(\sqrt{2})_{\mathbb{R}}$ et $\mathbb{Q}(\sqrt{3})_{\mathbb{R}}$ (qui s'évèrent être les mêmes en tant qu'élément de $\mathbb{R}$, mais ça n'a pas d'importance ici).

26. Lemme 3.88. Certes ce lemme ne parle que d'anneaux, mais à y bien penser, dans le passage de (6.109) à (6.110), nous ne considérons que les structures d'anneaux sur $\mathbb{K}[X]$ et $\mathbb{K}$.

27. Définition 6.52.

28. Définition 1.74. Oui, c'est un bête morphisme d'anneaux. Il n'y a pas plus de structure dans un corps que dans un anneau.

Soit $\alpha \in \mathbb{Q}(\sqrt{2})_{\mathbb{R}}$ tel que $\alpha^2 - 1 = 0$. Alors nous avons aussi

$$\psi(\alpha)^2 - 1 = \psi(\alpha^2) - \psi(1) = \psi(\alpha^2 - 1) = \psi(0) = 0. \tag{6.116}$$

Donc $\psi(\alpha)$ est un élément de $\mathbb{Q}(\sqrt{3})_{\mathbb{R}}$ qui est une racine de $X^2 - 1$.

Or un tel élément n'existe pas dans $\mathbb{Q}(\sqrt{3})_{\mathbb{R}}$ parce que nous savons que dans $\mathbb{R}$ entier, il n'y a que deux racines : $\pm\sqrt{2}$, et aucune des deux n'est dans $\mathbb{Q}(\sqrt{3})_{\mathbb{R}}$. □

Exemple 6.80.
Est-ce que $\mathbb{K}(\alpha)_{\mathbb{L}}$ dépend réellement de $\mathbb{L}$? Si $\mathbb{L}_2$ est une extension de $\mathbb{L}$ alors nous avons évidemment [29] $\mathbb{K}(\alpha)_{\mathbb{L}_2} = \mathbb{K}(\alpha)_{\mathbb{L}}$.

Nous commençons par construire un corps $\mathbb{K}$ un peu idiot qui, comme ensemble, est comme $\mathbb{Q}(\sqrt{2})_{\mathbb{R}}$, c'est-à-dire la partie

$$\{a + b\sqrt{2}\}_{a,b\in\mathbb{Q}}, \tag{6.117}$$

de $\mathbb{R}$.

Mais cette fois nous définissons la multiplication suivante :

$$(a + b\sqrt{2})(c + d\sqrt{2}) = ac + 3bd + (ad + bc)\sqrt{2}. \tag{6.118}$$

Cela est un corps parce que tout élément non nul est inversible. En effet, l'équation

$$(a + b\sqrt{2})(x + y\sqrt{2}) = 1 \tag{6.119}$$

donne

$$\begin{pmatrix} a & 3b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{6.120}$$

Ce système a une unique solution si et seulement si $\det \begin{pmatrix} a & 3b \\ b & a \end{pmatrix} = 0$. Cela survient si et seulement si

$$a^2 - 3b^2 = 0. \tag{6.121}$$

Les solutions de cela dans $\mathbb{R}$ sont $a = \pm\sqrt{3}|b|$. Dès que a ou b est non nul, cela ne peut pas satisfaire $a, b \in \mathbb{Q}$. Donc le déterminant est toujours non nul et il existe $x, y \in \mathbb{Q}$ tels que (6.119) soit satisfaite.

Tout cela nous a donné une corps $\mathbb{K}$ dont $\mathbb{Q}$ est un sous-corps et qui contient l'élément $\sqrt{2}$ de $\mathbb{R}$. Il n'est cependant pas un sous-corps de $\mathbb{R}$.

Ce corps est isomorphe à $\mathbb{Q}(\sqrt{3})_{\mathbb{R}}$. En effet, nous montrons que

$$\begin{aligned} \psi\colon \mathbb{K} &\to \mathbb{Q}(\sqrt{3})_{\mathbb{R}} \\ a + b\sqrt{2} &\mapsto q + b\sqrt{3} \end{aligned} \tag{6.122}$$

est un isomorphisme de corps. Pour le produit, nous avons

$$\begin{aligned} \psi\big((a + b\sqrt{2})(c + d\sqrt{2})\big) &= \psi\big(qc + 3bd + (ad + bc)\sqrt{2}\big) && \text{(6.123a)} \\ &= ac + 3bd + (ad + bc)\sqrt{3} && \text{(6.123b)} \\ &= (a + b\sqrt{3})(c + d\sqrt{3}) && \text{(6.123c)} \\ &= \psi(a + b\sqrt{2})\psi(c + d\sqrt{2}). && \text{(6.123d)} \end{aligned}$$

Remarques :

— L'application ψ est bien définie grâce au lemme 6.77 couplé au théorème 3.43 appliqué à $n = 2$ et $n = 3$.

— Dans le membre de gauche de (6.123a), $b\sqrt{2}$ est un produit dans $\mathbb{R}$ (d'où l'importance du lemme 6.77 qui permet de re-séparer les éléments de $\mathbb{R}$ partie rationnelle et partie multiple de $\sqrt{2}$), et le produit entre $(a + b\sqrt{2})$ et $(c + d\sqrt{2})$ est un produit dans $\mathbb{K}$.

29. Vérifiez-le tout de même.

— Dans (6.123b) et (6.123c), tous les produits sont dans $\mathbb{R}$.

En comparant avec le lemme 6.79, nous avons alors

$$\mathbb{Q}(\sqrt{2})_{\mathbb{K}} = \mathbb{Q}(\sqrt{3})_{\mathbb{R}} \neq \mathbb{Q}(\sqrt{2})_{\mathbb{R}} \tag{6.124}$$

△

6.81.
Nous allons encore enfoncer le clou sur le fait que $\mathbb{K}(\alpha)_{\mathbb{L}}$ dépend de $\mathbb{L}$.

Le fait est que si on y pense, l'objet $\sqrt{2}$ n'a aucun rapport avec $\mathbb{Q}$. En effet les objets de $\mathbb{Q}$ sont des classes d'équivalence de couples d'éléments de $\mathbb{Z}$, alors que l'élément $\sqrt{2}$ est une classe d'équivalence de suites de Cauchy dans $\mathbb{Q}$.

Lorsque nous écrivons $\mathbb{Q}(\sqrt{2})$, nous associons des objets de nature complètement différentes, et il n'y a aucune raison à priori de définir la multiplication entre eux d'une façon plutôt qu'une autre.

Plus généralement, dans ZF (que nous faisons du semblant de suivre tout en sachant que nous ne savons pas ce que c'est réellement [30]), tout est ensemble. Peut-on dire ce que serait $\mathbb{Q}(I)$ si I est un ensemble quelconque ? Attention : en écrivant $\mathbb{Q}(I)$, nous entendons un corps dont I est un élément, pas un corps qui contiendrait comme éléments tous les éléments de I.

Si I est juste un ensemble, quelle définition donner de I^2 ? Il y a plein de choix et rien ne se dégage clairement comme étant pertinent. Si par contre, en guise de I nous considérons l'ensemble $\sqrt{2}$ (oui, c'est un ensemble : un ensemble de suites de Cauchy dans $\mathbb{Q}$), alors tout de suite nous nous disons que la bonne façon de faire est $\sqrt{2}^2 = 2$. Ce réflexe est juste conditionné par le fait que nous connaissons déjà par ailleurs le corps $\mathbb{R}$. Rien de plus.

Donc oui, $\mathbb{K}(\alpha)_{\mathbb{L}}$ dépend de $\mathbb{L}$, mais dans les cas particuliers où $\mathbb{K}$ est un sous-corps de $\mathbb{C}$, il y a un implicite comme quoi $\mathbb{L} = \mathbb{C}$. Cela étant dit, il n'y a plus d'ambiguïtés en écrivant $\mathbb{Q}(\sqrt{2})$.

Dans l'énoncé suivant, la notation $R(\alpha)_{\mathbb{L}}$ signifie que l'évaluation de R sur α se fait en calculant dans le sur-corps $\mathbb{L}$ de $\mathbb{K}$. Cette proposition semble indiquer que $\mathbb{K}(\alpha)$ est donné en termes de $\mathbb{K}(X)$, lequel est défini de façon très intrinsèque sans faire appel à un corps ambiant de $\mathbb{K}$.

Proposition 6.82 ([1]).
Soit une extension $\mathbb{L}$ du corps $\mathbb{K}$ et $\alpha \in \mathbb{L}$. Alors nous avons les isomorphismes de corps suivants :

(1) $\mathbb{K}(\alpha)_{\mathbb{L}} = \operatorname{Frac}\big(\mathbb{K}[\alpha]_{\mathbb{L}}\big)$,

(2) $\mathbb{K}(\alpha)_{\mathbb{L}} = \{R(\alpha)_{\mathbb{L}}$ *tel que* $R \in \mathbb{K}(X)\}$.

Démonstration. Le corps $\mathbb{K}(\alpha)$ est un sous-corps de $\mathbb{L}$ contenant $\mathbb{K}[\alpha]$ comme sous-anneau. La proposition 1.104 nous dit alors que l'application suivante est un morphisme injectif de corps :

$$\begin{aligned} \epsilon\colon \operatorname{Frac}\big(\mathbb{K}[\alpha]\big) &\to \mathbb{K}(\alpha) \\ P/Q &\mapsto PQ^{-1}. \end{aligned} \tag{6.125}$$

Pour rappel, la notation P/Q est bien une notation pour la classe d'équivalence du couple (P, Q) pour la relation définie en 1.103.

Par ailleurs, la partie $\epsilon\Big(\operatorname{Frac}\big(\mathbb{K}[\alpha]\big)\Big)$ est inclue à $\mathbb{L}$ et est un corps contenant $\mathbb{K}$ et α. Donc le corps $\mathbb{L}$ fait partie des corps sur lesquels on prend l'intersection pour définir $\mathbb{K}(\alpha)$. Cela prouve que

$$\mathbb{K}(\alpha) \subset \epsilon\Big(\operatorname{Frac}\big(\mathbb{K}[\alpha]\big)\Big). \tag{6.126}$$

L'application ϵ est donc surjective sur $\mathbb{K}(\alpha)$. Vu qu'elle était déjà injective, elle est bijective.

Pour la seconde partie, veuillez lire la définition 1.105 de l'évaluation d'une fraction rationnelle sur un élément de l'anneau. Si $R = P/Q \in \mathbb{K}(X)$ et si $\alpha \in \mathbb{L}$, nous avons

$$R(\alpha) = P(\alpha)Q(\alpha)^{-1}. \tag{6.127}$$

30. En lisant quelques pages de Wikipédia, vous pourrez briller en société, mais ne tentez pas le coup à l'agrégation.

Tout sous-corps de $\mathbb{L}$ contenant $\mathbb{K}$ et α doit contenir en particulier $\{P(\alpha)$ tel que $P \in \mathbb{K}[X]\}$, les inverses $\{P(\alpha)^{-1}$ tel que $P \in \mathbb{K}[X]$, $P(\alpha) \neq 0\}$ et les produits d'iceux. Donc tout sous-corps de $\mathbb{L}$ contenant $\mathbb{K}$ et α contient $\{R(\alpha)$ tel que $R \in \mathbb{K}(X)\}$.

Nous avons donc

$$\{R(\alpha) \text{ tel que } R \in \mathbb{K}(X)\} \subset \mathbb{K}(\alpha). \tag{6.128}$$

Mais vu que $\mathbb{K}(\alpha)$ est lui-même un sous-corps de $\mathbb{L}$ contenant $\mathbb{K}$ et α, il est contenu dans $\{R(\alpha)$ tel que $R \in \mathbb{K}(X)\}$. D'où l'égalité. □

Pourquoi cela ne contredit pas l'exemple 6.80 ? Lorsque nous écrivons

$$\mathbb{K}(\alpha) = \{R(\alpha) \text{ tel que } R \in \mathbb{K}(X)\}, \tag{6.129}$$

certes $\mathbb{K}(X)$ est défini sans faire appel à un corps contenant $\mathbb{K}$. Mais l'évaluation $R(\alpha)$, oui. Pour calculer $R(\alpha)$, il faut écrire $R = P/Q$ et calculer $P(\alpha)Q(\alpha)^{-1}$. Tous les calculs de cette dernière expression doivent se faire dans un sur-corps de $\mathbb{K}$. Il suffit que le sur-corps en question soit un monceau de mauvaise foi comme celui de l'exemple 6.80, et en réalité $\mathbb{K}(\alpha)$ peut ne pas être ce que l'on croit.

Le corolaire suivant montre que les choses s'arrangent.

Corolaire 6.83.
Soient un corps $\mathbb{K}$, une extension $\mathbb{L}_1$ de $\mathbb{K}$, un élément $\alpha \in \mathbb{L}_1$ et une extension $\mathbb{L}_2$ de $\mathbb{L}_1$. Alors

$$\mathbb{K}(\alpha)_{\mathbb{L}_1} = \mathbb{K}(\alpha)_{\mathbb{L}_2}. \tag{6.130}$$

Démonstration. La proposition 6.82 nous dit que

$$\mathbb{K}(\alpha)_{\mathbb{L}_1} = \{R(\alpha)_{\mathbb{L}_1} \text{ tel que } R \in \mathbb{K}(X)\} \tag{6.131a}$$
$$\mathbb{K}(\alpha)_{\mathbb{L}_2} = \{R(\alpha)_{\mathbb{L}_2} \text{ tel que } R \in \mathbb{K}(X)\}. \tag{6.131b}$$

Mais lorsque $R \in \mathbb{K}(X)$, le calcul de $R(\alpha)$ est exactement le même dans $\mathbb{L}_1$ et dans $\mathbb{L}_2$ parce que $\mathbb{L}_2$ est un sur-corps de $\mathbb{L}_1$ et que les calculs effectifs de $R(\alpha) = P(\alpha)Q(\alpha)^{-1}$ ne font intervenir que des quantités de $\mathbb{K}$ et des puissances de α. □

Ce que ce corolaire nous dit est que si le contexte fixe une extension de $\mathbb{K}$, nous pouvons faire tous les calculs dans cette extension, même si il y a des piles d'extensions à côté.

Typiquement, à chaque fois que nous considérons des sous-corps de $\mathbb{C}$, les extensions se feront dans $\mathbb{C}$: pour tout $\alpha \in \mathbb{C}$, les corps $\mathbb{Q}(\alpha)$, $\mathbb{R}(\alpha)$ se calculent dans $\mathbb{C}$.

Proposition 6.84.
Soit un corps $\mathbb{K}$, une extension $\mathbb{L}$ et un élément $\alpha \in \mathbb{L}$. Nous considérons l'application

$$\begin{aligned} \varphi\colon \mathbb{K}[X] &\to \mathbb{L} \\ P &\mapsto P(\alpha). \end{aligned} \tag{6.132}$$

(1) Si α est transcendant, alors $\mathbb{K}[\alpha] = \mathbb{K}[X]$ (isomorphisme d'anneaux).

(2) Si α est transcendant, alors $\mathbb{K}(\alpha)_{\mathbb{L}} = \mathbb{K}(X)$ (isomorphisme de corps),

(3) Si α est algébrique, alors $\ker(\varphi)$ est un idéal possédant un unique générateur unitaire, lequel est le polynôme minimal[31] *de α sur $\mathbb{K}$.*

Démonstration. Point par point.

(1) Nous savons que $\mathbb{K}[\alpha] = \{Q(\alpha)$ tel que $Q \in \mathbb{K}[X]\}$ (c'est la proposition 3.171). Donc φ est surjective sur $\mathbb{K}[\alpha]$, et est donc bijective. Elle est un isomorphisme[32] parce que le lemme 6.62 dit déjà que c'est un morphisme.

31. Définition 6.57.

32. Les amateurs d'écriture inclusive ne seront, je l'espère, pas choqué par « *elle* est *un* isomorphisme » ; c'est une tournure que je propose ici sur le modèle de l'immonde « *elle* est *un* ministre » ou, à peine moins grave, « *il* est *une* sommité ».

(2) Nous supposons encore que α est transcendant et nous considérons l'application

$$\begin{aligned}\psi\colon \mathbb{K}(X) &\to \mathbb{K}(\alpha)\\ P &\mapsto R(\alpha).\end{aligned} \tag{6.133}$$

Note : cette application n'est pas φ. En effet φ n'est définie que sur $\mathbb{K}[X]$; le corps des fractions $\mathbb{K}(X)$ est nettement plus grand (classes d'équivalence de couples).

Le fait que cette application soit surjective est la proposition 6.82(2). Pour l'injectivité nous supposons que $\psi(R) = 0$, c'est-à-dire que $R(\alpha) = 0$. Nous considérons un représentant (P, Q) de R ; c'est-à-dire $R = P/Q$. L'égalité $R(\alpha) = 0$ signifie $P(\alpha)Q(\alpha)^{-1} = 0$ (égalité dans $\mathbb{L}$). Vu que $\mathbb{L}$ est un corps, c'est un anneau intègre et nous avons la règle du produit nul ; soit $P(\alpha) = 0$, soit $Q(\alpha)^{-1} = 0$. La seconde possibilité est impossible parce que zéro n'est pas inversible. Donc $P(\alpha) = 0$. Donc $\varphi(P) = 0$ et φ étant injective, $P = 0$.

Lorsque $P = 0$, la classe P/Q est nulle dans $\mathbb{K}(X) = \mathrm{Frac}\big(\mathbb{K}[X]\big)$.

(3) C'est le lemme-définition 6.57.

□

Proposition 6.85.
Soit un corps $\mathbb{K}$ et une extension $\mathbb{L}$. Soit $P \in \mathbb{K}[X]$ et $a \in \mathbb{L}$, une racine de P. Alors le polynôme minimal d'une racine divise[33] *tout polynôme annulateur.*

Autrement dit, l'idéal engendré par le polynôme minimal est l'idéal des polynômes annulateurs.

Démonstration. Nous considérons l'idéal

$$I = \{Q \in \mathbb{K}[X] \text{ tel que } Q(a) = 0\}. \tag{6.134}$$

Le fait que cela soit un idéal est simplement dû à la définition du produit : $(PQ)(a) = P(a)Q(a)$. Par le théorème 6.37, le polynôme minimal μ_a de a est dans I et qui plus est le génère : $I = (\mu_a)$. Par conséquent tout polynôme annulateur de a est divisé par μ_a. □

6.4.4.1 Extension algébrique, degré

Proposition 6.86.
Toute extension finie est algébrique.

Démonstration. Soient un corps $\mathbb{K}$, une extension $\mathbb{L}$ de degré[34] n de $\mathbb{K}$ et $a \in \mathbb{L}$. Nous devons montrer qu'il existe un polynôme annulateur de a à coefficients dans $\mathbb{K}$.

Soit la partie $S = \{1, a, a^2, \dots, a^n\}$ de $\mathbb{L}$. Si cette partie contient des éléments non distincts, alors c'est plié. En effet, si $a^k = a^l$, alors le polynôme X^{k-l} est un polynôme annulateur de a.

Nous supposons donc que S contienne exactement $n+1$ éléments distincts. Le lemme 4.9 nous assure que S est une partie liée : il existe des éléments $k_i \in \mathbb{K}$ tels que $\sum_{i=0}^{n} k_i a^i = 0$.

Donc le polynôme $\sum_i a_i X^i$ est un polynôme annulateur de a. □

Proposition 6.87 (Propriétés d'extensions algébriques[1])**.**
Soit $\mathbb{K}$ un corps commutatif[35] *et a un élément algébrique*[36] *sur $\mathbb{K}$, de polynôme minimal μ_a de degré n. Alors*

(1) En considérant l'application d'évaluation

$$\begin{aligned}\varphi_a\colon \mathbb{K}[X] &\to \mathbb{L}\\ Q &\mapsto Q(a),\end{aligned} \tag{6.135}$$

nous avons $\mathbb{K}[a] = \mathrm{Image}(\varphi_a)$.

33. Définition 3.176.
34. Définition 6.54.
35. Juste en passant nous rappelons que tous les corps considérés ici sont commutatifs
36. Définition 6.62.

(2) Une base de $\mathbb{K}[a]$ *comme espace vectoriel sur* $\mathbb{K}$ *est donnée par* $\{1, a, a^2, \dots, a^{n-1}\}$.

(3) Le degré de l'extension $\mathbb{K}[a]$ *est égal au degré du polynôme minimal :*

$$\big[\mathbb{K}[a] : \mathbb{K}\big] = n. \tag{6.136}$$

(4) L'anneau $\mathbb{K}[a]$ *est l'ensemble des polynômes en* a *de degré* $n-1$ *à coefficient dans* $\mathbb{K}$.

(5) $\mathbb{K}(a) = \mathbb{K}[a]$.

(6) Il existe un isomorphisme d'anneaux $\varphi\colon \mathbb{K}[a] \to \mathbb{K}[X]/(\mu_a)$ *tel que* $\varphi(k) = \bar{k}$ *pour tout* $k \in \mathbb{K}$. $\mathbb{K}[a] \simeq \mathbb{K}[X]/(\mu_a)$ *(isomorphisme d'anneau).*

L'intérêt de (6) est qu'il permet de caractériser $\mathbb{K}[a]$ sans avoir recours à un sur-corps de $\mathbb{K}$. Le point (3) indique que le degré d'une extension algébrique est égal au degré du polynôme minimal.

Démonstration. (1) Nous avons $\mathbb{K}[a] \subset \mathrm{Image}(\varphi_a)$ parce que $\mathrm{Image}(\varphi_a)$ est lui-même un sous-anneau de $\mathbb{L}$ contenant $\mathbb{K}$ et a. Pour rappel, $\mathbb{K}[a]$ est l'intersection de tous les tels sous-anneaux.

L'inclusion inverse est le fait que si $Q \in \mathbb{K}[X]$ alors $Q(a) \in \mathbb{K}[a]$ parce que $\mathbb{K}[a]$ est un anneau et contient donc tous les a^n.

(2) La partie $\{1, a, a^2, \dots, a^{n-1}\}$ est libre parce qu'une combinaison linéaire de ces éléments est un polynôme de degré $n-1$ en a. Un tel polynôme ne peut pas être nul parce que nous avons mis comme hypothèse que le polynôme minimal de a est n.

Rappelons qu'en vertu de la définition 6.57, le polynôme minimal μ_a est unitaire ; donc le polynôme $\mu_a(X) - X^n$ est un polynôme de degré $n-1$. Par conséquent en posant $S(X) = X^n - \mu_a(X)$, le polynôme S est de degré $n-1$ et vérifie $a^n = S(a)$.

En vertu du point (1), un élément de $\mathbb{K}[a]$ s'écrit $Q(a)$ pour un certain $Q \in \mathbb{K}[X]$. Supposons que Q soit de degré $p > n-1$; alors nous le décomposons en une partie contenant les termes de degré jusqu'à $n-1$ et une partie contenant les autres :

$$Q(X) = Q_1(X) + X^n Q_2(X) \tag{6.137}$$

où Q_1 est de degré $n-1$ et Q_2 de degré $p-n$. Nous évaluons cette égalité en a :

$$Q(a) = Q_1(a) + S(a)Q_2(a). \tag{6.138}$$

Donc $Q(a)$ est l'image de a par le polynôme $Q_1 + SQ_2$ qui est de degré $p-1$. Par récurrence, $Q(a)$ est l'image de a par un polynôme de degré $n-1$.

Notons que l'idée est très simple : il s'agit de remplacer récursivement tous les a^n par $S(a)$.

(3) Conséquence immédiate de (2).

(4) Conséquence immédiate de (2).

(5) Un élément général non nul de $\mathbb{K}[a]$ est de la forme $Q(a)$ avec $Q \in \mathbb{K}[X]$; il s'agit de lui trouver un inverse. Pour cela nous remarquons que les polynômes $\mu_a(X)$ et $Q(x)$ sont premiers entre eux, sinon μ_a ne serait pas un polynôme minimal (voir la proposition 6.59). Donc le théorème de Bézout 6.41 affirme l'existence d'éléments $U, V \in \mathbb{K}[X]$ tels que

$$U\mu_a + VQ = 1 \tag{6.139}$$

dans $\mathbb{K}[X]$. Nous évaluons cette égalité en a en tenant compte de $\mu_a(a) = 0$ dans $\mathbb{K}[a]$:

$$U(a)\mu_a(a) + V(a)Q(a) = 1 \tag{6.140}$$

dans $\mathbb{K}[a]$. Par conséquent $V(a)Q(a) = 1$, ce qui signifie que $V(a)$ est l'inverse de $Q(a)$.

(6) Nous considérons l'application

$$\begin{aligned}\psi\colon \mathbb{K}[X]/(\mu_a) &\to \mathbb{K}[a]\\ \bar{R} &\mapsto R(a)\end{aligned} \tag{6.141}$$

et nous montrons qu'elle convient. Pour cela, nous nous souvenons que la proposition 6.85 nous enseigne que (μ_a), l'idéal engendré par μ_a, est égal à l'idéal des polynômes annulateurs de a dans $\mathbb{K}[X]$. Le polynôme μ_a divise tous les éléments de cet idéal ; voir aussi la définition 3.48 de l'idéal (μ_a). Cela étant mis au point, nous passons à la preuve.

ψ est bien définie Si $\bar{R} = \bar{S}$ alors $R = S + Q$ avec $Q \in (\mu_a)$, et par conséquent $R(a) = S(a) + Q(a)$ avec $Q(a) = 0$.

Surjective Nous savons que $\mathbb{K}[a] = \text{Image}(\varphi_a)$. Si $x \in \mathbb{K}[a]$ alors il existe $Q \in \mathbb{K}[X]$ tel que $x = Q(a)$. Dans ce cas nous avons aussi $x = \psi(\bar{Q})$.

Injective Si $\psi(\bar{R}) = 0$ alors $R(a) = 0$, mais comme mentionné plus haut, μ_a engendre l'idéal est polynômes annulateurs de a. Donc $R \in (\mu_a)$ et nous avons $\bar{R} = 0$ dans $\mathbb{K}[X]/(\mu_a)$.

□

Exemple 6.88.
Un fait connu est que $\frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$. Donc l'inverse de $\sqrt{2}$ s'exprime bien comme un polynôme en $\sqrt{2}$ à coefficients dans $\mathbb{Q}$, ce qui confirme le point (5) de la proposition 6.87. Du point de vue de Bézout, $\mu_{\sqrt{2}}(X) = X^2 - 2$, et nous cherchons des polynômes U et V tels que

$$U(X^2 - 2) + VX = 1. \tag{6.142}$$

cette égalité est réalisée par $U = -\frac{1}{2}$ et $V = \frac{1}{2}X$. Et effectivement $V(\sqrt{2})$ est bien l'inverse de $\sqrt{2}$:

$$V(\sqrt{(2)}) = \frac{1}{2}\sqrt{2}. \tag{6.143}$$

△

Lemme 6.89 ([105])**.**
Un nombre complexe algébrique dont tous les conjugués sont de module 1 *est une racine de l'unité.*

Proposition 6.90 ([104])**.**
Soient un corps $\mathbb{K}$, un extension $\mathbb{L}$ de $\mathbb{K}$ et un élément α de $\mathbb{L}$. Il y a équivalence entre les trois points suivants :

(1) α est algébrique sur $\mathbb{K}$,

(2) $\mathbb{K}[\alpha] = \mathbb{K}(\alpha)$,

(3) $\mathbb{K}[\alpha]$ est un $\mathbb{K}$-espace vectoriel de dimension finie.

Si ces affirmation sont vraies, alors $[\mathbb{K}(\alpha) : \mathbb{K}]$ est le degré du polynôme minimal de α sur $\mathbb{K}$.

Démonstration. Démonstration décomposée en plusieurs implications.

(1) implique (2) Soit α algébrique sur $\mathbb{K}$. Nous considérons le polynôme minimal de α sur $\mathbb{K}$ (définition 6.57). Nous savons par le lemme 6.73 (qui fonctionne parce que α est algébrique) que $\mathbb{K}[\alpha] = \mathbb{K}[X]/(\mu)$ en tant qu'anneaux.

Mais $\mathbb{K}[X]$ est un anneau principal et μ en est un élément irréductible. Donc la proposition 3.124 dit que (μ) est un idéal maximum ; la proposition 3.126 avance encore un peu en disant que $\mathbb{K}[X]/(\mu)$ est un corps.

Donc $\mathbb{K}[X]/(\mu)$ est un corps isomorphe à $\mathbb{K}[\alpha]$ en tant qu'anneaux. En conséquence de quoi $\mathbb{K}[\alpha]$ est un corps.

Le corps $\mathbb{K}[\alpha]$ est un sous-corps de $\mathbb{L}$ contenant $\mathbb{K}$ et α ; par définition nous avons donc $\mathbb{K}(\alpha) \subset \mathbb{K}[\alpha]$. Mais d'autre part, $\mathbb{K}[\alpha]$ est contenu dans tout sous-corps de $\mathbb{L}$ contenant $\mathbb{K}$ et α, donc il est inclus dans l'intersection de tout ces corps, donc $\mathbb{K}[\alpha] \subset \mathbb{K}(\alpha)$.

Les deux inclusions sont prouvées.

(2) implique (1) Nous montrons que non-(1) implique non-(2). Nous disons donc que α est transcendant sur $\mathbb{K}$; cela implique par la proposition 6.84(1) que $\mathbb{K}[\alpha] = \mathbb{K}[X]$ en tant qu'anneaux. Donc $\mathbb{K}[\alpha]$ n'est pas un corps parce que $\mathbb{K}[X]$ ne l'est pas.

N'étant pas un corps, $\mathbb{K}[\alpha]$ ne peut pas être égal à $\mathbb{K}(\alpha)$ qui, lui, est un corps.

(1) implique (3) L'élément α est maintenant algébrique et nous considérons son polynôme minimal μ. Nous savons par le lemme 6.73 que $\mathbb{K}[\alpha] = \mathbb{K}[X]/(\mu)$ en tant qu'espaces vectoriels. Or $\mathbb{K}[X]/(\mu)$ est de dimension finie $\deg(\mu)$. Donc $\mathbb{K}[\alpha]$ est également de dimension finie.

(3) implique (1) Nous démontrons la contraposée. En supposant que α est transcendant nous avons $\mathbb{K}[\alpha] = \mathbb{K}[X]$ par la proposition 6.84. Or $\mathbb{K}[X]$ n'est pas de dimension finie sur $\mathbb{K}$, donc $\mathbb{K}[\alpha]$ non plus.

□

Lemme 6.91 ([106])**.**
Soit $\mathbb{L}$ un corps commutatif et $(\mathbb{K}_i)_{i\in I}$ une famille de sous-corps de $\mathbb{L}$. Alors $\bigcup_{i\in I}\mathbb{K}_i$ est un sous-corps de $\mathbb{L}$.

Définition 6.92.
Soit une extension[37] *de corps $j\colon \mathbb{K}\to\mathbb{L}$. Soit $A\subset\mathbb{L}$.*

(1) Nous notons $\mathbb{K}(A)_{\mathbb{L}}$ le plus petit sous corps de $\mathbb{L}$ contenant $j(\mathbb{K})$ et A. C'est l'intersection de tous les sous-corps de $\mathbb{L}$ contenant A et $j(\mathbb{K})$.

(2) Nous notons $\mathbb{K}[A]_{\mathbb{L}}$ le plus petit sous anneau de $\mathbb{L}$ contenant $j(\mathbb{K})$ et A. C'est l'intersection de tous les sous-anneaux de $\mathbb{L}$ contenant A et $j(\mathbb{K})$.

Le plus souvent, l'indice $\mathbb{L}$ dans $\mathbb{K}(A)_{\mathbb{L}}$ et $\mathbb{K}[A]_{\mathbb{L}}$ est omis parce que le contexte est clair[38]*, et nous avons même très souvent $\mathbb{K}\subset\mathbb{L}$ en tant que ensembles. Dans ce cas, l'application j est l'identité et elle sera omise.*

*Nous disons que l'extension $\mathbb{L}$ de $\mathbb{K}$ est **monogène** ou **simple** s'il existe $\theta\in\mathbb{L}$ tel que $\mathbb{L}=\mathbb{K}(\theta)$. Un tel élément θ est dit **élément primitif** de $\mathbb{L}$. Il n'est pas nécessairement unique.*

Remarque 6.93.
Les ensembles $\mathbb{K}(A)$ et $\mathbb{K}[A]$ sont aussi appelés respectivement corps **engendré** et anneau engendré par A. Cependant il faut bien remarquer que ce sont les parties de $\mathbb{L}$ engendrées par A. Il n'est pas question à priori de parler de corps engendré par A sans dire dans quel corps plus grand nous nous plaçons.

Exemple 6.94.
Nous savons que $\mathbb{R}$ est une extension de $\mathbb{Q}$. Si $a\in\mathbb{R}$ alors $\mathbb{Q}(a)$ est le plus petit corps contenant $\mathbb{Q}$ et a. △

Exemple 6.95.
Nous avons déjà vu à l'occasion de la définition 3.86 que $A[X]$ est l'anneau de tous les polynômes de degré fini en X. Cela rentre dans le cadre de la définition 6.92 parce un anneau contenant X doit contenir tous les X^n.

Notons que même si $\mathbb{K}$ est un corps, $\mathbb{K}[X]$ reste un anneau parce qu'un éventuel inverse de X n'est pas dedans[39]. Par contre, $\mathbb{K}(X)$ est un corps parce qu'il contient également les fractions rationnelles. △

Exemple 6.96.
Si nous prenons $\mathbb{F}_5$ et que nous l'étendons par i, nous obtenons le corps $\mathbb{K}=\mathbb{F}_5(i)$. Nous savons que tous les éléments $a\in\mathbb{F}_5$ sont racines de X^5-X. Mais étant donné que $i^5=i$, nous avons aussi

37. Définition 6.52.
38. Et je me demande si il est possible de trouver un cas tordu où $\mathbb{K}(A)_{\mathbb{L}}\neq\mathbb{K}(A)_{\mathbb{M}}$. Par exemple lorsque A est dans $\mathbb{L}$ et $\mathbb{M}$, mais que $\mathbb{L}$ n'est pas inclus à $\mathbb{M}$, ni $\mathbb{M}$ dans $\mathbb{L}$.
39. Lorsqu'on multiplie, les degrés montent toujours.

$x^5 = x$ pour tout $x \in \mathbb{F}_5(i)$. Pour le prouver, utiliser le morphisme de Frobenius. Le polynôme $X^5 - X$ est donc le polynôme nul dans $\mathbb{K}$.

Ceci est un cas très particulier parce que nous avons étendu $\mathbb{F}_p$ par un élément α tel que $\alpha^p = \alpha$. En général sur $\mathbb{F}_p(\alpha)$, le polynôme $X^p - X$ n'est pas identiquement nul, et possède donc au maximum p racines. Pour $x \in \mathbb{F}_p(\alpha)$, nous avons $x^p = x$ si et seulement si $x \in \mathbb{F}_p$. △

Lemme 6.97.
Soit $P \in \mathbb{K}[X]$ un polynôme unitaire irréductible de degré n. Il existe une extension $\mathbb{L}$ de $\mathbb{K}$ et $a \in \mathbb{L}$ telle que $\mathbb{L} = \mathbb{K}(a)$ et P est le polynôme minimal de a dans $\mathbb{L}$.

Démonstration. Nous prenons $\mathbb{L} = \mathbb{K}[X]/(P)$ où (P) est l'idéal dans $\mathbb{K}[X]$ généré par P. Cela est un corps par le corolaire 6.38. Nous identifions $\mathbb{K}$ avec $\phi(\mathbb{K})$ où

$$\phi\colon \mathbb{K}[X] \to \mathbb{L} \tag{6.144}$$

est la projection canonique. Nous considérons également $a = \phi(X)$.

Nous avons alors $P(a) = 0$ dans $\mathbb{L}$. En effet $P(a) = P\big(\phi(X)\big)$ est à voir comme l'application du polynôme P au polynôme X, le résultat étant encore un élément de $\mathbb{L}$. En l'occurrence le résultat est P qui vaut 0 dans $\mathbb{L}$.

Le polynôme P étant unitaire et irréductible, il est minimum dans $\mathbb{L}$.

Nous devons encore montrer que $\mathbb{L} = \mathbb{K}(a)$. Le fait que $\mathbb{K}(a) \subset \mathbb{L}$ est une tautologie parce qu'on calcule $\mathbb{K}(a)$ dans $\mathbb{L}$. Pour l'inclusion inverse soit $Q(X) = \sum_i Q_i X^i$ dans $\mathbb{K}[X]$. Dans $\mathbb{L}$ nous avons évidemment $Q = \sum_i Q_i a^i$. □

Proposition 6.98 ([41]).
Soit $\mathbb{K}$, un corps et $P \in \mathbb{K}[X]$ un polynôme. Soient a et b, deux racines de P dans (éventuellement) une extension $\mathbb{L}$ de $\mathbb{K}$. Si μ_a et μ_b sont les polynômes minimaux de a et b (dans $\mathbb{K}[X]$) et si $\mu_a \neq \mu_b$, alors $\mu_a\mu_b$ divise P dans $\mathbb{K}[X]$.

Démonstration. Nous considérons les idéaux

$$I_a = \{Q \in \mathbb{K}[X] \text{ tel que } Q(a) = 0\}; \tag{6.145a}$$
$$I_b = \{Q \in \mathbb{K}[X] \text{ tel que } Q(b) = 0\}. \tag{6.145b}$$

Même si $Q(a)$ est calculé dans $\mathbb{L}$, ce sont des idéaux de $\mathbb{K}[X]$. Le polynôme μ_a est par définition le générateur unitaire de I_a, et vu que a est une racine de P, nous avons $P \in I_a$ et il existe un polynôme $Q \in \mathbb{K}[X]$ tel que

$$P = \mu_a Q. \tag{6.146}$$

Montrons que $\mu_a(b) \neq 0$. Pour cela, nous supposons que $\mu_a(b) = 0$, c'est-à-dire que $\mu_a \in I_b$. Il existe alors $R \in \mathbb{K}[X]$ tel que $\mu_a = \mu_b R$. Mais par la proposition 6.59, le polynôme μ_a est irréductible, donc soit μ_b soit R est inversible. Vu que les inversibles sont les éléments de $\mathbb{K}$ (polynômes de degré zéro), μ_b n'est pas inversible (sinon il serait constant et ne pourait pas être annulateur de b). Donc R est inversible. Disons $R = k$.

Donc $\mu_a = k\mu_b$. Mais vu que μ_a et μ_b sont unitaires, nous avons obligatoirement $k = 1$. Cela donnerait $\mu_a = \mu_b$, ce qui est contraire aux hypothèses. Nous en déduisons que $\mu_a(b) \neq 0$.

Étant donné que $\mu_a(b) \neq 0$, l'évaluation de (6.146) en b montre que $Q(b) = 0$, de telle sorte que $Q \in I_b$ et il existe un polynôme S tel que $Q = \mu_b S$, c'est-à-dire tel que $P = \mu_a\mu_b S$, ce qui signifie que $\mu_a\mu_b$ divise P. □

Exemple 6.99.
Soit $P = (X^2 + 1)(X^2 + 2)$ dans $\mathbb{R}[X]$. Dans $\mathbb{C}$ nous avons les racines $a = i$ et $b = \sqrt{2}i$ dont les polynômes minimaux sont $\mu_a = X^2 + 1$ et $\mu_2 = X^2 + 2$. Nous avons effectivement $\mu_a\mu_b$ divise P dans $\mathbb{R}[X]$.

Si par contre nous considérions les racines $a = i$ et $b = -i$, nous aurions $\mu_a = \mu_b = X^2 + 1$, tandis que le polynôme μ_a^2 ne divise pas P. △

6.4.5 Racines de polynômes

Corolaire 6.100 (Factorisation d'une racine)**.**
Soit $P \in \mathbb{K}[X]$, un polynôme de degré n et $\alpha \in \mathbb{K}$ tel que $P(\alpha) = 0$. Alors il existe un polynôme Q de degré $n-1$ tel que $P(x) = (X-\alpha)Q$.

Démonstration. Il s'agit d'un cas particulier de la proposition 6.85 : si $\alpha \in \mathbb{K}$ alors son polynôme minimal dans $\mathbb{K}$ est $X - \alpha$; donc $X - \alpha$ divise P. Il existe un polynôme Q tel que $P = (X-\alpha)Q$. Le degré est alors immédiat. □

Avant de lire l'énoncé suivant, allez relire la définition 3.169 pour savoir ce qu'est un polynôme nul.

Théorème 6.101 (Polynôme qui a tellement de racines qu'il s'annule)**.**
Soit $\mathbb{K}$ un corps et $P \in \mathbb{K}[X]$ un polynôme de degré n possédant $n+1$ racines distinctes $\alpha_1, \dots, \alpha_{n+1}$, alors $P = 0$.

Démonstration. Si P est de degré 1, il s'écrit $P = aX + b$; s'il a comme racines α et β, nous avons le système

$$\begin{cases} a\alpha + b = 0 & \text{(6.147a)} \\ a\beta + b = 0. & \text{(6.147b)} \end{cases}$$

La différence entre les deux donne $a(\alpha - \beta) = 0$. Vu que $\alpha \neq \beta$, la règle du produit nul (lemme 1.100) nous donne $a = 0$. Maintenant que $a = 0$, l'annulation de b est alors immédiate.

Nous faisons maintenant la récurrence en supposant le théorème vrai pour le degré n et en considérant un polynôme P de degré $n+1$ possédant $n+2$ racines distinctes. Vu que $P(\alpha_1) = 0$, le corolaire 6.100 nous donne un polynôme Q de degré n tel que

$$P = (X - \alpha_1)Q. \tag{6.148}$$

Étant donne que pour tout $i \neq 1$ nous avons $\alpha_i \neq \alpha_1$,

$$0 = P(\alpha_i) = \underbrace{(\alpha_i - \alpha_1)}_{\neq 0} Q(\alpha_i), \tag{6.149}$$

et la règle du produit nul donne $Q(\alpha_i) = 0$. Par conséquent le polynôme Q est de degré n et possède $n+1$ racines distinctes ; tous ses coefficients sont alors nuls par hypothèse de récurrence. Tous les coefficients du produit (6.148) sont alors également nuls. □

Exemple 6.102.
Un polynôme à plusieurs variables peut s'annuler en une infinité de points sans être nul. Par exemple le polynôme $X^2 + Y^2 - 1 \in \mathbb{R}[X, Y]$ s'annule sur tout un cercle de $\mathbb{R}^2$ mais n'est pas nul, loin s'en faut.

Nous verrons dans la proposition 6.170 une condition pour qu'un polynôme à plusieurs variables s'annule du fait qu'il ait « trop » de racines. △

Remarque 6.103.
L'intérêt du théorème 6.101 est que si l'on prouve qu'un polynôme s'annule sur un corps infini, alors il s'annulera sur n'importe quel autre corps. Nous aurons un exemple d'utilisation de cela dans le théorème de Cayley-Hamilton 12.23.

6.4.6 Corps de rupture

Définition 6.104.
Soit $P \in \mathbb{K}[X]$ un polynôme irréductible. Une extension $\mathbb{L}$ de $\mathbb{K}$ est un ***corps de rupture*** *pour P s'il existe $a \in \mathbb{L}$ tel que $P(a) = 0$ et $\mathbb{L} = \mathbb{K}(a)$.*

6.105.
Nous insistons sur le fait que nous ne définissons le concept de corps de rupture pour un polynôme irréductible à coefficients dans un corps. Les deux points sont importants : irréductible et à coefficient dans un corps.

Nous discuterons brièvement le pourquoi de cela dans la section 6.4.11 et surtout dans la question (2) des questions difficiles d'algèbre.

Exemple 6.106.
Soit $\mathbb{K} = \mathbb{Q}$ et $P = X^2 - 2$. On pose $a = \sqrt{2}$ et $\mathbb{L} = \mathbb{Q}(\sqrt{2}) \subset \mathbb{R}$. De cette façon P est scindé dans $\mathbb{L}$:

$$P = (X - \sqrt{2})(X + \sqrt{2}). \tag{6.150}$$

Le corps $\mathbb{Q}(\sqrt{2})$ est donc un corps de rupture pour P. △

Exemple 6.107.
Dans l'exemple 6.106, nous avions un corps de rupture dans lequel le polynôme P était scindé. Il n'en est pas toujours ainsi. Prenons

$$P = X^3 - 2 \tag{6.151}$$

et $a = \sqrt[3]{2}$. Nous avons, certes, $P(a) = 0$ dans $\mathbb{Q}(\sqrt[3]{2})$, mais P n'est pas scindé parce qu'il y a deux racines complexes. △

Exemple 6.108.
Nous considérons le corps $\mathbb{Z}/p\mathbb{Z}$ où p est un nombre premier. Si $s \in \mathbb{Z}/p\mathbb{Z}$ n'est pas un carré, alors le polynôme $P = X^2 + s$ est irréductible et un corps de rupture de P sur $\mathbb{Z}/p\mathbb{Z}$ est donné par $(\mathbb{Z}/p\mathbb{Z})[X]/(X^2 + s)$, c'est-à-dire l'ensemble des polynômes de degré 1 en $\sqrt{s}$. Le cardinal en est p^2. △

Vu que nous allons abondamment parler du quotient $\mathbb{K}[X]/(P)$, nous nous permettons un petit lemme.

Lemme 6.109.
Soit un corps $\mathbb{K}$ et $P \in \mathbb{K}[X]$ non constant. Alors $\mathbb{K}[X]/(P)$ est un corps si et seulement si P est irréductible.

Démonstration. Nous utilisons le trio d'enfer dont il est question dans le thème 46. D'abord $\mathbb{K}[X]$ est un anneau principal par le lemme 3.178. Donc $\mathbb{K}[X]/(P)$ sera un corps si et seulement si (P) est un idéal maximum (proposition 3.56), et cela sera le cas si et seulement si (P) est engendré par un polynôme irréductible (proposition 3.124).

Il ne nous reste qu'à montrer que (P) est engendré par un irréductible si et seulement si P est irréductible. Il y a un sens dans lequel c'est évident.

Soit un irréductible μ tel que $(P) = (\mu)$. En particulier $\mu \in (P)$, c'est-à-dire qu'il existe Q tel que $\mu = PQ$. Vu que μ est irréductible, soit P soit Q est inversible. Si P est inversible, c'est-à-dire constant, ce que nous avons exclu par hypothèse. Si par contre Q est inversible, alors $P = k\mu$ pour un certain $k \in \mathbb{K}$, ce qui montre que P est irréductible autant que μ. □

Proposition 6.110 (Existence d'un corps de rupture)**.**
Soit un corps $\mathbb{K}$ et un polynôme irréductible non constant P. Alors

(1) Le corps $\mathbb{L} = \mathbb{K}[X]/(P)$ est un corps de rupture pour P.

(2) L'élément $\bar{X}$ de $\mathbb{L}$ est une racine de P.

(3) $\mathbb{L} = \mathbb{K}(\bar{X})_{\mathbb{L}}$

Démonstration. Commençons par nous convaincre que $\mathbb{K}[X]/(P)$ est une extension de $\mathbb{K}$ (définition 6.52). Le fait que ce soit un corps est le lemme 6.109. Le morphisme $j\colon \mathbb{K} \to \mathbb{K}[X]/(P)$ est simplement $k \mapsto \bar{k}$ où à droite, $\bar{k}$ voit k dans $\mathbb{K}[X]$ comme étant le polynôme constant. Notez qu'il est automatiquement injectif (lemme 1.102).

Il faut maintenant voir que $\mathbb{K}[X]/(P) = \mathbb{K}(\alpha)$ pour un certain $\alpha \in \mathbb{K}[X]/(P)$. Grâce à notre compréhension des notations acquise dans 3.8.2.2, nous savons que $X \in \mathbb{K}[X]$ et qu'il est donc parfaitement légitime de poser $\alpha = \bar{X}$ dans $\mathbb{K}[X]/(P)$. Il s'agit simplement de l'ensemble $\bar{X} = \{X + QP \text{ tel que } Q \in \mathbb{K}[X]\}$ où X est une notation pour la suite $(0, 1, 0, 0, \ldots)$.

Bref, nous notons $\alpha = \bar{X}$ et nous démontrons que $P(\alpha) = 0$ et que $\mathbb{K}[X]/(P) = \mathbb{K}(\alpha)$ (isomorphisme de corps).

<u>$P(\bar{X}) = 0$</u> C'est le moment de nous souvenir comment la notation des X fonctionne, et en particulier la pirouette autour de (3.125). D'abord la définition du produit sur $\mathbb{K}[X]/(P)$ est $\bar{P}\bar{Q} = \overline{PQ}$; en particulier si $P = \sum_k a_k X^k$, alors $P(\bar{X}) = \sum_k a_k \bar{X}^k = \sum_k a_k \overline{X^k}$, et

$$P(\bar{X}) = \overline{P(X)} = \bar{P} = 0. \tag{6.152}$$

<u>**L'égalité**</u> Nous montrons à présent que $\mathbb{K}(\bar{X})_{\mathbb{L}} = \mathbb{L}$. C'est-à-dire que $\mathbb{L}$ est bien engendrée par $\mathbb{K}$ et un seul élément. D'abord, $\mathbb{L} = \mathbb{K}[X]/(P)$ contient bien évidemment $\mathbb{K}$ et $\bar{X}$. Ensuite nous devons prouver que tout sous-corps de $\mathbb{L}$ contenant $\mathbb{K}$ et $\bar{X}$ est en réalité $\mathbb{L}$ entier.

Soit $Q \in \mathbb{K}[X]$, et montrons que $\bar{Q}$ est dans tout sous-corps de $\mathbb{L}$ contenant $\mathbb{K}$ et $\bar{X}$.

Par le lemme 3.167 nous avons $\bar{Q} = Q(\bar{X})$. Et si un corps contient $\mathbb{K}$ et $\bar{X}$, il doit contenir tous les polynômes en $\bar{X}$ à coefficients dans $\mathbb{K}$. Donc un tel corps doit contenir $Q(\bar{X})$ et donc $\bar{Q}$.

□

Exemple 6.111.
Soit le polynôme $P = X^2 + 1 \in \mathbb{Z}[X]$. Dans le quotient $\mathbb{Z}[X]/(P)$ nous avons $\bar{X}^2 + 1 = 0$ et donc $\bar{X}^2 = -1$. C'est-à-dire que $\mathbb{Z}[X]/(P)$ contient un élément dont le carré est -1. Avouez que c'est bien ce à quoi nous nous attendions.

Notons que $-\bar{X}$ est également une racine de P dans $\mathbb{Z}[X]/(P)$.

En calculant dans les polynômes à coefficients dans $\mathbb{Z}(\bar{X})$ nous avons :

$$(X + \bar{X})(X - \bar{X}) = X^2 - \bar{X}^2 = X^2 + 1, \tag{6.153}$$

c'est-à-dire que P est bien factorisé, et que nous avons retrouvé la multiplication $x^2 + 1 = (x + i)(x - i)$. △

6.112.
Il n'y a évidemment pas unicité d'un corps de rupture pour un polynôme donné. Une raison est qu'un polynôme peut accepter plusieurs racines complètement indépendantes. Le corps étendu par l'une ou l'autre racine donne deux corps de rupture différents. Par exemple dans $\mathbb{Q}[X]$, le polynôme

$$P = X^4 - X^2 - 2 \tag{6.154}$$

a pour racines (dans $\mathbb{C}$) les nombres $\sqrt{2}$ et i. Donc on a deux corps de rupture complètement différents : $\mathbb{Q}(\sqrt{2})$ et $\mathbb{Q}(i)$.

6.113.
La proposition suivante donne une unicité du corps de rupture dans le cas d'un polynôme irréductible. Et nous comprenons pourquoi : un polynôme irréductible n'a fondamentalement qu'une seule racine « indépendante ». Par exemple $X^2 - 2$ a pour racines $\pm\sqrt{2}$. Autre exemple, le polynôme $X^2 + 6X + 13$ a pour racines, dans $\mathbb{C}$, les nombres complexes conjugués $z = -3 + 2i$ et $\bar{z} = -3 - 2i$.

Proposition 6.114 ([99]).
Soient un corps $\mathbb{K}$ et un polynôme irréductible $P \in \mathbb{K}[X]$. Alors toute extension $\mathbb{L}$ contenant une racine α de P admet un unique morphisme de corps

$$\psi\colon \mathbb{K}[X]/(P) \to \mathbb{L} \tag{6.155}$$

tel que $\psi(\bar{X}) = \alpha$.

Dans un tel cas,

(1) l'image de ψ *est* $\mathbb{K}(\alpha)_{\mathbb{L}}$ *,*

(2) si $\mathbb{L} = \mathbb{K}(\alpha)_{\mathbb{L}}$ *alors* ψ *est un isomorphisme.*

Démonstration. L'idéal annulateur de α parmi les polynôme de $\mathbb{K}[X]$ n'est pas réduit à $\{0\}$ parce qu'il contient P. Le lemme 6.57 s'applique donc et nous avons le polynôme minimal μ de α dans $\mathbb{K}[X]$. Il divise P qui est irréductible, donc

$$P = \lambda\mu \tag{6.156}$$

pour un certain $\lambda \in \mathbb{K}$.

Nous posons

$$\begin{aligned} \psi\colon \mathbb{K}[X]/(P) &\to \mathbb{L} \\ \bar{Q} &\mapsto Q(\alpha). \end{aligned} \tag{6.157}$$

Bien définie Si $\bar{Q}_1 = \bar{Q}_2$ alors il existe un $R \in \mathbb{K}[X]$ tel que $Q_1 = Q_2 + RP$. Mais alors $\psi(\bar{Q}_1) = Q_1(\alpha) = Q_2(\alpha) + R(\alpha)P(\alpha) = Q_2(\alpha)$.

Injective Si $\psi(\bar{Q}_1) = \psi(\bar{Q}_2)$ alors $Q_1 - Q_2 = R$ pour un certain $R \in \mathbb{K}[X]$ vérifiant $R(\alpha) = 0$. Nous avons alors un polynôme S tel que $R = S\mu = \lambda^{-1}SP$. Donc $\bar{R} = 0$ et donc $\bar{Q}_1 = \bar{Q}_2$.

Morphisme Laissé comme exercice ; la paresse de l'auteur de ces lignes attend vos contributions.

La condition Le morphisme ψ respecte de plus la condition

$$\psi(\bar{X}) = X(\alpha) = \alpha. \tag{6.158}$$

En ce qui concerne l'unicité, fixer $\psi(\bar{X})$ est suffisant pour fixer un morphisme. En effet si $\psi(\bar{X}) = \alpha$, alors

$$\psi(\bar{Q}) = \psi\Big(\sum_k a_k \bar{X}^k\Big) = \sum_k a_k \psi(\bar{X})^k = \sum_k a_k \alpha^k. \tag{6.159}$$

Pour le second point de l'énoncé, il faut remarquer que α est algébrique et non transcendant. Donc en utilisant les propositions 3.171 et 6.87(5) nous trouvons

$$\text{Image}(\psi) = \{Q(\alpha) \text{ tel que } Q \in \mathbb{K}[X]\} = \mathbb{K}[\alpha] = \mathbb{K}(\alpha). \tag{6.160}$$

Et finalement pour le dernier point, un morphisme de corps est toujours injectif. Si il est également surjectif, il sera bijectif. □

6.4.7 Pile d'extensions

Lemme 6.115 ([1])**.**
Soient un corps $\mathbb{K}$*, des extensions* $\mathbb{L}_1, \dots, \mathbb{L}_n$ *et des éléments* $\alpha_i \in \mathbb{L}_i$ *tels que*

$$\mathbb{L}_1 = \mathbb{K}(\alpha_1)_{\mathbb{L}_1} \tag{6.161}$$

et

$$\mathbb{L}_k = \mathbb{L}_{k-1}(\alpha_k)_{\mathbb{L}_k}. \tag{6.162}$$

Alors

$$\mathbb{L}_n = \mathbb{K}(\alpha_1, \dots, \alpha_n)_{\mathbb{L}_n}. \tag{6.163}$$

Démonstration. Nous démontrons par récurrence sur n. Le cas $n = 1$ est simplement l'hypothèse (6.161).

Supposons donc que le lemme soit correct pour n, et étudions le cas $n + 1$. Nous avons, par définition et par hypothèse de récurrence :

$$\mathbb{L}_{n+1} = \mathbb{L}_n(\alpha_{n+1})_{\mathbb{L}_{n+1}} = \Big(\mathbb{K}(\alpha_1, \dots, \alpha_n)_{\mathbb{L}_n}\Big)(\alpha_{n+1})_{\mathbb{L}_{n+1}}. \tag{6.164}$$

Notre tâche sera donc de montrer que

$$\Big(\mathbb{K}(\alpha_1,\dots,\alpha_n)_{\mathbb{L}_n}\Big)(\alpha_{n+1}) = \mathbb{K}(\alpha_1,\dots,\alpha_{n+1}) \tag{6.165}$$

où nous n'écrivons plus les indices $\mathbb{L}_{n+1}$ partout.

Le membre de gauche est un sous-corps de $\mathbb{L}_{n+1}$ contenant à la fois $\mathbb{K}$ et tous les α_i, si bien que

$$\mathbb{K}(\alpha_1,\dots,\alpha_{n+1}) \subset \big(\mathbb{K}(\alpha_1,\dots,\alpha_n)_{\mathbb{L}_n}\big)(\alpha_{n+1})_{\mathbb{L}_{n+1}}. \tag{6.166}$$

Il faut donc prouver l'inclusion inverse ; c'est-à-dire montrer que tout élément x du corps $\big(\mathbb{K}(\alpha_1,\dots,\alpha_n)_{\mathbb{L}_n}\big)(\alpha_{n+1})$ est forcément dans tout sous-corps de $\mathbb{L}_{n+1}$ contenant $\mathbb{K}$ et les α_i. Un tel élément x est, par la proposition 6.82(2), de la forme $r(\alpha_{n+1})$ avec $r \in \mathbb{K}(\alpha_1,\dots,\alpha_n)(X)$, c'est-à-dire

$$P(\alpha_{n+1})Q(\alpha_{n+1})^{-1} \tag{6.167}$$

avec $P, Q \in \mathbb{K}(\alpha_1,\dots,\alpha_n)[X]$.

Prouvons d'abord que si $P \in \mathbb{K}(\alpha_1,\dots,\alpha_n)[X]$, alors $P(\alpha_{n+1})$ est dans tout sous-corps de $\mathbb{L}_{n+1}$ contenant $\mathbb{K}$ et les α_i. Nous pouvons écrire $P = \sum_i a_i X^i$ avec $a_i \in \mathbb{K}(\alpha_1,\dots,\alpha_n)$, et donc

$$P(\alpha_{n+1}) = \sum_i a_i \alpha_{n+1}^i. \tag{6.168}$$

Tout corps contenant $\mathbb{K}$ et les $\alpha_1,\dots,\alpha_n$ contient les a_i. Par produit, tout corps contenant $\mathbb{K}$, $\alpha_1,\dots,\alpha_{n+1}$ contient les termes $a_i\alpha_{n+1}^i$, et donc $P(\alpha_{n+1})$ par somme.

De la même façon, si un corps contient $\mathbb{K}$ et les α_i $(i = 1,\dots,n+1)$, alors il contient $Q(\alpha_{n+1})$. Comme c'est un corps, il contient aussi son inverse $Q(\alpha_{n+1})^{-1}$, et il contient aussi le produit

$$r(\alpha_{n+1}) = P(\alpha_{n+1})Q(\alpha_{n+1})^{-1}. \tag{6.169}$$

On vient ainsi de montrer que tout élément $x \in \big(\mathbb{K}(\alpha_1,\dots,\alpha_n)_{\mathbb{L}_n}\big)(\alpha_{n+1})$ était dans tout sous-corps de $\mathbb{L}_{n+1}$ qui contient $\mathbb{K}$ et les α_i $(i = 1,\dots,n+1)$; en d'autres termes :

$$\big(\mathbb{K}(\alpha_1,\dots,\alpha_n)_{\mathbb{L}_n}\big)(\alpha_{n+1})_{\mathbb{L}_{n+1}} \subset \mathbb{K}(\alpha_1,\dots,\alpha_{n+1}). \tag{6.170}$$

Les inclusions (6.166) et (6.170) prouvent l'égalité d'ensembles (6.165) que nous voulions montrer. □

6.4.8 Clôture algébrique

Le concept de clôture algébrique a été défini dans 6.66. Voici un lemme qui dit qu'une clôture algébrique est en quelque sorte une extension algébrique maximale.

Lemme 6.116 ([1]).
Soient un corps $\mathbb{K}$ et une extension algébrique $\mathbb{F}$ de $\mathbb{K}$. Nous supposons que pour toute extension algébrique de $\mathbb{L}$ nous avons $\mathbb{L} = \mathbb{F}$

Alors $\mathbb{F}$ est algébriquement clos[40].

Démonstration. Soit un polynôme $P \in \mathbb{K}[X]$. Nous voudrions prouver que P a des racines dans $\mathbb{F}$. Pour cela, nous voyons P comme un polynôme sur $\mathbb{F}[X]$ et, grace à la proposition 6.110 nous considérons un corps de rupture $\mathbb{L}$ pour P. Vu que $\mathbb{L}$ est une extension de $\mathbb{F}$, nous avons $\mathbb{L} = \mathbb{F}$. Donc $\mathbb{F}$ contient des racines de P. □

6.117.
Nous avons défini le concept d'extension algébrique en 6.66. Nous allons en construire un petit exemple très piéton.

D'abord la proposition 1.166 nous donne l'existence et l'unicité d'un réel $\sqrt{2}$ strictement positif dont le carré est 2. Ce réel est irrationnel par la proposition 1.123. Cela étant posé, nous y allons.

40. Définition 6.66(2).

Proposition 6.118 ([107]).
Soit $\mathbb{L} = \{a + b\sqrt{2}\}_{a,b\in\mathbb{Q}}$.

(1) C'est un sous-corps de $\mathbb{R}$.

(2) Tout sous-corps de $\mathbb{R}$ *contenant* $\mathbb{Q}$ *et* $\sqrt{2}$ *contient* $\mathbb{L}$.

Démonstration. Nous devons d'abord prouver que $\mathbb{L}$ est un corps en vérifiant d'une part que c'est un anneau (définition 1.73) et d'autre part le fait que tous les éléments non nuls sont inversibles.

— La partie $\mathbb{L}$ de $\mathbb{R}$ est stable pour l'addition : dès que $a, b, a', b' \in \mathbb{Q}$,

$$(a + b\sqrt{2}) + (a'b'\sqrt{2}) = (a + a') + (b + b')\sqrt{2} \in \mathbb{L}. \tag{6.171}$$

— Les neutres 0 et 1 sont dans $\mathbb{L}$.

— Si $\alpha \in \mathbb{L}$, alors $-\alpha \in \mathbb{L}$:

$$-(a + b\sqrt{2}) = -a - b\sqrt{2}. \tag{6.172}$$

— La partie $\mathbb{L}$ est stable pour le produit parce que

$$(a + b\sqrt{2})(a' + b'\sqrt{2}) = (aa' + 2bb') + (ab' + ba')\sqrt{2}. \tag{6.173}$$

— L'inverse d'un élément de $\mathbb{L}$ est dans $\mathbb{L}$. C'est le seul point pas tout à fait évident. D'abord, l'ensemble $\mathbb{R}$ est un corps par le théorème 1.131. Donc pour tout $a, b \in \mathbb{R}$, le nombre

$$\frac{1}{a + b\sqrt{2}} \tag{6.174}$$

existe dans $\mathbb{R}$.
D'abord $a - b\sqrt{2}$ n'est pas nul, parce que si il l'était, nous aurions $\sqrt{2} = -a/b \in \mathbb{Q}$ alors que $\sqrt{2}$ n'est pas rationnel par la proposition 1.123. Nous pouvons donc faire le coup de multiplier le numérateur et le dénominateur par le binôme conjugué :

$$\frac{1}{a + b\sqrt{2}} = \frac{a - b\sqrt{2}}{(a + b\sqrt{2})(a - b\sqrt{2})} = \frac{a}{a^2 - 2b^2} - \frac{b}{a^2 - 2b^2}\sqrt{2}. \tag{6.175}$$

Cela est un rationnel. Donc le éléments non nuls de $\mathbb{L}$ ont un inverse qui appartient également à $\mathbb{L}$.

Nous passons à la preuve du point (2). Si $\mathbb{L}'$ est un corps qui contient $\mathbb{Q}$ et $\sqrt{2}$, il doit contenir $b\sqrt{2}$ pour tout $b \in \mathbb{Q}$ et donc aussi tous les $a + b\sqrt{2}$ avec $a, b \in \mathbb{Q}$. En conséquence de quoi $\mathbb{L}'$ doit contenir au moins tout $\mathbb{L}$. □

Proposition 6.119.
Soit $\mathbb{L} = \{a + b\sqrt{2}\}_{a,b\in\mathbb{Q}}$.

(1) C'est un espace vectoriel de dimension 2 *sur* $\mathbb{Q}$.

(2) Si $\alpha \in \mathbb{L}$, *alors il existe un polynôme* $P \in \mathbb{L}[X]$ *de degré* 2 *ou moins tel que* $P(\alpha) = 0$.

(3) Le corps $\mathbb{L}$ *est une extension algébrique de* $\mathbb{Q}$.

Démonstration. En plusieurs parties.

(1) Pour la dimension, notez que $\{1, \sqrt{2}\}$ est une partie libre et génératrice de $\mathbb{L}$.

(2) Soit $\alpha \in \mathbb{L}$. La partie $\{1, \alpha, \alpha^2\}$ est de cardinal 1, 2 ou 3. Si c'est 1 ou 2, c'est que $1 = \alpha$ ou $1 = \alpha^2$ ou $\alpha = \alpha^2$. Si par exemple $1 = \alpha$ alors avec $P = X - 1$ nous avons $P(\alpha) = 0$.
Si au contraire $\{1, \alpha, \alpha^2\}$ est de cardinal 3, alors c'est une partie liée par la proposition 4.6. Il existe donc des rationnels a, b, c tels que $a + b\alpha + c\alpha^2 = 0$, c'est-à-dire $P(\alpha) = 0$ avec $P = cX^2 + bX + a$.

(3) Nous venons de voir que tous les éléments de $\mathbb{L}$ sont des racines de polynômes de $\mathbb{Q}[X]$.

□

Lemme 6.120.
Si $\mathbb{K}$ est un corps infini, alors $\mathbb{K}[X]$ est équipotent[41] *à $\mathbb{K}$.*

Démonstration. Notons provisoirement $\mathbb{K}_n[X]$ l'ensemble des polynômes de degré n. Nous avons une surjection

$$\begin{aligned}\varphi\colon \mathbb{K}^{n+1} &\to \mathbb{K}_n[X]\\ (k_0,\dots,k_n) &\mapsto \sum_{i=0}^{n} k_i X^i.\end{aligned} \tag{6.176}$$

Par récurrence sur le théorème[42] 1.67, nous avons $\mathbb{K}^{n+1} \approx \mathbb{K}$. La surjection (6.176) dit alors que

$$\mathbb{K}_n[X] \preceq \mathbb{K}^{n+1} \approx \mathbb{K}. \tag{6.177}$$

Mais vu qu'il y a une surjection $\mathbb{K} \to \mathbb{K}_n[X]$, nous avons aussi $\mathbb{K}_n[X] \succeq \mathbb{K}$. Le théorème 1.51 dit alors que $\mathbb{K}_n[X] \approx \mathbb{K}$.

Le lemme 1.68 nous permet alors de conclure que

$$\mathbb{K}[X] = \bigcup_{n=0}^{\infty} \mathbb{K}_n[X] \approx \mathbb{K}. \tag{6.178}$$

□

Proposition 6.121 ([1])**.**
Soit un corps $\mathbb{K}$. Une extension algébrique de $\mathbb{K}$ est

(1) au plus dénombrable si $\mathbb{K}$ est fini,

(2) équipotente à $\mathbb{K}$ si $\mathbb{K}$ est infini.

Démonstration. En deux parties.

Si $\mathbb{K}$ est fini Un polynôme non nul possède toujours au maximum un nombre fini de racines (éventuellement zéro) par la proposition 6.101. Par ailleurs, chaque degré de polynôme ayant seulement un nombre fini de possibilités, l'ensemble $\mathbb{K}[X]$ est au plus dénombrable (proposition 1.44).

Pour $P \in \mathbb{K}[X]$ nous avons une surjection de $\mathbb{N}$ vers l'ensemble des racines de P. Nous la notons $\varphi_P\colon \mathbb{N} \to \mathbb{L}$, en posant par exemple $\varphi_P(n) = 1$ si P n'a pas de racines. Enfin nous posons

$$\begin{aligned}\varphi\colon \mathbb{K}[X]\times\mathbb{N} &\to \mathbb{L}\\ (P,n) &\mapsto \varphi_P(n).\end{aligned} \tag{6.179}$$

C'est la fonction qui à un polynôme P et un nombre n fait correspondre la n^{e}racine de P.

Vu que $\mathbb{L}$ est une extension algébrique, φ est surjective.

En termes de cardinalité, que $\mathbb{K}[X]$ soit fini ou dénombrable, dans les deux cas, $\mathbb{K}[X]\times\mathbb{N}$ est dénombrable (proposition 1.41). Il existe donc une surjection d'un ensemble dénombrable vers $\mathbb{L}$. Le lemme 1.43 conclu que $\mathbb{L}$ est fini ou dénombrable.

Si $\mathbb{K}$ est infini Nous procédons de la même manière, mais nous devons faire appel à des résultats plus technologiques pour maitriser la cardinalité. Nous considérons à nouveau l'application

$$\begin{aligned}\varphi\colon \mathbb{K}[X]\times\mathbb{N} &\to \mathbb{L}\\ (P,n) &\mapsto \varphi_P(n).\end{aligned} \tag{6.180}$$

Cette application est encore surjective : $\mathbb{L} \preceq \mathbb{K}[X]\times\mathbb{N}$. Le lemme 6.120 nous assure que $\mathbb{K}[X] \approx \mathbb{K}$ parce que $\mathbb{K}$ est infini. Ensuite la proposition 1.61 nous dit que $\mathbb{K}[X]\times\mathbb{N} \approx \mathbb{K}[X]$. Donc

$$\mathbb{K} \approx \mathbb{K}[X] \approx \mathbb{K}[X]\times\mathbb{N} \succeq \mathbb{F}. \tag{6.181}$$

41. Définition 1.27.

42. J'ai quand même du mal à croire qu'il faut vraiment le lemme de Zorn pour prouver que $\mathbb{K}[X]$ est équipotent à $\mathbb{K}$. Si vous connaissez un moyen plus direct, écrivez-moi.

Mais $\mathbb{F}$ est une extension de $\mathbb{K}$. Donc il existe une injection $\mathbb{K} \to \mathbb{F}$, c'est à dire $\mathbb{K} \preceq \mathbb{F}$.

Vu que $\mathbb{K} \preceq \mathbb{F} \preceq \mathbb{K}$, le théorème 1.51 implique que $\mathbb{K} \approx \mathbb{F}$.

□

Lemme 6.122 ([1]).
Soient des corps $\mathbb{K}$ et $\mathbb{L}$ ainsi qu'un homomorphisme de corps $\rho\colon \mathbb{K} \to \mathbb{L}$. Si $P \in \mathbb{K}[X]$ a une racine dans $\mathbb{K}$, alors le polynôme $\rho(P)$ a une racine dans $\mathbb{L}$.

Démonstration. Nous notons $P = \sum_k P_k X^k$. Si $a \in \mathbb{K}$ est une racine de P, alors $\sum_k P_k a^k = 0$. Nous appliquons ρ à cette égalité : $\sum_k \rho(P_k)\rho(a)^k = 0$, c'est à dire $\rho(P)\big(\rho(a)\big) = 0$. Donc $\rho(a) \in \mathbb{L}$ est une racine de $\rho(P)$. □

Lemme 6.123 ([108]).
Nous considérons un triple $(\mathbb{K}, \mathbb{L}, \mathbb{F})$ où

(1) $\mathbb{K}$, $\mathbb{L}$ et $\mathbb{F}$ sont des corps

(2) il existe $a \in \mathbb{L}$ algébrique sur $\mathbb{K}$ tel que $\mathbb{L} = \mathbb{K}(a)$ et un morphisme de corps $\alpha\colon \mathbb{K} \to \mathbb{L}$.

(3) $\mathbb{F}$ est une extension algébriquement close de $\mathbb{K}$: il existe un morphisme $\beta\colon \mathbb{K} \to \mathbb{F}$.

Alors il existe un morphisme de corps $\sigma\colon \mathbb{L} \to \mathbb{F}$ tel que $\sigma|_{\alpha(\mathbb{L})} = \beta$.

Note : en pratique, les corps $\mathbb{L}$ et $\mathbb{F}$ sont le plus souvent des sur-corps de $\mathbb{K}$, de telle sorte que les applications α et β sont l'identité. En particulier, la conclusion de ce lemme s'écrit le plus souvent $\sigma|_{\mathbb{K}} = \mathrm{Id}$. Il faut juste savoir que le Frido est un névrosé des notations précises.

Démonstration. Vu que $\mathbb{L}$ est monogène, si $\mu_a \in \mathbb{K}[X]$ est le polynôme minimal de $a \in \mathbb{L}$, alors les points (5) et (6) de la proposition 6.87 disent que $\mathbb{L} \simeq \mathbb{K}[a] \simeq \mathbb{K}[X]/(\mu_a)$.

Les coefficients de μ_a sont dans $\mathbb{K}$, donc nous pouvons voir $\mu_a \in \mathbb{F}[X]$. Plus précisément, si $\mu_a = \sum_k a_k X^k$, nous définissons

$$\mu'_a = \sum_k \beta(a_k) X^k \in \mathbb{F}[X]. \tag{6.182}$$

Comme $\mathbb{F}$ est algébriquement clos, le polynôme μ'_a possède une racine (au moins) $b \in \mathbb{F}$: $\mu'_a(b) = 0$.

Nous posons

$$\begin{aligned} \sigma'\colon \mathbb{K}[X]/(\mu_a) &\to \mathbb{F} \\ \overline{\sum_k s_k X^k} &\mapsto \sum_k \beta(s_k) b^k. \end{aligned} \tag{6.183}$$

Montrons que cela est bien défini. Si $P = \sum_k s_k X^k$ et $\mu_a = \sum_k a_k X^k$ $(a_k, s_k \in \mathbb{K})$, alors

$$\sigma'(\overline{P + \mu_a}) = \sigma'\big(\overline{\sum_k (a_k + s_k) X^k}\big) = \sum_k \beta(a_k) b^k + \sum_k s_k b^k = \mu'_a(b) + \sigma'(\overline{P}) = \sigma'(\overline{P}). \tag{6.184}$$

L'application σ' est un morphisme de corps.

Si $\varphi\colon \mathbb{L} \to \mathbb{K}[X]/(\mu_a)$ est l'isomorphisme dont nous parlions plus haut, il vérifie $\varphi\big(\alpha(k)\big) = \bar{k}$ (la classe du polynôme de degré zéro). En posant $\sigma = \sigma' \circ \varphi$ nous avons, pour tout $k \in \mathbb{K}$,

$$(\sigma' \varphi \alpha)(k) = \sigma'(\bar{k}) = \beta(k), \tag{6.185}$$

c'est ce qu'il fallait. □

Lemme 6.124 ([108]).
Soit un corps $\mathbb{K}$ muni de deux extensions $\alpha\colon \mathbb{K} \to \mathbb{L}$ et $\beta\colon \mathbb{K} \to \mathbb{F}$ où $\mathbb{L}$ est algébrique sur $\mathbb{K}$ et $\mathbb{F}$ est algébriquement clos.

Alors il existe un morphisme de corps $\sigma\colon \mathbb{L} \to \mathbb{F}$ tel que $\sigma \circ \alpha = \beta$.

Démonstration. Nous allons utiliser le lemme de Zorn 1.15 sur l'ensemble

$$\mathcal{A} = \left\{ (\mathbb{M}, \varphi) \text{ tel que } \begin{cases} \mathbb{M} \text{ est un sous-corps de } \mathbb{L} \\ \alpha(\mathbb{K}) \subset \mathbb{M} \\ \varphi \colon \mathbb{M} \to \mathbb{F} \text{ est une extension de corps} \\ \varphi \circ \alpha = \beta \end{cases} \right\}. \tag{6.186}$$

Nous ordonnons (partiellement) cet ensemble en disant que $(\mathbb{M}_1, \varphi_1) < (\mathbb{M}_2, \varphi_2)$ si $\mathbb{M}_1 \subset \mathbb{M}_2$ et $\varphi_2|_{\mathbb{M}_1} = \varphi_1$. Il se fait que $\mathcal{A}$ est un ensemble inductif et que nous pouvons donc lui appliquer le lemme de Zorn.

Soit un élément maximum $(\mathbb{M}, \varphi)$. Nous allons montrer que $\mathbb{M} = \mathbb{L}$.

Soit $l \in \mathbb{L}$. Vu que $\mathbb{L}$ est une extension algébrique de $\mathbb{K}$, il existe un polynôme P à coefficients dans $\alpha(\mathbb{K})$ tel que $P(l) = 0$. Mais vu que $\alpha(\mathbb{K}) \subset \mathbb{M}$, ce polynôme est également à coefficients dans $\mathbb{M}$. Donc l est un élément algébrique sur $\mathbb{M}$.

Nous pouvons donc considérer le triple $\big(\mathbb{M}, \mathbb{M}(l), \mathbb{F}\big)$ qui vérifie les hypothèses du lemme 6.123. Il existe donc un morphisme de corps $\sigma \colon \mathbb{M}(l) \to \mathbb{F}$ tel que $\sigma|_{\mathbb{M}} = \varphi$.

Nous avons

$$\sigma \circ \alpha = \sigma|_{\sigma(\mathbb{K})} \circ \alpha = \sigma|_{\mathbb{M}} \circ \alpha = \varphi \circ \alpha = \beta. \tag{6.187}$$

Donc l'élément $\big(\mathbb{M}(l), \sigma\big)$ majore $(\mathbb{M}, \varphi)$ dans $\mathcal{A}$.

Par maximalité, nous déduisons que $\mathbb{M} = \mathbb{L}$. Donc le morphisme $\varphi \colon \mathbb{L} \to \mathbb{F}$ vérifie $\varphi \circ \alpha = \beta$, ce qu'il nous fallait. □

Théorème 6.125 (Steinitz[109, 110]).
À propos de clôture algébrique.

(1) Tout corps possède une clôture algébrique.

(2) Si $\alpha_1 \colon \mathbb{K} \to \mathbb{F}_1$ et $\alpha_2 \colon \mathbb{K} \to \mathbb{F}_2$ sont deux clôtures algébriques du même corps $\mathbb{K}$, alors il existe un isomorphisme de corps $\varphi \colon \mathbb{F}_1 \to \mathbb{F}_2$ tel que $\varphi \circ \alpha_1 = \alpha_2$.

Démonstration. Nous commençons par l'existence, en plusieurs points.

Un ensemble Nous considérons un ensemble Ω qui contient $\mathbb{K}$, qui est strictement surpotent[43] à $\mathbb{K}$ et qui est infini non dénombrable si $\mathbb{K}$ est fini. Par exemple $\mathcal{P}(\mathbb{K}) \cup \mathbb{K}$ si $\mathbb{K}$ est infini et $\mathbb{R} \cup \mathbb{K}$ si $\mathbb{K}$ est fini (voir le théorème de Cantor 1.54).

L'ensemble pour Zorn Nous considérons l'ensemble des extensions algébriques de $\mathbb{K}$ contenues dans Ω, c'est à dire

$$\mathcal{A} = \left\{ (\mathbb{L}, +_{\mathbb{L}}, \times_{\mathbb{L}}) \text{ tel que } \begin{cases} \mathbb{K} \subset \mathbb{L} \subset \Omega \\ (\mathbb{L}, +_{\mathbb{L}}, \times_{\mathbb{L}}) \text{ est une extension algébrique de } (\mathbb{K}, +, \times). \end{cases} \right\} \tag{6.188}$$

Nous ordonnons $\mathcal{A}$ par l'inclusion : nous disons que

$$(\mathbb{L}_1, +_1, \times_1) < (\mathbb{L}_2, +_2, \times_2) \tag{6.189}$$

lorsque $(\mathbb{L}_2, +_2, \times_2)$ est un sur-coprs de $(\mathbb{L}_1, +_1, \times_1)$ (en particulier $\mathbb{L}_1 \subset \mathbb{L}_2$).

$\mathcal{A}$ est inductif Soit une partie $\mathcal{F} = \{(\mathbb{L}_i, +_i, \times_i)\}_{i \in I}$ de $\mathcal{A}$ que nous supposons être totalement ordonnées. Nous allons lui trouver un majorant dans $\mathcal{A}$. Nous posons $\mathbb{L} = \bigcup_{i \in I} \mathbb{L}_i$, et si $a \in \mathbb{L}_1$, $b \in \mathbb{L}_j$, alors nous définissons

$$\begin{cases} a +_{\mathbb{L}} b = a +_k b & \text{(6.190a)} \\ a \times_{\mathbb{L}} b = a \times_k b & \text{(6.190b)} \end{cases}$$

où $k \in I$ est sélectionné de telle façon à avoir $(\mathbb{L}_i, +_i, \times_i) < (\mathbb{L}_k, +_k, \times_k)$ et $(\mathbb{L}_j, +_j, \times_j) < (\mathbb{L}_k, +_k, \times_k)$. Vu que tous les corps L_i sont des sous-corps les uns des autres, cela est une bonne définition.

43. Définition 1.27.

Lemme de Zorn Nous utilisons le lemme de Zorn 1.15. Nous notons $(\mathbb{F}, +, \times)$ un élément maximal de $\mathcal{A}$. Vu que $\mathbb{K}$ en est un sous-corps, il n'y a pas d'ambiguïté de noter $+$ et $\times$ ses opérations.

Stratégie pour la suite Nous allons montrer que si $\mathbb{E}$ est une extension algébrique de $\mathbb{F}$, alors $\mathbb{E} = \mathbb{F}$ (le but est d'utiliser le lemme 6.116).

Un peu de cardinalité D'abord, vu que $\mathbb{F}$ est algébrique sur $\mathbb{K}$, l'ensemble $\mathbb{F}$ est équipotent à $\mathbb{K}$ si $\mathbb{K}$ est infini, et au plus dénombrable sur $\mathbb{K}$ est fini ; c'est la proposition 6.121. En bref :

— Si $\mathbb{K}$ est infini, $\mathbb{K} \approx \mathbb{F} \approx \mathbb{E} \prec \Omega$.

— Si $\mathbb{K}$ est fini, $\mathbb{K} \preceq \mathbb{F} \preceq \mathbb{E} \prec \Omega$ où $\mathbb{E}$ est au maximum dénombrable et Ω est indénombrable.

Dans tous les cas, Ω est strictement surpotent à $\mathbb{F}$, et le lemme 1.63 permet de dire

$$\mathbb{E}\backslash\mathbb{F} \preceq \mathbb{E} \prec \Omega \approx \Omega\backslash\mathbb{F}. \tag{6.191}$$

Quelques injections Il existe donc une injection $\varphi\colon \mathbb{E}\backslash\mathbb{F} \to \Omega\backslash\mathbb{F}$. Nous posons

$$\begin{aligned} f\colon \mathbb{E} &\to \Omega \\ x &\mapsto \begin{cases} x & \text{si } x \in \mathbb{F} \\ \varphi(x) & \text{si } x \notin \mathbb{F}. \end{cases} \end{aligned} \tag{6.192}$$

Nous montrons que f est injective. Soient $x, y \in \mathbb{E}$ tels que $f(x) = f(y)$. Si $x, y \in \mathbb{F}$, alors $x = f(x) = f(y) = y$. Si $\in \mathbb{F}$ et $y \notin \mathbb{F}$, alors $x = \varphi(y)$ alors que $x \in \mathbb{F}$ et $\varphi(y) \in \Omega\backslash F$; ce cas est impossible. Enfin si x et y sont hors de $\mathbb{F}$, alors $f(x) = \varphi(x)$ et $f(y) = \varphi(y)$; donc $\varphi(x) = \varphi(y)$ et $x = y$ par injectivité de φ.

Nous avons donc bien une injection $f\colon \mathbb{E} \to \Omega$.

La maximalité Nous pouvons mettre sur $f(\mathbb{E}) \subset \Omega$ la structure de corps venant de $\mathbb{E}$. Vu que $f(\mathbb{F}) = \mathbb{F}$, le corps $f(\mathbb{E})$ est une extension algébrique de $\mathbb{F}$. Par maximalité, $f(\mathbb{E}) = \mathbb{F}$.

Mais si $x \in \mathbb{E}\backslash\mathbb{F}$, alors $f(x) \in \Omega\backslash\mathbb{F}$. Donc en réalité nous avons aussi $\mathbb{E} \subset \mathbb{F}$.

Conclusion En conclusion $\mathbb{E} = \mathbb{F}$ et le lemme 6.116 termine en disant que $\mathbb{F}$ est une clôture algébrique de $\mathbb{K}$.

Nous passons à la partie « unicité » de la clôture algébrique. Étant donné que $\mathbb{F}_1$ est une extension algébrique de $\mathbb{K}$ et que $\mathbb{F}_2$ est algébriquement clos, le lemme 6.124 nous donne un morphisme de corps $\sigma\colon \mathbb{F}_1 \to \mathbb{F}_2$ tel que $\sigma \circ \alpha_1 = \alpha_2$.

Nous sommes donc dans la situation où $\sigma\colon \mathbb{F}_1 \to \mathbb{F}_2$ est une extension de corps où $\mathbb{F}_1$ est algébriquement clos et $\mathbb{F}_2$ est algébrique. Le lemme 6.70 conclu que $\sigma(\mathbb{F}_1) = \mathbb{F}_2$, c'est à dire que σ est surjectif. En tant que morphisme de corps, σ était déjà injective ; elle est donc bijective.

Donc $\sigma\colon \mathbb{F}_1 \to \mathbb{F}_2$ est un isomorphisme de corps vérifiant $\sigma \circ \alpha_1 = \alpha_2$. □

6.126.
Bien que $\mathbb{C}$ soit une extension algébriquement close de $\mathbb{Q}$, l'ensemble $\mathbb{C}$ n'est pas une clôture algébrique de $\mathbb{Q}$. C'est ce que nous montrons maintenant.

Lemme 6.127.
Le corps $\mathbb{C}$ n'est pas une clôture algébrique[44] *de $\mathbb{Q}$.*

Démonstration. Nous montrons qu'il existe des éléments de $\mathbb{C}$ qui ne sont pas des racines de polynômes à coefficients rationnels. L'ensemble $\mathbb{Q}$ est dénombrable par la proposition 1.118. L'ensemble des polynômes de degré n à coefficients dans $\mathbb{Q}$ est en bijection avec les n-uples de rationnels, c'est-à-dire avec $\mathbb{Q}^n$ qui est également dénombrable par la proposition 1.46. Enfin l'ensemble des polynômes à coefficients sur $\mathbb{Q}$ est l'union des polynômes de degré fixés, donc dénombrable par la proposition 1.44.

44. Clôture algébrique, définition 6.66.

Jusqu'ici nous avons prouvé que l'ensemble des polynômes à coefficients rationnels était dénombrable. Or chaque polynôme possède une quantité finie de racines par le corolaire 3.195. Donc la partie de $\mathbb{C}$ constituée des nombres qui sont des racines de polynômes à coefficients dans $\mathbb{Q}$ est dénombrable. Mais $\mathbb{C}$ n'est pas dénombrable, donc possède des éléments qui ne sont pas des racines de polynômes. □

6.4.9 Polynômes à plusieurs variables

Nous avons déjà vu $A[X, Y]$ lorsque A est un anneau en la définition 3.91.

Définition 6.128.
Soit un corps $\mathbb{K}$. Le corps $\mathbb{K}(X_1, \dots, X_n)$ est le corps des fractions de l'anneau $\mathbb{K}[X_1, \dots, X_n]$.

Définition 6.129.
Soient un corps $\mathbb{K}$ et une extension $\mathbb{L}$ de $\mathbb{K}$ contenant les éléments $\alpha_1, \dots, \alpha_n$ de $\mathbb{K}$. Nous définissons $\mathbb{K}(\alpha_1, \dots, \alpha_n)$ comme étant l'intersection de tous les sous-corps de $\mathbb{L}$ contenant $\mathbb{K}$ et les α_i.

La proposition suivante est analogue à 6.82(2).

Lemme 6.130 ([1])**.**
Soient un corps $\mathbb{K}$, une extension $\mathbb{L}$ et des éléments $\alpha_1, \dots, \alpha_n$ dans $\mathbb{L}$. Alors

$$\mathbb{K}(\alpha_1, \dots, \alpha_n) = \{r(\alpha_1, \dots, \alpha_n) \text{ tel que } r \in \mathbb{K}(X_1, \dots, X_n)\}. \tag{6.193}$$

Démonstration. Ce que nous avons à droite est un corps : par exemple pour l'inverse, si $r = P/Q$ alors $r(\alpha_1, \dots, \alpha_n) = P(\alpha_1, \dots, \alpha_n)Q(\alpha_1, \dots, \alpha_n)^{-1}$. Cet élément a un inverse en la personne de $(Q/P)(\alpha_1, \dots, \alpha_n)$.

Donc à droite nous avons un sous-corps de $\mathbb{L}$ contenant $\mathbb{K}$ ainsi que les α_i. Donc

$$\mathbb{K}(\alpha_1, \dots, \alpha_n) \subset \big\{r(\alpha_1, \dots, \alpha_n) \text{ tel que } r \in \mathbb{K}(X_1, \dots, X_n)\big\}. \tag{6.194}$$

D'autre part, tout corps contenant $\mathbb{K}$ et les α_i doit contenir tous les $P(\alpha_1, \dots, \alpha_n)$ ($P \in \mathbb{K}[X_1, \dots, X_n]$), leurs inverses ainsi que leurs produits ; bref doit contenir tous les $r(\alpha_1, \dots, \alpha_n)$ avec $r \in \mathbb{K}[X_1, \dots, X_n]$. □

6.4.10 Corps de décomposition

Définition 6.131.
*Soit $\mathbb{K}$ un corps commutatif et $F = (P_i)_{i \in I}$ une famille d'éléments non constants de $\mathbb{K}[X]$. Un **corps de décomposition** de F est une extension $\mathbb{L}$ de $\mathbb{K}$ telle que*

(1) les P_i sont scindés sur $\mathbb{L}$,

(2) $\mathbb{L} = \mathbb{K}(R)$ où $R = \bigcup_{i \in I} \{x \in \mathbb{L} \text{ tel que } P_i(x) = 0\}$.

C'est-à-dire que $\mathbb{L}$ étend $\mathbb{K}$ par toutes les racines de tous les polynômes de F.

Proposition 6.132 ([111])**.**
Tout polynôme admet un corps de décomposition. Plus précisément, soit un corps $\mathbb{K}$ et un polynôme $P \in \mathbb{K}[X]$ de degré n. Il existe un corps de décomposition $\mathbb{D}$ de la forme $\mathbb{D} = \mathbb{K}(\alpha_1, \dots, \alpha_n)$ où les α_i sont des racines de P dans $\mathbb{D}$.

Notons que rien dans l'énoncé ne prétend que les α_i soient tous distincts, ni même que certains (voire tous) ne seraient pas dans $\mathbb{K}$.

Démonstration. Soient un corps $\mathbb{K}$ et un polynôme $P \in \mathbb{K}[X]$. Si le degré de P est 0 ou 1, alors $\mathbb{K}$ est un corps de décomposition pour P. Pour le reste nous faisons une récurrence sur le degré de P.

Il y a deux possibilités, soit il existe $\alpha \in \mathbb{K}$ tel que $P(\alpha) = 0$, soit non.

Si racine dans $\mathbb{K}$ Alors le corolaire 6.100 nous permet de factoriser $X - \alpha$:

$$P = (X - \alpha)Q \tag{6.195}$$

avec $\deg(Q) = \deg(P) - 1$. Dans ce cas, $\mathbb{K}$ est un corps de rupture de P.

Si pas de racines dans $\mathbb{K}$ Nous prenons alors un corps de rupture $\mathbb{L} = \mathbb{K}(\alpha)$ avec $P(\alpha) = 0$ (c'est la proposition 6.110 qui donne l'existence d'un corps de rupture). Dans $\mathbb{L}_1$ nous avons

$$P = (X - \alpha)Q \tag{6.196}$$

avec $Q \in \mathbb{L}_1[X]$ et $\deg(Q) = \deg(P) - 1$.

Dans les deux cas Dans les deux cas, par hypothèse de récurrence nous avons un corps de décomposition pour Q qui se présente sous la forme

$$\mathbb{L} = \mathbb{K}(\alpha_1, \dots, \alpha_{n-1}). \tag{6.197}$$

De plus, $\mathbb{L}$ est une extension de $\mathbb{L}_1$ parce que c'est une extension du corps sur lequel est Q.

Pour terminer la preuve nous prouvons que

$$\mathbb{D} = \mathbb{K}(\alpha_1, \dots, \alpha_{n-1}, \alpha) \tag{6.198}$$

est un corps de décomposition de P. Vu que $\mathbb{D}$ contient $\mathbb{K}(\alpha)$ (comme cas particulier du lemme 6.130), dans $\mathbb{D}$ nous avons l'égalité $P = (X-\alpha)Q$. Et vu que $\mathbb{D}$ contient également $\mathbb{K}(\alpha_1, \dots, \alpha_{n-1})$, toujours dans $\mathbb{D}$ nous avons aussi

$$Q = (X - \alpha_1) \dots (X - \alpha_{n-1}). \tag{6.199}$$

Donc nous avons dans $\mathbb{D}$ l'égalité

$$P = (X - \alpha)(X - \alpha_1) \dots (X - \alpha_{n-1}). \tag{6.200}$$

□

Lemme 6.133 ([1]).
Soit un polynôme $P \in \mathbb{K}[X]$, et un corps $\mathbb{L}$ dans lequel P est scindé. Si $P = P_1 \dots P_r$ est la décomposition de P en irréductibles dans $\mathbb{K}$, alors chacun des P_i est scindé dans $\mathbb{L}$.

Démonstration. Juste pour le mentionner, le fait que P ait une décomposition en irréductibles est le fait que $\mathbb{K}[X]$ soit factoriel, c'est-à-dire la proposition 6.28.

Le polynôme P est scindé dans $\mathbb{L}$, c'est-à-dire que, en notant n le degré de P,

$$P = \prod_{i=1}^{n} (X - \lambda_i) \tag{6.201}$$

avec $\lambda_i \in \mathbb{L}$.

Soit $\mathbb{L}_1$, une extension de $\mathbb{L}$ dans laquelle P_1 est scindé. Ensuite, $\mathbb{L}_2$ une extension de $\mathbb{L}_1$ dans laquelle P_2 est scindé et ainsi de suite. Nous avons construit $\mathbb{L}_r$, une extension de $\mathbb{L}$ dans laquelle tous les P_i sont scindés ainsi que P lui-même. Dans ce corps nous avons l'égalité

$$P = \prod_{k=1}^{n} (X - \mu_k) \tag{6.202}$$

où les μ_k sont des éléments des diverses extensions $\mathbb{L}_i$, et sont les racines des P_i. En tout cas, tous sont dans $\mathbb{L}_r$.

Les deux décompositions (6.201) et (6.202) sont des décompositions dans $\mathbb{L}_r[X]$ du polynôme P. Vu que ce dernier est factoriel, en réalité les deux décompositions sont identiques (se souvenir de la définition 3.114), et nous avons $\mu_k \in \mathbb{L}$ pour tout k. Toutes les extensions $\mathbb{L}_i$ sont en réalité triviales, et nous avons $\mathbb{L}_r = \mathbb{L}$.

Bref, tout cela pour dire que les P_i ont toutes leurs racines dans $\mathbb{L}$. □

Théorème 6.134 ([99]).
Soient :

— *un isomorphisme de corps* $\tau\colon \mathbb{K} \to \mathbb{K}'$ *;*

— *un polynôme non constant* $P \in \mathbb{K}[X]$ *de degré* n *;*

— *un corps de décomposition* $\mathbb{L}$ *de* P *sur* $\mathbb{K}$ *;*

— *un corps de décomposition* $\mathbb{L}'$ *de* P *sur* $\mathbb{K}'$ *;*

Alors τ *se prolonge en un isomorphisme* $\sigma\colon \mathbb{L} \to \mathbb{L}'$.

Démonstration. Soit m le nombre de racines de P dans $\mathbb{L}\backslash\mathbb{K}$. Nous faisons une récurrence sur m.

Si $m = 0$ alors $\mathbb{K}$ est un corps de rupture de P ; nous avons

$$P = (X - \lambda_1)\dots(X - \lambda_n) \tag{6.203}$$

avec $\lambda_i \in \mathbb{K}$. Dans ce cas nous avons aussi

$$\tau(P) = \big(X - \tau(\lambda_1)\big)\dots\big(X - \tau(\lambda_n)\big) \tag{6.204}$$

avec $\tau(\lambda_i) \in \mathbb{K}'$. Nous avons donc $\mathbb{L}' = \mathbb{K}'$ et prendre $\sigma = \tau$ fonctionne.

Nous supposons à présent que $m > 0$. Plus précisément nous considérons un polynôme possédant exactement m racines dans $\mathbb{L}\backslash\mathbb{K}$. Soit

$$P = P_1 \dots P_r \tag{6.205}$$

sa décomposition en irréductibles dans $\mathbb{K}[X]$ (notons que $r \leqslant n$ parce que chacun des facteurs est de degré au moins 1). Au moins un des P_i est de degré plus grand ou égal à 2. Nous savons de la proposition 6.28 que $\mathbb{K}[X]$ est un anneau factoriel. Le lemme 6.133 nous assure que les polynômes P_i sont également scindés dans $\mathbb{L}$. Et l'unicité de la décomposition fait en sorte que les racines des P_i sont celles de P.

Soit $\alpha \in \mathbb{L}$, une racine de P_1. Vu que P_1 est irréductible sur $\mathbb{K}$, l'application suivante est un isomorphisme de corps par le lemme 6.73 :

$$\begin{aligned} \psi\colon \mathbb{K}[X]/(P_1) &\to \mathbb{K}[\alpha] \\ \bar{Q} &\mapsto Q(\alpha). \end{aligned} \tag{6.206}$$

Notons que le lemme parle du quotient par le polynôme minimal, mais ici nous avons un irréductible. Un polynôme annulateur irréductible est multiple du polynôme minimal, et l'idéal engendré est le même.

Nous avons aussi la décomposition

$$\tau(P) = \tau(P_1)\dots\tau(P_r), \tag{6.207}$$

et chacun des $\tau(P_i)$ a ses racines dans $\mathbb{L}'$. Soit β, une racine de $\tau(P_1)$ dans $\mathbb{L}'$. Alors nous avons l'isomorphisme

$$\psi'\colon \mathbb{K}'[X]/\big(\tau(P_1)\big) \to \mathbb{K}'[\beta]. \tag{6.208}$$

De plus, par le lemme 6.40, nous savons que τ passe aux classes :

$$\phi_\tau\colon \mathbb{K}[X]/(P_1) \to \mathbb{K}'[X]/\big(\tau(P_1)\big) \tag{6.209}$$

est un isomorphisme d'anneaux. Et enfin, dernier résultat externe à invoquer, la proposition 6.87 nous assure que $\mathbb{K}[\alpha] = \mathbb{K}(\alpha)$ et $\mathbb{K}'[\beta] = \mathbb{K}'(\beta)$. Posons pour l'occasion $\mathbb{K}_1 = \mathbb{K}(\alpha)$ et $\mathbb{K}'_1 = \mathbb{K}'(\beta)$.

Nous avons l'enchainement suivant d'isomorphismes de corps [45] :

$$\tau_1 = \psi' \circ \phi_\tau \circ \psi^{-1}\colon \mathbb{K}_1 \to \mathbb{K}[X]/(P_1) \to \mathbb{K}'[X]/\big(\tau(P_1)\big) \to \mathbb{K}'_1. \tag{6.210}$$

45. En réalité il est plus exact de dire « isomorphisme d'anneaux », parce que la structure de corps n'est en réalité aucune nouvelle structure par rapport à l'anneau.

Cet isomorphisme $\tau_1 \colon \mathbb{K}_1 \to \mathbb{K}_2$ prolonge τ. Si vous aimez les diagrammes, en voici un sur lequel les i représentent des inclusions et où τ et τ_1 sont des isomorphismes

$$\begin{tikzcd} \mathbb{K} \arrow[r, "i"] \arrow[d, "\tau"'] & \mathbb{K}_1 \arrow[r, "i"] \arrow[d, "\tau_1"] & \mathbb{L} \\ \mathbb{K}' \arrow[r, "i"'] & \mathbb{K}'_1 \arrow[r, "i"] & \mathbb{L}' \end{tikzcd} \tag{6.211}$$

Le corps $\mathbb{L}$ est un corps de décomposition de P sur $\mathbb{K}_1$, et le nombre de racines de P dans $\mathbb{L}\backslash\mathbb{K}_1$ est strictement plus petit que m parce qu'il y en a exactement m dans $\mathbb{L}\backslash\mathbb{K}$ et que $\mathbb{K}_1$ en a au moins une de plus que $\mathbb{K}$. Même raisonnement pour $\mathbb{K}'$, $\mathbb{K}'_1$ et $\mathbb{L}'$.

Résumons la situation :

— $\tau_1 \colon \mathbb{K}_1 \to \mathbb{K}'_1$ est un isomorphisme de corps ;
— $P \in \mathbb{K}_1[X]$ est un polynôme non constant ;
— $\mathbb{L}$ est un corps de décomposition de P sur $\mathbb{K}_1$;
— $\mathbb{L}'$ est un corps de décomposition de P sur $\mathbb{K}'_1$;
— le nombre de racines de P dans $\mathbb{L}\backslash\mathbb{K}_1$ est strictement inférieur à m.

Donc, par hypothèse de récurrence sur m, il existe un isomorphisme $\sigma \colon \mathbb{L} \to \mathbb{L}'$ qui prolonge τ_1. Vu que τ_1 prolonge τ, nous avons également σ qui prolonge τ. □

L'énoncé le plus compact pour l'unicité du corps de décomposition (à isomorphisme près) est le suivant :

Proposition 6.135.
Soit $\mathbb{K}$ un corps et $P \in \mathbb{K}[X]$. Soient $\mathbb{L}$ et $\mathbb{F}$ deux corps de décomposition de P. Alors il existe un isomorphisme $f \colon \mathbb{L} \to \mathbb{F}$ tel que $f|_{\mathbb{K}} = \mathrm{Id}$.

Démonstration. C'est un cas particulier du théorème 6.134, où nous considérons $\mathbb{K} = \mathbb{K}'$ muni de l'isomorphisme identité. □

Cependant le passage par l'énoncé plus compliqué 6.134 est nécessaire pour les besoins de la récurrence.

6.136.
À propos de terminologie. Lorsque nous disons « *un* corps de décomposition » nous référons à la définition 6.131 et il n'y a pas vraiment unicité. Si nous disons « *le* corps de décomposition » nous référons en général à celui construit dans la proposition 6.132 qui n'est en réalité même pas très explicite.

Quoi qu'il en soit, nous nous permettons de dire « le » corps de décomposition lorsque nous parlons de propriétés invariantes par isomorphisme.

6.137.
La construction du corps de décomposition d'un polynôme fonctionne en prenant successivement le corps de rupture des facteurs irréductibles. Nous insistons sur le fait que cette opération se fait sur chaque facteur irréductible séparément.

L'exemple suivant montre dans quel ordre se passent les choses.

Exemple 6.138.
Soit le polynôme $P = X^4 - 5X^2 + 6$. Sa factorisation en irréductibles est :

$$P = (X^2 - 2)(X^2 - 3). \tag{6.212}$$

Ce polynôme n'est pas irréductible sur $\mathbb{Q}$ et il ne s'agit donc pas de prendre brutalement un corps de rupture pour P. Il s'agit de poser $P = P_1P_2$ avec

$$P_1 = X^2 - 2 \tag{6.213a}$$
$$P_2 = X^2 - 3, \tag{6.213b}$$

de remarquer que P_1 et P_2 sont irréductibles sur $\mathbb{Q}$ et de chercher des corps de rupture pour eux. On commence par P_1. Nous avons le corps de rupture $\mathbb{L}_1 = \mathbb{Q}(\sqrt{2})$ et la factorisation

$$P_1 = (X + \sqrt{2})(X - \sqrt{2}). \tag{6.214}$$

Ensuite nous considérons P_2 dans $\mathbb{L}_1[X]$. Ce P_2 est encore irréductible. Nous lui cherchons un corps de rupture, et c'est $\mathbb{L}_2 = \mathbb{L}_1(\sqrt{3})$ dans lequel nous avons

$$P_2 = (X - \sqrt{3})(X + \sqrt{3}). \tag{6.215}$$

Nous savons (par le lemme 6.115) que

$$\mathbb{L}_2 = \mathbb{L}_1(\sqrt{3}) = \big(\mathbb{Q}(\sqrt{2})\big)(\sqrt{3}) = \mathbb{Q}(\sqrt{2}, \sqrt{3}). \tag{6.216}$$

Nous pouvons donc écrire en toute confiance, dans $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ la factorisation

$$P = (X + \sqrt{2})(X - \sqrt{2})(X + \sqrt{3})(X - \sqrt{3}). \tag{6.217}$$

Et nous notons que si nous avions commencé par P_2 au lieu de P_1, nous aurions eu le même résultat final. △

Corolaire 6.139 ([1]).
Le corps de décomposition d'un polynôme est une extension finie.

Démonstration. Soient un corps $\mathbb{K}$, un polynôme $P \in \mathbb{K}[X]$ et un corps de décomposition $\mathbb{D}$ de P de la forme $\mathbb{D} = \mathbb{K}(\alpha_1, \dots, \alpha_n)$ où les α_i sont les racines de P dans $\mathbb{D}$. Cela existe par la proposition 6.132.

Vu que le lemme 6.115 donne

$$\mathbb{K}(\alpha_1, \dots, \alpha_n) = \big(\mathbb{K}(\alpha_1, \dots, \alpha_{n-1})\big)(\alpha_n), \tag{6.218}$$

le corps $\mathbb{D}$ se construit comme une pile d'extensions finies. Les degrés se composant par le lemme 6.56, au final ce corps de décomposition est une extension finie.

Soit maintenant un corps de décomposition quelconque $\mathbb{L}$. La proposition 6.135 donne un isomorphisme de corps[46] $f\colon \mathbb{L} \to \mathbb{D}$ tel que f soit l'identité sur $\mathbb{K}$.

Si $\{v_i\}_{i\in I}$ est une base de $\mathbb{D}$ comme espace vectoriel sur $\mathbb{K}$, êtes-vous prêts à parier que $\{f(v_i)\}_{i\in I}$ est une base de $\mathbb{L}$ comme espace vectoriel sur $\mathbb{K}$[47] ? □

6.4.11 Non irréductible ou pas corps ?

Nous avons déjà mentionné que nous ne définissons le corps de rupture d'un polynôme que dans le cas de polynôme irréductible à coefficients dans un corps.

D'abord si P n'est pas irréductible, la question de chercher un corps de rupture n'a pas beaucoup de sens.

Si A est un anneau intègre et si P est un polynôme irréductible sur A, nous pouvons considérer le corps des fractions de A, dire $P \in \mathrm{Frac}(A)[X]$ et continuer. Étendre la définition de corps de rupture de cette façon aux polynômes à coefficients dans un anneau intègre n'est pas une grande révolution.

Au lieu de cela, nous pouvons nous demander dans quel cas nous aurions que $A[X]/(P)$ est directement un corps.

Exemple 6.140.
Soit le polynôme constant $P = 2$ dans $\mathbb{Z}[X]$. Il y est irréductible parce qu'il ne peut pas être écrit comme produit de deux non inversibles. Ce polynôme a deux propriétés ennuyeuses :

— Il n'est plus irréductible sur $\mathbb{Q}$,

— Il n'existe aucun corps contenant une racine de P tout en contenant $\mathbb{Z}$ comme sous-anneau.

△

46. Un isomorphisme de corps est juste un isomorphisme d'anneaux.
47. Personnellement, je n'ai pas vérifié. Vérifiez vous-même et écrivez-moi pour dire si c'est bon ou non.

6.4.12 Clôture algébrique

Théorème 6.141.
Tout corps $\mathbb{K}$ possède une clôture algébrique[48] *Ω. De plus si $\mathbb{L}$ est une extension de $\mathbb{K}$, alors $\mathbb{L}$ est $\mathbb{K}$-isomorphe à un sous corps de Ω.*

Les deux parties de ce théorème utilisent l'axiome du choix.

Notons en particulier que si Ω' est une autre clôture algébrique de $\mathbb{K}$, alors Ω et Ω' sont des sous corps l'un de l'autre et sont donc $\mathbb{K}$-isomorphes.

Lemme 6.142.
Les polynômes $P, Q \in \mathbb{K}[X]$ ne sont pas premiers entre eux si et seulement s'ils ont une racine commune dans la clôture algébrique Ω de $\mathbb{K}$.

Démonstration. Soit A un polynôme non inversible divisant P et Q. Par définition de Ω, ce polynôme A a une racine dans Ω qui est alors une racine commune à P et Q dans Ω.

Pour le sens inverse, si α est une racine commune de P et Q, alors le polynôme $X - \alpha$ divise P et Q et donc P et Q ne sont pas premiers entre eux. □

Exemple 6.143.
Soit p un nombre premier. Montrons que le polynôme

$$Q(X) = X^p - X + 1 \tag{6.219}$$

est irréductible dans $\mathbb{F}_p$.

Nous supposons qu'il n'est pas irréductible, c'est-à-dire que

$$Q(X) = R(X)S(X) \tag{6.220}$$

avec R et S, des polynômes de degrés $\geqslant 1$ dans $\mathbb{F}_p[X]$

Soit $\bar{\mathbb{F}}_p$ une clôture algébrique[49] de $\mathbb{F}_p$ et $\alpha \in \bar{\mathbb{F}}_p$ tel que $R(\alpha) = 0$. Pour tout $a \in \mathbb{F}_p$, nous avons

$$\begin{aligned}
Q(\alpha + a) &= (\alpha + a)^p - (\alpha + a) + 1 && (6.221a)\\
&= \alpha^p + a^p - \alpha - a + 1 && (6.221b)\\
&= \alpha^p - \alpha + 1 && (6.221c)\\
&= Q(\alpha) && (6.221d)\\
&= 0 && (6.221e)
\end{aligned}$$

où nous avons utilisé le fait que $a^p = a$ et que α était une racine de Q. Ce que nous venons de prouver est que l'ensemble des racines de Q dans $\bar{\mathbb{F}}_p$ est donné par $\{\alpha + a \text{ tel que } a \in \mathbb{F}_p\}$.

Les polynômes R et S sont donc formés de produits de termes $X - (\alpha + a)$ avec $a \in \mathbb{F}_p$. L'un des deux –disons R pour fixer les idées– doit bien en avoir plus que 1. Nous avons alors

$$R(X) = \prod_{i=1}^{k} \big(X - (\alpha + a_i)\big) \tag{6.222}$$

où les a_i sont les éléments de $\mathbb{F}_p$. En développant un peu,

$$R(X) = X^k - \sum_{i=1}^{k} (\alpha + a_i^{k-1}) + \text{termes de degré plus bas en } X. \tag{6.223}$$

Le coefficient devant X^{k-1} n'est autre que $k\alpha + \sum_i a_i$. Étant donné que $k \neq 0$ et que $R \in \mathbb{F}_p[X]$, nous devons avoir $\alpha \in \mathbb{F}_p$. Par conséquent nous avons $\alpha^p = \alpha$ et une contradiction :

$$Q(\alpha) = \alpha^p - \alpha + 1 = 1 \neq 0. \tag{6.224}$$

Le polynôme $X^p - X + 1$ est donc irréductible sur $\mathbb{F}_p$. △

48. Définition 6.66.
49. Définition 6.66. Pour l'existence c'est le théorème 6.141.

6.4.13 Extensions séparables

Notons que dans ce qui va suivre nous allons parler de $\mathbb{K}[X]$, l'ensemble des polynômes sur un corps. Cela ne s'applique donc pas à $\mathbb{Z}[X]$ par exemple.

Une des choses intéressantes avec les extensions séparables c'est qu'elles vérifient le théorème de l'élément primitif 6.156.

Définition 6.144.
Soit $\mathbb{K}$ un corps. Un polynôme irréductible *$P \in \mathbb{K}[X]$ est **séparable** sur $\mathbb{K}$ si dans un corps de décomposition, ses racines sont distinctes, c'est-à-dire que si P est de degré n, alors il possède n racines distinctes dans un corps de décomposition.*

*Si P est un polynôme non constant dont la décomposition en irréductibles est $P = P_1 \dots P_r$, nous disons qu'il est **séparable** si tous les P_i le sont.*

La proposition suivante donne un sens à la définition de polynôme irréductible séparable.

Proposition 6.145.
Soit P irréductible dans $\mathbb{K}[X]$ ayant des racines distinctes dans le corps de décomposition $\mathbb{L}$. Si $\mathbb{L}'$ est un autre corps de décomposition pour P, alors P a aussi ses racines distinctes dans $\mathbb{L}'$.

Démonstration. L'ingrédient est la proposition 6.135 qui donne l'unicité du corps de décomposition à $\mathbb{K}$-isomorphisme près. Soit donc $\psi\colon \mathbb{L} \to \mathbb{L}'$ un isomorphisme laissant invariant les éléments de $\mathbb{K}$. D'une part, étant donné que P est à coefficients dans $\mathbb{K}$, nous avons $\psi(P) = P$. D'autre part dans $\mathbb{L}$ le polynôme P s'écrit

$$P = a(X - \alpha_1) \dots (X - \alpha_n) \tag{6.225}$$

avec $a \in \mathbb{K}$ et $\alpha_i \in \mathbb{L}$. Nous avons donc

$$P = \psi(P) = a(X - \psi(\alpha_1)) \dots (X - \psi(\alpha_n)). \tag{6.226}$$

Donc les racines de P dans $\mathbb{L}'$ sont les éléments $\psi(\alpha_i)$ qui sont distincts. □

Exemple 6.146.
Un polynôme peut être séparable sur un corps, mais non séparable sur un autre. Soit $\mathbb{L} = \mathbb{F}_p(T)$ et $\mathbb{K} = \mathbb{F}_p(T^p)$. Nous considérons le polynôme

$$P = X^p - T^p \tag{6.227}$$

dans $\mathbb{K}[X]$. Par le morphisme de Frobenius nous avons

$$P = (X - T)^p \tag{6.228}$$

dans $\mathbb{L}[X]$. Le polynôme P est irréductible sur $\mathbb{K}[X]$ parce que ses diviseurs sont de la forme $(X - T)^k$ qui contiennent T^k qui n'est pas dans $\mathbb{K}$ (sauf si $k = n$ ou $k = 0$).

Ce polynôme n'est pas séparable sur $\mathbb{K}$ parce que dans le corps de décomposition $\mathbb{L}$, la racine T est multiple. Notons bien le raisonnement : P étant irréductible, pour savoir s'il est séparable, on le regarde dans un corps de décomposition.

Par contre si nous regardons P dans $\mathbb{L}[X]$ alors P n'est plus irréductible parce que ses facteurs irréductibles sont $(X - T)$. N'étant pas irréductible, nous regardons les racines de *ses facteurs irréductibles*. Or chacun des facteurs irréductibles étant $X - T$, les racines sont simples. △

Exemple 6.147.
Le polynôme $(X - 1)^3$ est séparable sur $\mathbb{Q}$ parce que ses facteurs irréductibles dans $\mathbb{Q}[X]$ sont $X - 1$ et $X^2 + X + 1$, et ces deux polynômes ont des racines simples (dans $\mathbb{Q}(i)$). △

Exemple 6.148.
Le polynôme $(X^2 + 1)^2$ est séparable dans $\mathbb{Q}[X]$. En effet, il a pour facteurs irréductibles le polynôme $X^2 + 1$ (en deux exemplaires), et ce polynôme a pour racines $\pm i$ dans l'extension $\mathbb{Q}(i)$, racines qui sont simples pour ce polynôme. △

Proposition 6.149 ([112]).
Soit $P \in \mathbb{K}[X]$ un polynôme non constant. Les propriétés suivantes sont équivalentes.

(1) P a une racine multiple dans une extension de $\mathbb{K}$. C'est-à-dire qu'il existe une extension de $\mathbb{K}$ dans laquelle P a une racine multiple.

(2) P a une racine multiple dans tout corps de décomposition .

(3) P et P' ont une racine commune dans une extension de $\mathbb{K}$.

(4) le degré de $\operatorname{pgcd}(P, P')$ est $\geqslant 1$.

Démonstration. (1)⇒(2) Soit a, une racine multiple de P dans une extension $\mathbb{L}$ de $\mathbb{K}$, et $\mathbb{E}$, un corps de décomposition de P. Alors nous voulons prouver que P ait une racine multiple dans $\mathbb{E}$.

Nous pouvons voir $P \in \mathbb{L}[X]$, et construire une corps de décomposition $\mathbb{E}'$ qui est une extension de $\mathbb{L}$. Vu que $\mathbb{E}$ et $\mathbb{E}'$ sont deux corps de décomposition de P nous avons un isomorphisme $\psi\colon \mathbb{E}' \to \mathbb{E}$. Si $a \in \mathbb{E}$ est une racine multiple de P, alors $\psi(a)$ est une racine multiple de P dans $\mathbb{E}'$ parce que

$$P\big(\psi(a)\big) = \psi\big(P(a)\big). \tag{6.229}$$

(2)⇒(3) Soit $\mathbb{L}$ un corps de décomposition de P sur $\mathbb{K}$ et $a \in \mathbb{L}$, une racine multiple de P. On a alors $P = (X - a)^2 Q$ avec $Q \in \mathbb{L}[X]$. En dérivant,

$$P' = 2(X - a)Q + (X - a)^2 Q', \tag{6.230}$$

et donc a est également une racine de P'.

(3)⇒(4) Soit D un pgcd de P et P'. D'après le théorème de Bézout il existe $A, B \in \mathbb{K}[X]$ tels que

$$AP + BP' = D. \tag{6.231}$$

Si a est une racine commune de P et P' dans une extension $\mathbb{L}$, alors c'est aussi une racine de D et donc $\deg(D) \geqslant 1$.

(4)⇒(1) Si le degré de D est plus grand ou égal à 1, alors nous considérons une racine a de D dans $\mathbb{L}$ (une extension de $\mathbb{K}$). Étant donné que D divise P et P', l'élément a est une racine commune de P et P'. Nous montrons maintenant que a est alors une racine multiple de P. Vu que $P(a) = 0$ nous avons

$$P = (X - a)Q, \tag{6.232}$$

et $P' = Q + (X - a)Q'$. Mais alors $P'(a) = Q(a)$ et donc $Q(a) = 0$ et donc a est une racine double de P. Par conséquent a est une racine multiple de P dans $\mathbb{K}$.

□

Notons que si P est irréductible, cette proposition donne des conditions pour que P ne soit pas séparable.

Proposition 6.150.
Soit $P \in \mathbb{K}[X]$ irréductible. Le polynôme P est séparable si et seulement si $P' \neq 0$.

Démonstration. Soit $D = \operatorname{pgcd}(P, P')$ et nous voudrions prouver que $\deg(D) \geqslant 1$ si et seulement si $P' = 0$. Si $P' = 0$, alors $\operatorname{pgcd}(P, P') = P$ est donc $\deg(D) \geqslant 1$.

Dans l'autre sens, si P est irréductible, il est associé à D parce qu'il n'a pas d'autres diviseurs que lui-même et le polynôme constant 1. Ainsi, $D \in \mathbb{K}$, ou bien $P = \lambda D$ avec $\lambda \in \mathbb{K}$. et donc $\deg(P) \geqslant 1$. Dans les deux cas, P' est nécessairement non-nul. □

Corolaire 6.151.
Si $\mathbb{K}$ est de caractéristique nulle, alors tout polynôme de $\mathbb{K}[X]$ est séparable.

Démonstration. Il suffit de montrer que les irréductibles sont séparables. Soit P un polynôme irréductible et unitaire de degré d. Le terme de plus haut degré de P' est alors dX^{d-1} qui est non nul parce que $d \neq 0$ en caractéristique nulle. Donc $P' \neq 0$ et donc P est séparable par la proposition 6.149. □

Définition 6.152.
Soit $\mathbb{L}$ une extension algébrique de $\mathbb{K}$.

(1) *On dit que l'élément $a \in \mathbb{L}$ est **séparable** sur $\mathbb{K}$ si son polynôme minimal dans $\mathbb{K}[X]$ est séparable sur $\mathbb{K}$ (définition 6.144).*
(2) *L'extension $\mathbb{L}$ est **séparable** si tous ses éléments sont séparables.*

Proposition 6.153.
Soit $\mathbb{K}$ un corps. Les conditions suivantes sont équivalentes :

(1) *toutes les extensions algébriques de $\mathbb{K}$ sont séparables ;*
(2) *tout polynôme irréductible de $\mathbb{K}[X]$ est séparable.*

En particulier les extensions algébriques des corps de caractéristique nulle sont toutes séparables.

Démonstration. En plusieurs parties.

(1) implique (2) Soit un polynôme irréductible P de $\mathbb{K}[X]$, et un corps de décomposition $\mathbb{L}$ de P. Cela est une extension algébrique par le corolaire 6.139. Elle est donc séparable par hypothèse.

Voilà une première chose de dite.

Maintenant, nous voudrions montrer que P est un polynôme séparable. Dans $\mathbb{L}$ nous avons

$$P = \prod_{i=1}^{n}(X - a_i), \tag{6.233}$$

et tout le défi est de prouver que les a_i sont tous distincts.

Soient deux racines $a, b \in \mathbb{L}$ de P. Nous considérons les polynômes minimaux μ_a et μ_b dans $\mathbb{K}[X]$. Ces deux polynômes divisent P parce que P est à la fois dans l'idéal annulateur de a et de b. Mais comme P est irréductible, il existe $k_a, k_n \in \mathbb{K}$ tels que $P = k_a\mu_a$ et $P = k_b\mu_b$. Donc les polynômes μ_a, μ_b et P sont multiples les uns des autres. Vu que μ_a et μ_n sont unitaires, $\mu_a = \mu_b$.

Nous avons :

$$P = k\mu = \prod_{i=1}^{n}(X - a_i). \tag{6.234}$$

Or le polynôme μ est irréductible par la proposition 6.59(1), et l'extension $\mathbb{L}$ est séparable, donc μ n'a que des racines simples, Donc tous les a_i sont distincts.

(2) implique (1) Soit une extension algébrique $\mathbb{L}$ de $\mathbb{K}$. Soit $a \in \mathbb{L}$. Nous devons prouver que le polynôme minimal de a dans $\mathbb{K}$ est séparable, c'est-à-dire qu'il n'a que des racines simples.

Le polynôme minimal $\mu_a \in \mathbb{K}[X]$ de a est irréductible et donc séparable. Donc $\mathbb{L}$ est séparable.

La dernière phrase est une conséquence du corolaire 6.151. □

Corolaire 6.154.
Toute les extensions algébriques de $\mathbb{Q}$ sont séparables.

Démonstration. Le corps $\mathbb{Q}$ est de caractéristique nulle (définition 3.59). Le corolaire 6.151 dit alors que tout polynôme sur $\mathbb{Q}$ est séparable. La proposition 6.153 conclut en disant que toutes les extensions algébriques de $\mathbb{Q}$ sont séparables. □

Théorème 6.155 ([51])**.**
Soit $\mathbb{K}$ un corps (pas spécialement fini). Tout sous-groupe fini de $\mathbb{K}^*$ est cyclique.

Démonstration. Soit G un sous-groupe fini de $\mathbb{K}^*$ et ω son exposant (qui est le PPCM des ordres des éléments de G). Étant donné que $|G|$ est divisé par tous les ordres, il est divisé par le PPCM des ordres. Bref, nous avons

$$x^\omega = 1 \tag{6.235}$$

pour tout $x \in G$. Mais ce polynôme possède au plus ω racines dans $\mathbb{K}$. Du coup $|G| \leqslant \omega$. Et comme on avait déjà vu que $\omega \mid |G|$, on a $\omega = |G|$. Il suffit plus que trouver un élément d'ordre effectivement ω. Cela est fait par le lemme 3.40. □

Théorème 6.156 (Théorème de l'élément primitif[51])**.**
Toute extension de corps séparable finie admet un élément primitif[50]*.*

Autrement dit, soient des éléments algébriques $\alpha_1, \dots, \alpha_n$ séparables[51] *sur $\mathbb{K}$, et soit l'extension engendrée $\mathbb{L} = \mathbb{K}(\alpha_1, \dots, \alpha_n)$. Alors $\mathbb{L}$ admet un élément primitif, c'est-à-dire un élément θ tel que $\mathbb{L} = \mathbb{K}(\theta)$.*

Démonstration. Si le corps $\mathbb{K}$ est fini, alors $\mathbb{L}$ est également fini. Donc $\mathbb{L}^*$ est cyclique par le théorème 6.155. Si θ est un générateur de $\mathbb{L}^*$, alors $\mathbb{L} = \mathbb{K}(\theta)$.

Passons au cas où $\mathbb{K}$ est infini. Il suffit d'examiner le cas $n = 2$; en effet pour $n = 1$ c'est trivial et si $n > 2$, alors

$$\mathbb{K}(\alpha_1, \dots, \alpha_n) = \mathbb{K}(\alpha_1, \dots, \alpha_{n-1})(\alpha_n), \tag{6.236}$$

et donc si $\mathbb{K}(\alpha_1, \dots, \alpha_{n-1}) = \mathbb{K}(\theta)$, nous avons

$$\mathbb{K}(\alpha_1, \dots, \alpha_n) = \mathbb{K}(\theta, \alpha_n) \tag{6.237}$$

et nous sommes réduit au cas $n = 2$ par récurrence.

Soit donc $\mathbb{L} = \mathbb{K}(\alpha, \beta)$; soit P le polynôme minimal de α sur $\mathbb{K}$ et Q celui de β. Nous nommons $\mathbb{E}$, un corps de décomposition de PQ. Nous avons $\mathbb{L} \subset \mathbb{E}$. Vu que P et Q sont polynômes minimaux d'éléments qui sont par hypothèse séparables, les polynômes P et Q sont séparables. Donc dans $\mathbb{E}$ les racines de P sont distinctes parce que P est irréductible (et idem pour Q). Soient les racines

$$\alpha_1 = \alpha, \alpha_2, \dots, \alpha_r \tag{6.238}$$

de P dans $\mathbb{E}$ et les racines

$$\beta_1 = \beta, \beta_2, \dots, \beta_s \tag{6.239}$$

de Q dans $\mathbb{E}$. Ici r et s sont les degrés de P et Q.

Si $s = 1$ alors $Q = X - \beta$ et donc $\beta \in \mathbb{K}$ (parce que $Q \in \mathbb{K}[X]$). Du coup nous avons $\mathbb{L} = \mathbb{K}(\alpha)$ et le théorème est démontré. Nous supposons donc maintenant que $s \geqslant 2$.

Pour chaque $(i, j) \in [\![1, r]\!] \times [\![2, s]\!]$, l'équation $\alpha_i + x\beta_k = \alpha_1 + x\beta_1$ pour $x \in \mathbb{K}$ a au plus[52] une solution donnée le cas échéant par

$$x = (\alpha_i - \alpha_1)(\beta_1 - \beta_k)^{-1} \tag{6.240}$$

Notons que cela est de toutes façons dans $\mathbb{L}$ et qu'étant donné que $\beta_1 \neq \beta_k$, cette solution a un sens (ici on utilise l'hypothèse de séparabilité). Étant donné que $\mathbb{K}$ est infini nous pouvons donc trouver un $c \in \mathbb{K}$ qui ne résout aucune des équations (6.240) :

$$\alpha_i + c\beta_k \neq \alpha_1 + c\beta_1. \tag{6.241}$$

Nous posons $\theta = \alpha_1 + c\beta_1$ et nous prétendons que $\mathbb{L} = \mathbb{K}(\theta)$.

Pour cela, commençons par montrer que $\beta_1 \in \mathbb{K}(\theta)$. On considère, dans $\mathbb{K}(\theta)[T]$, les polynômes $Q(T)$ et $S(T) = P(\theta - cT)$, et on nomme R le PGCD de ces deux polynômes. Alors, une racine

50. Définition 6.92.
51. Définition 6.152(1).
52. La solution (6.240) peut être dans $\mathbb{L}$ et non dans $\mathbb{K}$. L'équation peut donc très bien ne pas avoir de solutions $x \in \mathbb{K}$.

de R doit être une racine de Q, et est donc un des β_i. Or, d'une part, le choix de θ fait que β_1 est une racine de R parce que

$$S(\beta_1) = P(\theta - c\beta_1) = P(\alpha_1 + c\beta_1 - c\beta_1) = P(\alpha_1) = 0. \tag{6.242}$$

D'autre part, si $k \geqslant 2$, alors

$$S(\beta_k) = P(\alpha_1 + c\beta_1 - c\beta_k) = P\big(\alpha_1 + c(\beta_1 - \beta_k)\big) \neq 0 \tag{6.243}$$

parce que $\alpha_1 + c(\beta_1 - \beta_k)$ ne vaut ni α_1 (le second terme est non-nul), ni un autre α_i (à cause de (6.241)).

Il s'ensuit que Q et S n'ont qu'une racine commune $\beta_1 = \beta$, qui est donc l'unique racine de R. Ainsi,

$$R = X - \beta \in \mathbb{K}(\theta)[T], \tag{6.244}$$

et donc $\beta \in \mathbb{K}(\theta)$.

Dès lors $\alpha = \alpha_1 = \theta - c\beta$ est alors immédiatement dans $\mathbb{K}(\theta)$; puisque les deux éléments α et β sont dans $\mathbb{K}(\theta)$, nous avons obtenu $\mathbb{L} = \mathbb{K}(\alpha, \beta) = \mathbb{K}(\theta)$.

□

Exemple 6.157.
Le théorème de l'élément primitif 6.156 ne tient pas pour les corps non commutatifs. Considérons par exemple le corps $\mathbb{K}$ des quaternions et le groupe à 8 éléments $G = \{\pm 1, \pm i, \pm j, \pm k\}$. Ce dernier groupe n'est pas cyclique alors qu'il est un groupe fini dans $\mathbb{K}^*$. △

Exemple 6.158.
Il est aussi possible pour un groupe fini d'avoir $\omega(G) = |G|$ sans pour autant que G soit cyclique. Par exemple pour $G = S_3$, nous avons $|S_3| = 6$ alors que les éléments de S_3 sont soit d'ordre 2 soit d'ordre 3 et $\omega(G) = \text{ppcm}(2, 3) = 6$. Pourtant S_3 n'est pas cyclique. △

6.5 Idéal maximum

6.5.1 Idéal maximum

Définition 6.159.
*Une $\mathbb{K}$-algèbre est de **type fini** si elle est le quotient de $\mathbb{K}[X_1, \dots, X_n]$ par un idéal (pour un certain n).*

Théorème 6.160 ([113])**.**
Soit $\mathbb{K}$ un corps et B, une $\mathbb{K}$-algèbre de type fini. Si B est un corps, alors c'est une extension algébrique finie de $\mathbb{K}$.

Théorème 6.161 ([113])**.**
Si $\mathbb{K}$ est un corps algébriquement clos, les idéaux maximaux de $\mathbb{K}[X_1, \dots, X_n]$ sont de la forme

$$(X_1 - a_1, \dots, X_n - a_n) \tag{6.245}$$

où les a_i sont des éléments de $\mathbb{K}$.

Démonstration. Nous commençons par montrer que

$$J = (X_1 - a_1, \dots, X_n - a_n) \tag{6.246}$$

est un idéal maximum. Pour cela nous considérons le morphisme surjectif d'anneaux

$$\begin{aligned} \phi\colon \mathbb{K}[X_1, \dots, X_n] &\to \mathbb{K} \\ P &\mapsto P(a_1, \dots, a_n). \end{aligned} \tag{6.247}$$

Soit $P \in \ker(\phi)$; nous écrivons la division euclidienne de P par $X - a_1$ puis celle du reste par $X - a_2$ et ainsi de suite :

$$P = (X - a_1)Q_1 + \cdots + (X_n - a_n)Q_n + R \tag{6.248}$$

où R doit être une constante parce que le premier reste est de degré zéro en X_1, le second est de degré zéro en X_1 et X_2, etc. Afin d'identifier cette constante, nous appliquons l'égalité (6.248) à $(a_1, \ldots, a_n)$ et en nous rappelant que $P \in \ker(\phi)$ nous obtenons

$$0 = P(a_1, \ldots, a_n) = R, \tag{6.249}$$

donc $R = 0$ et $P = (X_1 - a_1)Q_1 + \cdots + (X_n - a_n)Q_n$, c'est-à-dire $P \in J$. Nous avons donc $\ker(\phi) \subset J$. Par ailleurs $J \subset \ker(\phi)$ est évident, donc $J = \ker(\phi)$.

Vu que J est le noyau de l'application $\mathbb{K}[X_1, \ldots, X_n] \to \mathbb{K}$, nous avons

$$\frac{\mathbb{K}[X_1, \ldots, X_n]}{J} = \mathbb{K}. \tag{6.250}$$

Donc J est un idéal maximal parce que tout polynôme n'étant pas dans J doit avoir un terme indépendant non nul et donc être dans $\mathbb{K}$ vis à vis du quotient $\mathbb{K}[X_1, \ldots, X_n]/J$.

Nous montrons maintenant l'implication inverse. Nous supposons que I est un idéal maximum et nous montrons qu'il doit être égal à J (pour un certain choix de $a_1, \ldots, a_n$).

Le quotient

$$\frac{\mathbb{K}[X_1, \ldots, X_n]}{I} \tag{6.251}$$

est une $\mathbb{K}$-algèbre de type fini (définition 6.159). De plus c'est un corps par la proposition 3.56. C'est donc une extension algébrique finie de $\mathbb{K}$ par le théorème 6.160. Mais $\mathbb{K}$ étant algébriquement clos, il est sa propre et unique extension algébrique ; nous en déduisons que

$$\frac{\mathbb{K}[X_1, \ldots, X_n]}{I} = \mathbb{K}. \tag{6.252}$$

Donc pour tout $1 \leqslant i \leqslant n$, il existe $a_i \in \mathbb{K}$ tel que $X_i - a_i \in I$, sinon le monôme X_i ne se projetterait pas sur un élément dans $\mathbb{K}$ dans le quotient. Cela prouve que J est contenu dans I ; par maximalité nous avons donc $I = J$. □

Corolaire 6.162.
Soit $\mathbb{K}$ un corps algébriquement clos et I, un idéal de $\mathbb{K}[X_1, \ldots, X_n]$. Si nous notons

$$V(I) = \{x \in \mathbb{K}^n \text{ tel que } P(x_1, \ldots, x_n) = 0\} \tag{6.253}$$

l'ensemble des racines communes à tous les éléments de I, on a $V(I) = \varnothing$ si et seulement si $I = \mathbb{K}[X_1, \ldots, X_n]$.

Démonstration. Si $I = \mathbb{K}[X_1, \ldots, X_n]$ en particulier $1 \in I$ et nous avons évidemment $V(I) = \varnothing$. Le sens difficile est l'autre sens.

Supposons que $I \neq \mathbb{K}[X_1, \ldots, X_n]$ et que K est un idéal maximum contenu dans I. Nous savons déjà par le théorème 6.161 que K est de la forme $K = (X_1 - a_1, \ldots, X_n - a_n)$. Un élément de I est dans K, donc si $P \in I$ nous avons

$$P(a_1, \ldots, a_n) = 0, \tag{6.254}$$

c'est-à-dire que $(a_1, \ldots, a_n) \in V(I)$ et donc que $V(I) \neq \varnothing$. □

6.6 Polynômes symétriques et alternés

6.6.1 Polynômes symétriques, alternés ou semi-symétriques

Nous rappelons que le groupe symétrique S_n agit sur l'anneau des polynômes de n variables sur l'anneau A par le lemme 3.95.

Définition 6.163.
Un polynôme Q en n indéterminées est

(1) ***symétrique*** *si $Q = \sigma \cdot Q$ pour tout $\sigma \in S_n$;*

(2) ***alterné*** *si $\sigma \cdot Q = \epsilon(\sigma)Q$ pour tout $\sigma \in S_n$;*

(3) ***semi-symétrique*** *si $\sigma \cdot Q = Q$ pour tout $\sigma \in A_n$*

Le polynôme $T_1 + T_2$ est symétrique ; le polynôme $T_1 + T_2^2$ ne l'est pas.

6.6.2 Polynôme symétrique élémentaire

Définition 6.164.
Le k-ième ***polynôme symétrique élémentaire*** *à n inconnues est le polynôme*

$$\sigma_k(T_1, \dots, T_n) = \sum_{s \in F_k} \prod_{i=1}^{k} T_{s(i)} \tag{6.255}$$

où F_k est l'ensemble des fonctions strictement croissantes $\{1, 2, \dots, k\} \to \{1, 2, \dots, n\}$.

Une autre façon de décrire ces polynômes élémentaires est

$$\sigma_k = \sum_{1 \leqslant i_1 < \dots < i_k \leqslant n} X_{i_1} \dots X_{i_k}. \tag{6.256}$$

Par exemple

$$\sigma_1(T_1, \dots, T_n) = T_1 + T_2 + \dots + T_n \tag{6.257a}$$
$$\sigma_2(T_1, \dots, T_n) = T_1T_2 + \dots + T_1T_n + T_2T_3 + \dots + T_2T_n + \dots + T_{n-1}T_n \tag{6.257b}$$
$$\sigma_n(T_1, \dots, T_n) = T_1 \dots T_n. \tag{6.257c}$$

En particulier, $\sigma_2(x, y, z) = xy + yz + xz$.

Théorème 6.165 ([114])**.**
Si Q est un polynôme symétrique en $T_1, \dots, T_n$, alors il existe un et un seul polynôme P en n indéterminées tel que

$$Q(T_1, \dots, T_n) = P\big(\sigma_1(T_1, \dots, T_n), \dots, \sigma_n(T_1, \dots, T_n)\big). \tag{6.258}$$

Exemple 6.166.
Nous voulons décomposer $P(x, y, z) = x^3 + y^3 + z^3$ en polynômes symétriques élémentaires, c'est-à-dire en

$$\begin{cases} \sigma_1 = x + y + z & (6.259a) \\ \sigma_2 = xy + xz + yz & (6.259b) \\ \sigma_3 = xyz. & (6.259c) \end{cases}$$

Étant donné que P est de degré 3, les seules combinaisons des σ_i qui peuvent intervenir sont σ_1^3, $\sigma_1\sigma_2$ et σ_3. Étant donné que dans P le coefficient de x^3 est un, il est obligatoire d'avoir un coefficient 1 devant σ_1^3. Nous le calculons :

```
----------------------------------------------------------------------
| Sage Version 4.8, Release Date: 2012-01-20                         |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: var('x,y,z')
(x, y, z)
sage: P=x**3+y**3+z**3
sage: S1=x+y+z
sage: S2=x*y+x*z+y*z
sage: S3=x*y*z
sage: (S1**3).expand()
x^3 + 3*x^2*y + 3*x^2*z + 3*x*y^2 + 6*x*y*z + 3*x*z^2 + y^3
                + 3*y^2*z + 3*y*z^2 + z^3
sage: (S1**3-P).expand()
3*x^2*y + 3*x^2*z + 3*x*y^2 + 6*x*y*z + 3*x*z^2 + 3*y^2*z + 3*y*z^2
x^3 + 3*x^2*y + 3*x^2*z + 3*x*y^2 + 6*x*y*z + 3*x*z^2
            + y^3 + 3*y^2*z + 3*y*z^2 + z^3
```

Dans la différence $\sigma_1^3 - P$ nous voyons que le terme en xyz est $6xyz$; par conséquent nous savons que le coefficient de σ_3 sera -6. Il nous reste :

```
sage: (S1**3+6*S3-P).expand()
3*x^2*y + 3*x^2*z + 3*x*y^2 + 12*x*y*z + 3*x*z^2 + 3*y^2*z + 3*y*z^2
```

que nous identifions facilement avec $3\sigma_1\sigma_2$. Nous avons donc

$$P = \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3. \tag{6.260}$$

△

Lemme 6.167 ([71]).
Soit $\mathbb{K}$ une extension de degré δ de $\mathbb{Q}$ et $P \in \mathbb{K}[T_1, \dots, T_m]$. Alors il existe $\bar{P} \in \mathbb{Q}[T_1, \dots, T_m]$ tel que

(1) $\deg \bar{P} = \delta \deg(P)$

(2) pour tout $(z_1, \dots, z_m) \in \mathbb{C}^m$ tel que $P(z_1, \dots, z_m) = 0$, on a $\bar{P}(z_1, \dots, z_m) = 0$.

Démonstration. En vertu de la proposition 6.153 et du corolaire 6.154, $\mathbb{K}$ est une extension séparable de $\mathbb{Q}$, et donc vérifie le théorème de l'élément primitif (6.156). Il existe $\theta \in \mathbb{K}$ tel que $\mathbb{K} = \mathbb{Q}(\theta)$. Soit $P_\theta \in \mathbb{Q}[X]$ le polynôme minimal de θ. L'extension $\mathbb{K}$ étant de degré δ, et θ étant un générateur, une base de $\mathbb{K}$ comme espace vectoriel sur $\mathbb{Q}$ est

$$\{1, \theta, \dots, \theta^{\delta-1}\}. \tag{6.261}$$

Mais par ailleurs la proposition 6.87(2) nous indique qu'une base de $\mathbb{Q}(\theta)$ sur $\mathbb{Q}$ est donnée par

$$\{1, \theta, \dots, \theta^{n-1}\} \tag{6.262}$$

où n est le degré de P_θ. Donc P_θ est de degré δ. Nous nommons $\theta_1, \dots, \theta_\delta$ les racines de P_θ dans un corps de décomposition. Ici nous notons $\theta = \theta_1$ et nous ne prétendons pas que $\theta_k \in \mathbb{K}$. Notons que ces θ_i sont toutes des racines simples de P_θ, sinon nous aurions un facteur irréductible $(X - \theta_k)^2$, et P_θ ne serait pas irréductible sur $\mathbb{Q}$.

Soit σ_k le morphisme canonique

$$\begin{aligned} \sigma_k \colon \mathbb{Q}(\theta) &\to \mathbb{Q}(\theta_k) \\ \sum_i q_i\theta^i &\mapsto \sum_i q_i\theta_k^i \end{aligned} \tag{6.263}$$

Nous avons $\sigma_1 : \mathbb{K} \to \mathbb{K}$ qui est l'identité.

Notons N le degré du polynôme $P \in \mathbb{K}[T_1, \dots, T_m]$ dont il est question dans l'énoncé. Nous le décomposons alors en

$$P = \sum_{l=0}^{N} \sum_{i=1}^{m} c_{il} T_i^l \tag{6.264}$$

avec $c_{il} \in \mathbb{K}$. Nous voyons $c_{i,\cdot}$ comme un élément de $\mathbb{K}^m$ et donc nous écrivons [53]

$$P = \sum_{l=0}^{N} \sum_{i=1}^{m} c_l(\theta)_i T_i^l \tag{6.265}$$

où $c_l \in \mathbb{Q}[X]^m$. Nous pouvons choisir $\deg(c_l) < \delta$ parce que les puissances plus grandes de θ ne génèrent rien de nouveau.

Nous posons aussi

$$P^{\sigma_k} = \sum_{l,i} c_l(\theta_k)_i T_i^l \in \mathbb{Q}(\theta_k)[T_1, \dots, T_m], \tag{6.266}$$

et $\bar{P} = PP^{\sigma_2} \dots P^{\sigma_k}$. Le coefficient de T_i^l dans $\bar{P}$ est

$$\bar{c}_l(\theta_1, \dots, \theta_\delta)_i = \sum_{l_1 + \dots + l_\delta = l} c_{l_1}(\theta_1)_i \dots c_{l_\delta}(\theta_\delta)_i. \tag{6.267}$$

Ce dernier est un polynôme en les θ_k à coefficients dans $\mathbb{Q}$. Qui plus est, c'est un polynôme symétrique. En effet un terme contenant $\theta_k^a \theta_l^b$ provenant de $c_{l_i}(\theta_k) c_{l_j}(\theta_l)$ a un terme correspondant $\theta_k^b \theta_l^a$ provenant de $c_{l_j}(\theta_k) c_{l_i}(\theta_l)$.

C'est donc le moment d'utiliser le théorème 6.165 à propos des polynômes symétriques élémentaires qui nous dit que les coefficients de $\bar{P}$ sont en réalité des polynômes en ceux de P_θ qui sont dans $\mathbb{Q}$. Donc $\bar{P} \in \mathbb{Q}[T_1, \dots, T_m]$. Par ailleurs nous avons que

$$\deg(\bar{P}) = \delta \deg(P) \tag{6.268}$$

parce que $\bar{P}$ est le produit de δ « copies » de P. De plus $P = P^{\sigma_1}$ divise $\bar{P}$ donc on a bien que si $P(z) = 0$ alors $\bar{P}(z) = 0$. Le polynôme $\bar{P}$ est celui que nous cherchions. □

6.6.3 Relations coefficients racines

Théorème 6.168 (Relations coeffitients-racines)**.**
Soit le polynôme $P = a_n X^n + \dots + a_1 X + a_0$ et r_i ses n racines. Alors nous avons pour chaque $1 \leqslant k \leqslant n$ la relation

$$\sigma_k(r_1, \dots, r_n) = (-1)^k \frac{a_{n-k}}{a_n} \tag{6.269}$$

où σ_k est le k^epolynôme symétrique défini en 6.164.

Exemple 6.169.
Soit le polynôme

$$P(x) = x^3 + 2x^2 + 3x + 4 \tag{6.270}$$

et ses racines que nous nommons a, b, c. Nous voudrions calculer $a^2 + b^2 + c^2$. D'abord nous décomposons $Q(a, b, c) = a^2 + b^2 + c^2$ en polynômes symétriques élémentaires : $Q(a, b, c) = \sigma_1(a, b, c)^2 - 2\sigma_2(a, b, c)$.

Mais les relations coefficients-racines [54] nous donnent $\sigma_1(a, b, c) = -2$ et $\sigma_2(a, b, c) = 3$, donc

$$a^2 + b^2 + c^2 = (-2)^2 - 2 \cdot 3 = -2. \tag{6.271}$$

Cela nous assure déjà qu'au moins une des solutions n'est pas réelle.

Nous pouvons en avoir une vérification directe en calculant explicitement les racines (ce qui est possible pour le degré 3) :

53. Il me semble qu'il manque la somme sur i dans [71].
54. Théorème 6.168

```
sage: P(x)=x**3+2*x**2+3*x+4
sage: S=solve(  P(x)==0,x  )
sage: sols=[ s.rhs() for s in S  ]
sage: Q=[  s**2 for s in sols ]
sage: s=sum(Q)
sage: s.simplify_full()
-2
```

tex/frido/VAYVmNRpolynomeSym.py

Notez qu'il faut un peu chipoter pour isoler les solutions depuis la réponse de la fonction `solve`. △

En suivant le même cheminement que dans l'exemple, si P est un polynôme de degré n et si r_i sont ses racines, il est facile de calculer $Q(r_1, \dots, r_n)$ pour n'importe quel polynôme symétrique Q

Proposition 6.170 (Annulation de fonctions polynomiales[115])**.**
Soit $\mathbb{K}$ un corps et P un polynôme à n indéterminées. Nous supposons que P s'annule sur un ensemble de la forme $A_1 \times \cdots \times A_n$ avec $\mathrm{Card}(A_j) > \deg_{X_j}(P)$ pour tout j. Alors $P = 0$.

De plus si $P = 0$ alors tous ses coefficients sont nuls[55].

Démonstration. Nous prouvons le résultat par récurrence sur le nombre n d'indéterminées. Si $n = 1$, cela est le théorème 6.101. Nous classons les monômes du polynôme P par ordre de puissance de X_n et nous le factorisons :

$$P = \sum_{i=1}^{m} P_i X_n^i \tag{6.272}$$

avec $P_i \in \mathbb{K}[X_1, \dots, X_{n-1}]$. Soit $(a_1, \dots, a_{n-1}) \in A_1 \times \dots \times A_{n-1}$ et posons

$$Q(T) = P(a_1, \dots, a_{n-1}, T) = \sum_{i=1}^{m} P_i(a_1, \dots, a_{n-1}) T^i. \tag{6.273}$$

Le polynôme Q s'annule sur A_n avec $\deg(Q) = \deg_{X_n}(P) < \mathrm{Card}(A_n)$ et le théorème 6.101 nous donne $Q = 0$. Or les coefficients des différentes puissances de T dans $Q(T)$ sont les $P_i(a_1, \dots, a_{n-1})$; ils sont donc nuls.

Nous avons montré que le polynôme P_i s'annule pour tout élément de $A_1 \times \dots \times A_{n-1}$, mais nous avons

$$\deg_{X_j}(P_i) \leqslant \deg_{X_j} P < \mathrm{Card}(A_j), \tag{6.274}$$

donc l'hypothèse de récurrence donne $P_i = 0$. Par suite, $P = 0$ également. □

6.7 Minuscule morceau sur la théorie de Galois

Vous trouverez des détails et des preuves à propos de la théorie de Galois dans [116, 51].

Définition 6.171.
Soit $\mathbb{K}$, un corps.

*Le **groupe de Galois** d'une extension $\mathbb{L}$ de $\mathbb{K}$ est le groupe des automorphismes de $\mathbb{L}$ laissant $\mathbb{K}$ invariant.*

Le groupe de Galois d'un polynôme sur $\mathbb{K}$ est le groupe de Galois de son corps de décomposition sur $\mathbb{K}$.

55. L'intérêt de cela est qu'un polynôme de $\mathbb{Z}[X_1, \dots, X_n]$ peut s'évaluer sur un élément de n'importe quel corps ; il restera le polynôme nul.

Définition 6.172.
Des éléments $b_1, \ldots, b_n$ d'une extension de $\mathbb{K}$ sont ***algébriquement indépendants*** *si ils ne satisfont à aucune relation du type*

$$\sum \alpha_{i_1 \ldots i_n} b_1^{i_1} \ldots b_n^{i_n} = 0 \tag{6.275}$$

avec $\alpha_{i_1 \ldots i_n} \in \mathbb{K}$.

Nous disons que l'équation

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0 \tag{6.276}$$

est l'**équation générale** de degré n si les coefficients a_i sont algébriquement indépendants sur $\mathbb{K}$.

Théorème 6.173.
Le groupe de Galois d'un polynôme de degré n est isomorphe au groupe symétrique S_n.

Corolaire 6.174 ([117])**.**
L'équation générale de degré n est résoluble par radicaux si et seulement si $n \leqslant 5$.

Chapitre 7

Topologie générale

7.1 Éléments généraux de topologie

7.1.1 Définitions et propriétés de base

Définition 7.1 ([118]).
Soit X, un ensemble et $\mathcal{T}$, une partie de l'ensemble de ses parties qui vérifie les propriétés suivantes.

(1) Les ensembles $\emptyset$ et X sont dans $\mathcal{T}$,

(2) Une union quelconque[1] *d'éléments de $\mathcal{T}$ est dans $\mathcal{T}$.*

(3) Une intersection finie *d'éléments de $\mathcal{T}$ est dans $\mathcal{T}$.*

Un tel choix $\mathcal{T}$ de sous-ensembles de X est une ***topologie*** *sur X, et les éléments de $\mathcal{T}$ sont appelés des* ***ouverts****. On dit aussi que $(X, \mathcal{T})$ (voire simplement X lorsqu'il n'y a pas d'ambiguïté) est un* ***espace topologique****.*

Définition 7.2.
Si X est un espace topologique, un sous-ensemble F de X est dit ***fermé*** *si son complémentaire, F^c, est ouvert.*

Si $a \in X$, on dit que $V \subset X$ est un ***voisinage*** *de a s'il existe un ouvert $\mathcal{O} \in \mathcal{T}$ tel que $a \in \mathcal{O}$ et $\mathcal{O} \subset V$.*

Lemme 7.3.
Union et intersection de fermés.

(1) Une intersection quelconque de fermés est fermée.

(2) Une union finie de fermés est fermée.

Démonstration. Soit $\{F_i\}_{i \in I}$ un ensemble de fermés ; nous avons

$$\left(\bigcap_{i \in I} F_i \right)^c = \bigcup_{i \in I} F_i^c. \tag{7.1}$$

Le membre de droite est une union d'ouverts, c'est donc un ouvert ; donc l'intersection qui apparaît dans le membre de gauche est le complémentaire d'un ouvert : c'est donc un fermé.

De la même manière, le complémentaire d'une union finie de fermés est une intersection finie de complémentaires de fermés, et est donc ouvert[2]. □

Dans un espace topologique, nous avons une caractérisation très importante des ouverts.

1. Par « quelconque » nous entendons vraiment quelconque : c'est-à-dire indicée par un ensemble qui peut autant être $\mathbb{N}$ que $\mathbb{R}$ qu'un ensemble encore considérablement plus grand.
2. Un bon exercice est d'écrire ces unions et intersections, pour se convaincre que ça fonctionne.

Théorème 7.4.
Une partie d'un espace topologique est ouverte si et seulement si elle contient un voisinage[3] *ouvert de chacun de ses éléments.*

Démonstration. Soit X un espace topologique et $A \subset X$. Le sens direct est évident : A lui-même est un ouvert autour de $x \in A$, qui est inclus dans A.

Pour le sens inverse, nous supposons que A contienne un ouvert autour de chacun de ses points. Pour chaque $x \in A$, choisissons $\mathcal{O}_x \subset A$ un ouvert autour de x. Alors,

$$A = \bigcup_{x \in A} \mathcal{O}_x \tag{7.2}$$

en effet, d'une part, $A \subset \bigcup_{x \in A} \mathcal{O}_x$ parce que chaque élément x de A est dans le $\mathcal{O}_x$ corrrespondant, par construction ; et d'autre part, $\bigcup_{x \in A} \mathcal{O}_x \subset A$ parce que chacun des $\mathcal{O}_x$ est inclus dans A.

Ainsi, A est égal à une union d'ouverts, cela prouve que A est un ouvert. □

Le lemme 7.235 est une version particulière de celui-ci, pour l'espace topologique $\mathbb{R}$. Une autre application typique est la proposition 7.60 et le théorème 7.74.

7.1.2 Quelques exemples

7.1.2.1 Une première vague

Exemple 7.5.
Pour un ensemble X quelconque, on considère l'ensemble $\mathcal{T} = \{\emptyset; X\}$. Avec cet ensemble, on confère à X une structure d'espace topologique - même si elle nous apprend peu de choses... La topologie ainsi posée sur X est appelée **topologie grossière**. △

Exemple 7.6.
Pour un ensemble X quelconque, on considère l'ensemble $\mathcal{T}$ constitué de toutes les parties de X. Avec cet ensemble, on confère à nouveau une structure d'espace topologique à X ; toutes les parties sont des ouverts, et aussi des fermés. La topologie ainsi posée sur X est appelée **topologie discrète**. △

Exemple 7.7 (Toutes les topologies d'un ensemble à 3 éléments)**.**
On pose $X = \{1, 2, 3\}$. Alors on peut munir X de 29 topologies différentes[119] ; saurez-vous les retrouver toutes ? △

7.1.2.2 Topologie engendrée

Proposition-définition 7.8 (Topologie engendrée, prébase[120])**.**
Soient un ensemble X et $\mathcal{T}_0$ un ensemble de parties de X. Nous définissons $\tau(\mathcal{T}_0)$ comme étant l'union quelconque d'intersections finies d'éléments de $\mathcal{T}_0$.

Plus précisément, nous faisons les constructions suivantes :

(1) Nous notons $\{\mathcal{O}_i\}_{i \in I}$ les éléments de T_0 indexés par l'ensemble I.

(2) Soit $B(\mathcal{T}_0)$ l'ensemble des intersections finies d'éléments de $\mathcal{T}_0$:

$$B(\mathcal{T}_0) = \{\bigcap_{j \in J} \mathcal{O}_j\}_{J \text{ fini dans } I} \tag{7.3}$$

où nous convenons que $\bigcap_{j \in \emptyset} \mathcal{O}_j = X$[4]*.*

(3) Soit A un ensemble qui indexe $B(\mathcal{T}_0)$:

$$B(\mathcal{T}_0) = \{B_\alpha\}_{\alpha \in A}. \tag{7.4}$$

3. Définition 7.2.
4. Bref, nous mettons X dans $B(\mathcal{T}_0)$.

(4) Nous posons

$$\tau(\mathcal{T}_0) = \{ \bigcup_{\alpha\in S} B_\alpha \}_{S\subset A}. \tag{7.5}$$

Alors $\tau(\mathcal{T}_0)$ est une topologie sur X, qu'on appelle ***topologie engendrée*** *par $\mathcal{T}_0$. La partie $\mathcal{T}_0$ est appelée* ***prébase*** *de la topologie $\tau(\mathcal{T}_0)$.*

Démonstration. Nous devons montrer les différents points de la définition 7.1 d'une topologie.

(1) L'ensemble vide est dans $\tau(\mathcal{T}_0)$ parce que $\emptyset = \bigcup_{\alpha\in\emptyset} B_\alpha$. L'ensemble X est également dans $\tau(\mathcal{T}_0)$ parce que $X \in B(\mathcal{T}_0)$.

(2) Soient $\{D_l\}_{l\in L}$ des éléments de $\tau(\mathcal{T}_0)$ indexés par un ensemble L. Pour chaque l nous avons un ensemble $S \subset A$ tel que $D_l = \bigcup_{\alpha\in S_l} B_\alpha$. En posant $S = \bigcup_{l\in L} S_l$ nous avons

$$\bigcup_{l\in L} D_l = \bigcup_{\alpha\in S} B_\alpha \in \tau(\mathcal{T}_0). \tag{7.6}$$

Donc $\tau(\mathcal{T}_0)$ est stable par union quelconque.

(3) Soient D_1 et D_2 des éléments de $\tau(\mathcal{T}_0)$. Nous posons $D_i = \bigcup_{\alpha\in S_i} B_\alpha$. Alors nous avons

$$\bigcup_{\alpha\in S_1} B_\alpha \cap \bigcup_{\beta ß S_2} B_\beta = \bigcup_{\alpha,\beta\in S_1\times S_2} (B_\alpha \cap B_\beta). \tag{7.7}$$

Mais B_α et B_β sont dans $B(\mathcal{T}_0)$. Donc $B_\alpha \cap B_\beta \in B(\mathcal{T}_0)$. Donc (7.7) est un union d'éléments de $B(\mathcal{T}_0)$.

Au final nous avons prouvé que $\tau(\mathcal{T}_0)$ est une topologie sur X. □

Lemme 7.9.
Soient un ensemble X et un ensemble $\mathcal{T}_0$ de parties de X. Toute topologie sur X contenant $\mathcal{T}_0$ contient $\tau(\mathcal{T}_0)$.

Démonstration. Soit une topologie μ sur X contenant $\tau(\mathcal{T}_0)$. Vu que μ est une topologie, les intersections finies d'éléments de μ sont dans μ, donc, en suivant les notations de 7.8, $B(\mathcal{T}_0) \subset \mu$.

Vu que toutes les unions d'éléments de μ sont dans μ, l'inclusion de $B(\mathcal{T}_0)$ dans μ implique celle de $\tau(\mathcal{T}_0)$. □

Dès que nous avons une topologie nous avons une notion de convergence de suite.

Définition 7.10 (Convergence de suite)**.**
Une suite (x_n) d'éléments de E ***converge*** *vers un élément y de E si pour tout ouvert $\mathcal{O}$ contenant y, il existe un $K \in \mathbb{N}$ tel que $k > K$ implique $x_k \in \mathcal{O}$.*

La proposition suivante montre que vérifier la convergence d'une suite sur une prébase suffit pour vérifier la convergence.

Proposition 7.11.
Soit $\mathcal{T}_0$ un ensemble de parties de l'ensemble X. Soient une suite (x_n) dans X ainsi que $x \in X$. Nous supposons que la suite (x_n) satisfasse la propriété suivante : pour tout $A \in \mathcal{T}_0$ tel que $x \in A$, il existe $K \in \mathbb{N}$ tel que $k \geqslant K$ implique $x_k \in A$.

Alors nous avons la convergence de suite[5]

$$x_n \xrightarrow{(X,\tau(\mathcal{T}_0))} x. \tag{7.8}$$

Démonstration. Nous considérons la topologie $\tau(\mathcal{T}_0)$ sur X. Soit un ouvert $\mathcal{O}$ contenant x. Nous le décomposons en suivant (à l'envers) la construction de la définition 7.8 :

$$\mathcal{O} = \bigcup_{\alpha\in S} B_\alpha \tag{7.9}$$

5. Définition 7.10.

avec $B_\alpha \in B(\mathcal{T}_0)$. Donc pour chaque α, il existe un ensemble fini J_α tel que

$$B_\alpha = \bigcap_{j \in J_\alpha} A_j \tag{7.10}$$

avec $A_j \in \mathcal{T}_0$. Vu que $x \in \mathcal{O}$, nous avons un α_0 tel que $x \in B_{\alpha_0}$. Donc $x \in A_j$ pour tous les $j \in J_{\alpha_0}$.

Pour chaque $j \in J_{\alpha_0}$, il existe $K_j \in \mathbb{N}$ tel que $k \geqslant K_j$ implique $x_k \in A_j$. Vu que J_{α_0} est un ensemble fini, nous pouvons poser $K = \max_{j \in J_{\alpha_0}} K_j$.

Maintenant, si $k \geqslant K$, nous avons $x_k \in A_j$ pour tout j, et donc $x_k \in B_{\alpha_0}$. Par conséquent aussi $x_k \in \mathcal{O}$. □

7.1.2.3 Topologie produit

Définition 7.12 (Produit d'espaces topologiques, thème 9)**.**
*Soient $X_1, \dots, X_n$ des espaces topologiques. Leur **produit** est l'ensemble*

$$X = \prod_{i=1}^{n} X_i \tag{7.11}$$

muni de la topologie engendrée[6] *par les produits $A_1 \times \cdots \times A_n$, avec $A_i \in X_i$ ouverts de chacun des ensembles.*

Lemme 7.13.
Si $\mathcal{B}_i$ est une base de la topologie de X_i, alors $\{A_1 \times \dots A_n$ tel que $A_i \in \mathcal{B}_i\}$ est une prébase de $X_1 \times \dots X_n$.

Proposition 7.14 (Convergence composante par composante)**.**
Soient des espaces topologiques X_i $(i = 1, \dots, n)$ et une suite $(a_k^{(1)}, \dots, a_k^{(n)})$ dans $X_1 \times \dots \times X_n$. Nous avons la convergence

$$(a_k^{(1)}, \dots, a_k^{(n)}) \xrightarrow{X_1 \times \dots \times X_n} (a^{(1)}, \dots, a^{(n)}) \tag{7.12}$$

si et seulement si $a_k^{(i)} \to a^{(i)}$ pour chaque i.

Démonstration. En deux parties.

Sens direct Soient des ouverts $\mathcal{O}_i$ autour de $a^{(i)}$ dans X_i. Vu que $\mathcal{O}_1 \times \dots \times \mathcal{O}_n$ est un ouvert autour de $(a^{(1)}, \dots, a^{(n)})$, il existe $K \in \mathbb{N}$ tel que si $k \geqslant K$ nous avons $(a_k^{(1)}, \dots, a_k^{(n)}) \in \mathcal{O}_1 \times \dots \times \mathcal{O}_n$. Pour ce K nous avons séparément $a_k^{(i)} \in \mathcal{O}_i$ pour chaque i.

Sens inverse Une prébase de la topologie sur $X_1 \times \dots \times X_n$ est donné par les $\mathcal{O}_1 \times \dots \times \mathcal{O}_n$ où $\mathcal{O}_i$ est un ouvert de X_i. Voir la définition 7.12 de la topologie produit et la définition 7.8 de ce qu'est une prébase.

La proposition 7.11 nous permet de ne vérifier la convergence de $(a_k^{(1)}, \dots, a_k^{(n)})$ que sur la prébase. Soit donc $\mathcal{O} = \mathcal{O}_1 \times \dots \mathcal{O}_n$ avec $(a^{(1)}, \dots, a^{(n)}) \in \mathcal{O}$. Vu que $(a_k^{(i)})_{k \in \mathbb{N}} \to a^{(i)}$, pour chaque i, il existe $K_i \in \mathbb{N}$ tel que si $k \geqslant K_i$ alors $a_k^{(i)} \in \mathcal{O}_i$.

En posant $K = \max_i(K_i)$, nous avons $(a_k^{(1)}, \dots, a_k^{(n)}) \in \mathcal{O}_1 \times \dots \mathcal{O}_n$ pour tout $k \geqslant K$.

La proposition 7.11 permet de conclure.

□

7.1.2.4 Topologie induite

Définition 7.15 (Topologie induite)**.**
*Soit un espace topologique $(X, \mathcal{T})$, et soit $Y \subset X$. Alors on peut munir Y de la topologie constituée des $Y \cap \mathcal{O}$, pour $\mathcal{O} \in \mathcal{T}$: c'est ce qu'on appelle la **topologie induite**.*

6. Définition 7.8.

Lemme 7.16 ([1]).
Soit (X, τ_X) un espace topologique et $S \subset X$, un fermé de X sur lequel nous considérons la topologie induite τ_S. Si F est un fermé de (S, τ_S) alors F est fermé de (X, τ_X).

Démonstration. Nous savons que le complémentaire de F dans S est un ouvert de (S, τ_S) : il existe un ouvert $\Omega \in \tau_X$ tel que $S \backslash F = S \cap \Omega$. Si maintenant nous considérons le complémentaire de S dans X nous avons

$$F^c = (S \backslash F) \cup (X \backslash S) = (S \cap \Omega) \cup S^c = (S \cap \Omega) \cup (S^c \cap \Omega) \cup S^c = \Omega \cup S^c. \tag{7.13}$$

Vu que Ω et S^c sont des ouverts de X, l'union est un ouvert. Donc $F^c \in \tau_X$ et F est un fermé de X. □

Lemme 7.17.
Si $B \subset A$ alors la fermeture de B pour la topologie de A (induite de X) que nous noterons $\tilde{B}$ est

$$\tilde{B} = \bar{B} \cap A \tag{7.14}$$

où $\bar{B}$ est la fermeture de B pour la topologie de X.

Démonstration. Si $a \in \bar{B} \cap A$, un ouvert de A autour de a est un ensemble de la forme $\mathcal{O} \cap A$ où $\mathcal{O}$ est un ouvert de X. Vu que $a \in \bar{B}$, l'ensemble $\mathcal{O}$ intersecte B et donc $(\mathcal{O} \cap A) \cap B \neq \emptyset$. Donc a est bien dans l'adhérence de B au sens de la topologie de A.

Pour l'inclusion inverse, soit $a \in \tilde{B}$, et montrons que $a \in \bar{B} \cap A$. Par définition $a \in A$, parce que $\tilde{B}$ est une fermeture dans l'espace topologique A. Il faut donc seulement montrer que $a \in \bar{B}$. Soit donc $\mathcal{O}$ un ouvert de X contenant a ; par hypothèse $\mathcal{O} \cap A$ intersecte B (parce que tout ouvert de A contenant a intersecte B). Donc $\mathcal{O}$ intersecte B. Cela signifie que tout ouvert (de X) contenant a intersecte B, ou encore que $a \in \bar{B}$. □

Si A est un ouvert de X, on pourrait croire que la topologie induite n'a rien de spécial. Il est vrai que B sera ouvert dans A si et seulement s'il est ouvert dans X, mais certaines choses surprenantes se produisent tout de même.

Exemple 7.18.
Prenons $X = \mathbb{R}$ et $A =]0, 1[$. Si $B =]\frac{1}{2}, 1[$, alors la fermeture de B dans A sera $\tilde{B} = [\frac{1}{2}, 1[$ et non $[\frac{1}{2}, 1]$ comme on l'aurait dans $\mathbb{R}$. △

Prendre la topologie induite de $\mathbb{R}$ vers un fermé de $\mathbb{R}$ donne des boules un peu spéciales comme le montre l'exemple suivant.

Exemple 7.19.
Quid de la boule ouverte $B(1, \epsilon)$ dans le compact $[0, 1]$? Par définition c'est

$$B(1, \epsilon) = \{x \in [0, 1] \text{ tel que } |x - 1| < \epsilon\} =]1 - \epsilon, 1]. \tag{7.15}$$

Oui, cela est *ouvert* dans $[0, 1]$. C'est d'ailleurs un des ouverts de la topologie induite de $\mathbb{R}$ sur $[0, 1]$.

Donc pour la topologie de $[0, 1]$, toutes les boules ouvertes $B(x, \delta)$ avec $x \in [0, 1]$ sont incluses à $[0, 1]$. △

7.1.3 Adhérence, fermeture, intérieur, point d'accumulation et isolé

7.1.3.1 Intérieur

Définition 7.20.
Soient un espace topologique X et une partie A de X.

*(1) Un point $x \in X$ est **intérieur** à A s'il est contenu dans un ouvert inclus dans A. L'ensemble des points intérieurs de A est noté* $\mathrm{Int}(A)$.

*(2) L'**intérieur** de A, notée $\mathring{A}$, est l'union de tous les ouverts de X contenus dans A.*

Remarque 7.21.
Quelques remarques en vrac.

(1) Pour tout $A \subset X$, l'ensemble $\mathring{A}$ est un ouvert, comme union quelconque d'ouverts.

(2) Par ailleurs, on a $\mathring{A} = \operatorname{Int} A$: en effet, $x \in \operatorname{Int} A$ si et seulement s'il existe un ouvert contenant x et inclus dans A, si et seulement si x est dans l'union de tous les ouverts contenus dans A, si et seulement si $x \in \mathring{A}$.

(3) On a $\mathring{A} \subset A$, et $\mathring{A} = A$ si et seulement si A est un ouvert : en sens direct, c'est clair par égalité d'ensembles ; en sens inverse, c'est aussi clair puisque A est alors un ouvert contenu à A, donc $A \subset \mathring{A}$.

7.1.3.2 Adhérence et fermeture

Disons-le tout de suite : « adhérence » et « fermeture » sont synonymes. Dans le Frido, nous allons nous évertuer à utiliser le mot « adhérance » et la notation $\operatorname{Adh}(A)$ au lieu de $\bar{A}$ que l'on rencontre assez souvent. Le fait que est $\bar{z}$ est le conjugué complexe de z. Dans certains cas, ça peut mener à des confusions.

Définition 7.22.
*Soient un espace topologique X et une partie A de X. Un point $x \in X$ est **adhérent** à A si tout ouvert de X contenant x a une intersection non vide avec A. L'ensemble des points d'adhérence de A est noté* $\operatorname{Adh}(A)$.

Lemme 7.23.
L'adhérence de A est l'intersection de tous les fermés de X contenant A.

Par ailleurs, nous avons le lien

$$(\operatorname{Int}(A))^c = \operatorname{Adh}(A^c). \tag{7.16}$$

Démonstration. Commençons par prouver la dernière égalité d'ensembles. On a les équivalences entre les éléments suivants, pour tout $x \in X$:

— x n'est pas dans $\mathring{A}$;

— il n'y a aucun ouvert contenant x et inclus dans A ;

— tout ouvert contenant x a une intersection non-vide avec A^c ;

— x est dans $\overline{A^c}$.

Nous allons à présent montrer l'égalité d'ensembles $\operatorname{Adh}(A) = \bar{A}$ en prouvant la double inclusion par contraposée.

Si $x \in \bar{A}$ alors $x \in \operatorname{Adh}(A)$ Si x n'est pas dans $\bar{A}$ alors nous avons un fermé F contenant A et pas x. Le complémentaire F^c est un ouvert qui contient x et dont l'intersection avec A est vide. Donc x n'est pas dans $\operatorname{Adh}(A)$.

Si $x \in \bar{A}$ alors $x \in \operatorname{Adh}(A)$ Si x n'est pas dans $\operatorname{Adh}(A)$ alors il existe un ouvert $\mathcal{O}$ contenant x et n'intersectant pas A. Le complémentaire $\mathcal{O}^c$ est un fermé qui contient A et qui ne contient pas x.

Vu que $\bar{A}$ est l'intersection de tous les fermés contenant A, nous avons $\bar{A} \subset \mathcal{O}^c$ et donc x n'est pas dans $\bar{A}$.

□

Remarque 7.24.
Comme corolaire du lemme précédent, et grâce aux remarques faites pour les intérieurs, on obtient que pour $A \subset X$:

(1) l'ensemble $\bar{A}$ est fermé : c'est en effet le complémentaire d'un ouvert, précisément l'intérieur de A^c ;

(2) A est fermé si et seulement si $\bar{A} = A$: en effet, A est fermé si et seulement si A^c est ouvert, si et seulement si l'intérieur de A^c est A^c lui-même ; or, l'intérieur de A^c est le complémentaire de $\bar{A}$ par le lemme 7.23, si bien que A est fermé si et seulement si $(\bar{A})^c = A^c$, ou encore... ce qu'on affirmait au début.

Définition 7.25.
*Soit X un espace topologique. Un sous-ensemble A de X est **dense** dans X si $\bar{A} = X$.*

7.1.3.3 Frontière

Définition 7.26.
*Soit X un espace topologique, et $A \subset X$. La **frontière** de A, notée ∂A, est l'ensemble des points adhérents de A qui ne sont pas intérieurs à A. Ainsi,*

$$\partial A = \bar{A} \backslash \mathring{A}. \tag{7.17}$$

7.1.3.4 Points d'accumulation et isolés

Définition 7.27.
*Soient un espace topologique X et une partie A de X. Un point $s \in X$ est un **point d'accumulation** de A si tout ouvert contenant s contient au moins un élément de $A \backslash \{s\}$.*

Quelle est la différence entre un point d'accumulation et un point d'adhérence ? La différence est que tous les points de A sont des points d'adhérence de A, parce que tout voisinage de $a \in A$ contient au moins a lui-même, alors que certains points de A peuvent ne pas être des points d'accumulation de A. Voir l'exemple 7.251.

Notons qu'un point d'accumulation de A dans X n'est pas spécialement dans A.

Définition 7.28.
*Soient un espace topologique X et une partie A de X. Un point $s \in A$ est un **point isolé** de A si il existe un voisinage ouvert $\mathcal{O}$ de s dans X tel que $A \cap \mathcal{O} = \{s\}$.*

7.2 Suites et convergence

7.29.
À propos de notations. La pire notation possible pour une suite est $(a_n)_n$. Mais que vient faire le second indice n ? Il peut être raisonnable d'écrire $(a_n)_{n \in I}$ lorsqu'on veut dire dans quel ensemble se déplace n. Si nous parlons de *suite*, il faut une sérieuse raison de prendre autre chose que $\mathbb{N}$ comme ensemble d'indices.

Une suite étant une fonction, de la même façon qu'on ne devrait pas dire « la fonction $f(x)$ », mais « la fonction f » ou « la fonction $x \mapsto f(x)$ », nous devrions simplement écrire a pour désigner la suite dont les éléments sont a_n.

Par conséquent, il est parfaitement légal, et même conseillé, d'écrire « $a + b$ » pour la somme des suites a et b. Et il est tout aussi légal d'écrire « $\lim a$ » au lieu de $\lim_{n \to \infty} a_n$.

Le hic est que nous écrivons souvent x la limite de la suite $n \mapsto x_n$. Dans ce cas, nous sommes évidemment obligé d'écrire l'indice n pour parler de la suite.

Tout cela pour dire qu'il faut être souple avec les notations.

7.2.1 Convergence dans un fermé

Proposition 7.30 ([1])**.**
Une suite contenue dans un fermé ne peut converger que vers un élément de ce fermé.

Démonstration. Soient un espace topologique X et un fermé F dans X. Nous supposons que la suite (x_k) soit contenue dans F. Nous allons prouver qu'aucun élément de F^c ne peut être limite.

Soit $a \in F^c$. Vu que le complémentaire de F est un ouvert, et vu le théorème 7.4, il existe un ouvert $\mathcal{O}_a$ contenant a, et contenu dans F^c. Le voisinage $\mathcal{O}_a$ de a ne contient donc aucun élément de la suite (x_k), qui ne peut donc pas converger vers a. □

Corolaire 7.31.
Soit A un sous-ensemble d'un espace topologique X. Toute suite d'éléments de A qui converge, admet pour limite un élément de $\bar{A}$.

Démonstration. Une fois la suite (x_n) fixée, il suffit de remarquer que tous les x_n sont dans $\bar{A}$, et puis d'appliquer la proposition 7.30. □

Lemme 7.32.
Soit $A \subset X$ muni de la topologie induite de X et (x_n) une suite dans A. Si (x_n) converge vers un élément x dans A, alors elle converge aussi vers x dans X.

Démonstration. Soit $\mathcal{O}$ un ouvert autour de x dans X. Alors $A \cap \mathcal{O}$ est un ouvert autour de x dans A et il existe $N \in \mathbb{N}$ tel que si $n \geqslant N$, alors $x_n \in A \cap \mathcal{O} \subset \mathcal{O}$. □

7.2.2 Pour des limites uniques : séparabilité

Notons que l'on a parlé d'*une* limite de suite jusqu'à présent : en effet, s'il existe deux éléments distincts x et y tels que tout ouvert contenant x contient y, alors la définition 7.10 dit que toute suite convergeant vers y converge aussi vers x...

Exemple 7.33.
Oui, il y a moyen de converger vers plusieurs points distincts si l'espace n'est pas super cool. Nous pouvons par exemple [121] considérer la droite réelle munie de sa topologie usuelle et y ajouter un point $0'$ (qui clone le réel 0) dont les voisinages sont les voisinages de 0 dans lesquels nous remplaçons 0 par $0'$. Dans cet espace, la suite $(1/n)$ converge à la fois vers 0 et $0'$.

En fait, on « voit » le problème : on ne peut pas distinguer d'un point de vue topologique le 0 et le $0'$. △

Nous posons la définition suivante, qui nous permettra de donner une assez grande classe d'espaces topologiques dans lesquels nous avons unicité de la limite[7].

Définition 7.34 (Espace topologique séparé)**.**
Si deux points distincts admettent toujours deux voisinages disjoints[8], nous disons que l'espace est ***séparé*** *ou* ***Hausdorff****.*

Attention, cette notion est à ne pas confondre avec :

Définition 7.35 (Espace topologique séparable)**.**
Un espace topologique est ***séparable*** *s'il possède une partie dénombrable[9] dense[10].*

Proposition 7.36.
Dans un espace séparé, si une suite converge, alors sa limite est unique.

Démonstration. Supposons que la suite (x_k) converge vers deux éléments distincts x et y. L'espace étant séparé, il existe deux ouverts $\mathcal{O}_x$ et $\mathcal{O}_y$, disjoints, contenant respectivement x et y. La suite convergeant à la fois vers x et y, il existe k_x et k_y, tels que, si $k \geqslant \max\{k_x, k_y\}$, l'élément x_k est (à la fois) dans $\mathcal{O}_x$ et $\mathcal{O}_y$. Cela est en contradiction avec le fait que ces deux ensembles sont disjoints. □

7. Voir la proposition 7.70.
8. Définition 1.2.
9. Définition 1.33.
10. Définition 7.25.

7.37.
Donc, on pourra parler, avec des espaces séparés, de « la limite d'une suite ». On notera $x_n \to a$, ou $\lim_{n\to\infty} x_n = a$, pour signifier que la suite (x_n) converge vers a.

Proposition 7.38 ([1]).
La convergence de suite pour la topologie de l'espace produit[11] *est équivalente à la convergence des suites « composante par composante ».*

Démonstration. En deux parties

Sens direct Pour simplifier les notations, nous allons considérer le produit de deux espaces. Soit donc $(x_k, y_k) \xrightarrow{X\times Y} (x, y)$ et des ouverts $\mathcal{O}_1$ dans X autour de x et $\mathcal{O}_2$ autour de y dans Y.
La partie $\mathcal{O}_1 \times \mathcal{O}_2$ est ouverte dans $X \times Y$. Donc il existe K tel que $k > K$ implique $(x_k, y_k) \in \mathcal{O}_1 \times \mathcal{O}_2$.
Nous avons prouvé que pour tout ouvert $\mathcal{O}_1$ autour de x il existe K tel que $k > K$ implique $x_k \in \mathcal{O}_1$. Donc $x_k \xrightarrow{X} x$. Idem pour y.

Dans l'autre sens klm

□

Lemme 7.39 ([1]).
Soit un espace topologique X. Soient dans X une suite (x_n) et un élément x tels que toute sous-suite de (x_n) contient une sous-suite convergente vers x. Alors $x_n \to x$.

Démonstration. Supposons que (x_n) ne converge pas vers x. Il existe alors un ouvert $\mathcal{O}$ autour de x tel que pour tout $N > 0$, il existe $n \geqslant N$ tel que x_n n'est pas dans $\mathcal{O}$.

Cela nous permet de construire une sous-suite de (x_n) composée d'éléments hors de $\mathcal{O}$. Aucune sous-suite de cette sous-suite ne peut converger vers x. □

7.2.3 Fonctions équivalentes

Proposition-définition 7.40 ([122]).
Soit un espace topologique X et $D \subset X$. Soient encore des fonctions $f, g\colon D \to \mathbb{C}$ et un point $a \in \mathrm{Adh}(D)$[12].

Nous définissons sur $\mathrm{Fun}(D, \mathbb{C})$ la relation $f \sim_a g$ lorsque qu'il existe un voisinage V de a dans X et une fonction $\alpha\colon V \to \mathbb{R}$ telles que

(1) $\lim_{x\to a} \alpha(x) = 0$,

(2) pour tout $x \in (V \cap D)\backslash\{a\}$,
$$f(x) = \big(1 + \alpha(x)\big)g(x). \tag{7.18}$$

Cette relation est une relation d'équivalence.

Lorsque $f \sim_a g$, nous disons que f et g sont ***équivalentes*** *en a.*

Notons que la notion d'équivalence de fonctions, de même que la notion de limite, ne dépend pas des valeurs exactes atteintes par les fonctions au point.

Lemme 7.41.
Si f et g sont équivalentes en a, et si g ne s'annule pas sur un voisinage de a, alors pour tout $\epsilon > 0$, il existe r tel que
$$\frac{f(x)}{g(x)} \in B(1, \epsilon) \tag{7.19}$$
pour tout $x \in B(a, r)$.

Démonstration. Nous considérons un voisinage V de a sur lequel en même temps :

11. Définition 7.12.
12. Adhérence ou fermeture, c'est la même chose. Voir la définition 7.22 et le lemme 7.23.

— la fonction α de la définition d'équivalence est définie,

— $|\alpha(x)| < \epsilon$ pour tout $x \in V$,

— $g(x) \neq 0$, pour tout $x \in V$.

Ensuite nous considérons $r > 0$ tel que $B(a, r) \subset V$. En divisant la condition (7.18) par $g(x)$ nous trouvons

$$\frac{f(x)}{g(x)} = 1 + \alpha(x). \tag{7.20}$$

Donc

$$|\frac{f(x)}{g(x)} - 1| = |\alpha(x)| \leqslant \epsilon, \tag{7.21}$$

ce qu'il fallait prouver. □

7.3 Connexité

L'idée de la connexité, c'est de s'assurer qu'un ensemble est « d'un seul tenant ».

Définition 7.42.
*Lorsque E est un espace topologique, nous disons qu'un sous-ensemble A est **non connexe** quand on peut trouver des ouverts O_1 et O_2 disjoints tels que*

$$A = (A \cap O_1) \cup (A \cap O_2), \tag{7.22}$$

*et tels que $A \cap O_1 \neq \emptyset$, et $A \cap O_2 \neq \emptyset$. Si un sous-ensemble n'est pas non-connexe, alors on dit qu'il est **connexe**.*

Une autre façon d'exprimer la condition (7.22) est de dire que A n'est pas connexe quand il est contenu dans la réunion de deux ouverts disjoints qui intersectent tous les deux A.

Proposition-définition 7.43 ([123])**.**
Soient un espace topologique E et un point $x \in E$.

(1) La réunion de toutes les parties connexes de E contenant x est connexe.

(2) Cette réunion est la plus grande (au sens de la relation d'inclusion) de toutes les parties connexes de E contenant x.

*La réunion de toutes les parties connexes de E contenant x est nommée **composante connexe** de x dans E.*

Démonstration. Pas de preuve pour l'instant. Si vous en connaissez une vous pouvez soit l'ajouter, soit me l'envoyer. □

Proposition 7.44.
Soit X un espace topologique. Les conditions suivantes sont équivalentes.

(1) L'espace X est connexe.

(2) Si $X = A \sqcup B$ avec A et B fermés disjoints dans X, alors $A = \emptyset$ ou $B = \emptyset$.

(3) Si $A \subset X$ avec A ouvert et fermé en même temps, alors $A = \emptyset$ ou $A = X$.

Nous verrons plus tard (proposition 7.89) une autre caractérisation de la connexité.

Proposition 7.45.
Si $A \subset X$ est connexe et si $A \subset B \subset \bar{A}$, alors B est connexe.

Proposition 7.46.
Stabilité de la connexité par union.

(1) Une union quelconque de connexes ayant une intersection non vide est connexe.

(2) *Pour tout $n \in \mathbb{N}, n > 0$, si $A_1, \dots, A_n$ sont des connexes de X avec $A_i \cap A_{i+1} \neq \varnothing$, alors l'union $\bigcup_{i=1}^n A_i$ est connexe.*

Démonstration. Point par point.

(1) Soient $\{C_i\}_{i\in I}$ un ensemble de connexes et un point p dans l'intersection : $p \in \bigcap_{i\in I} C_i$. Supposons que l'union ne soit pas connexe. Alors nous considérons A et B, deux ouverts disjoints recouvrant tous les C_i et ayant chacun une intersection non vide avec l'union.

Supposons pour fixer les idées que $p \in A$ et prenons $x \in B \cap \bigcup_{i\in I} C_i$. Il existe un $j \in I$ tel que $x \in C_j$. Avec tout cela nous avons

(1a) $C_j \subset A \cup B$ parce que $A \cup B$ recouvre tous les C_i,

(1b) $C_j \cap A \neq \varnothing$ parce que p est dans l'intersection,

(1c) $C_j \cap B \neq \varnothing$ parce que x est dans cette intersection.

Cela contredit le fait que C_j soit connexe.

(2) Pour la seconde partie nous procédons de proche en proche [13]. D'abord $A_1 \cup A_2$ est connexe par la première partie, ensuite $(A_1 \cup A_2) \cup A_3$ est connexe parce que les connexes $A_1 \cup A_2$ et A_3 ont un point d'intersection par hypothèse, et ainsi de suite.

□

7.4 Compacité

La compacité est le thème 12.

7.4.1 Définition et notions connexes

Soit E, un sous-ensemble de $\mathbb{R}$. Nous pouvons considérer les ouverts suivants :

$$\mathcal{O}_x = B(x, 1) \tag{7.23}$$

pour chaque $x \in E$. Évidemment,

$$E \subseteq \bigcup_{x\in E} \mathcal{O}_x. \tag{7.24}$$

Cette union contient en général de nombreuses redondances. Si par exemple $E = [-10, 10]$, l'élément $3 \in E$ est contenu dans $\mathcal{O}_{3.5}$, $\mathcal{O}_{2.7}$ et bien d'autres. Pire : même si on enlève par exemple $\mathcal{O}_2$ de la liste des ouverts, l'union de ce qui reste continue à être tout E. La question est : *est-ce qu'on peut en enlever suffisamment pour qu'il n'en reste qu'un nombre fini ?*

Définition 7.47.
Soit E, un sous-ensemble de $\mathbb{R}$. Une collection d'ouverts $\mathcal{O}_i$ est un ***recouvrement*** *de E si $E \subseteq \bigcup_i \mathcal{O}_i$.*

Définition 7.48.
Une partie A d'un espace topologique est ***compacte*** *s'il vérifie la propriété de Borel-Lebesgue : pour tout recouvrement de A par des ouverts (c'est-à-dire une collection d'ouverts dont la réunion contient A) on peut extraire un recouvrement fini.*

Remarque 7.49.
Certaines sources (dont wikipédia) disent que pour être compact il faut aussi être séparé [14]. Pour ces sources, un espace qui ne vérifie que la propriété de Borel-Lebesgue est alors dit **quasi-compact**.

13. Parce qu'on a la flemme de faire une preuve par récurrence !
14. Définition 7.34.

7.50.
La définition 7.48 en cache deux. En effet, si la partie A est l'espace topologique lui-même, cela définit un espace topologique compact. Un espace topologique est compact *en soi* lorsque de tout recouvrement par des ouverts, nous pouvons extraire un sous-recouvrement fini. Dans ce cas, si X est l'espace et si $\{A_i\}_{i\in I}$ est le recouvrement, nous avons $X = \bigcup_{i\in I} A_i$ et non une simple inclusion $X \subset \bigcup_{i\in I} A_i$.

Lemme 7.51.
Si K est une partie compacte de l'espace topologie X, alors K est un espace topologique compact pour la topologie induite[15] *de X.*

Démonstration. Nous notons τ la topologie de X et τ_K la topologie induite de X vers K, c'est-à-dire

$$\tau_K = \{\mathcal{O} \cap K \text{ tel que } \mathcal{O} \in \tau\}. \tag{7.25}$$

Soient des ouverts $A_i \in \tau_K$ ($i \in I$ où I est un ensemble quelconque) tels que $\bigcup_i A_i = K$. Pour chaque $i \in I$, il existe un $\mathcal{O}_i \in \tau$ tel que $A_i = K \cap \mathcal{O}_i$. Nous avons

$$K = \bigcup_{i\in I}(K \cap \mathcal{O}_i) \subset \bigcup_{i\in I} \mathcal{O}_i. \tag{7.26}$$

Donc les $\mathcal{O}_i$ forment un recouvrement de K par des ouverts de X. Vu que K est une partie compacte de X, il existe un sous-ensemble fini J de I tel que

$$K \subset \bigcup_{j\in J} \mathcal{O}_i. \tag{7.27}$$

Nous avons donc aussi

$$K \subset \bigcup_{j\in J} K \cap \mathcal{O}_i = \bigcup_{j\in J} A_j. \tag{7.28}$$

Nous avons prouvé que $\{A_j\}_{j\in J}$ est un recouvrement fini de K par des ouverts de K. Donc K est un espace topologique compact. □

Définition 7.52.
Une partie d'un espace topologique est ***relativement compact*** *si son adhérence est compacte.*

Définition 7.53.
Un espace topologique est ***localement compact*** *si tout élément possède un voisinage compact.*

Définition 7.54 (Séquentiellement compact)**.**
Nous disons qu'un espace topologique est ***séquentiellement compact*** *si toute suite admet une sous-suite convergente.*

Définition 7.55.
Un espace topologique est ***dénombrable à l'infini*** *s'il est réunion dénombrable de compacts.*

Définition 7.56.
Une famille $\mathcal{A}$ de parties de X a la ***propriété d'intersection finie non vide*** *si tout sous-ensemble fini de $\mathcal{A}$ a une intersection non vide.*

Proposition 7.57.
Soient X un espace topologique et $K \subset X$. Les propriétés suivantes sont équivalentes :

(1) K est compact.

(2) Si $\{F_i\}$ est une famille de fermés telle que $\bigcap_{i\in I} F_i \cap K = \emptyset$, alors il existe une partie finie non vide A de I tel que $\bigcap_{i\in A} F_i \cap K = \emptyset$.

15. Définition 7.15.

(3) *Si $\{F_i\}_{i\in I}$ est une famille de fermés telle que pour tout choix de A fini dans I, $\bigcap_{i\in A} F_i \bigcap K \neq \emptyset$, alors l'intersection complète est non vide : $\bigcap_{i\in I} F_i \bigcap K \neq \emptyset$.*

(4) *Toute famille de fermés de X, à laquelle K est joint, et qui a la propriété d'intersection finie non vide, a une intersection non vide.*

Démonstration. Les propriétés (3) et (2) sont équivalentes par contraposition. De plus le point (4) est une simple[16] reformulation en français de la propriété (3).

Prouvons (1) $\Rightarrow$ (2). Soit $\{F_i\}_{i\in I}$ une famille de fermés tels que $K \bigcap_{i\in I} F_i = \emptyset$. Les complémentaires $\mathcal{O}_i$ de F_i dans X recouvrent K et donc on peut en extraire un sous-recouvrement fini :

$$K \subset \bigcup_{i\in A} \mathcal{O}_i \tag{7.29}$$

pour un certain sous-ensemble fini A de I. Pour ce même choix A, nous avons alors aussi

$$\bigcap_{i\in A} F_i \cap K = \emptyset. \tag{7.30}$$

L'implication (2) $\Rightarrow$ (1) est la même histoire de passage aux complémentaires. □

Le théorème 7.304 est en général celui qu'on nomme « théorème des fermés emboîtés », mais le corolaire suivant en mériterait également le nom.

Corolaire 7.58 ([1])**.**
Soient un espace topologique compact X et une suite $(F_i)_{i\in\mathbb{N}}$ de fermés emboîtés[17] *dans X telle que*

$$\bigcap_{i\in\mathbb{N}} F_i = \emptyset. \tag{7.31}$$

Alors il existe $j_0 \in \mathbb{N}$ tel que $F_i = 0$ pour tout $i \geqslant j_0$.

Démonstration. La proposition 7.57 nous dit qu'il existe une partie finie non vide J de $\mathbb{N}$ telle que $\bigcup_{j\in J} F_j = \emptyset$. Si $j_0 = \min(J)$, alors $F_j \subset F_{j_0}$ pour tout $j \in J$ et nous avons

$$\emptyset = \bigcap_{j\in J} F_j = F_{j_0}. \tag{7.32}$$

Dès que $F_{j_0} = \emptyset$, tous les suivants sont également vides. □

7.4.2 Base de topologie

Définition 7.59 (Base de topologie[124])**.**
Une famille $\mathcal{B}$ d'ouverts de X est une ***base de la topologie*** *de X si pour tout $x \in X$ et pour tout voisinage V de x, il existe $A \in \mathcal{B}$ tel que $x \in A \subset V$.*

Proposition 7.60.
Si $\mathcal{B}$ est une base de la topologie de X alors tout ouvert de X est une union d'éléments de $\mathcal{B}$.

Démonstration. Soit $\mathcal{O}$ un ouvert de X ; pour chaque $x \in \mathcal{O}$ nous considérons un ouvert $U(x)$ tel que $x \in U(x) \subset \mathcal{O}$ (possible par le théorème 7.4). Nous prenons alors $B(x) \in \mathcal{B}$ tel que

$$x \in B(x) \subset U(x) \subset \mathcal{O}. \tag{7.33}$$

Alors nous avons $\mathcal{O} = \bigcup_{x\in\mathcal{O}} B(x)$. □

Notons toutefois que nous sommes loin d'avoir une union dénombrable en général.

16. Enfin, simple... il faut remarquer que dans la formulation de (4), les intersections peuvent ne pas faire intervenir K, mais, au final, on s'en moque.
17. C'est-à-dire que $F_{i+1} \subset F_i$.

7.4.3 Quelques propriétés

Lemme 7.61.
Une partie K d'un espace topologique est compacte si et seulement si de tout recouvrement par des ouverts d'une base de topologie nous pouvons extraire un sous-recouvrement fini.

Remarquons que la partie qui est réellement à prouver est que, si « ça marche » pour des ouverts d'une base de topologie, alors « ça marche » pour tous types d'ouverts.

Démonstration. Soit K une partie d'un espace topologique et $\{\mathcal{O}_i\}_{i\in I}$ un recouvrement de K par des ouverts. Chacun des $\mathcal{O}_i$ est une union d'éléments de la base de topologie par la proposition 7.60 : disons $\mathcal{O}_i = \bigcup_{j\in J_i} A_{(i,j)}$. Soit $J = \{j = (i, j_i) | i \in I, j_i \in J_i\}$; alors nous obtenons $\bigcup_{j\in J} A_j = \bigcup_{i\in I} \mathcal{O}_i$.

Par hypothèse nous pouvons extraire un ensemble fini $J_0 \subset J$ tel que $K \subset \bigcup_{j\in J_0} A_j$. Par construction chacun des A_j est inclus dans (au moins) un des $\mathcal{O}_i$. Le choix d'un élément de I pour chacun des éléments de J_0 donne une partie finie I_0 de I telle que $K \subset \bigcup_{j\in J_0} A_j \subset \bigcup_{i\in I_0} \mathcal{O}_i$. □

Exemple 7.62 (Un compact non fermé)**.**
En général, un compact n'est pas toujours fermé. Si nous prenons par exemple un ensemble X de plus de deux points muni de la topologie grossière $\{\emptyset, X\}$. Toutes les parties de cet espace sont compactes, mais les seuls fermés sont $\{\emptyset, X\}$. Toutes les autres parties sont alors compactes et non fermées. △

Proposition 7.63 ([125])**.**
Tout compact d'un espace topologique séparé est fermé.

Démonstration. Soient X un espace séparé et K compact dans X. Nous considérons $y \in \complement K$ et, par hypothèse de séparation, pour chaque $x \in K$ nous considérons un voisinage ouvert V_x de x et un voisinage ouvert [18] W_x de y tels que $V_x \cap W_x = \emptyset$. Bien entendu les V_x forment un recouvrement de K par des ouverts dont nous pouvons extraire un sous-recouvrement fini : soit S fini dans K tel que

$$K \subset \bigcup_{x\in S} V_x. \tag{7.34}$$

L'ensemble $W = \bigcap_{x\in S} W_x$ est une intersection finie d'ouverts autour de y et est donc un ouvert autour de y.

Montrons que $W \cap K = \emptyset$. Soit $a \in K$; par définition de S, il existe $s \in S$ tel que $a \in V_s$. Par conséquent, a n'est pas dans W_s et donc pas non plus dans W.

L'ouvert W prouve que y est dans l'intérieur du complémentaire de K, et comme y est arbitraire, nous concluons que le complémentaire de K est ouvert (théorème 7.4), en d'autres termes, que K est fermé. □

Lemme 7.64 ([126])**.**
Une partie fermée d'une compact est elle-même compacte.

Démonstration. Soient F fermé dans un compact K et $\{\mathcal{O}_i\}_{i\in I}$ un recouvrement de F par des ouverts. Vu que F est fermé, F^c est ouvert et $\{\mathcal{O}_i\}_{i\in I} \cup \{K\backslash F\}$ est un recouvrement de K par des ouverts. Si nous en extrayons un sous-recouvrement fini, c'est un recouvrement de F, et en supprimant éventuellement l'ouvert $K\backslash F$, ça reste un sous-recouvrement fini de F tout en étant extrait de $\{\mathcal{O}_i\}_{i\in I}$. □

Proposition 7.65.
Si V est une partie de l'espace topologique X muni de la topologie induite [19] *τ_V de celle de X, et si K est un compact de (V, τ_V) alors K est un compact de (X, τ_X).*

18. Oui, la notation du voisinage peut surprendre, mais elle est quand même pratique pour ce qu'on veut en faire.
19. Définition 7.15.

Démonstration. Soient $(\mathcal{O}_\alpha)_{\alpha\in A}$ des ouverts de X recouvrant K. Alors les ensembles $V\cap\mathcal{O}_\alpha$ recouvrent également K, mais sont des ouverts de V. Donc il en existe un sous-recouvrement fini. Soient donc $(V\cap\mathcal{O}_i)_{i\in I}$ recouvrant K avec I un sous-ensemble fini de A. Les ensembles $(\mathcal{O}_i)_{i\in I}$ recouvrent encore K et sont des ouverts de X. □

7.4.4 Compactifié d'Alexandrov

Proposition-définition 7.66 ([127]).
Soit un espace topologique localement compact[20] *X. Nous considérons un élément $\omega\notin X$ et l'ensemble $\hat{X}=X\cup\{\omega\}$. Nous nommons « ouverts de $\hat{X}$ » les parties suivantes :*

- *les ouverts de X,*
- *les parties de la forme $K^c\cup\{\omega\}$ où K est compact de X. Ici, le complémentaire de K est pris dans X, pas dans $\hat{X}$.*

Alors $\hat{X}$ est un espace topologique compact (cela justifie le nom « ouvert » donné aux parties sus-définies).

Oh bien entendu les plus férus de questions embarrassantes demanderont, si X est l'espace considéré, où prendre ce ω ? Quel « objet » exist en-dehors de X ? Qui m'assure que X n'est pas tellement grand que tout est dedans ? Je vous laisse dormir sur ces questions ; sachez que X lui-même n'est certainement pas un élément de X.

En ce qui concerne $\mathbb{R}$ auquel nous pouvons attacher deux infinis ($+\infty$ et $-\infty$), ce sera la définition 11.24.

Pour $\mathbb{C}$, nous donnerons une caractérisation de la limite en ∞ dans le lemme 11.93.

7.5 Limites et continuité de fonctions

7.5.1 Limites

Définition 7.67 (Limite d'une fonction, thème 17).
*Soient X et Y des espaces topologiques, et un point d'accumulation a de X. Soit encore une fonction $f\colon X\to Y$. L'élément $y\in Y$ est une **limite** de f en a si pour tout voisinage W de y (pour la topologie de Y), il existe un voisinage V de a dans X tel que*

$$f\big(V\backslash\{a\}\big)\subset W. \tag{7.35}$$

Si un tel élément est unique[21]*, alors nous disons que cet élément est la **limite** de f et nous notons*

$$\lim_{x\to a} f(x)=y. \tag{7.36}$$

7.68.
Souvent, nous considérons une fonction $f\colon D\to\mathbb{R}$ avec $D\subset\mathbb{R}$. Dans ce cas, le X de la définition de la limite est D muni de la topologie induite de $\mathbb{R}$ vers D. Dans ce cas, la condition (7.35) s'écrit sous la forme

$$f\big(V\cap D\backslash\{a\}\big)\subset W. \tag{7.37}$$

où V est un voisinage de a dans $\mathbb{R}$ et non un voisinage de D dans D.

Remarque 7.69.
Nous ne saurions trop insister sur le fait que la valeur de f en a n'intervient pas dans la définition de la limite de f en a. Il n'est même pas pas nécessaire que f soit définie en a pour que l'on puisse parler de limite de f en a. Par exemple nous avons

$$\lim_{x\to 1}\frac{x^2-1}{x-1}=2, \tag{7.38}$$

20. Définition 7.53.
21. Rappelons que ce n'est pas toujours le cas, mais que ça l'est si l'espace topologique est séparé – définition 7.34.

alors que la fonction n'est pas définie en $x = 1$.

Plus généralement, un peu par principe, toutes les fois que la notion de limite apporte une information, le point où l'on prend la limite est spécial. Sinon on ne calculerait pas la limite, mais on regarderait directement la valeur de la fonction. Cela est typiquement le cas lorsque nous verrons les dérivées. En effet, regardons (en faisans du semblant d'anticiper) la définition (11.105). Dans la formule

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}, \tag{7.39}$$

la fonction sur laquelle nous prenons la limite n'est *jamais* définie en $x = a$.

Cela est intimement lié à ce qu'on raconte dans 7.7.17.

Proposition 7.70 (Unicité de la limite pour un espace séparé)**.**
Soient X un espace topologique, A une partie de X et Y un espace topologique séparé[22]*. Nous considérons une fonction $f\colon A \to Y$. Si $a \in \bar{A}$, alors f admet au plus une limite en a.*

Démonstration. Soient y et y' des limites de f en a, ainsi que des voisinages V et V' de y et y'. Nous prenons également les voisinages W et W' correspondants :

$$\begin{cases} f(W \cap A) \subset V & \text{(7.40a)} \\ f(W' \cap A) \subset V'. & \text{(7.40b)} \end{cases}$$

Quitte à prendre des sous-ensembles nous pouvons supposer que W et W' sont ouverts. Il s'ensuit alors que :

— l'ensemble $W \cap W'$ est un ouvert contenant a et intersecte donc A ;

— l'ensemble $(W \cap W') \cap A$ est donc non vide ;

— et donc, $f(W \cap W' \cap A)$ est aussi non vide.

Mais

$$f(W \cap W' \cap A) \subset f(W \cap A) \subset V, \tag{7.41}$$

et

$$f(W \cap W' \cap A) \subset f(W' \cap A) \subset V', \tag{7.42}$$

d'où V et V' ont une intersection. Puisque ces ensembles sont arbitraires, nous avons prouvé que tout voisinage de y et tout voisinage de y' ont une intersection non vide ; étant donné que Y est séparé, nous devons avoir $y = y'$. □

7.5.2 Continuité

7.5.2.1 Définitions et propriétés

La définition suivante est *la* définition de la continuité dans tous les cas.

Définition 7.71 (Fonction continue[128])**.**
Deux définitions :

(1) Soient une fonction $f\colon X \to Y$ entre les espaces topologiques X et Y et un point $a \in X$. Nous disons que f est ***continue*** *en a si pour tout ouvert W contenant $f(a)$, il existe un voisinage V de a dans X tel que $f(V) \subset W$.*

(2) Une fonction $f\colon X \to Y$ est ***continue*** *sur X si pour tout ouvert $\mathcal{O}$ de Y, l'ensemble*

$$f^{-1}(\mathcal{O}) = \{x \in X \text{ tel que } f(x) \in \mathcal{O}\} \tag{7.43}$$

est ouvert dans X.

22. Définition 7.34.

7.72.
Lorsque nous écrivons $f\colon X \to Y$, nous entendons que f est définie sur tout X, mais pas qu'elle soit surjective sur Y. En particulier, pour que f soit continue en a, il faut que a soit dans le domaine de f.

Dans le cas de fonctions $\mathbb{R} \to \mathbb{R}$, l'espace X sera la partie de $\mathbb{R}$ sur laquelle f sera définie, et la topologie sera la topologie induite de $\mathbb{R}$.

Exemple 7.73 ([1])**.**
Un truc bien avec la définition 7.71(1) est que la continuité de f en un point est définie pour tout point du domaine ; pas seulement les points d'accumulation. Soit par exemple une fonction simple

$$\begin{aligned} f\colon \{a\} &\to \mathbb{R} \\ a &\mapsto 4. \end{aligned} \tag{7.44}$$

Si W est un ouvert de $\mathbb{R}$ contenant 4, nous avons l'ouvert $V = \{a\}$ tel que $f(V) \subset W$. Donc f est continue au point 4.

Mais f est également continue sur $\{4\}$ en tant qu'espace topologique. En effet, si W est un ouvert de $\mathbb{R}$, l'ensemble $f^{-1}(W)$ est soit $\emptyset$ soit $\{a\}$. Dans les deux cas c'est un ouvert. △

La proposition 7.303 donnera des détails sur ce qu'il se passe lorsque l'espace est métrique.

Théorème 7.74.
Une fonction $f\colon X \to Y$ est une fonction continue si et seulement si elle est continue en chacun des points de X.

Démonstration. En deux parties.

Sens direct Nous supposons que f est une fonction continue. Soient $a \in X$ et W un voisinage de $f(a)$. Nous considérons $\mathcal{O}$, un voisinage ouvert de $f(a)$ contenu dans W ; l'ensemble $f^{-1}(\mathcal{O})$ est alors un ouvert contenant a, et l'image de $f^{-1}(\mathcal{O})$ par f est bien entendu contenue dans W.

Sens inverse Soit $\mathcal{O}$ un ouvert de Y. Pour prouver que $f^{-1}(\mathcal{O})$ est un ouvert de X, nous allons considérer un élément $a \in f^{-1}(\mathcal{O})$ et montrer qu'il existe un voisinage ouvert de a contenu dans $f^{-1}(\mathcal{O})$; le théorème 7.4 nous assurera alors que $f^{-1}(\mathcal{O})$ est ouvert.

L'ensemble $\mathcal{O}$ est un voisinage ouvert de $f(a)$ parce que a a été choisi dans $f^{-1}(\mathcal{O})$. Donc la continuité de f en a nous assure qu'il existe un voisinage W de a tel que $f(W) \subset \mathcal{O}$. En prenant un ouvert contenant a à l'intérieur de W nous avons un voisinage ouvert de a contenu dans $f^{-1}(\mathcal{O})$.

□

Remarque 7.75.
À cause de l'éventuelle non unicité de la limite, deux fonctions continues et égales sur un sous-ensemble dense ne sont pas spécialement égales. Ce sera vrai sur les espaces métriques et plus généralement pour les espaces séparés. Voir l'exemple 7.33 et la proposition 7.70.

Lemme 7.76 ([1])**.**
Soient une fonction $f\colon X \to Y$, et un point d'accumulation $a \in X$[23]*. La fonction f est continue en a si et seulement si $f(a)$ est une limite de f en a.*

Démonstration. En deux parties.

Sens direct Nous supposons que f est continue en $a \in X$. Soit un voisinage W de $f(a)$ dans Y. Par continuité de f en a, il existe un voisinage V de A tel que $f(V) \subset W$. A forciori, $f\big(V \backslash a\big) \subset W$ comme le demande la définition de la limite.

23. Un point d'accumulation de X n'est pas spécialement dans X, si X est un sous-espace d'un autre. Par exemple 0 est un point d'accumulation de $]0,1[$ dans $\mathbb{R}$. Ici nous supposons que $a \in X$, sinon il n'y a de toutes façons pas de continuité en a.

Sens inverse Nous supposons que $f(a)$ est une limite de $f(x)$ lorsque x tend vers a. Si W est un ouvert de Y contenant $f(a)$, il existe un voisinage V de a dans X tel que $f\big(V\backslash a\big) \subset W$. Mais vu que $f(a) \in W$, nous avons $f(V) \subset W$.

□

7.5.2.2 Continuité séquentielle

Définition 7.77.
*Si X et Y sont deux espaces topologiques, une fonction $f\colon X \to \mathbb{R}$ est **séquentiellement continue** en un point a si pour toute suite convergente $x_n \to a$ dans X nous avons $f(x_n) \to f(x)$ dans Y.*

7.78.
Nous allons maintenant voir deux résultats disant que si une fonction est continue, alors elle peut être permutée avec une limite de suite. Dans le cas des espaces métriques, la proposition 7.268 montrera la réciproque : si pour toute suite $x_n \to a$, nous avons $\lim_{n\to\infty} f(x_n) = y$, alors f a une limite e a qui vaut y.

Proposition 7.79 (Permuter limite et fonction continue[1])**.**
Soient deux espaces topologiques X et Y ainsi qu'une fonction $f\colon X \to Y$. Soit $a \in X$ et $\ell \in Y$. Si

$$\lim_{x\to a} f(x) = \ell, \tag{7.45}$$

alors, pour toute suite (x_k) telle que $x_k \to a$, on a

$$\lim f(x_k) = \ell. \tag{7.46}$$

Démonstration. Nous considérons une suite (x_k) qui converge vers a dans X. Soient V un voisinage de ℓ et W un voisinage de a tels que $f(W) \subset V$ (définition 7.67 de la continuité en un point). Par la convergence $a_k \to a$, il existe N tel que pour tout $k > N$, $a_k \in W$, et donc tel que $f(a_k) \in V$, ce qui donne la continuité séquentielle de f. □

Corolaire 7.80 (Caractérisation séquentielle de la continuité en un point[1])**.**
Un application entre deux espaces topologiques est continue en un point y est séquentiellement continue.

Démonstration. Soit une application $f\colon X \to Y$ entre les espaces topologies X et Y. Nous supposons que f est continue en $a \in X$. Soit une suite convergente $x_k \xrightarrow{X} a$. Nous devons prouver que $f(x_k) \to f(a)$.

Soit un voisinage V de $f(a)$ dans Y. Le fait que f soit continue en a signifie[24] que $f(a)$ est une limite de f en a, c'est-à-dire[25] qu'il existe un voisinage W de a tel que $f(W\backslash\{a\}) \subset V$.

Vu que $x_k \to a$, il existe N tel que $x_k \in W$ pour tout $k \geqslant N$. Pour ces valeurs de k, nous avons $f(x_k) \in V$.

Nous avons prouvé que pour tout voisinage V de $f(a)$ dans Y, il existe N tel que $f(x_k) \in V$ dès que $k \geqslant N$. Cela signifie exactement que $f(x_k) \to f(a)$. □

7.5.2.3 Application réciproque

Définition 7.81 (injection, surjection, bijection)**.**
Soient des ensembles A et B ainsi qu'une application $f\colon A \to B$.

*(1) La fonction f est **injective** si $f(x_1) = f(x_2)$, implique $x_1 = x_2$.*

*(2) La fonction f est **surjective** si tous les éléments de B sont atteints, c'est-à-dire si pour tout $y \in B$ il existe $x \in A$ tel que $f(x) = y$.*

24. C'est la définition 7.71 de la continuité en un point.
25. Définition 7.67 d'être une limite.

(3) *La fonction f est une* ***bijection*** *entre A et B si elle est injective et surjective, c'est-à-dire si pour tout $y \in B$ il existe un unique $x \in A$ tel que $f(x) = y$.*

La surjection et l'injection sont des propriétés bien différentes qu'il convient de prouver séparément. De plus une même « formule » peut définir une application injective, surjective, bijective ou non selon le domaine sur laquelle nous la considérons.

Définition 7.82.
*Soit $f\colon A \to B$ une bijection. L'**application réciproque** de f est la fonction*

$$\begin{aligned} f^{-1}\colon B &\to A \\ y &\mapsto \text{le } x \in A \text{ tel que } f(x) = y. \end{aligned} \tag{7.47}$$

Plus généralement si $f\colon X \to Y$ est une application quelconque et si $S \subset Y$ nous notons

$$f^{-1}(S) = \{x \in X \text{ tel que } f(x) \in S\}, \tag{7.48}$$

et dans le cas où S est réduit à un unique élément y, nous notons $f^{-1}(y)$ au lieu de $f^{-1}\big(\{y\}\big)$. Si de plus $f^{-1}(S)$ est un singleton x, nous noterons $f^{-1}(S) = x$ et non $f^{-1}(S) = \{x\}$.

Les plus acharnés parmi les lecteurs se rendront compte de la différence ontologique fondamentale entre x et $\{x\}$.

Proposition 7.83.
Soit $f\colon A \subset \mathbb{R}^n \to B \subset \mathbb{R}^m$ une bijection continue. Si A est compact, alors $f^{-1}\colon B \to A$ est continue.

Proposition 7.84.
Soient I un intervalle dans $\mathbb{R}$ et $f\colon I \to \mathbb{R}$ une fonction continue strictement monotone. Alors la fonction réciproque $f^{-1}\colon f(I) \to \mathbb{R}$ est continue sur l'intervalle $f(I)$.

7.5.2.4 Homéomorphisme

Définition 7.85.
*Un **homéomorphisme** est une application bijective continue entre deux espaces topologiques dont la réciproque est continue. Deux espaces topologiques X et Y pour lesquels il existe un homéomorphisme entre X et Y, sont dits **isomorphes**.*

7.5.3 Continuité et topologie induite

Proposition 7.86 ([1])**.**
Soit une fonction $f\colon X \to Y$, continue sur l'ouvert A de X au sens où elle est continue en chaque point de A. Alors la fonction restriction $\tilde{f}\colon A \to Y$ est également continue pour la topologie sur A, induite[26] *de X.*

Démonstration. Soit $a \in A$, et montrons que $\tilde{f}$ est continue en a, c'est-à-dire que $\tilde{f}(a) = f(a)$ soit une limite de $\tilde{f}$ en a. Soit un voisinage V de $\tilde{f}(a)$ dans Y. Par la continuité de f, nous avons un ouvert W de X tel que

$$f\big(W \backslash \{a\}\big) \subset V. \tag{7.49}$$

La partie $W \cap A$ est un voisinage de a pour la topologie de A, et vérifie

$$f\big(W \cap A \backslash \{a\}\big) \subset V. \tag{7.50}$$

donc $f(a)$ est une limite de $\tilde{f}$ pour $x \to a$. La fonction $\tilde{f}\colon A \to Y$ est continue en chaque point de A. □

26. Définition 7.15.

Au niveau de la notion de continuité, il n'y a pas trop de changements en passant de $\mathbb{R}$ à $\mathbb{Q}$ muni de la topologie induite.

Exemple 7.87.
Que signifie d'être continue pour une fonction $f\colon \mathbb{Q} \to \mathbb{R}$? D'après le théorème 7.74, il s'agit d'être continue en chaque point de $\mathbb{Q}$. Il s'agit donc, par la définition 7.71 que pour tout $q \in \mathbb{Q}$, le nombre $f(q)$ soit une limite de f pour $x \to q$.

L'espace d'arrivée étant $\mathbb{R}$, un voisinage de $f(q)$ est pris comme une boule de taille ϵ. La continuité de f exige qu'il y ait un voisinage W de q dans $\mathbb{Q}$ tel que pour tout $q' \in W$ (différent que q), $|f(q) - f(q')| < \epsilon$.

Qu'est-ce qu'un ouvert dans $\mathbb{Q}$? D'après la définition 7.15 de la topologie induite, ce sont les ensembles $\mathbb{Q} \cap \mathcal{O}$ avec $\mathcal{O}$ ouvert dans $\mathbb{R}$. Tout cela pour dire que pour tout $\epsilon > 0$, il doit exister $\delta > 0$ tel que pour tout $q' \in \mathbb{Q}$ tel que $0 < |q - q'| < \delta$, nous ayons $|f(q) - f(q')|$.

Bref, c'est exactement le mécanisme usuel de la continuité sur $\mathbb{R}$, sauf qu'il faut seulement considérer les rationnels. △

Lemme 7.88 (Application partielle[1])**.**
Soient trois espaces topologiques X_1, X_2 et Y. Nous considérons une fonction continue $f\colon X_1 \times X_2 \to Y$ ainsi que $x_1 \in X_1$. Alors l'application

$$\begin{aligned} g\colon X_2 &\to Y \\ x_2 &\mapsto f(x_1, x_2) \end{aligned} \tag{7.51}$$

est continue.

Démonstration. Soit un ouvert $\mathcal{O}$ de Y ; par hypothèse sur f, la partie $f^{-1}(\mathcal{O})$ est ouverte dans $X_1 \times X_2$. Notre but est de prouver que $g^{-1}(\mathcal{O})$ est un ouvert de X_2. Nous avons :

$$g^{-1}(\mathcal{O}) = \{x_2 \in X_2 \text{ tel que } (x_1, x_2) \in f^{-1}(\mathcal{O})\}. \tag{7.52}$$

Nous considérons $x_2 \in g^{-1}(\mathcal{O})$ et nous prouvons qu'il existe dans X_2 un voisinage de x_2 entièrement contenu dans $g^{-1}(\mathcal{O})$.

Étant donné que (x_1, x_2) est dans $f^{-1}(\mathcal{O})$ qui est ouvert, la définition 7.12 de la topologie sur $X_1 \times X_2$ nous donne des ouverts A_1 dans X_1 et A_2 dans X_2 tels que

$$(x_1, x_2) \in A_1 \times A_2 \subset f^{-1}(\mathcal{O}). \tag{7.53}$$

Nous montrons à présent que $A_2 \subset g^{-1}(\mathcal{O})$. Soit $y_2 \in A_2$. Par construction $(x_1, y_2) \in A_1 \times A_2 \subset f^{-1}(\mathcal{O})$, donc

$$g(y_2) = f(x_1, x_2) \in \mathcal{O}. \tag{7.54}$$

Cela termine la démonstration. □

7.5.4 Continuité et connexité

Proposition 7.89.
Un espace topologique X est connexe si et seulement si toute application continue $X \to \mathbb{Z}$ est constante.

Proposition 7.90.
L'image d'un ensemble connexe par une fonction continue est connexe.

Démonstration. Soit $f\colon X \to Y$ une application continue entre deux espaces topologiques, et E une partie connexe de X. Nous devons montrer que $f(E)$ est connexe dans Y.

Par l'absurde nous considérons A et B, deux ouverts de Y disjoints recouvrant $f(E)$. Étant donné que f est continue, les ensembles $f^{-1}(A)$ et $f^{-1}(B)$ sont ouverts dans X. De plus ces deux ensembles recouvrent E.

Si x est un élément de $f^{-1}(A) \cap f^{-1}(B)$, alors $f(x) \in A \cap B$, ce qui est impossible parce que nous avons supposé que A et B étaient disjoints. Par conséquent $f^{-1}(A)$ et $f^{-1}(B)$ sont deux ouverts disjoints recouvrant E. Contradiction avec la connexité de E. Nous concluons que $f(E)$ est connexe. □

Une application de ce théorème sera le théorème de valeurs intermédiaires 11.50.

Exemple 7.91.
Les espaces topologiques $\mathbb{R}$ et $\mathbb{R}^2$ ne sont pas homéomorphes. △

Démonstration. Supposons par l'absurde que $f\colon \mathbb{R} \to \mathbb{R}^2$ soit un homéomorphisme. Nous posons $E = f\big(\mathbb{R}\backslash\{0\}\big)$ et $z_0 = f(0)$. Vu que f est bijective nous avons

$$E = \mathbb{R}^2\backslash\{z_0\}, \tag{7.55}$$

qui est connexe.

Vu que E est connexe et que f^{-1} est continue, la proposition 7.90 nous dit que $f^{-1}(E)$ est connexe. Mais par définition, $f^{-1}(E) = \mathbb{R}\backslash\{0\}$ qui n'est pas connexe. □

7.5.5 Continuité et compacité

Théorème 7.92.
L'image d'un compact[27] *par une fonction continue est un compact.*

Dans le cadre des espaces vectoriels normés, ce théorème est démontré en la proposition 7.299.

Démonstration. Soit $K \subset X$, un ensemble compact, et regardons $f(K)$; en particulier, nous considérons Ω, un recouvrement de $f(K)$ par des ouverts. Nous avons que

$$f(K) \subseteq \bigcup_{\mathcal{O}\in\Omega} \mathcal{O}. \tag{7.56}$$

Par construction, nous avons aussi

$$K \subseteq \bigcup_{\mathcal{O}\in\Omega} f^{-1}(\mathcal{O}), \tag{7.57}$$

en effet, si $x \in K$, alors $f(x)$ est dans un des ouverts de Ω, disons $f(x) \in \mathcal{O}$, et évidemment, $x \in f^{-1}(\mathcal{O})$. Les $f^{-1}(\mathcal{O})$ recouvrent le compact K, et donc on peut en choisir un sous-recouvrement fini, c'est-à-dire un choix de $\{f^{-1}(\mathcal{O}_1), \dots, f^{-1}(\mathcal{O}_n)\}$ tels que

$$K \subseteq \bigcup_{i=1}^{n} f^{-1}(\mathcal{O}_i). \tag{7.58}$$

Dans ce cas, nous avons que

$$f(K) \subseteq \bigcup_{i=1}^{n} \mathcal{O}_i, \tag{7.59}$$

ce qui prouve la compacité de $f(K)$. □

7.6 Topologie, distances et normes

Certains ensembles ont plus de structures qu'une topologie. Nous fixons quelques bases maintenant, et nous détaillerons certains résultats plus tard.

27. Définition 7.48.

7.6.1 Distance et topologie métrique

Définition 7.93.
Si E est un ensemble, une ***distance*** *sur E est une application $d\colon E \times E \to \mathbb{R}$ telle que pour tout $x, y \in E$,*

(1) $d(x, y) \geqslant 0$

(2) $d(x, y) = 0$ *si et seulement si* $x = y$,

(3) $d(x, y) = d(y, x)$

(4) $d(x, y) \leqslant d(x, z) + d(z, y)$.

*La dernière condition est l'****inégalité triangulaire****.*

Un couple (E, d) formé d'un ensemble et d'une distance est un ***espace métrique****.*

La définition-théorème suivante donne une topologie sur les espaces métriques en partant des boules.

Théorème-définition 7.94.
Soit (E, d) un espace métrique. Nous définissons les ***boules ouvertes*** *par*

$$B(x, r) = \{y \in E \text{ tel que } d(x, y) < r\}. \tag{7.60}$$

pout tout $x \in E$ et $r > 0$. Alors en posant

$$\mathcal{T} = \big\{\mathcal{O} \subset E \text{ tel que } \forall x \in \mathcal{O}, \exists r > 0 \text{ tel que } B(x, r) \subset \mathcal{O}\big\} \tag{7.61}$$

nous définissons une topologie sur E.

Cette topologie sur E est la ***topologie métrique*** *de (E, d). En présence d'une distance, sauf mention explicite du contraire, c'est toujours cette topologie-là que nous utiliserons.*

Démonstration. D'abord $\varnothing \in \mathcal{T}$ parce que tout élément de l'ensemble vide ... heu ... enfin parce que d'accord hein [28]. Ensuite si $(A_i)_{i\in I}$ sont des éléments de $\mathcal{T}$ et si $x \in \bigcup_{i\in I} A_i$ alors il existe $k \in I$ tel que $x \in A_k$. Par hypothèse il existe une boule $B(x, r) \subset A_k \subset \bigcup_{i\in I} A_i$.

Enfin si $(A_i)_{i\in\{1,\dots,n\}}$ sont des éléments de $\mathcal{T}$ alors pour tout i il existe $r_i > 0$ tel que $B(x, r_i) \subset A_i$. En prenant $r = \min\{r_i\}_{i=1,\dots,n}$ nous avons $B(x, r) \subset \bigcap_{i=1}^n A_i$. □

Remarque 7.95.
Quatre remarques à propos de cette définition.

(1) Cette définition est faite exprès pour respecter le théorème 7.4. Même si, à priori, on aurait dû utiliser la topologie engendrée faite à l'exemple 7.8... mais on peut montrer que les deux topologies sont les mêmes.

(2) Par construction, les boules ouvertes sont une base de la topologie (définition 7.59) des espaces métriques.

(3) Si V est un voisinage de x, alors il existe r tel que $B(x, r) \subset V$.

(4) Tout espace métrique est séparé. En effet, si deux éléments x et y sont distincts, alors en posant $r = d(x, y)/3 > 0$, les boules $B(x, r)$ et $B(y, r)$ sont disjointes. Très pratique pour les limites : elles sont uniques, grâce aux propositions 7.36 et 7.70 !

7.96.
Si vous avez un peu de temps, vous pouvez vérifier que si $\mathbb{K}$ est un corps totalement ordonné, alors avec toutes les définitions de 1.109, en posant $d(x, y) = |x - y|$ nous avons une distance sur $\mathbb{K}$.

De plus, les boules définies en 1.109 sont alors les mêmes que celles définies en (7.60), ce qui donne à tout corps totalement ordonné une structure d'espace topologique.

28. Pour qui ne seraient pas d'accord, allez ajouter $\varnothing$ dans la définition des ouverts et puis c'est tout.

7.6.1.1 Les boules, une base de topologie

Proposition 7.97.
Un espace métrique séparable[29] *accepte une base de topologie dénombrable.*

Soit A dense et dénombrable dans l'espace métrique séparable (E, d). Si $\{a_i\}_{i\in\mathbb{N}}$ est une énumération de A et $\{r_i\}_{i\in\mathbb{N}}$ une énumération de $\mathbb{Q}$, alors

$$\mathcal{B} = \{B(a_i, r_j)\}_{i,j\in\mathbb{N}} \tag{7.62}$$

est une base de la topologie[30] *de E.*

Démonstration. Soient $x \in E$ et V un voisinage de x. Ce dernier contient une boule $B(x, r)$ et quitte à prendre r un peu plus petit nous supposons que $r \in \mathbb{Q}$ (existence d'un tel rationnel par le lemme 1.144).

Soit $a \in A$ avec $\|a - x\| < \frac{r}{3}$ (existe par densité de A dans E) ; nous avons $B(a, \frac{2r}{3}) \subset B(x, r)$ parce que si $y \in B(a, \frac{2r}{3})$ alors

$$\|y - x\| \leqslant \|y - a\| + \|a - x\| < \frac{2}{3}r + \frac{1}{3}r = r. \tag{7.63}$$

La seconde inégalité est stricte parce que les boules sont ouvertes. Le tout montre que $y \in B(x, r)$. Par ailleurs $x \in B(a, \frac{2}{3}r)$ et nous avons trouvé un élément de $\mathcal{B}$ contenant x tout en étant inclus dans V. Cela prouve que $\mathcal{B}$ est bien une base de la topologie de E. □

Remarque 7.98.
Il est vite vu que les cubes ouverts forment aussi une base de la topologie de $\mathbb{R}^n$. Cela est à mettre en rapport avec le fait que toutes les normes sont équivalentes sur $\mathbb{R}^n$ (proposition 7.224).

Voir aussi le corolaire 13.215 qui donnera tout ouvert comme union de pavés presque disjoints.

Définition 7.99.
Soit (X, d) un espace métrique. Un sous-ensemble $A \subset X$ est ***borné*** *s'il existe une boule de X contenant A.*

Proposition 7.100.
Toute réunion finie d'ensembles bornés est un ensemble borné. Toute partie d'un ensemble borné est un ensemble borné.

7.6.1.2 Continuité et compacité

Un résultat important dans la théorie des fonctions sur les espaces vectoriels normés est qu'une fonction continue sur un compact est bornée et atteint ses bornes. Ce résultat sera (dans d'autres cours) énormément utilisé pour trouver des maximums et minimums de fonctions. Le théorème exact est le suivant.

Lemme 7.101 (de Lebesgue[129])**.**
Soit (X, d) un espace métrique tel que toute suite ait une sous-suite convergente à l'intérieur de l'espace. Si $\{V_i\}$ est un recouvrement par des ouverts de X, alors il existe ϵ tel que pour tout $x \in X$, nous ayons $B(x, \epsilon) \subset V_i$ pour un certain i.

Démonstration. Par l'absurde, nous supposons que pour tout n, il existe un $x_n \in X$ tel que la boule $B(x_n, \frac{1}{n})$ n'est contenue dans aucun des V_i. Ce des x_n nous extrayons une sous-suite convergente (que nous nommons encore (x_n)) et nous posons $x_n \to x$. Pour n assez grand ($\frac{1}{n} < \epsilon$) nous avons $x_n \in B(x, \epsilon)$, donc tous les x_n suivants sont dans le V_i qui contient x. □

29. Qui possède une partie dense dénombrable, définition 7.35.
30. Définition 7.59.

Lemme 7.102 ([129]).
Soit (X, d) un espace métrique tel que toute suite possède une sous-suite convergente. Pour tout $\epsilon > 0$, il existe un ensemble fini $\{x_i\}_{i\in I}$ tel que les boules $B(x_i, \epsilon)$ recouvrent X.

Démonstration. Soit par l'absurde un $\epsilon > 0$ contredisant le lemme. Il n'existe pas d'ensemble finis autour des points duquel les boules de taille ϵ recouvrent X.

Nous construisons par récurrence une suite ne possédant pas de sous-suites convergente. Le premier terme, x_0 est pris arbitrairement dans X. Ensuite si nous en avons N termes, nous savons que les boules de rayon ϵ et centrées en les points $\{x_i\}_{i=1,\dots,N}$ ne recouvrent pas X. Donc nous prenons x_{N+1} hors de l'union de ces boules.

Ainsi nous avons une suite (x_n) dont tous les termes sont à distance plus grande que ϵ les uns des autres. Une telle suite ne peut pas contenir de sous-suite convergente. Contradiction. □

Théorème 7.103 (Bolzano-Weierstrass[129], thème 12).
Un espace métrique est compact si et seulement si toute suite admet une sous-suite qui converge à l'intérieur de l'espace.

Démonstration. Soient X un espace métrique compact et (x_n) une suite dans X. Nous considérons la suite de fermés emboîtés

$$X_n = \overline{\{x_k \text{ tel que } k > n\}}. \tag{7.64}$$

Ce sont des fermés ayant la propriété d'intersection finie non vide, et donc la proposition 7.57 nous dit qu'ils ont une intersection non vide. Un élément de cette intersection est automatiquement un point d'accumulation de la suite[31].

Nous passons à l'autre sens. Nous supposons que toute suite dans X contient une sous-suite convergente, et nous considérons $\{V_i\}_{i\in I}$, un recouvrement de X par des ouverts. Par le lemme 7.101, nous considérons un ϵ tel que pour tout x, il existe un $i \in I$ avec $B(x, \epsilon) \subset V_i$. Par le lemme 7.102, nous considérons un ensemble fini $\{y_i\}_{i\in A}$ tel que le boules $B(y_i, \epsilon)$ recouvrent X.

Par construction, chacune de ces boules $B(y_i, \epsilon)$ est contenue dans un des ouverts V_i. Nous sélectionnons donc parmi les V_i le nombre fini qu'il faut pour recouvrir les $B(y_i, \epsilon)$ et donc pour recouvrir X. □

Exemple 7.104 (Non compacité de la boule unité en dimension infinie).
Le théorème de Bolzano-Weierstrass permet de voir tout de suite que la boule unité n'est pas compacte dans un espace vectoriel de dimension infinie : la suite des vecteurs de base ne possède pas de sous-suites convergentes. △

Le théorème de Bolzano–Weierstrass 7.103 a l'importante conséquence suivante.

Théorème 7.105 (Weierstrass).
Une fonction continue à valeurs réelles définie sur un compact est bornée et atteint ses bornes.

Démonstration. Soient K un compact et $f\colon K \to \mathbb{R}$ une fonction continue. Nous désignons par A l'ensemble des valeurs prises par f sur K :

$$A = f(K) = \{f(x) \text{ tel que } x \in K\}. \tag{7.65}$$

Nous considérons le supremum $M = \sup A = \sup_{x\in K} f(x)$ avec la convention comme quoi si A n'est pas borné supérieurement, nous posons $M = \infty$ (voir définition 1.157).

Nous allons maintenant construire une suite (x_n) de deux façons différentes suivant que $M = \infty$ ou non.

(1) Si $M = \infty$, nous choisissons, pour chaque $n \in \mathbb{N}$, un $x_n \in K$ tel que $f(x_n) > n$. Cela est certainement possible parce que si A n'est pas borné, nous pouvons y trouver des nombres aussi grands que nous voulons.

31. Définition 7.27.

(2) Si $M < \infty$, nous savons que pour tout ε, il existe un $y \in A$ tel que $y > M - \varepsilon$. Pour chaque n, nous choisissons donc $x_n \in K$ tel que $f(x_n) > M - \frac{1}{n}$.

Quel que soit le cas dans lequel nous sommes, la suite (x_n) est une suite dans K qui est compact, et donc nous pouvons en extraire une sous-suite convergente à l'intérieur de K par le théorème de Bolzano-Weierstrass 7.103. Afin d'alléger la notation, nous allons noter (x_n) la sous-suite convergente. Nous avons donc

$$x_n \to x \in K. \tag{7.66}$$

Par la proposition 7.80, nous avons que f prend en x la valeur

$$f(x) = \lim_{n\to\infty} f(x_n). \tag{7.67}$$

Donc $f(x) < \infty$. Évidemment, si nous avions été dans le cas où $M = \infty$, la suite x_n aurait été choisie pour avoir $f(x_n) > n$ et donc il n'aurait pas été possible d'avoir $\lim_{n\to\infty} f(x_n) < \infty$. Nous en concluons que $M < \infty$, et donc que f est bornée sur K.

Afin de prouver que f atteint sa borne, c'est-à-dire que $M \in A$, nous considérons les inégalités

$$M - \frac{1}{n} < f(x_n) \leqslant M. \tag{7.68}$$

En passant à la limite $n \to \infty$, ces inégalités deviennent

$$M \leqslant f(x) \leqslant M, \tag{7.69}$$

et donc $f(x) = M$, ce qui prouve que f atteint sa borne M au point $x \in K$. □

Lemme 7.106 ([1]).
Soient des compacts A, B et une fonction continue $f\colon A \times B \to \mathbb{R}$. Alors

$$\sup_{(x,y)\in A\times B} |f(x,y)| = \sup_{x\in A}\big(\sup_{y\in B}|f(x,y)|\big). \tag{7.70}$$

Démonstration. Pour chaque $x \in A$, la fonction $f_x\colon B \to \mathbb{R}$ donnée par $f_x(y) = |f(x,y)|$ est continue et atteint donc sa borne[32] en $y_M(x)$. Notons que cela ne définit pas univoquement $y_M(x)$ parce que f_x peut atteindre son maximum en plusieurs points. L'important est que pour tout x, le nombre $|f\big(x, y_M(x)\big)|$ ne dépend pas du choix de $y_M(x)$ parmi les y qui réalisent le maximum.

Notons (x_0, y_0) le point de $A \times B$ sur lequel $|f|$ réalise son maximum[33] :

$$\sup_{(x,y)\in A\times B} |f(x,y)| = |f(x_0,y_0)|. \tag{7.71}$$

Nous avons d'une part

$$\sup_{x\in A}\big(\sup_{y\in B}|f(x,y)|\big) = \sup_{x\in A}|f\big(x, y_M(x)\big)| \leqslant |f(x_0,y_0)| \tag{7.72}$$

Et d'autre part,

$$\begin{aligned}
\sup_{x\in A}\big(\sup_{y\in B}|f(x,y)|\big) &\leqslant \sup_{x\in A}\sup_{y\in B}|f(x_0,y_0)| && (7.73a)\\
&= |f(x_0,y_0)| && (7.73b)\\
&\leqslant |f\big(x_0, y_M(x_0)\big)| && (7.73c)\\
&\leqslant \sup_{x\in A}|f(x), y_M(x)| && (7.73d)\\
&\leqslant \sup_{x\in A}\big(\sup_{y\in B}|f(x,y)|\big). && (7.73e)
\end{aligned}$$

Vu que les premiers et derniers termes des inégalités (7.73) sont égaux, toutes les inégalités sont en réalité des égalités. En particulier, en en reprenant (7.71),

$$\sup_{(x,y)\in A\times B} |f(x,y)| = |f(x_0,y_0)| = \sup_{x\in A}\big(\sup_{y\in B}|f(x,y)|\big). \tag{7.74}$$

□

32. Théorème 7.105.
33. Encore une fois, ce point n'est pas déterminé de façon unique par cette propriété.

7.6.2 Distance à un ensemble

Définition 7.107.
Si A est une partie de l'espace métrique (X, d), et si $b \in X$, nous définissons

$$d(b, A) = \inf_{y \in A} d(b, y). \tag{7.75}$$

Lemme 7.108 ([1]).
Si A est fermé dans (X, d), et si $b \in X$ vérifie $d(b, A) = 0$, alors $b \in A$.

Démonstration. Vu que A est fermé, le complémentaire A^c est ouvert (c'est la définition 7.2). Supposons que $b \in A^c$. Alors il existe $r > 0$ tel que $B(b, r) \subset A^c$. Si $a \in A$ nous avons alors $d(b, A) \geqslant r$ et donc $d(b, A) \geqslant r > 0$. Cela contredit l'hypothèse $d(b, A) = 0$.

Nous en déduisons que b n'est pas dans A^c et qu'il est donc dans A. □

Exemple 7.109 (Pas avec un ouvert).
En prenant l'ouvert $A =]0, 1[$ dans $\mathbb{R}$ nous avons $d(0, A) = 0$, alors que 0 n'est pas dans A. △

Lemme 7.110 ([1]).
Soient un espace métrique (X, d) ainsi qu'une partie $A \subset X$. Soit $r > 0$. La partie

$$\mathcal{O} = \{x \in X \text{ tel que } d(x, A)\} \tag{7.76}$$

est ouverte.

Démonstration. Soit $y \in \mathcal{O}$; nous avons $d(y, A) < r$. Autrement dit,

$$\inf_{a \in A} d(y, a) < r \tag{7.77}$$

et donc il existe $a \in A$ tel que $d(y, a) < r$. Soit $\delta = d(y, a) < r$. Nous montrons à présent que $B(y, r - \delta)$ est dans $\mathcal{O}$. En effet si $z \in B(y, r - \delta)$, alors

$$d(z, a) \leqslant d(z, y) + d(y, a) < r - \delta + \delta = r. \tag{7.78}$$

□

Lemme 7.111 ([1]).
Si F est un fermé dans (X, d) et si x n'est pas dans F, alors $d(x, F) > 0$.

Lemme 7.112 ([1]).
Si A est une partie de (X, d), alors la fonction

$$\begin{aligned} f \colon \Omega &\to [0, \infty[\\ x &\mapsto d(x, A) \end{aligned} \tag{7.79}$$

est continue.

7.6.3 Norme

Définition 7.113 ([130], thème 7).
Soit E un espace vectoriel (pas spécialement de dimension finie) sur le corps $\mathbb{K}$ ($= \mathbb{R}$ ou $\mathbb{C}$). Une ***norme*** *sur E est une application $N \colon E \to \mathbb{R}^+$ telle que*

(1) $N(x) = 0$ si et seulement si $x = 0$;

(2) $N(\lambda x) = |\lambda| N(x)$ pour tout $\lambda \in \mathbb{R}$ et $x \in E$;

(3) $N(x + y) \leqslant N(x) + N(y)$

pour tout $x, y \in E$ et pour tout $\lambda \in \mathbb{K}$.

La propriété (3) est appelée ***inégalité triangulaire****.*

Un espace vectoriel muni d'une norme est un ***espace vectoriel normé****.*

En prenant $\lambda = -1$ dans la propriété (2), nous trouvons immédiatement que $N(-x) = N(x)$.

Proposition 7.114.
Toute norme N sur l'espace vectoriel E vérifie l'inégalité

$$\big|N(x) - N(y)\big| \leqslant N(x-y) \tag{7.80}$$

pour tout $x, y \in E$.

Démonstration. Nous avons, en utilisant le point (3) de la définition 7.113,

$$N(x) = N(x-y+y) \leqslant N(x-y) + N(y), \tag{7.81a}$$
$$N(y) = N(y-x+x) \leqslant N(y-x) + N(x). \tag{7.81b}$$

Supposons d'abord que $N(x) \geqslant N(y)$. Dans ce cas, en utilisant (7.81a),

$$\big|N(x) - N(y)\big| = N(x) - N(y) \leqslant N(x-y) + N(y) - N(y) = N(x-y). \tag{7.82}$$

Si par contre $N(x) \leqslant N(y)$, alors nous utilisons (7.81b) et nous trouvons

$$\big|N(x) - N(y)\big| = N(y) - N(x) \leqslant N(y-x) + N(x) - N(x) = N(y-x). \tag{7.83}$$

Dans les deux cas, nous avons retrouvé l'inégalité annoncée. □

Cette proposition signifie aussi que

$$-N(x-y) \leqslant N(x) - N(y) \leqslant N(x-y). \tag{7.84}$$

7.115.
Afin de suivre une notation proche de celle de la valeur absolue, à partir de maintenant, la norme d'un vecteur v sera notée $\|v\|$ au lieu de $N(v)$. La proposition 7.114 s'énoncera donc

$$\big|\|x\| - \|y\|\big| \leqslant \|x-y\|. \tag{7.85}$$

Un espace vectoriel E muni d'une norme est, on l'a déjà dit, un **espace vectoriel normé** ; on le notera $(E, \|.\|)$ pour distinguer la norme fixée.

Une autre inégalité utile de temps en temps.

Corolaire 7.116.
Si a et b sont dans un espace vectoriel normé, alors

$$\big|\|a-b\| - \|b\|\big| \leqslant \|a\|. \tag{7.86}$$

Démonstration. Il s'agit seulement de la proposition 7.114 avec $x = a - b$ et $y = -b$. □

Lemme-définition 7.117 (Distance induite par une norme)**.**
Soit un espace vectoriel normé $(E, \|.\|)$. Nous posons

$$d(x,y) = \|x-y\|. \tag{7.87}$$

Alors

(1) d est invariante par translations : $d(a,b) = d(a+u, b+u)$

(2) d est une distance[34] *sur E.*

*C'est la **distance induite** par la norme.*

34. Définition 7.93.

Démonstration. Le fait que la formule (7.87) soit invariante par translations est immédiat. En ce qui concerne le fait que ce soit une distance, le seul point délicat à vérifier est l'inégalité triangulaire. Mais, pour tous $x, y, z \in E$, on a

$$d(x,y) = \|x-y\| = \|x-z+z-y\| \leqslant \|x-z\| + \|z-y\| = d(x,z) + d(z,y). \tag{7.88}$$

□

Corolaire 7.118.
Un espace vectoriel normé est un espace vectoriel topologique : en d'autres mots, l'addition et la multiplication par un élément du corps sont continues.

Nous étudierons plus en détail les espaces vectoriels topologiques à partir de la définition 7.283.

7.7 Topologie sur l'ensemble des réels

Nous allons à présent donner la topologie sur $\mathbb{R}$ et ainsi résoudre les questions laissées en suspens lors de la construction des réels, voir 1.129.

Afin de pouvoir étudier la topologie des espaces métriques, il faut savoir quelques propriétés des réels parce que nous allons étudier la fonction distance qui est une fonction continue à valeurs dans les réels.

La valeur absolue de la définition 1.109(2) permet de définir une norme sur $\mathbb{R}$.

Lemme 7.119.
L'application

$$x \mapsto |x| \tag{7.89}$$

est une norme[35] *sur* $\mathbb{R}$.

Démonstration. Grâce au lemme 1.112 et à la remarque 1.113, on a, pour tous $x,\ y,\ \lambda \in \mathbb{R}$:

(1) $|x| = 0$ implique $x = 0$,

(2) $|\lambda x| = |\lambda||x|$,

(3) $|x+y| \leqslant |x| + |y|$,

et donc, les conditions de la définition 7.113 sont immédiatement vérifiées. □

Définition 7.120 (Topologie sur $\mathbb{R}$ et sur $\mathbb{Q}$)**.**
Le lemme 7.119 donne une norme sur $\mathbb{R}$ et $\mathbb{Q}$ à partir de la valeur absolue. La définition 7.94 donne alors une structure d'espace topologique. Hors cas rarissimes qui seront signalés, nous utiliserons toujours cette topologie sur $\mathbb{R}$ et sur $\mathbb{Q}$.

Proposition 7.121.
Les rationnels sont denses dans les réels[36].

Démonstration. Soient $r \in \mathbb{R}$ et $\epsilon \in \mathbb{R}^+$. Nous devons prouver l'existence d'un rationnel dans $B(x,\epsilon)$. Le lemme 1.144 dit qu'il existe un rationnel dans $]x-\epsilon/2, x+\epsilon/2[$ et donc dans $B(x,\epsilon)$. □

Proposition 7.122 ([1])**.**
Quel que soit le réel r, il existe une suite croissante de rationnels convergente vers r.

Démonstration. Soient $x \in \mathbb{R}$ et $\delta \in \mathbb{R}$; vu que $x - \delta$ et x sont des réels, le lemme 1.144 donne un élément $x_\delta \in \mathbb{Q}$ tel que

$$x - \delta < x_\delta < x. \tag{7.90}$$

35. Définition 7.113.
36. Pour les topologies usuelles données en la définition 7.120.

Il suffit alors de pêcher parmi ces x_δ pour trouver une suite croissante, et on montrera que cette suite converge vers x.

Soit x_0 un rationnel plus petit que x. Nous posons $\delta_0 = x - x_0$ et ensuite :

$$\begin{cases} \delta_i = x - x_i & (7.91a) \\ x_{i+1} = x_{\delta_i/2} \in \mathbb{Q}. & (7.91b) \end{cases}$$

Ainsi nous avons pour tout i les inégalités

$$x_i = x - \delta_i < x - \frac{\delta_i}{2} < x_{i+1} < x. \tag{7.92}$$

La suite (x_i) est donc une suite de rationnels, croissante et toujours plus petite que x. Mais nous avons à chaque étape $\delta_{i+1} < \frac{\delta_i}{2}$, ce qui implique que la suite des δ_i converge vers 0. Soit $\epsilon > 0$. Il existe k_0 tel que pour tout $k > k_0$, $\delta_k < \epsilon$. Pour un tel k, nous avons alors

$$x_{k+1} \in B(x, \frac{\delta_k}{2}) \subset B(x, \epsilon). \tag{7.93}$$

Tous les x_k, pour $k > k_0 + 1$, sont tels que $|x - x_k| < \epsilon$: la suite des x_k converge donc vers x. □

7.7.1 Compacité pour les réels

Pour la définition générale d'un compact, c'est 7.48.

Proposition 7.123.
Les parties compactes de $\mathbb{R}$ sont fermées et bornées.

Démonstration. Prouvons d'abord qu'un ensemble compact est borné. Pour cela, supposons que K est un compact non borné vers le haut [37]. Donc il existe une suite infinie de nombres strictement croissante $x_1 < x_2 < \dots$ tels que $x_i \in K$. Prenons n'importe quel recouvrement ouvert de la partie de K plus petite ou égale à x_1, et complétons ce recouvrement par les ouverts $\mathcal{O}_i =]x_{i-1}, x_i[$. Le tout forme bien un recouvrement de K par des ouverts.

Il n'y a cependant pas moyen d'en tirer un sous recouvrement fini parce que si on ne prend qu'un nombre fini parmi les $\mathcal{O}_i$, on en aura fatalement un maximum, disons $\mathcal{O}_k$. Dans ce cas, les points x_{k+1}, x_{k+1},... ne seront pas dans le choix fini d'ouverts.

Cela prouve que K doit être borné.

Pour prouver que K est fermé, nous allons prouver que le complémentaire est ouvert. Et pour cela, nous allons prouver que si le complémentaire n'est pas ouvert, alors nous pouvons construire un recouvrement de K dont on ne peut pas extraire de sous recouvrement fini.

Si $\mathbb{R}\backslash K$ n'est pas ouvert, il possède un point, disons x, tel que tout voisinage de x intersecte K. Soit $B(x, \epsilon_1)$, un de ces voisinages, et prenons $k_1 \in K \cap B(x, \epsilon_1)$. Ensuite, nous prenons ϵ_2 tel que k_1 n'est pas dans $B(x, \epsilon_1)$, et nous choisissons $k_2 \in K \cap B(x, \epsilon_2)$. De cette manière, nous construisons une suite de $k_i \in K$ tous différents et de plus en plus proches de x. Prenons un recouvrement quelconque par des ouverts de la partie de K qui n'est pas dans $B(x, \epsilon_1)$. Les nombres k_i ne sont pas dans ce recouvrement.

Nous ajoutons à ce recouvrement les ensembles $\mathcal{O} =]k_i, k_{i+1}[$. Le tout forme un recouvrement (infini) par des ouverts dont il n'y a pas moyen de tirer un sous recouvrement fini, pour exactement la même raison que la première fois. □

Théorème 7.124 (Borel-Lebesgue)**.**
Un intervalle de $\mathbb{R}$ est compact si et seulement si il est de la forme $[a, b]$.

Démonstration. Tous les intervalles de $\mathbb{R}$ sont listés dans la proposition 1.158. Un compact est fermé et borné (proposition 7.123). Donc les intervalles dont une borne est $\pm\infty$ ne sont pas compacts. Parmi les intervalles $]a, b[$, $]a, b]$, $[a, b[$ et $[a, b]$, seul le dernier est fermé. Nous avons prouvé que si un intervalle est compact, alors il est de la forme $[a, b]$.

37. Nous laissons à titre d'exercice le cas où K est borné par le haut et pas par le bas.

Nous prouvons à présent l'implication inverse : tous les intervalles de la forme $[a, b]$ sont compacts.

Soit Ω, un recouvrement du segment $[a, b]$ par des ouverts, c'est-à-dire que

$$[a, b] \subseteq \bigcup_{\mathcal{O} \in \Omega} \mathcal{O}. \tag{7.94}$$

Nous notons par M le sous-ensemble de $[a, b]$ des points m tels que l'intervalle $[a, m]$ peut être recouvert par un sous-ensemble fini de Ω. C'est-à-dire que M est le sous-ensemble de $[a, b]$ sur lequel le théorème est vrai. Le but est maintenant de prouver que $M = [a, b]$.

M est non vide En effet, $a \in M$ parce que il existe un ouvert $\mathcal{O} \in \Omega$ tel que $a \in \mathcal{O}$. Donc $\mathcal{O}$ tout seul recouvre l'intervalle $[a, a]$.

M est un intervalle Soient m_1, $m_2 \in M$. Le but est de montrer que si $m' \in [m_1, m_2]$, alors $m' \in M$. Il y a un sous recouvrement fini de l'intervalle $[a, m_2]$ (par définition de $m_2 \in M$). Ce sous recouvrement fini recouvre évidemment aussi $[a, m']$ parce que $[a, m'] \subseteq [a, m_2]$, donc $m' \in M$.

M est une ensemble ouvert Soit $m \in M$. Le but est de prouver qu'il y a un ouvert autour de m qui est contenu dans M. Mettons que Ω' soit un sous recouvrement fini qui contienne l'intervalle $[a, m]$. Dans ce cas, on a un ouvert $\mathcal{O} \in \Omega'$ tel que $m \in \mathcal{O}$. Tous les points de $\mathcal{O}$ sont dans M, vu qu'ils sont tous recouverts par Ω'. Donc $\mathcal{O}$ est un voisinage de m contenu dans M.

M est un ensemble fermé M est un intervalle qui commence en a, en contenant a, et qui finit on ne sait pas encore où. Il est donc soit de la forme $[a, m]$, soit de la forme $[a, m[$. Nous allons montrer que M est de la première forme en démontrant que M contient son supremum s. Ce supremum est un élément de $[a, b]$, et donc il est contenu dans un des ouverts de Ω. Disons $s \in \mathcal{O}_s$. Soit c, un élément de $\mathcal{O}_s$ strictement plus petit que c ; étant donné que s est supremum de M, cet élément c est dans M, et donc on a un sous recouvrement fini Ω' qui recouvre $[a, c]$. Maintenant, le sous recouvrement constitué de Ω' et de $\mathcal{O}_s$ est fini et recouvre $[a, s]$.

Nous pouvons maintenant conclure : le seul intervalle non vide de $[a, b]$ qui soit à la fois ouvert et fermé est $[a, b]$ lui-même (proposition 7.44), ce qui prouve que $M = [a, b]$, et donc que $[a, b]$ est compact [38]. □

Lemme 7.125 ([131]).
Si $a < b \in \mathbb{R}$ alors le segment $[a, b]$ est compact [39].

Démonstration. Soit $\{\mathcal{O}_i\}_{i \in I}$ un recouvrement de $[a, b]$ par des ouverts. Nous posons

$$M = \{x \in [a, b] \text{ tel que } [a, x] \text{ admet un sous-recouvrement fini extrait de } \{\mathcal{O}_i\}_{i \in I}\}. \tag{7.95}$$

Notre but est de prouver que $b \in M$.

a est dans M Le point a est naturellement dans un des $\mathcal{O}_i$. L'intervalle $[a, a]$ est donc recouvert par un seul des $\mathcal{O}_i$.

M est un intervalle Soient $m \in M$ et $m' \in [a, m[$. Le sous-recouvrement fini qui recouvre $[a, m]$ recouvre a fortiori $[a, m']$.

Les trois possibilités restantes À ce niveau de la preuve, il reste trois possibilités pour M soit il est de la forme $[a, c]$ ou $[a, c[$ avec $c < b$, soit il est de la forme $[a, b]$. Nous allons maintenant éliminer les deux premiers cas.

38. Si vous n'aimez pas le coup du fermé et ouvert, le lemme 7.125 donne une autre preuve.
39. Définition 7.48

Ce que M n'est pas D'abord M n'est pas de la forme $[a, c[$ avec $c < b$. Par l'absurde, commençons par considérer $\mathcal{O}_{i_0}$ un ouvert du recouvrement qui contient c ; choisissons $m \in \mathcal{O}_{i_0}$ tel que $m < c$. Alors $m \in M$, et, si nous joignons $\mathcal{O}_{i_0}$ à un recouvrement fini de $[a, m]$ alors nous avons un recouvrement fini de $[a, c]$. On en déduit $c \in M$.

Ensuite M n'est pas de la forme $[a, c]$ avec $c < b$. En effet si on a un recouvrement fini de $[a, c]$ par des ouverts, alors un de ces ouverts contient c et donc contient des éléments de $[a, b]$ plus grands que c.

Nous déduisons que $M = [a, b]$ et qu'il est possible d'extraire un sous-recouvrement fini recouvrant $[a, b]$. □

Lemme 7.126 ([1]).
Si K_1 et K_2 sont des compacts dans $\mathbb{R}$ alors $K_1 \times K_2$ est compact dans $\mathbb{R}^2$.

Démonstration. Soit $\{\mathcal{O}_i\}_{i\in I}$ un recouvrement de $K_1 \times K_2$ par des ouverts ; grâce au lemme 7.61 nous pouvons supposer que ce sont des carrés. Pour chaque $x \in K_1$, l'ensemble $\{x\} \times K_2$ est compact et donc recouvert par un nombre fini des $\mathcal{O}_i$. Soit R_x un ensemble fini des $\mathcal{O}_i$ recouvrant $\{x\} \times K_2$.

Vu que R_x est une collection finie de carrés nous pouvons considérer m_x, le minimum des rayons. L'ensemble K_1 est recouvert par les boules $B(x, m_x)$ et il existe donc une collection finie de $\{x_i\}_{i\in A}$ tels que $B(x_i, m_{x_i})$ recouvre K_1.

Alors $\{R_{x_i}\}_{i\in A}$ recouvre $K_1 \times K_2$ parce que R_{x_i} recouvre l'ensemble $B(x_i, m_{x_i}) \times \{K_2\}$. □

7.7.2 Conséquence : les fermés bornés sont compacts

Théorème 7.127 (Théorème de Borel-Lebesgue).
Une partie d'un espace vectoriel normé réel de dimension finie est compacte si et seulement si elle est fermée et bornée.

Démonstration. Sens direct.

Compact implique borné En effet si K est non borné dans E alors K contient une suite (x_n) avec $\|x_n\| > n$. Les boules $B_i(x_i, \frac{1}{3})$ sont disjointes. On pose $\mathcal{O}_0 = \complement \bigcup_i \overline{B(x_i, \frac{1}{5})}$, qui est ouvert comme complément d'un fermé. Pour $i \geqslant 1$ nous posons $\mathcal{O}_i = B(x_i, \frac{1}{4})$. Nous avons

$$K \subset \bigcup_{i\in\mathbb{N}} \mathcal{O}_i \tag{7.96}$$

mais vu que x_i est uniquement dans $\mathcal{O}_i$, nous ne pouvons pas extraire de sous-recouvrement fini.

Compact implique fermé Cela est la proposition 7.63.

Sens réciproque.

Un intervalle fermé et borné est compact dans $\mathbb{R}$ C'est le lemme 7.125.

Un produit de segments est compact Le produit de deux compacts de $\mathbb{R}$ est un compact dans $\mathbb{R}^2$ par le lemme 7.126.

Un fermé et borné est compact Soit K fermé et borné. Vu que K est borné, il est contenu dans un produit de segments. L'ensemble K est donc compact parce que fermé dans un compact, lemme 7.64.

□

Exemple 7.128 (Compacité de la boule unité).
La boule unité fermée $\overline{B(0, 1)}$ d'un espace vectoriel normé de dimension finie est compacte parce que fermée et bornée. En dimension infinie, cela n'est plus le cas. Certes la boule unité est encore fermée et bornée, mais elle n'est plus compacte. En effet nous allons donner un recouvrement par des ouverts duquel il ne sera pas possible d'extraire un sous-recouvrement fini.

Autour de chacune des extrémités des vecteurs de base, nous considérons la boule $A_i = B(e_i, \frac{1}{3})$. Ensuite aussi l'ouvert

$$B(0,1)\backslash\bigcup_i \overline{B(e_i, \frac{1}{4})}. \tag{7.97}$$

Le tout recouvre $B(0,1)$ mais toutes les premières boules sont nécessaires. △

Le théorème de Bolzano-Weierstrass 7.138 nous permettra de prouver plus simplement la non compacité en dimension infinie. Voir l'exemple 7.104.

7.7.3 Suites et limites dans les réels

7.7.3.1 Limites, convergence

Dans le cas de suites réelles, nous avons la caractérisation suivante qui est souvent donnée comme une définition lorsque seule la topologie sur $\mathbb{R}$ est considérée.

Proposition 7.129 (Limite d'une suite numérique)**.**
La suite (x_n) est convergente si et seulement s'il existe un réel ℓ tel que

$$\forall \epsilon > 0, \exists N \in \mathbb{N} \text{ tel que } \forall n \geqslant N, |x_n - \ell| < \epsilon. \tag{7.98}$$

Dans ce cas, le nombre ℓ est la limite de la suite (x_n).

Proposition-définition 7.130.
Une suite (x_n) dans un espace vectoriel normé V est convergente[40] *si et seulement si il existe un élément $\ell \in V$ tel que*

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ tel que } n \geqslant N \Rightarrow \|x_n - l\| < \varepsilon. \tag{7.99}$$

Dans ce cas, ℓ est la limite de la suite (x_n).

Démonstration. En deux parties.

Sens direct Si $x_n \to \ell$ et si $\epsilon > 0$ il existe N_ϵ tel que pour tout $n \geqslant N$ nous avons $x_n \in B(\ell, \epsilon)$ (parce que cette boule est un ouvert contenant ℓ). Vue la définition d'une boule, cette condition s'écrit bien $\|x_n - \ell\| < \epsilon$.

Sens inverse Dans l'autre sens, soit $\mathcal{O}$ un ouvert contenant ℓ. Par définition de la topologie, il existe $\epsilon > 0$ tel que $B(\ell, \epsilon) \subset \mathcal{O}$. La condition (7.98) nous assure qu'il existe N_ϵ tel que pour tout $n \geqslant N_\epsilon$ nous ayons

$$x_n \in B(\ell, \epsilon) \subset \mathcal{O}, \tag{7.100}$$

ce qui assure que la suite (x_n) converge vers ℓ pour la topologie métrique de V.

□

Une façon équivalente d'exprimer le critère (7.98) est de dire que pour tout ϵ positif, il existe un rang $N \in \mathbb{R}$ tel que l'intervalle $[\ell - \epsilon, \ell + \epsilon]$ contient tous les termes x_n au-delà de N.

Il est à noter que le rang N dont il est question dans la définition de suite convergente dépend de ϵ.

Définition 7.131.
Nous disons qu'une suite réelle (x_n) converge[41] *vers ℓ lorsque pour tout ε, il existe un N tel que*

$$n > N \Rightarrow |x_n - \ell| \leqslant \varepsilon. \tag{7.101}$$

Le concept fondamental de cette définition est la notion de valeur absolue qui permet de donner la « distance » entre deux réels. Dans un espace vectoriel normé quelconque, cette notion est généralisée par la distance associée à la norme (définition 7.113). Nous pouvons donc facilement définir le concept de convergence d'une suite dans un espace vectoriel normé.

40. Définition 7.10.
41. Voir la définition 7.129 pour plus de détail.

7.7.4 Opérations sur les limites

Proposition 7.132 ([1]).
Soient des suites à valeurs réelles (a_i) et (b_j) si elles sont convergentes, alors la suite ab est convergente et

$$\big(\lim_i a_i\big)\big(\lim_j b_j\big) = \lim_i (a_i b_i). \tag{7.102}$$

Démonstration. Nous nommons a et b les limites des suites (a_i) et (b_j). Soit $\epsilon > 0$ ainsi que $i \in \mathbb{N}$. Nous avons la majoration

$$|a_i b_i - ab| \leqslant |a_i b_i - a_i b| + |a_i b - ab| \tag{7.103a}$$
$$\leqslant |a_i||b_i - b| + b|a_i - a|. \tag{7.103b}$$

Vu que la suite (a_i) est convergente, elle est bornée. Nous pouvons donc majorer $|a_i|$ par $R > 0$ qui ne dépend pas de i. Soit $\eta > 0$ tel que $(R + b)\eta < \epsilon$. Alors en prenant i assez grand pour que $|b_i - b| < \eta$ et $|a_i - a| < \eta$, nous avons bien

$$|a_i b_i - ab| \leqslant (R + b)\eta < \epsilon. \tag{7.104}$$

□

Proposition 7.133.
Soient des suites (x_n) et (y_n) dans un espace vectoriel normé V. Si $x_n \xrightarrow{V} x$ et $y_n \xrightarrow{V} y$, alors

$$x_n + y_n \xrightarrow{V} x + y. \tag{7.105}$$

Démonstration. Soit $\epsilon > 0$. Nous considérons N tel que si $n \geqslant N$, alors $\|x_n - x\| \leqslant \epsilon$ et $\|y_n - y\| \leqslant \epsilon$. En utilisant l'inégalité 7.113(3),

$$\|x_y + y_n - (x + y)\| \leqslant \|x_n - x\| + \|y_n - y\| \leqslant 2\epsilon. \tag{7.106}$$

Donc la suite $(x_n + y_n)$ converge vers $x + y$. □

7.7.5 Exemples

Exemple 7.134.
Quelques suites usuelles.

(1) La suite $x_n = \frac{1}{n}$ converge vers 0.
(2) La suite $x_n = (-1)^n$ ne converge pas.

△

Deux limites pour voir comment ça fonctionne.

Lemme 7.135.
Si $r > 1$ nous avons :

(1) $\lim_{n\to\infty} r^n = \infty$.
(2) $\lim_{n\to\infty} \frac{r^n}{n} = \infty$.

Démonstration. Vu que $r > 1$ nous pouvons écrire $r = 1 + \delta$ avec $\delta > 0$. La formule du binôme de Newton (3.77) nous donne

$$(1 + \delta)^n = \sum_{k=0}^{n} \binom{k}{n} \delta^k > \binom{1}{n} \delta = n\delta. \tag{7.107}$$

La proposition 1.143 ($\mathbb{R}$ est archimédien) nous indique que $n\delta$ est arbitrairement grand lorsque n est grand, quelle que soit $\delta > 0$. Cela finit la preuve de la première limite.

Pour la seconde, nous posons $a_n = \frac{r^n}{n}$. Nous avons

$$\frac{a_{n+1}}{a_n} = \frac{n}{n+1}r. \tag{7.108}$$

Vu que $\frac{n}{n+1} \to 1$, la suite $\frac{n}{n+1}r$ tend vers $r > 0$, et en particulier pour tout $\delta > 0$ tel que $r > 1 + \delta$, il existe $N \in \mathbb{N}$ tel que, pour tout $n > N$,

$$\frac{n}{n+1}r > 1 + \delta. \tag{7.109}$$

Soit maintenant $k \in \mathbb{N}$. En utilisant un produit télescopique,

$$a_{N+k} = a_N \frac{a_{N+1}}{a_N}\frac{a_{N+2}}{a_{N+1}} \cdots \frac{a_{N+k}}{a_{N+k-1}} > a_N(1+\delta)^{k-1}. \tag{7.110}$$

Or $(1+\delta)^{k-1}$ tend vers ∞ lorsque $k \to \infty$ par le premier point. Donc nous avons $\lim_{n\to\infty} r^n/n = \infty$. □

Définition 7.136.
*Nous disons que deux suites (u_n) et (v_n) sont **équivalentes** s'il existe une fonction $\alpha\colon \mathbb{N} \to \mathbb{R}$ telle que*

(1) pour tout n à partir d'un certain rang, $u_n = v_n\alpha(n)$

(2) $\alpha(n) \to 1$.

7.7.6 Suites croissantes et bornées

Une suite est dite **contenue** dans un ensemble A si $x_n \in A$ pour tout n. Une suite est **bornée supérieurement** s'il existe un M tel que $x_n \leqslant M$ pour tout n. De la même manière, la suite est bornée inférieurement s'il existe un m tel que $x_n \geqslant m$ pour tout n.

Le lemme suivant est souvent utilisé pour prouver qu'une suite est convergente.

Lemme 7.137.
Une suite croissante et bornée supérieurement converge. Une suite décroissante bornée inférieurement est convergente.

Une erreur courante est de croire que la borne est la limite : le lemme n'affirme pas ça. Par contre il est vrai que la borne donne ... hum ... une borne inférieure (ou supérieure) pour la limite.

Théorème 7.138 (Bolzano-Weierstrass, thème 12)**.**
Toute suite contenue dans un compact admet une sous-suite convergente.

Démonstration. Nous faisons la preuve par l'absurde en supposant que (x_k) n'admette pas de sous-suite convergente. Soit $a \in K$; aucune sous-suite de (x_k) ne converge vers a. En particulier, il existe un voisinage ouvert $\mathcal{O}_a$ de a et une partie finie I_a de $\mathbb{N}$ tel que $x_k \in \mathcal{O}_a$ seulement pour $k \in I_a$.

Les ouverts $\mathcal{O}_a$ recouvrent K ; nous pouvons en extraire un sous-recouvrement fini (c'est la définition 7.48 de la compactié). Nous avons donc des points $a_1, \dots, a_n$ tels que

$$K \subset \bigcup_{i=1}^{n} \mathcal{O}_{a_i} \tag{7.111}$$

et tels que pour chaque $\mathcal{O}_{a_i}$, nous avons $x_k \in \mathcal{O}_{a_i}$ seulement pour $k \in I_{a_i}$. Bien entendu, toute la suite est dans K et donc dans l'union.

En conclusion, nous avons $\mathbb{N} = \bigcup_{i=1}^{n} I_{a_i}$, ce qui prouve que $\mathbb{N}$ est un ensemble fini. Contradiction avec la proposition 1.31 qui dit que $\mathbb{N}$ est infini. □

Proposition 7.139.
Une suite (x_n) dans $\mathbb{R}^m$ est convergente dans $\mathbb{R}^m$ si et seulement si les suites de chaque composante sont convergentes dans $\mathbb{R}$. Dans ce cas nous avons

$$\lim x_n = \Big(\lim(x_n)_1, \lim(x_n)_2, \dots, \lim(x_n)_m \Big) \tag{7.112}$$

où $(x_n)_k$ dénote la k-ième composante de (x_n).

Exemple 7.140.
La suite $x_n = \left(\frac{1}{n}, 1 - \frac{1}{n}\right)$ converge vers $(0,1)$ dans $\mathbb{R}^2$. En effet, en utilisant la proposition 7.139, nous devons calculer séparément les limites

$$\begin{aligned} \lim \frac{1}{n} &= 0 \\ \lim \left(1 - \frac{1}{n}\right) &= 1. \end{aligned} \tag{7.113}$$

△

Exemple 7.141.
Étant donné que la suite $(-1)^n$ n'est pas convergente, la suite $x_n = \left((-1)^n, \frac{1}{n}\right)$ n'est pas convergente dans $\mathbb{R}^2$. △

7.7.7 Suites adjacentes

Définition 7.142 ([132]).
*Les suites (a_n) et (b_n) sont **adjacentes** si l'une est croissante, l'autre décroissante et si $a_n - b_n \to 0$.*

Théorème 7.143 (Théorème des suites adjaentes).
Nous considérons des suites adjacentes (a_n) et (b_n) avec (a_n) croissante et (b_n) décroissante. Alors

(1) $b_n \geqslant a_n$ pour tout n,

(2) $a_n \leqslant b_q$ pour tout n et q. C'est-à-dire que toute la suite a est plus petite que toute la suite b.

(3) les suites a et b sont convergentes,

(4) les suites a et b convergent vers la même limite, notée ℓ,

(5) nous avons $a_n \leqslant \ell \leqslant b_n$ pour tout n.

Démonstration. La suite $n \mapsto b_n - a_n$ est décroissante parce que $b_n - a_n \geqslant b_{n+1} - a_{n+1}$. Comme en plus $b_a - a_n \to 0$ nous avons

$$b_n - a_n \geqslant 0 \tag{7.114}$$

pour tout $n \in \mathbb{N}$. De plus $a_n \leqslant b_0$ pour tout n parce que si $a_N > b_0$ alors, b étant décroissante, $a_N > b_0 \geqslant b_N$ qui est contraire à ce que nous venons de prouver. La suite a étant croissante et majorée, elle est convergente [42] ; notons ℓ sa limite.

La suite b peut maintenant être écrite par

$$b_n = (b_n - a_n) + a_n \tag{7.115}$$

qui est une somme de deux suites convergentes. Elle est donc convergente et sa limite est la somme des limites [43], donc

$$\lim_{n\to\infty} b_n = \lim_{n\to\infty} (b_n - a_n) + a_n = 0 + \ell = \ell. \tag{7.116}$$

Voila. Donc les suites a et b convergent et ont la même limite.

Pour tout $n, q \in \mathbb{N}$ nous avons l'inégalité $a_n \leqslant b_q$. En prenant la limite $n \to \infty$ nous trouvons

$$\ell \leqslant b_q \tag{7.117}$$

42. Proposition 7.137.
43. Proposition 7.133.

pour tout q. Et de la même façon, $b_n \geqslant a_q$ donne $\ell \geqslant a_q$. L'un avec l'autre donne

$$a_q \leqslant \ell \leqslant b_q \tag{7.118}$$

pour tout $q \in \mathbb{N}$. □

Proposition 7.144 ([133])**.**
Soit une suite (a_n) dans $\mathbb{R}$. Nous supposons que les suites extraites (a_{2n}) et (a_{2n+1}) convergent vers la même limite notée ℓ.

Alors $a_n \to \ell$.

Démonstration. Soit $\epsilon > 0$. Il existe N_1 tel que $|a_{2n} - \ell| \leqslant \epsilon$ dès que $n \geqslant N_1$. Il existe également N_2 dès que $|a_{2n+1} - \ell| \leqslant \epsilon$ dès que $n \geqslant N_2$.

Nous posons $N = \max\{2N_1, 2N_2 + 2\}$ et nous avons, pour tout $n \geqslant N$:

$$|a_n - \ell| \leqslant \epsilon, \tag{7.119}$$

c'est-à-dire que $a \to \ell$. □

7.7.8 Limite supérieure et inférieure

Lemme-définition 7.145.
Soit (a_n) une suite dans $\bar{\mathbb{R}}$. Les limites suivantes existent dans $\bar{\mathbb{R}}$

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \big(\sup_{k\geqslant n} a_k\big) \tag{7.120}$$

et

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} \big(\inf_{k\geqslant n} a_k\big). \tag{7.121}$$

*Elles sont nommées **limite supérieure** et la **limite inférieure** de la suite (a_k).*

Démonstration. Pour la limite supérieure, l'ensemble des $k \geqslant n$ est de plus en plus petit lorsque n grandit. Donc les ensembles $A_n = \{a_k \text{ tel que } k \geqslant n\}$ sont emboîtés et la suite $n \to \sup A_n$ est une suite décroissante. Elle a donc une limite dans $\bar{\mathbb{R}}$. □

7.146.
En ce qui concerne les suites d'ensembles, utiles en théorie des probabilités, nous définissons de même. Si les A_n sont des parties de Ω, nous définissons la **limite supérieure** et la **limite inférieure** de la suite A_n par

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geqslant 1} \bigcup_{k\geqslant n} A_k \tag{7.122}$$

et

$$\liminf_{n\to\infty} A_n = \bigcup_{n\geqslant 1} \bigcap_{k\geqslant n} A_k \tag{7.123}$$

Nous avons

$$\limsup A_n = \{\omega \in \Omega \text{ tel que } \omega \in A_n \text{pour une infinité de } n\}. \tag{7.124}$$

Lemme 7.147.
Nous avons les formules pratiques suivantes :

$$\limsup a_n = \inf_{n\geqslant 1} \big(\sup_{k\geqslant n} a_k\big) \tag{7.125a}$$

$$\liminf a_n = \sup_{n\geqslant 1} \big(\inf_{k\geqslant n} a_k\big). \tag{7.125b}$$

Démonstration. La suite $n \mapsto \sup_{k\geqslant n} a_k$ est une suite décroissante, donc la limite est l'infimum. Même argument pour l'autre. □

Lemme 7.148.
La suite (a_n) dans $\mathbb{R}$ converge si et seulement si

$$\limsup a_n = \liminf a_n. \tag{7.126}$$

Dans ce cas, $\lim a_n = \limsup a_n = \liminf a_n$.

Démonstration. Nous commençons par supposer que $\limsup a_n = \liminf a_n = l$, et nous prouvons que $\lim a_n$ existe et vaut l. Soit $\epsilon > 0$. Il existe N tel que si $n \geqslant N$ nous avons

$$\big|\sup_{k\geqslant n} a_k - l\big| < \epsilon \tag{7.127}$$

et

$$\big|\inf_{k\geqslant n} a_k - l\big| < \epsilon. \tag{7.128}$$

Pour tout $k \geqslant N$ nous avons alors $a_k \leqslant l + \epsilon$ et $a_k \geqslant l - \epsilon$. Cela donne $a_n \in B(l, \epsilon)$, c'est-à-dire $a_k \to l$ par la proposition 7.129.

Dans l'autre sens, nous supposons que $\lim_n a_n = l$ et nous prouvons que les limites supérieures et inférieures sont toutes deux égales à l. Soit $\epsilon > 0$ et N_ϵ tel que $|a_n - l| < \epsilon$ pour tout $n \geqslant N_\epsilon$. Si $n \geqslant N_\epsilon$ nous avons

$$\big|\sup_{k\geqslant n} a_k - l\big| \leqslant \epsilon \tag{7.129}$$

et donc la limite de $\sup_{k\geqslant n} a_k$ lorsque $n \to \infty$ est bien l. □

7.7.9 Ouverts, voisinage, topologie

Lorsque $x \in E$, nous rappelons qu'un voisinage[44] de x est n'importe quel sous-ensemble de E qui contient une boule ouverte centrée en x. La proposition 7.4 nous dit qu'un ensemble est ouvert s'il contient un voisinage de chacun de ses points. Au passage, rappelons que l'ensemble vide est ouvert.

Pour rappel, la remarque 7.95(2) dit que l'ensemble des boules ouvertes d'un espace métrique génère la topologie de l'espace.

Nous rappelons qu'une partie A d'un espace métrique est dite bornée[45] s'il existe une boule[46] qui contient A.

Mais revenons à $\mathbb{R}$...

Lemme 7.149.
Une partie ouverte de $\mathbb{R}$ ne contient pas son supremum.

Démonstration. Soit $\mathcal{O}$, un ensemble ouvert et s, son supremum. Si s était dans $\mathcal{O}$, on aurait un voisinage $B = B(s, r)$ de s contenu dans $\mathcal{O}$. Le point $s + r/2$ est alors à la fois dans $\mathcal{O}$ et plus grand que s, ce qui contredit le fait que s soit un supremum de $\mathcal{O}$. □

Par le même genre de raisonnements, on montre que l'union et l'intersection de deux ouverts sont encore des ouverts.

Remarque 7.150.
L'intersection d'une *infinité* d'ouverts n'est pas spécialement un ouvert comme le montre l'exemple suivant :

$$\mathcal{O}_i =]1, 2 + \frac{1}{i}[.$$

Tous les ensembles $\mathcal{O}_i$ contiennent le point 2 qui est donc dans l'intersection. Mais quel que soit le $\epsilon > 0$ que l'on choisisse, le point $2 + \epsilon$ n'est pas dans $\mathcal{O}_{(1/\epsilon)+1}$. Donc aucun point au-delà de 2 n'est dans l'intersection, ce qui prouve que 2 ne possède pas de voisinages contenus dans $\bigcap_{i=1}^{\infty} \mathcal{O}_i$.

44. Définition 7.2.
45. Définition 7.99.
46. À titre d'exercice, convainquez-vous que l'on peut dire boule *ouverte* ou *fermée* au choix sans changer la définition.

Proposition 7.151.
Quels que soient les ensembles A et B dans $\mathbb{R}$, nous avons

$$\sup(A \cap B) \leqslant \sup A \leqslant \sup(A \cup B).$$

7.7.10 Intervalles et connexité

Nous allons déterminer tous les sous-ensembles connexes[47] de $\mathbb{R}$. Pour cela nous relisons d'abord la notion d'intervalle donnée en 1.13 ainsi que la proposition 1.158 qui liste tous les intervalles de $\mathbb{R}$. La partie $I \subset \mathbb{R}$ est un intervalle si pour tout $a, b \in I$, tout nombre entre a et b est également dans I. Cette définition englobe tous les exemples connus d'intervalles ouverts, fermés avec ou sans infini : $[a, b]$, $[a, b[$, $]-\infty, a]$, ... L'ensemble $\mathbb{R}$ lui-même est un intervalle.

Si I est un intervalle, les nombres $\inf(I)$ et $\sup(I)$[48] sont les **extrémités** de I.

Définition 7.152.
*Étant donnés deux points a et b dans $\mathbb{R}^p$ on appelle **segment** d'extrémités a et b, et on note $[a, b]$, l'image de $[0, 1]$ par l'application $s : [0, 1] \to \mathbb{R}^p$, $s(t) = (1 - t)a + tb$. On pose $]a, b[= s\left(]0, 1[\right)$, et $]a, b] = s\left(]0, 1]\right)$.*

Il faut observer que le segment $[a, b]$ est une courbe orientée : certes en tant que ensembles, $[a, b] = [b, a]$, mais si nous regardons la fonction de t correspondante à $[b, a]$, nous voyons qu'elle va dans le sens inverse de celle qui correspond à $[a, b]$. Nous approfondirons ces questions lorsque nous parlerons d'arcs paramétrés autour de la section 20.7.

Le segment $[b, a]$ est l'image de l'application $r\colon [0, 1] \to \mathbb{R}^p$ donnée par $r(t) = (1 - t)b + ta$.

Proposition 7.153.
Une partie de $\mathbb{R}$ est connexe si et seulement si c'est un intervalle[49].

Démonstration. La preuve est en deux parties. D'abord nous démontrons que si un sous-ensemble de $\mathbb{R}$ est connexe, alors c'est un intervalle ; et ensuite nous démontrons que tout intervalle est connexe.

Afin de prouver qu'un ensemble connexe est toujours un intervalle, nous allons prouver que si un ensemble n'est pas un intervalle, alors il n'est pas connexe. Prenons A, une partie de $\mathbb{R}$ qui n'est pas un intervalle. Il existe donc a, $b \in A$ et un x_0 entre a et b qui n'est pas dans A. Comme le but est de prouver que A n'est pas connexe, il faut couper A en deux ouverts disjoints. L'élément x_0 qui n'est pas dans A est le bon candidat pour effectuer cette coupure. Prenons M, un majorant de A et m, un minorant de A, et définissons

$$\begin{aligned}\mathcal{O}_1 &=]m, x_0[\\ \mathcal{O}_2 &=]x_0, M[.\end{aligned}$$

Si A n'a pas de minorant, nous remplaçons la définition de $\mathcal{O}_1$ par $]-\infty, x_0[$, et si A n'a pas de majorant, nous remplaçons la définition de $\mathcal{O}_2$ par $]x_0, \infty[$. Dans tous les cas, ce sont deux ensembles ouverts dont l'union recouvre tout A. En effet, $\mathcal{O}_1 \cup \mathcal{O}_2$ contient tous les nombres entre un minorant de A et un majorant sauf x_0, mais on sait que x_0 n'est pas dans A. Cela prouve que A n'est pas connexe.

Jusqu'à présent nous avons prouvé que si un ensemble n'est pas un intervalle, alors il ne peut pas être connexe. Pour remettre les choses à l'endroit, prenons un ensemble connexe, et demandons-nous s'il peut être autre chose qu'un intervalle ? La réponse est *non* parce que s'il était autre chose, il ne serait pas connexe.

Prouvons à présent que tout intervalle est connexe. Pour cela, nous refaisons le coup de la contraposée. Nous allons donc prendre une partie A de $\mathbb{R}$, supposer qu'elle n'est pas connexe et

47. Définition 7.42.
48. Qui existent par la proposition 1.157, quitte à poser $\pm\infty$ comme infimum et supremum lorsque I n'est pas borné.
49. Définition 1.13.

puis prouver qu'elle n'est alors pas un intervalle. Nous avons deux ouverts disjoints $\mathcal{O}_1$ et $\mathcal{O}_2$ tels que $A \subset \mathcal{O}_1 \cup \mathcal{O}_2$. Notons $A_1 = A \cap \mathcal{O}_1$ et $A_2 = A \cap \mathcal{O}_2$; et prenons $a \in A_1$ et $b \in A_2$. Pour fixer les idées, on suppose que $a < b$. Maintenant, le jeu est de montrer qu'il existe une point x_0 entre a et b qui ne soit pas dans A (cela montrerait que A n'est pas un intervalle). Nous allons prouver que c'est le cas du point

$$x_0 = \sup\{x \in \mathcal{O}_1 \text{ tel que } x < b\}.$$

Étant donné que l'ensemble $\mathcal{A} = \{x \in \mathcal{O}_1 \text{ tel que } x < b\}$ est ouvert[50], le point x_0 n'est pas dans l'ensemble par le lemme 7.149. Nous avons donc

— soit x_0 n'est pas dans $\mathcal{O}_1$,

— soit $x_0 \leqslant b$,

— soit les deux en même temps.

Nous allons montrer qu'un tel x_0 ne peut pas être dans A. D'abord, remarquons que $\sup \mathcal{A} \leqslant \sup \mathcal{O}$ parce que $\mathcal{A}$ est une intersection de $\mathcal{O}$ avec quelque chose. Ensuite, il n'est pas possible que x_0 soit dans $\mathcal{O}_2$ parce que tout élément de $\mathcal{O}_2$ possède un voisinage contenu dans $\mathcal{O}_2$. Un point de $\mathcal{O}_2$ est donc toujours strictement plus grand que le supremum de $\mathcal{O}_1$.

Maintenant, remarque que si $x_0 \leqslant b$, alors $x_0 = b$, sinon b serait un majorant de $\mathcal{A}$ plus petit que x_0, ce qui n'est pas possible vu que x_0 est le supremum de $\mathcal{A}$ et donc le plus petit majorant. Oui mais si $x_0 = b$, c'est que $x_0 \in \mathcal{O}_2$, ce qu'on vient de montrer être impossible. Nous voilà déjà débarrassé des deuxièmes et troisièmes possibilités.

Si la première possibilité est vraie, alors x_0 n'est pas dans A parce qu'on a aussi prouvé que $x_0 \notin \mathcal{O}_2$. Or n'être ni dans $\mathcal{O}_1$ ni dans $\mathcal{O}_2$ implique de ne pas être dans A. Ce point $x_0 = \sup \mathcal{A}$ est donc hors de A.

Oui, mais comme $a \in \mathcal{A}$, on a obligatoirement que $x_0 \geqslant a$. Mais par construction, on a aussi que $x_0 \leqslant b$ (ici, l'inégalité est même stricte, mais ce n'est pas important). Donc

$$a \leqslant x_0 \leqslant b$$

avec a, $b \in A$, et $x_0 \notin A$. Cela finit de prouver que A n'est pas un intervalle. □

Théorème 7.154 (Théorème des bornes atteintes)**.**
Une fonction à valeurs réelles continue sur un compact est bornée et atteint ses bornes.

C'est-à-dire qu'il existe $x_0 \in K$ tel que $f(x_0) = \inf\{f(x) \text{ tel que } x \in K\}$ ainsi que x_1 tel que $f(x_1) = \sup\{f(x) \text{ tel que } x \in K\}$.

Démonstration. Soient un espace topologique compact K et une fonction continue $f\colon K \to \mathbb{R}$. Alors le théorème 7.92 indique que $f(K)$ est compact. Par conséquent $f(K)$ est un fermé borné de $\mathbb{R}$ par le théorème de Borel-Lebesgue 7.127. Vu que $f(K)$ est borné, la fonction f est bornée.

De plus $f(K)$ étant fermé, son infimum est un minimum et son supremum est un maximum : il existe $x \in K$ tel que $f(x) = \sup f(K)$ et il existe $y \in K$ tel que $f(y) = \inf f(K)$. □

Le théorème suivant est essentiellement inutile pour les raisons suivantes :

— Il est un cas particulier du théorème 7.103 qui donne pour tout espace métrique, l'équivalence entre la compacité et la compacité séquentielle.

— Il est un cas particulier du théorème 7.138 qui le donne pour tous les espaces compacts.

— Il utilise le cas particulier de $\mathbb{R}$, qui n'est pas démontré directement dans le Frido.

Bref, nous ne le laissons que pour le lecteur qui n'aurait pas en tête d'autres définitions de « compact » à part « fermé borné ».

Théorème 7.155 (Théorème de Bolzano-Weierstrass)**.**
Toute suite contenue dans un compact de $\mathbb{R}^m$ admet une sous-suite convergente.

50. C'est l'intersection entre l'ouvert $\mathcal{O}_1$ et l'ouvert $\{x \text{ tel que } x < b\}$.

Démonstration. Nous rappelons qu'une partie compacte de $\mathbb{R}^n$ est fermée et bornée par le théorème de Borel-Lebesgue 7.127.

Soit (x_n) une suite contenue dans une partie bornée de $\mathbb{R}^m$. Considérons (a_n), la suite réelle des premières composantes des éléments de (x_n) : pour chaque $n \in \mathbb{N}$, le nombre a_n est la première composante de x_n. Étant donné que la suite (x_n) est bornée, il existe un M tel que $\|x_n\| < M$. La croissance de la fonction racine carrée donne

$$|a_n| \leqslant \|x_n\| \leqslant M. \tag{7.130}$$

La suite (a_n) est donc une suite réelle bornée et donc contient une sous-suite convergente par le théorème correspondant dans $\mathbb{R}$: 7.103. Soit a_{I_1} une sous-suite convergente de (a_n). Nous considérons maintenant x_{I_1}, c'est-à-dire la suite de départ dont on a enlevé tous les éléments qu'il faut pour qu'elle converge en ce qui concerne la première composante.

Si nous considérons la suite b_{I_1} des *secondes* composantes de x_{I_1}, nous en extrayons, de la même façon que précédemment, une sous-suite convergente, c'est-à-dire que nous avons un $I_2 \subset I_1$ tel que b_{I_2} est convergent. Notons que a_{I_2} est une sous-suite de la (sous) suite convergente x_{I_1}, et donc a_{I_2} est encore convergente.

En continuant ainsi, nous construisons une sous-sous-sous-suite x_{I_3} telle que la suite des *troisièmes* composantes est convergente. Lorsque nous avons effectué cette procédure m fois, la suite x_{I_m} est une suite dont toutes les composantes convergent, et donc est une suite convergente par la proposition 7.139.

Le tableau suivant donne un petit schéma de la façon dont nous procédons. Les $\bullet$ sont les éléments de la suite que nous gardons, et les $\times$ sont ceux que nous « jetons ».

$$\begin{array}{lcccccccccc}
x_{\mathbb{N}} & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \dots \\
x_{I_1} & \times & \bullet & \bullet & \times & \bullet & \times & \times & \bullet & \bullet & \bullet & \dots \\
x_{I_2} & \times & \bullet & \times & \times & \bullet & \times & \times & \bullet & \bullet & \times & \dots \\
\vdots & & & & & & & & & & & \\
x_{I_m} & \times & \times & \times & \times & \bullet & \times & \times & \times & \bullet & \times & \dots
\end{array} \tag{7.131}$$

La première ligne, $x_{\mathbb{N}}$, est la suite de départ. □

Corolaire 7.156.
Si un suite est croissante et bornée alors elle est convergente.

Démonstration. Nous nommons (x_n) la suite et nous prenons un majorant M. Toute la suite est alors contenue dans le compact $[x_0, M]$, ce qui donne une sous-suite $(x_{\alpha(n)})$ convergente par le théorème de Bolzano-Weierstrass 7.138. Si ℓ est la limite de cette sous-suite alors nous avons $\ell \geqslant x_n$ pour tout n.

Pour tout $\epsilon > 0$ il existe K tel que si $n > K$ alors $|\ell - x_{\alpha(n)}| < \epsilon$. Vu que ℓ majore la suite nous avons même

$$x_{\alpha(n)} + \epsilon > \ell. \tag{7.132}$$

Vu que la suite est croissante pour tout $m > \alpha(K)$ nous avons $x_m + \epsilon > l$, ce qui signifie $|x_m - \ell| < \epsilon$. □

Nous aurons une version pour les fonctions croissantes et bornées en la proposition 11.66.

La proposition suivante dit que la notion d'ensemble non dénombrable ne prend pas réellement de force entre $\mathbb{R}$ et $\mathbb{R}^n$: il n'y a pas moyen de caser $\mathbb{R}$ dans $\mathbb{R}^n$ de façon à ce qu'il y tienne à son aise.

Proposition 7.157.
Une partie non dénombrable de $\mathbb{R}^n$ possède un point d'accumulation[51].

51. Définition 7.27.

Démonstration. Soit une partie $A \subset \mathbb{R}^n$ sans point d'accumulation. Nous allons prouver que A est dénombrable.

Soient les compacts $K_n = \overline{B(0,n)}$. La partie $A \cap K_n$ est finie ; sinon elle aurait une partie en bijection avec $\mathbb{N}$ (proposition 1.47) et donc une suite. Or une suite dans un compact possède un point d'accumulation par le théorème 7.138.

Donc tous les $A \cap K_n$ sont finis. Vu que $A = \bigcup_n A \cap K_n$, l'ensemble A est une réunion dénombrable d'ensembles finis. Il est donc dénombrable. □

7.7.11 Recouvrement par des intervalles ouverts

Soit un ensemble E et un ensemble $\mathcal{A}$ de parties de E. Soit $A \in \mathcal{A}$. Nous aimerions savoir quelles sont les éléments de $\mathcal{A}$ qui sont atteignables en partant de A et en ne « sautant » que d'intersection en intersection.

Nous notons $\mathcal{A} = \{B_i\}_{i\in I}$ où I est un ensemble d'indices (un ensemble quelconque).

$$s_1(A) = \{i \in I \text{ tel que } B_i \cap A \neq \varnothing\} \tag{7.133a}$$

$$\sigma_1(A) = \bigcup_{B\in s_1(A)} B. \tag{7.133b}$$

Et ensuite :

$$s_{k+1}(A) = \{i \in I \text{ tel que } B_i \cap \sigma_k(A) \neq \varnothing\} \tag{7.134a}$$

$$\sigma_{k+1}(A) = \bigcup_{B\in s_{k+1}(A)} B \tag{7.134b}$$

Lemme 7.158.
Soient un intervalle A de $\mathbb{R}$ et $\mathcal{A} = \{I_i\}_{i=1,\dots,N}$ un recouvrement de A par des intervalles ouverts. Si $I_1 \cap A \neq \varnothing$ alors

(1) $\sigma_N = \sigma_{N+1}$

(2) $A \subset \sigma_N(I_1)$.

Démonstration. Si $\sigma_{k+1} = \sigma_k$, alors tous les σ_{k+l} sont identiques. De plus si $\sigma_{k+1} \neq \sigma_k$, alors σ_{k+1} contient au moins un élément de plus que σ_k. Donc $\mathrm{Card}(\sigma_k) \geqslant k$ et en particulier $N \leqslant \mathrm{Card}(\sigma_N) \leqslant N$. Cela prouve le premier point.

L'ensemble $\sigma_N(I_1)$ est une union d'ouverts et est donc un ouvert. Quitte à renuméroter nous écrivons

$$\sigma_N(I_1) = I_1 \cup \dots \cup I_n. \tag{7.135}$$

L'ensemble

$$\tau = \bigcup_{k=n+1}^{N} I_k \tag{7.136}$$

est ouvert et est disjoint de $\sigma_N(I_1)$ parce que si I_l ($I_l \geqslant n+1$) intersectait $\sigma_N(I_1)$, nous aurions $l \in s_{N+1}$ ou encore $I_l \subset \sigma_{N+1} \backslash \sigma_N$.

Donc τ et σ_N sont deux ouverts disjoints qui recouvrent A. Vu que A est un intervalle, il est connexe[52]. Donc soit $A \subset \tau$ soit $A \subset \sigma_N$. Comme $I_1 \cap A \neq \varnothing$ nous sommes dans le cas $A \subset \sigma_N$. □

Lemme 7.159.
Soit $x \in \mathbb{R}$. Si $\mathcal{A} = \{I_s\}_{s\in S}$ est un ensemble d'intervalles contenant x, alors $I = \bigcup_{s\in S} I_s$ est un intervalle[53].

52. Définition 7.42 et proposition 7.153.
53. Définition 1.13.

Démonstration. Soient $a, b \in I$ (nous supposons $a < b$). Nous devons prouver que $[a, b] \subset I$. Pour cela nous considérons $y \in [a, b]$; il y a deux possibilités : soit $y < x$ soit $y > x$ (si $y = x$ alors $y \in I_s$).

Si $y < x$, alors $a \leqslant y < x$ et donc $y =\in I_s$. Si $y > x$, alors $x < y \leqslant b$ et $y \in I_t$. □

Proposition 7.160 ([134, 135])**.**
Un ouvert de $\mathbb{R}$ peut s'écrire comme union au plus dénombrable d'intervalles ouverts disjoints.

Plus précisément, si $\mathcal{O}$ est un ouvert de $\mathbb{R}$, il existe un ensemble $\mathcal{F} = \{I_s\}_{s\in S}$ où

(1) Chaque I_s est un intervalle ouvert contenu dans $\mathcal{O}$,

(2) Pour $s, t \in S$, si $I_s \neq I_t$, alors $I_s \cap I_t = \emptyset$.

(3) S est dénombrable,

Démonstration. Pour $x \in \mathcal{O}$, nous définissons J_x comme étant l'union de tous les intervalles ouverts contenus dans $\mathcal{O}$ et contenant x. Les J_x ne sont pas vides parce qu'ils contiennent toujours une boule centrée en x [54].

En tant que union d'intervalles, J_x est un intervalle par le lemme 7.159. De plus, J_x est ouvert parce que toute union d'ouverts est ouverte.

Nous notons $\mathcal{A}$ l'ensemble des intervalles ouverts contenus dans $\mathcal{O}$, et

$$\mathcal{A}_x = \{I \in \mathcal{A} \text{ tel que } x \in I\}. \tag{7.137}$$

Si $y \in J_x$, alors $J_x = J_y$ Vu que $y \in J_x$, nous pouvons considérer $J =\in \mathcal{A}_x \cap \mathcal{A}_y$. Nous avons

$$J_y = \bigcup_{I\in\mathcal{A}_y} I \subset \bigcup_{I\in\mathcal{A}_y} \underbrace{(I \cup J)}_{\in\mathcal{A}_x} \subset \bigcup_{I\in\mathcal{A}_x} I = J_x. \tag{7.138}$$

L'inclusion dans l'autre sens s'obtient en écrivant la même chose en remplaçant x par y et vice-versa.

Les J_x sont disjoints Nous prouvons à présent que pour $x, y \in mO$, nous avons $J_x = J_y$ ou $J_x \cap J_y = \emptyset$. En effet si $a \in J_x \cap J_y$, alors $J_a = J_x$ et $J_a = J - y$. Donc $J_x = J_y$.

Dénombrable C'est le moment d'écrire $\mathcal{F} = \{J_x\}_{x\in\mathcal{O}}$. Vu que tout intervalle contient au moins un rationnel (proposition 7.121), nous avons aussi

$$\mathcal{F} = \{J_x\}_{x\in\mathcal{O}} = \{J_q\}_{q\in\mathbb{Q}\cap\mathcal{O}}. \tag{7.139}$$

Cet ensemble $\mathcal{F}$ vérifie les conditions demandées.

□

7.7.12 Connexité par arcs

Définition 7.161.
*Une partie A d'un espace topologique est **connexe par arcs** si pour tout $a, b \in A$, il existe une application continue $\gamma\colon [0, 1] \to A$ telle que $\gamma(0) = a$ et $\gamma(1) = b$.*

7.162.
Un exemple d'ensemble connexe mais pas connexe par arcs est donné par la proposition 20.56. L'idée de cet exemple est de construire un ensemble en deux parties reliées par un chemin de longueur infinie.

Un espoir fou nous prend alors de croire que nous pouvons produire un exemple plus simple avec $\mathbb{R} \cup \{\infty\}$ parce que, dans cet ensemble, 1 et ∞ sont reliés par un chemin de longueur infinie. La proposition 11.69 nous montrera que non.

54. C'est la définition 7.61 de la topologie métrique.

7.7.13 Un peu de connexité par arcs

Principe. L'idée est de considérer une propriété topologique (invariante par homéomorphisme) et de voir qu'elle est vérifiée par les ouverts de $\mathbb{R}^2$ mais pas ceux du cône.

Lemme 7.163.
Si V est un voisinage de 0 sur le cône C, alors $V\backslash\{0\}$ n'est pas connexe, donc n'est pas connexe par arc.

Démonstration. Le cône C est la réunion de $C^+ = C \cap \left(\mathbb{R}^2 \times \mathbb{R}_0^+\right)$ et $C^- = C \cap \left(\mathbb{R}^2 \times \mathbb{R}_0^-\right)$ car le seul point à cote nulle est la singularité 0. Dès lors, V s'écrit comme l'union disjointe de $V \cap C^+$ et $V \cap C^-$, qui sont non-vides. Donc V n'est pas connexe. □

On procède en deux étapes, en montrant d'abord qu'il existe des points en « dessous » et au « dessus » de 0, puis en essayant de les relier. Comme V est un voisinage de 0, il existe un ouvert U du cône centré en 0 inclut à V. Donc par définition de la topologie induite, et puisque les boules forment une base de la topologie de $\mathbb{R}^3$, il existe une boule B centrée en 0 dont U est la trace sur C, telle que $0 \in (B \cap C) \subset V$. On choisit $p = (p_x, p_y, p_z) \in (B \cap C)$, et en considérant $p' = (p_x, p_y, -p_z)$ on a ainsi trouvé deux points qui vérifient $p_z > 0$ et $p'_z < 0$ (au besoin, on les échange).

(1) Supposons que $V\backslash\{0\}$ soit connexe par arc. Donc il existe un chemin

$$\gamma : [0;1] \to V\backslash\{0\} : t \mapsto (\gamma_x(t), \gamma_y(t), \gamma_z(t))$$

qui relie p à p' et qui vérifie $\gamma_z(0) = p_z > 0$ et $\gamma_z(1) = -p_z < 0$. Or $\gamma_z(t)$ est une fonction continue (car γ est continu), donc par le théorème des valeurs intermédiaires, il existe $\bar{t}$ qui vérifie $\gamma_z(\bar{t}) = 0$. Or le seul point de C dont la cote (coordonnée en z) soit nulle est le sommet 0 qui n'est pas dans $V\backslash\{0\}$, d'où la contradiction.

remSoient deux espaces topologiques E et F, et $f : E \to F$ un homéomorphisme. Pour toute partie A de E, l'espace $E\backslash A$ est homéomorphe au sous-espace $F\backslash f(A)$ via la restriction $f_{|E\backslash A}$.

Lemme 7.164.
Soient deux espaces topologiques E et F, et $f : E \to F$ un homéomorphisme. E est connexe par arc si et seulement si F l'est.

Démonstration. On montre en réalité que l'image d'un connexe par arc par une application continue est un connexe par arc, ce qui implique chaque sens de l'équivalence de l'énoncé.

Soient p et q des points de F. Il existe un chemin reliant un antécédent de p et un antécédent de q (dans E). L'image de ce chemin est un chemin reliant p et q (dans F) puisque composé d'applications continues. □

Lemme 7.165.
Une sphère de $\mathbb{R}^n$ est connexe par arc si $n > 1$

Démonstration. On voit qu'un cercle est connexe par arc car on a un paramétrage en sinus et cosinus. Pour une sphère S de centre a en dimension $n > 2$, on se donne p et q sur S et on définit P le plan affine passant par a, p et q. Alors $P \cap S$ est un cercle, donc on peut relier p à q par un chemin dans cette intersection.

Pour voir sur une formule que $P \cap S$ est un cercle, on peut écrire $x - a = \lambda(a - p) + \mu(a - q)$ l'équation (en x) du plan P, et $|x - a|^2 = R^2$ l'équation (en x) de la sphère. En injectant, on obtient une équation du second degré en λ, μ qui se révèle être l'équation d'un cercle à une transformation affine près. □

Lemme 7.166.
Un ouvert connexe par arc dans $\mathbb{R}^n$ ($n \geqslant 2$) reste connexe par arc même si on lui enlève un point.

Démonstration. En effet, soit U un tel ouvert connexe par arc, et p un point de U. Soient x et y sur $U\backslash\{p\}$. Il existe un chemin γ de x à y. Si le chemin ne passe pas par p, c'est gagné. Si il passe par p, on choisit une boule B fermée (de rayon non-nul) centrée en p qui ne contient ni x ni y. On note

$$E = \gamma^{-1}(B) \subset [0;1]$$

c'est un ensemble compact (fermé, par continuité de γ, et borné) dont on regarde le maximum $\bar{t}$ et le minimum $\underline{t}$.

Il reste enfin à définir un chemin entre p et q par morceaux

(1) Les points p et $\gamma(\underline{t})$ sont reliés par γ,

(2) Par connexité par arc, il existe un chemin sur la sphère qui relie $\gamma(\underline{t})$ à $\gamma(\bar{t})$,

(3) et enfin $\gamma(\bar{t})$ et q sont reliés via γ ;

ce qui achève la construction d'un chemin continu entre p et q. □

Pour conclure l'exercice, par l'absurde, on prend un voisinage connexe et ouvert V de 0 dans le cône, homéomorphe à un ouvert connexe U de $\mathbb{R}^2$. Or $V\backslash\{0\}$ n'est pas connexe par arc, alors que l'ouvert dont on retire un point reste connexe par arc. C'est impossible, donc l'homéomorphisme n'existe pas, et le cône n'est pas une variété de dimension 2.

7.7.14 Des exemples

Exemple 7.167.
Nous étudions l'exemple suivant :

$$A_1 = \{(x,y) \in \mathbb{R}^2 \mid 2y^2 + 4y + 2 < x \leqslant \sqrt{4-y^2},\ y \in [-1.5, 0.5[\}. \tag{7.140}$$

On commence par tracer la parabole $x = 2y^2+4y+2$, la circonférence $x^2+y^2 = 4$ et les droites $y = -1.5$ et $y = 1/2$. On voit tout de suite que l'aire délimitée par les quatre courbes est donnée par l'union de deux parties. Dans la première $\sqrt{4-y^2} \leqslant x \leqslant 2y^2+4y+2$, $y \in [0, 0.5]$ et dans l'autre $2y^2+4y+2 \leqslant x \leqslant \sqrt{4-y^2}$, $y \in [-1.5, 0]$. L'ensemble A_1 est contenu dans la deuxième, 7.1. L'intérieur de A_1 est donné par $\mathrm{Int}(A_1) = \{(x,y) \in \mathbb{R}^2 \mid 2y^2+4y+2 < x < \sqrt{4-y^2},\ y \in]-1.5, 0[\}$, et sa frontière est l'union de 3 morceaux de courbe ℓ_1, ℓ_2, ℓ_3 :

$$\begin{aligned}
\ell_1 &= \{(x,y) \,|\, x = 2y^2+4y+2,\ y \in [-1.5, 0]\} \\
\ell_2 &= \{(x,y) \,|\, x = \sqrt{4-y^2},\ y \in [-1.5, 0]\} \\
\ell_3 &= \{(x,y) \,|\, x \in [0.5, \sqrt{7/4}]\ y = -1.5\}.
\end{aligned} \tag{7.141}$$

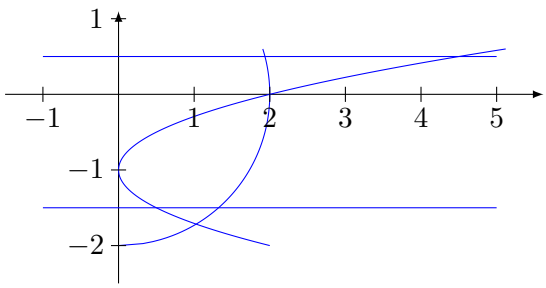


FIGURE 7.1 –

△

Exemple 7.168.
Nous étudions

$$A_3 = \mathbb{N} \times \mathbb{Q} = \{(x,y) \in \mathbb{R}^2 \mid x \in \mathbb{N}, y \in \mathbb{Q}\}. \tag{7.142}$$

L'ensemble A_3 n'est pas ouvert, ni fermé, ni borné dans la topologie de $\mathbb{R}^2$. En fait, comme on a vu dans les exercices du cours, $\mathbb{Q}$ a intérieur vide et sa fermeture est $\mathbb{R}$. L'ensemble $\mathbb{N}$, par contre est fermé et non borné. On peut remarquer que tous les points de $\mathbb{N}$ sont points isolés. La fermeture de A_3 est alors $\mathbb{N} \times \mathbb{R}$ et son intérieur est vide. On peut dessiner la fermeture de cet ensemble comme une famille de droites verticales $x = n$, pour tout n dans $\mathbb{N}$. △

Exemple 7.169.
Nous étudions l'ensemble

$$A_3 = \{(t, 2t) \in \mathbb{R}^2 \mid t \in [0, 1]\ \}. \tag{7.143}$$

L'ensemble A_3 est un petit segment de droite. Son intérieur est vide parce que toute boule centrée en un point de la droite intersecte l'extérieur de la droite. Son adhérence et sa frontière sont A_3 lui-même parce que nous considérons les valeurs de t dans $[0, 1]$ qui est un intervalle fermé. Si l'intervalle avait été ouvert, l'adhérence et la frontière auraient été trouvés en fermant :

$$\overline{\{(t, 2t) \text{ tels que } t \in [0, 1[\ \}} = \{(t, 2t) \text{ tels que } t \in [0, 1]\} \tag{7.144}$$

Étant donné que son adhérence est égal à lui-même, cet ensemble est fermé (et donc pas ouvert). Il est également borné parce qu'il est contenu dans une boule de rayon 3. △

Exemple 7.170.
Nous étudions l'ensemble

$$A_4 = \mathbb{Q} \times \mathbb{Q} = \{(x, y) \in \mathbb{R}^2 \mid x \in \mathbb{Q}, y \in \mathbb{Q}\}. \tag{7.145}$$

Dans $\mathbb{R}$ nous savons que $\bar{\mathbb{Q}} = \mathbb{R}$, $\mathrm{Int}(\mathbb{Q}) = \varnothing$ et $\partial\mathbb{Q} = \mathbb{R}$ parce que toute boule centrée en un rationnel contient un irrationnel, et inversement, toute boule centrée en un irrationnel contient un rationnel. Dans $\mathbb{R}^2$ nous avons le même phénomène parce dans la boule $B\big((p, q), r\big)$ avec $(p, q) \in \mathbb{Q} \times \mathbb{Q}$, se trouvent en particulier les points de la forme (p, x) avec $x \in B(q, r) \subset \mathbb{R}$. Évidement, certains de ces x ne sont pas dans $\mathbb{Q}$ et par conséquent, la boule $B\big((p, q), r\big)$ contient les points $(p, x) \notin \mathbb{Q} \times \mathbb{Q}$.

De la même manière, si (x, y) est un point de $\mathbb{R}^2$, dans toute boule centrée en (x, y), il y aura un élément de $\mathbb{Q}^2$.

Par conséquent, $\mathrm{Int}(\mathbb{Q} \times \mathbb{Q}) = \varnothing$, $\overline{\mathbb{Q} \times \mathbb{Q}} = \mathbb{R} \times \mathbb{R}$ et $\partial(\mathbb{Q} \times \mathbb{Q}) = \mathbb{R}^2$.

Il n'est ni ouvert ni fermé (parce qu'il n'est égal ni à son intérieur ni à sa fermeture). Il n'est pas borné non plus parce qu'il existe des nombres rationnels arbitrairement grands. △

Exemple 7.171.
Nous étudions l'ensemble

$$A_5 = \{(x, y) \in \mathbb{R}^2 \mid x \in]0, 1[, \sin \frac{1}{x} < y < 3\}. \tag{7.146}$$

La fonction $x \mapsto \sin(\frac{1}{x})$ est une des fonctions dont le graphe doit être connu. La figure 7.2 montre la situation. Comme d'habitude, il est fortement recommandé de refaire le dessin soi-même.

L'ensemble A_5 est ouvert parce que les conditions $x \in]0, 1[$ et $\sin \frac{1}{x} < y < 3$ sont des conditions « ouvertes » au sens où si un point les vérifient, alors on peut trouver une boule dans lequel ces conditions restent vérifiées. Cela prouve que $\mathrm{Int}(A_5) = A_5$.

La fermeture de A_5 contient en outre les points tels que $\sin \frac{1}{x} = y$ entre $x = 0$ et $x = 1$ (les bornes étant incluses) ainsi que les points des trois segments de droites suivants :

$$\begin{aligned} \{(0, y) \text{ tels que } y \in [-1, 3]\} \\ \{(x, 3) \text{ tels que } x \in [0, 1]\} \\ \{(1, y) \text{ tels que } y \in [\sin(1), 3]\}. \end{aligned} \tag{7.147}$$

La frontière est composée de ces trois segments et du graphe de la fonction $\sin \frac{1}{x}$ entre 0 et 1.

L'ensemble A_5 est borné parce qu'il est contenu par exemple dans la boule centrée en $(0, 0)$ et de rayon 10. Il est ouvert et donc pas fermé. △

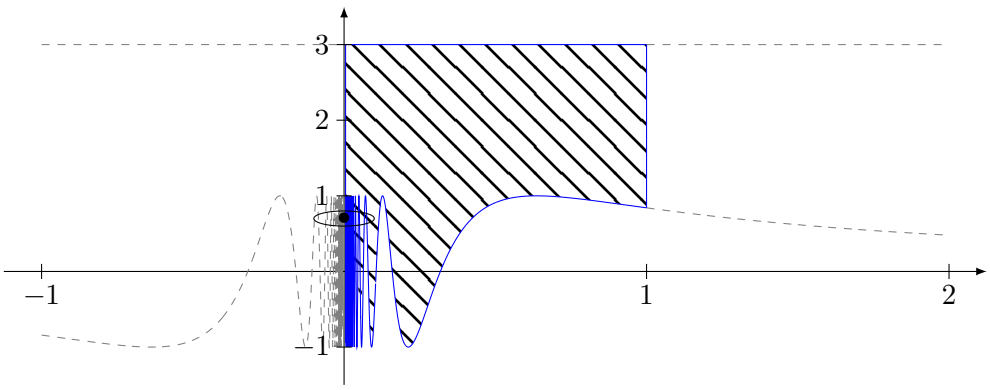


FIGURE 7.2 – Les points qui dont sur l'axe vertical entre 0 et 3 sont sur la frontière, mais pas dans l'ensemble A_5.

Exemple 7.172.
Nous étudions l'ensemble

$$A_6 = \bigcup_{n\in\mathbb{N}_0} \{(\frac{1}{n}, y) \mid y \in [0,1]\ \}. \tag{7.148}$$

L'ensemble A_6 est une union infinie de segments de droites verticaux, voir figure 7.3

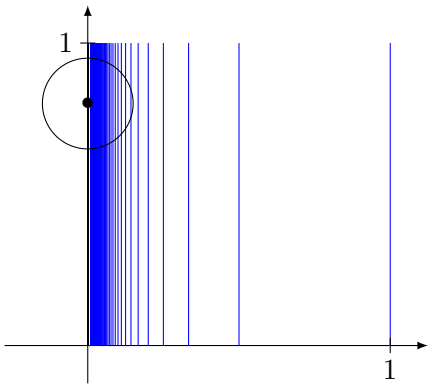


FIGURE 7.3 – Le segment sur l'axe vertical entre $y = 0$ et $y = 1$ fait partie de l'adhérence et de la frontière, mais pas de l'ensemble A_6 lui-même.

L'intérieur est vide parce qu'autour de tout réel de la forme $\frac{1}{n}$, il y a un réel qui n'est pas de cette forme. En ce qui concerne la frontière et l'adhérence, il s'agit de l'union de tous ces segments plus le segment en $x = 0$.

En effet, la boule de rayon r autour du point $(0, y)$ le point $(\frac{1}{n}, y)$ avec n assez grand pour que $\frac{1}{n} < r$. △

7.7.15 Topologie de la droite réelle complétée

Nous introduisons l'ensemble $\bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. À présent les symboles $+\infty$ et $-\infty$ n'ont aucune signification particulière ; il s'agit seulement de deux éléments que nous ajoutons à $\mathbb{R}$ pour former un ensemble que nous notons $\bar{\mathbb{R}}$.

Pas plus tard qu'immédiatement nous leur donnons une signification en définissant une topologie sur $\bar{\mathbb{R}}$. Les ouverts sur $\bar{\mathbb{R}}$ sont

(1) tous les ouverts de $\mathbb{R}$,

(2) les intervalles de la forme $]-\infty, a[$ pour tous les $a \in \mathbb{R}$,

(3) les intervalles de la forme $]a, +\infty[$ pour tous les $a \in \mathbb{R}$,

(4) la topologie engendrée par toutes ces parties de $\bar{\mathbb{R}}$.

Par construction, les boules de $\mathbb{R}$ et les intervalles $]-\infty, a[$ et $]a, +\infty[$ forment une base de topologie pour $\bar{\mathbb{R}}$.

Si f est une fonction $f\colon \mathbb{R} \to \mathbb{R}$, que signifie $\lim_{x\to\infty} f(x)$? Il s'agit de considérer la fonction élargie

$$\begin{aligned} \tilde{f}\colon \bar{\mathbb{R}} &\to \bar{\mathbb{R}} \\ x &\mapsto \begin{cases} f(x) & \text{si } x \in \mathbb{R} \\ 0 & \text{si } x = \pm\infty. \end{cases} \end{aligned} \tag{7.149}$$

Ensuite, c'est la définition topologie usuelle de la limite. Notons que les limites en a ne dépendent pas de la valeur effective de f en a, donc le prolongement par 0 est sans conséquences. Nous pouvions tout aussi bien prolonger par 4.

Le même raisonnement tient pour donner un sens à $\lim_{x\to a} f(x) = \pm\infty$.

7.7.16 Quelques mots à propos de la droite réelle complétée

Définition 7.173.
*La **droite réelle complétée** est l'ensemble $\mathbb{R} \cup \{\pm\infty\}$ où $\pm\infty$ sont deux nouveaux éléments. Nous la notons $\overline{\mathbb{R}}$ pour des raisons que nous verrons à peine plus bas.*

Cette définition ne servirait à rien si nous n'y mettions pas une topologie pour positionner les éléments $\pm\infty$ par rapport à ceux qui existaient déjà dans $\mathbb{R}$.

Définition 7.174 (Topologie sur $\bar{\mathbb{R}}$)**.**
La topologie sur $\bar{\mathbb{R}}$ est celle sur $\mathbb{R}$ à laquelle nous ajoutons les voisinages de $\pm\infty$ de la façon suivante. Une partie V de $\bar{\mathbb{R}}$ est un voisinage de $+\infty$ s'il existe $m > 0$ tel que $]m, +\infty] \subset V$.

Le lemme suivant justifie la notation $\overline{\mathbb{R}}$ pour la droite réelle complétée [55].

Lemme 7.175.
L'adhérence de $\mathbb{R}$ dans $\overline{\mathbb{R}}$ est $\overline{\mathbb{R}}$.

Pour la suite nous utilisons la notation (pratique en probabilité)

$$\{f < a\} = \{x \in S \text{ tel que } f(x) < a\}. \tag{7.150}$$

7.7.17 Limite pointée ou épointée ?

Si vous êtes dans l'enseignement en France [56], vous devriez lire ceci à propos de limite pointée. Dans tous les autres cas, la limite pointée est une notion qui ne vous intéresse à priori pas.

Définition 7.176 ([136])**.**
*Soient X et Y deux espaces topologiques, A une partie de X, f une application de A dans Y, a un point de X adhérent à A et ℓ un point de Y. On dit que ℓ est une **limite pointée** de f au point a si pour tout voisinage V de ℓ, il existe un voisinage W de a tel que pour tout point x de $W \cap A$, l'image $f(x)$ appartient à V.*

La notion de limite pointée ne diffère de la limite que du fait que pour calculer la limite pointée en a, nous tenons compte des valeurs de f sur *tout* le voisinage de a, y compris le point a lui-même.

Le choix entre la limite pointée ou épointée a été discuté en de nombreuses occasions [137, 138, 139, 140].

(1) Dans la majorité des cas, la limite pointée donne le même résultat que la limite parce que, fondamentalement, si nous voulons calculer une limite de f au point a, c'est que f n'est pas définie en a. C'est en particulier toujours le cas pour les limites en l'infini ou les limites définissant les dérivées.

55. Mais ne justifie pas le qualificatif « complété » parce que l'espace métrique $\mathbb{R}$ était déjà complet.
56. En particulier si vous voulez passer l'agrégation.

(2) La limite pointée est un peu plus simple au départ.
(3) La limite épointée est un peu plus riche. Par exemple si on dit « la limite de f en a existe » , ça donne une régularité pour f autour de 0 que la limite pointée ne parvient pas à exprimer.
(4) La limite pointée n'est connue qu'en France.

Le point (1) est le plus important parce qu'il explique pourquoi il y a moyen de finir l'agrégation, et même de faire de la recherche en ne tombant jamais sur un cas où la différence est importante.

Exemple 7.177.
La fonction f donnée par

$$f(x) = \begin{cases} 0 & \text{si } x \neq 0 \\ 4 & \text{si } x = 0 \end{cases} \tag{7.151}$$

a une limite épointée pour $x \to 0$ qui vaut 0. Elle n'a par contre pas de limite pointée en 0.

Cette fonction est l'exemple-type de différence entre limite usuelle et limite pointée. △

Exemple 7.178.
Si vous voulez un cas dans lequel la différence se voit de façon macroscopique, aller lire le lemme 11.153, sa démonstration et l'exemple 11.154. △

Dans le Frido, nous choisissons de prendre la limite épointée comme définition de limite. Nous donnons ici quelques raisons pour ce choix.

(1) C'est la définition unanimement acceptée dans la communauté mathématique hors France.
(2) La limite pointée ne donne à peu près rien de nouveau par rapport à la continuité.
Ce que le concept de limite apporte est la possibilité d'étudier le comportement de f pour les points « proches » de a, sans regarder la valeur en a lui-même. Si l'idée est de regarder le comportement « proche » de a y compris au point a lui-même, c'est la notion de continuité qui fait le travail.
Donc les contextes dans lesquels le concept de limite est intéressant sont justement les contextes dans lesquels la fonction étudiée n'existe pas au point étudié. Dans ce cas, les limites pointées et épointées coïncident.

Que devez-vous faire ?

Enseignement en France La notion de limite pointée est celle nommée « limite » dans les programmes, et ce que nous nommons ici « limite » est nommé « limite épointée ». Peut-être pour induire en erreur tout le reste de la planète ?

Recherche Si vous faites de la recherche où que ce soit y compris en France, la seule définition de limite est la limite dite « épointée », celle qui sera toujours utilisée dans le Frido.

Doctorat Vous commencez un doctorat en math, et vous avez vu la limite pointée comme seule définition de limite durant vos études ? Oubliez-la. Ou alors attendez-vous à vous à de sérieux quiproquos lorsque vous discuterez de mathématique avec des étrangers.
Disons clairement que si vous utilisez la limite pointée devant des non Français, ils se diront juste que vous devriez relire vos cours de base. Et si vous leur expliquez, il y a de bonnes chance qu'ils ne vous croient pas.

7.7.17.1 Le théorème de composition

Une des raisons fréquentes pour utiliser la limite pointée est que le théorème de composition est plus simple[139].

Le voici avec des limites pointées. Pour faire la différence, j'adopte la notation LP pour la limite pointée et LE pour la limite épointée.

Proposition 7.179.
Soient f et g des fonctions $\mathbb{R} \to \mathbb{R}$. Si $LP_{x\to a} g(x) = \ell$ et si $LP_{y\to \ell} f(y) = b$, alors

$$LP_{x\to a}(f \circ g)(x) = LP_{y\to \ell} f(y) = b. \tag{7.152}$$

Démonstration. Soit $\epsilon > 0$. L'hypothèse de limite pour f donne $\eta > 0$ tel que

$$|y - \ell| < \eta \Rightarrow |f(y) - b| < \epsilon. \tag{7.153}$$

Soit $\delta > 0$ tel que $|x - a| < \delta$ implique $|g(x) - \ell| < \eta$.

Avec tout ça, si $|x - a| < \delta$ nous avons $|g(x) - \ell| < \eta$ et en appliquant l'implication (7.153) à $y = g(x)$ nous trouvons $|f\big(g(x)\big) - b| < \epsilon$. □

Avant d'énoncer et de démontrer le résultat correspondant pour les limites épointées, nous avons besoin d'un lemme, pour comprendre la différence d'hypothèse.

Lemme 7.180.
Si $LP_{x\to a} f(x) = b$ alors il y a deux possibilités :

(1) Soit f est définie en a et alors f y est continue,

(2) soit f n'est pas définie en a et alors poser $f(a) = b$ donne un prolongement continu.

Proposition 7.181.
Soient f et g des fonctions $\mathbb{R} \to \mathbb{R}$. Nous supposons $LE_{x\to a} g(x) = \ell$ et que f admette le prolongement continu b en ℓ[57] *Alors*

$$LE_{x\to a}(f \circ g)(x) = b. \tag{7.154}$$

Bon. Woaw. La différence est énorme.

Démonstration. Soit $\epsilon > 0$. Par hypothèse de prolongement continu, il existe $\eta > 0$ tel que

$$|t - \ell| < \eta \Rightarrow |f(y) - b| < \epsilon. \tag{7.155}$$

Soit $\delta > 0$ tel que

$$0 < |x - a| < \delta \Rightarrow |g(x) - \ell| \leqslant \eta. \tag{7.156}$$

Avec ce δ nous avons que $|0 < |x - a| < \delta|$ implique $|f\big(g(x)\big) - b| < \epsilon$. □

Cela est surement une raison de présenter la limite pointée chez les petits. Mais ce n'est pas une raison pour les grands, que du contraire. L'énoncé de la proposition 7.179 est à peine plus compliquée que celui de 7.179, mais elle dit un peu plus alors que la démonstration est la même.

Donc non, la proposition 7.181 n'ajoute pas d'hypothèse par rapport à 7.179. Au contraire, elle en enlève : la proposition 7.181 demande pour g une limite épointée au lieu d'une limite pointée. En ce qui concerne f, les hypothèses sont les mêmes. La proposition 7.181 concerne une classe de fonctions un peu plus grande.

Au final, la proposition « épointée » 7.181 est un poil plus compliquée, mais elle a une hypothèse un peu plus faible et une conclusion un peu plus faible (existence d'une limite pointée). Bref, il s'agit d'un résultat différent. Mais comme maintenant nous sommes grands, nous sommes prêts à avoir des énoncés plus compliquée pour avoir des résultats plus complets.

7.7.17.2 Et les filtres ?

Si vous ne savez pas ce qu'est un filtre, vous pouvez sauter ces paragraphes. Sinon, vous pouvez vous dire que le débat « limite pointée » contre « limite épointée » n'a aucun sens parce que de toutes façons, la bonne façon de définir une limite passe par des filtres.

Alors le mieux est de se demander comment on construit, à partir de la notion de filtre, le nombre $\lim_{x\to a} f(a)$?

Pas de bol, ça dépend du filtre choisi. Le premier filtre auquel on pense pour trouver une définition raisonnable de la limite de $f(x)$ quand $x \to a$ est le filtre des voisinages de a. La notion de limite associée est la limite pointée. En ce sens la limite pointée est plus naturelle que la limite

57. Cette hypothèse est équivalente à dire que f a une limite pointée b en ℓ, c'est-à-dire la même hypothèse que dans la proposition 7.181.

épointée. Cependant « naturel » signifie souvent « le premier qui nous tombe sous la main », ce qui ne signifie pas spécialement « le plus intéressant à utiliser ».

La notion de limite épointée est la limite associée au filtre des voisinages épointés. Ce n'est, certes, pas le premier filtre qui nous tombe sous la main, mais il est, au moins dans le cadre de l'étude des fonctions sur $\mathbb{R}^n$, le plus efficace ; celui qui donne le plus de nouvelles informations par rapport à la continuité.

7.8 Formes bilinéaires et quadratiques

Plus à propos de formes bilinéaires dans le thème 49.

Définition 7.182 ([141]).
*Soient trois espaces vectoriels E, F et V sur le même corps commutatif $\mathbb{K}$. Une application $b\colon E \times F \to V$ est **bilinéaire** si elle est séparément linéaire en ses deux variables, c'est-à-dire si*

(1) $b(u_1 + u_2, v) = b(u_1, v) + b(u_2, v)$,

(2) $b(u, v_1 + v_2) = b(u, v_1) + b(u, v_2)$

(3) $b(\lambda u, v) = b(u, \lambda v) = \lambda b(u, v)$

pour tout $u, u_1, u_2 \in E$, $v, v_1, v_2 \in F$ et pour tout $\lambda \in \mathbb{K}$.

*Dans le cas $E = F$ et $V = \mathbb{K}$, nous parlons de **forme bilinéaire** sur E.*

*Nous parlons de forme bilinéaire **symétrique** si de plus $b(u, v) = b(v, u)$.*

7.183.
Une application bilinéaire $E \times E \to \mathbb{K}$ n'est pas une application linéaire ; la distinction est importante. La linéarité est

$$b(\lambda u, \lambda v) = b\big(\lambda(u, v)\big) = \lambda b(u, v) \tag{7.157}$$

et la bilinéarité est

$$b(\lambda u, v) = b(u, \lambda v) = \lambda b(u, v). \tag{7.158}$$

En réalité la seule forme qui soit à la fois linéaire et bilinéaire est la forme identiquement nulle : la condition

$$b(\lambda u, \lambda v) = \lambda^2 b(u, v) = \lambda b(u, v) \tag{7.159}$$

pour tout $\lambda \in \mathbb{K}$ implique $b(u, v) = 0$.

Exemple 7.184 ([142]).
L'application

$$\begin{aligned} b\colon \mathrm{M}(n, \mathbb{K}) \times \mathrm{M}(n, \mathbb{K}) &\to \mathbb{K} \\ (A, B) &\mapsto \mathrm{Tr}(AB) \end{aligned} \tag{7.160}$$

est une forme bilinéaire symétrique.

La vérification est un calcul :

$$\mathrm{Tr}(BA) = \sum_i (BA)_{ii} = \sum_{ik} B_{ik} A_{ki} = \sum_{ik} A_{ki} A_{ik} = \sum_k (AB)_{kk} = \mathrm{Tr}(AB). \tag{7.161}$$

△

7.9 Produit scalaire, produit hermitien

Définition 7.185 (Définie positive, thème 42).
Si g est une application bilinéaire[58] *sur un espace vectoriel E nous disons qu'elle est*

(1) ***définie positive*** *si $g(x, x) \geqslant 0$ pour tout $x \in E$ et $g(x, x) = 0$ si et seulement si $x = 0$.*

58. Définition 7.182.

(2) ***semi-définie positive*** *si* $g(x,x) \geqslant 0$ *pour tout* $x \in E$. *Nous dirons aussi parfois qu'elle est simplement « positive ».*

Cela est évidemment à lier à la définition 9.152 et la proposition 9.156 : une application bilinéaires est définie positive si et seulement si sa matrice symétrique associée l'est.

Définition 7.186.
Un ***produit scalaire*** *sur un espace vectoriel réel est une forme bilinéaire*[59] *symétrique strictement définie positive*[60].

La définition suivante est utile pour celles qui veulent faire de la relativité[61].

Définition 7.187.
Un ***produit pseudo-scalaire*** *sur un espace vectoriel réel est une forme bilinéaire et symétrique.*

Vu que nous allons voir un pâté d'espaces avec des produits scalaires, nous leur donnons un nom.

Définition 7.188.
Un espace vectoriel ***euclidien*** *est un espace vectoriel de dimension finie muni d'un produit scalaire (définition 7.186).*

Avouez que c'est drôle qu'un espace vectoriel est euclidien lorsqu'il possède une *multiplication* alors qu'un anneau est euclidien lorsqu'il possède une *division* (voir la définition 3.151). C'est pas très profond, mais si ça peut vous servir de moyen mnémotechnique...

Définition 7.189 ([143])**.**
Soit E *est un espace vectoriel sur* $\mathbb{C}$. *Une application* $\langle .,. \rangle \colon E \times E \to \mathbb{C}$ *est* ***sesquilinéaire à droite*** *si pour tout* $x, y \in E$ *et pour tout* $\lambda \in \mathbb{C}$,

(1) $\langle \lambda x, y \rangle = \lambda \langle x, y, \rangle = \langle x, \bar{\lambda} y \rangle$,

(2) $\langle x + y, z \rangle = \langle x, y \rangle + \langle y, z \rangle$,

(3) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$.

Cette forme est ***hermitienne*** *si de plus*

$$\langle x, y \rangle = \overline{\langle y, x \rangle}. \tag{7.162}$$

Un ***produit hermitien*** *est une forme hermitienne strictement définie positive, c'est-à-dire telle que* $\langle x, x \rangle \geqslant 0$ *pour tout* $x \in E$ *et* $\langle x, x \rangle = 0$ *si et seulement si* $x = 0$.

Exemple 7.190.
L'ensemble $E = \mathbb{C}^n$ vu comme espace vectoriel de dimension n sur $\mathbb{C}$ est muni d'une forme sesquilinéaire

$$\langle x, y \rangle = \sum_{k=1}^{n} x_k \bar{y}_k \tag{7.163}$$

pour tout $x, y \in \mathbb{C}^n$. Cela est un espace vectoriel hermitien. △

7.9.1 Norme, produit scalaire et Cauchy-Schwarz (cas réel)

Dans la suite, le produit scalaire de x et y pourra être noté indifféremment par $x \cdot y$, $\langle x, y \rangle$ ou $b(x, y)$ lorsque une forme bilinéaire est donnée.

Nous rappelons au passage que les espaces vectoriels réels sont susceptibles de recevoir un produit scalaire, alors que les espaces vectoriels complexes sont susceptibles de recevoir un produit

59. Définition 7.182.
60. Définition 7.185.
61. Voir le théorème 17.168 qui établit les transformations de Lorentz.

hermitien. Bien que de nombreux résultats soient identiques ou très similaires, ces deux notions sont à ne pas confondre.

Nous commençons par prouver qu'un produit scalaire étant donné, nous pouvons définir une norme par la formule $\|x\|^2 = \langle x, x\rangle$. Pour cela nous aurons besoin de l'inégalité de Cauchy-Schwarz.

Théorème 7.191 (Inégalité de Cauchy-Schwarz, cas réel)**.**
Soit un espace vectoriel muni d'un produit scalaire $(x, y) \mapsto x \cdot y$. En posant[62]

$$\|x\| = \sqrt{x \cdot x}, \tag{7.164}$$

nous avons

$$|x \cdot y| \leqslant \|x\|\|y\|. \tag{7.165}$$

Nous avons une égalité si et seulement si x et y sont multiples l'un de l'autre.

Démonstration. Étant donné que les deux membres de l'inéquation sont positifs, nous allons travailler en passant au carré afin d'éviter les racines carrés dans le second membre.

Nous considérons le polynôme

$$P(t) = \|x + ty\|^2 = (x + ty) \cdot (x + ty) = x \cdot x + x \cdot ty + ty \cdot x + t^2 y \cdot y. \tag{7.166}$$

En utilisant la bilinéarité (pour sortir les t) et la symétrique du produit scalaire, puis en ordonnant les termes selon les puissances de t,

$$P(t) = \|y\|^2 t^2 + 2(x \cdot y)t + \|x\|^2. \tag{7.167}$$

Cela est un polynôme du second degré en t dont le signe est toujours positif (ou nul). Par conséquent le discriminant[63] doit être négatif ou nul. Nous avons donc

$$\Delta = 4(x \cdot y)^2 - 4\|x\|^2\|y\|^2 \leqslant 0, \tag{7.168}$$

ce qui donne immédiatement

$$(x \cdot y)^2 \leqslant \|x\|^2\|y\|^2. \tag{7.169}$$

En ce qui concerne le cas d'égalité, si nous avons $x \cdot y = \|x\|\|y\|$, alors le discriminant Δ ci-dessus est nul et le polynôme P admet une racine double t_0. Pour cette valeur nous avons

$$P(t_0) = |x + t_0 y| = 0, \tag{7.170}$$

ce qui implique $x + t_0 y = 0$ et donc que x et y sont liés. □

La proposition suivante montre que toute norme dérivant d'un produit scalaire vérifie l'identité du parallélogramme. Ce résultat sert souvent à prouver que des normes ne dérivent pas d'un produit scalaire. C'est le cas de la norme $N(x, y) = |x| + |y|$ du lemme 7.194 ainsi que du théorème de Weinersmith 26.42.

Proposition 7.192 (Norme dérivant d'un produit scalaire)**.**
Si $x, y \mapsto x \cdot y$ est un produit scalaire sur un espace vectoriel réel E. Nous posons $\|x\| = \sqrt{x \cdot x}$. Alors

(1) L'opération $\|.\|$ est une norme[64]*.*

(2) Cette norme vérifie l'identité du parallélogramme :

$$\|x - y\|^2 + \|x + y\|^2 = 2\|x\|^2 + 2\|y\|^2. \tag{7.171}$$

62. Attention à la notation : pour l'instant nous ne savons pas que c'est une norme. Ce sera justifié dans la proposition 7.192.
63. Le fameux $b^2 - 4ac$.
64. Définition 7.113.

Démonstration. En deux parties.

C'est une norme Nous allons nous contenter de prouver l'inégalité triangulaire. Si $x, y \in E$ nous avons

$$\|x+y\| = \sqrt{\|x\|^2 + \|y\|^2 + 2x \cdot y}. \tag{7.172}$$

Par l'inégalité de Cauchy-Schwarz, théorème 7.191 nous avons aussi

$$2x \cdot y \leqslant 2\|x\|\|y\|. \tag{7.173}$$

Nous pouvons donc majorer ce qui est dans la racine carrée :

$$\|x\|^2 + \|y\|^2 + 2x \cdot y \leqslant \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| = \big(\|x\| + \|y\|\big)^2. \tag{7.174}$$

En remettant les bouts ensemble,

$$\|x+y\| = \sqrt{\|x\|^2 + \|y\|^2 + 2x \cdot y} \leqslant \sqrt{\big(\|x\| + \|y\|\big)^2} = \|x\| + \|y\|. \tag{7.175}$$

Inégalité du parallélogramme Cette assertion est seulement un calcul :

$$\begin{aligned}\|x-y\|^2 + \|x+y\|^2 &= (x-y) \cdot (x-y) + (x+y) \cdot (x+y) \\ &= x \cdot x - x \cdot y - y \cdot x + y \cdot y \\ &\quad + x \cdot x + x \cdot y + y \cdot x + y \cdot y \\ &= 2x \cdot x + 2y \cdot y \\ &= 2\|x\|^2 + 2\|y\|^2.\end{aligned} \tag{7.176}$$

□

7.193.
Un produit scalaire fourni donc toujours une norme et donc une topologie. Il ne faudrait cependant pas croire que toute norme dérive d'un produit scalaire, même pas en dimension finie. Et ce, malgré l'équivalence de toutes les normes du théorème 7.224 dont vous avez déjà peut-être entendu parler.

L'intérêt du lemme suivant sera apparent en 7.226.

Lemme 7.194.
Sur $\mathbb{R}^2$, l'application $N(x,y) = |x| + |y|$ est une norme[65] *qui ne dérive pas d'un produit scalaire*[66].

Démonstration. Nous commençons par montrer que N est une norme. Il faut vérifier les trois conditions de la définition 7.113.

(1) Il faut utiliser le lemme 1.112(1) dans les deux sens. Si $(x,y) = (0,0)$, alors évidemment $N(x,y) = 0$. Dans l'autre sens, si $N(x,y) = 0$ nous avons

$$0 = |x| + |y| \geqslant |x|. \tag{7.177}$$

Donc $|x| \leqslant 0$, mais comme $|x| \geqslant 0$, nous avons $|x| = 0$ et donc $x = 0$. Le même raisonnement tient pour y.

(2) En tenant compte du fait que $|\lambda x| = |\lambda||x|$, nous avons

$$N\big(\lambda(x,y)\big) = N(\lambda x, \lambda y) = |\lambda||x| + |\lambda||y| = |\lambda|(|x| + |y|) = |\lambda| N(x,y). \tag{7.178}$$

(3) Nous avons le calcul

$$\begin{aligned}N\big((x,y) + (a,b)\big) &= N(x+a, y+b) &(7.179a)\\ &= |x+a| + |y+b| &(7.179b)\\ &\leqslant |x| + |a| + |y| + |b| &(7.179c)\\ &= N(x,y) + N(a,b) &(7.179d)\end{aligned}$$

Justification : pour (7.179c) nous avons utilité $|a+b| \leqslant |a| + |b|$, du lemme 1.112.

65. Définition 7.113.
66. La norme d'un produit scalaire est la proposition 7.192.

Pour voir qu'elle ne dérive pas d'un produit scalaire, nous montrons qu'elle ne vérifie pas l'identité du parallélogramme de la proposition 7.192.

Voici un petit bout de code qui nous permet de ne pas faire de recherches à la main :

```
# Dans un cas réel, vous avez nettement intérêt à
# créer une classe 'Vecteur' qui implémente somme, différence
# et norme.
def N(v):
    return abs( v[0] )+abs(v[1])

def parall(v,w):
    # La différence v-w
    d=( v[0]-w[0],v[1]-w[1] )
    # La somme v+w
    s=( v[0]+w[0],v[1]+w[1] )

    return N(d)**2+N(s)**2-2*N(v)**2-2*N(w)**2
```

tex/sage/sageSnip018.sage

Il est vite vu qu'avec $v = (-1, 1)$ et $w = (1, 1)$, l'identité du parallélogramme n'est pas vérifiée. □

Lemme 7.195 ([54]).
Soit V un espace vectoriel muni d'un produit scalaire et de la norme associée. Si $x, y \in V$ satisfont à $\|x + y\| = \|x\| + \|y\|$, alors il existe $\lambda \geqslant 0$ tel que $x = \lambda y$.

Démonstration. Quitte à raisonner avec $x/\|x\|$ et $y/\|y\|$, nous supposons que $\|x\| = \|y\| = 1$. Dans ce cas l'hypothèse signifie que $\|x+y\|^2 = 4$. D'autre part en écrivant la norme en termes de produit scalaire,

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle, \tag{7.180}$$

ce qui nous mène à affirmer que $\langle x, y\rangle = 1 = \|x\|\|y\|$. Nous sommes donc dans le cas d'égalité de l'inégalité de Cauchy-Schwarz[67], ce qui nous donne un λ tel que $x = \lambda y$. Étant donné que $\|x\| = \|y\| = 1$ nous avons obligatoirement $\lambda = \pm 1$, mais si $\lambda = -1$ alors $\langle x, y\rangle = -1$, ce qui est le contraire de ce qu'on a prétendu plus haut. Par souci de cohérence, nous allons donc croire que $\lambda = 1$. □

Proposition 7.196.
si $v_1, \cdots, v_k$ sont des vecteurs non nuls, orthogonaux deux à deux, alors ces vecteurs forment une famille libre.

Lemme 7.197.
Une isométrie d'un espace euclidien fixe l'origine.

Démonstration. Soit une isométrie f d'un espace euclidien : $f(x) \cdot f(y) = x \cdot y$ pour tout $x, y \in E$. En particulier pour $x = 0$ nous avons

$$f(0) \cdot f(y) = 0 \tag{7.181}$$

pour tout y. Vu que f est une bijection, nous avons $f(0) \cdot x = 0$ pour tout x. Comme le produit scalaire est non dégénéré cela implique que $f(0) = 0$. □

67. Théorème 7.191.

7.9.2 Cauchy-Schwarz etc. cas complexe

Théorème 7.198 (Inégalité de Cauchy-Schwarz, cas complexe[144])**.**
Soit un espace vectoriel complexe muni d'un produit hermitien $\langle .,. \rangle$. Alors pour tout vecteurs x, y nous avons

$$|\langle x, y\rangle| \leqslant \|x\|\|y\| \tag{7.182}$$

où nous avons posé $\|x\| = \sqrt{\langle x, x\rangle}$.

Démonstration. Si $\langle x, y\rangle = 0$, le résultat est évident ; nous supposons que non. Nous posons

$$\theta = \frac{\langle x, y\rangle}{|\langle x, y\rangle|}. \tag{7.183}$$

C'est un élément de $\mathbb{C}$ de norme 1. Nous avons

$$\langle \frac{1}{\theta}x, y\rangle = \frac{|\langle x, y\rangle|}{\langle x, y\rangle}\langle x, y\rangle = |\langle x, y\rangle| \geqslant 0 \tag{7.184}$$

où le symbole « $\geqslant$ » signifie « est réel et positif ». Nous posons $x' = \frac{1}{\theta}x$ et nous considérons $t \in \mathbb{R}$. Remarquons que $\|x'\|^2 = \|x\|^2$:

$$\|x'\|^2 = \langle x', x'\rangle = \frac{1}{\theta\bar{\theta}}\langle x, x\rangle = \|x\|^2 \tag{7.185}$$

parce que $|\theta| = 1$.

En utilisant le fait que $\langle a, b\rangle + \langle b, a\rangle = \mathrm{Re}(\langle a, b\rangle)$ nous avons :

$$0 \leqslant \|x' + ty\|^2 = \|x'\|^2 + t\langle x', y\rangle + t\langle y, x'\rangle + t^2\|y\|^2 \tag{7.186a}$$

$$= \|y\|^2t^2 + 2\,\mathrm{Re}(\langle x', y\rangle)t + \|x'\|^2. \tag{7.186b}$$

Cela est un polynôme de degré 2 en t qui n'est jamais strictement négatif. Autrement dit, il a au maximum une seule racine, ce qui signifie que son discriminant est négatif ou nul :

$$\mathrm{Re}(\langle x', y\rangle)^2 - \|y\|^2\|x'\|^2 \leqslant 0. \tag{7.187}$$

Mais nous avons choisi x' de telle sorte que $\langle x', y\rangle = |\langle x, y\rangle| \in \mathbb{R}$ et $\|x'\|^2 = \|x\|^2$; nous avons donc

$$|\langle x, y\rangle|^2 \leqslant \|x\|^2\|y\|^2, \tag{7.188}$$

comme il se devait. □

Proposition 7.199 (Identité du parallélogramme[145])**.**
Soit une espace vectoriel complexe E muni d'un produit hermitien $\langle .,. \rangle$. Nous posons $\|x\| = \sqrt{\langle x, x\rangle}$. Nous avons

(1) $\|.\|$ est une norme.

(2) Elle vérifie l'identité du parallélogramme :

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|b\|^2 \tag{7.189}$$

pour tout $x, y \in E$.

Démonstration. En ce qui concerne le fait que $\|.\|$ soit une norme, tout est essentiellement dans la définition 7.189 d'un produit hermitien. Voyons tout de même l'inégalité triangulaire. Nous avons :

$$\|x + y\|^2 = \langle x + y, x + y\rangle \tag{7.190a}$$

$$= \|x\|^2 + \|y\|^2 + \langle x, y\rangle + \langle y, x\rangle \tag{7.190b}$$

$$= \|x\|^2 + \|y\|^2 + 2\Re\big(\langle x, y\rangle\big) \tag{7.190c}$$

$$\leqslant \|x\|^2 + \|y\|^2 + 2|\Re\big(\langle x, y\rangle\big)| \tag{7.190d}$$

$$\leqslant \|x\|^2 + \|y\|^2 + 2|\langle x, y\rangle| \tag{7.190e}$$

$$\leqslant \|x\|^2 + \|y\|^2 + 2\|x\|\|y\| \tag{7.190f}$$

$$= \big(\|x\| + \|y\|\big)^2. \tag{7.190g}$$

Pour (7.190f) nous avons utilisé Cauchy-Schwarz 7.198. □

7.9.3 Projection et orthogonalité

La définition du produit scalaire dans $\mathbb{R}^n$ est 4.122 et le lien avec la matrice d'une application linéaire est la proposition 4.123.

Remarque 7.200.
Outre l'orthogonalité, le produit scalaire permet de savoir l'angle entre deux vecteurs à travers la définition 17.41. D'autres interprétations géométriques du déterminant sont listées dans le thème 56.

Nous sommes maintenant en mesure de déterminer, pour deux vecteurs quelconques u et v, la projection orthogonale de u sur v. Ce sera le vecteur $\bar{u}$ parallèle à v tel que $u - \bar{u}$ est orthogonal à v. Nous avons donc

$$\bar{u} = \lambda v \tag{7.191}$$

et

$$(u - \lambda v) \cdot v = 0. \tag{7.192}$$

La seconde équation donne $u \cdot v - \lambda v \cdot v = 0$, ce qui fournit λ en fonction de u et v :

$$\lambda = \frac{u \cdot v}{\|v\|^2}. \tag{7.193}$$

Nous avons par conséquent

$$\bar{u} = \frac{u \cdot v}{\|v\|^2} v. \tag{7.194}$$

Armés de cette interprétation graphique du produit scalaire, nous comprenons pourquoi nous disons que deux vecteurs sont orthogonaux lorsque leur produit scalaire est nul.

Nous pouvons maintenant savoir quel est le coefficient directeur d'une droite orthogonale à une droite donnée. En effet, supposons que la première droite soit parallèle au vecteur X et la seconde au vecteur Y. Les droites seront perpendiculaires si $X \cdot Y = 0$, c'est-à-dire si

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 0. \tag{7.195}$$

Cette équation se développe en

$$x_1 y_1 = -x_2 y_2. \tag{7.196}$$

Le coefficient directeur de la première droite est $\frac{x_2}{x_1}$. Isolons cette quantité dans l'équation (7.196) :

$$\frac{x_2}{x_1} = -\frac{y_1}{y_2}. \tag{7.197}$$

Donc le coefficient directeur de la première est l'inverse et l'opposé du coefficient directeur de la seconde.

Exemple 7.201.
Soit la droite $d \equiv y = 2x + 3$. Le coefficient directeur de cette droite est 2. Donc le coefficient directeur d'une droite perpendiculaires doit être $-\frac{1}{2}$. △

Preuve alternative. La preuve peut également être donnée en ne faisant pas référence au produit scalaire. Il suffit d'écrire toutes les quantités en termes des coordonnées de X et Y. Si nous posons

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_2 \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \tag{7.198}$$

l'inégalité à prouver devient

$$(x_1 y_1 + x_2 y_2 + x_3 y_3)^2 \leqslant (x_1^2 + x_2^2 + x_3^2)(y_1^2 + y_2^2 + y_3^2). \tag{7.199}$$

Nous considérons la fonction

$$\varphi(t) = (x_1 + ty_1)^2 + (x_2 + ty_2)^2 + (x_3 + ty_3)^2 \tag{7.200}$$

En tant que norme, cette fonction est évidemment positive pour tout t. En regroupant les termes de chaque puissance de t, nous avons

$$\varphi(t) = (y_1^2 + y_2^2 + y_3^2)t^2 + 2(x_1y_1 + x_2y_2 + x_3y_3)t + (x_1^2 + x_2^2 + x_3^2). \tag{7.201}$$

Cela est un polynôme du second degré en t. Par conséquent le discriminant doit être négatif. Nous avons donc

$$4(x_1y_1 + x_2y_2 + x_3y_3)^2 - (x_1^2 + x_2^2 + x_3^2)(y_1^2 + y_2^2 + y_3^2) \leqslant 0. \tag{7.202}$$

La thèse en découle aussitôt. □

Proposition 7.202.
La norme euclidienne a les propriétés suivantes :

(1) Pour tout vecteur X et réel λ, $\|\lambda X\| = |\lambda|\|X\|$. Attention à ne pas oublier la valeur absolue !

(2) Pour tout vecteurs X et Y, $\|X + Y\| \leqslant \|X\| + \|Y\|$.

Démonstration. Pour le second point, nous avons les inégalités suivantes :

$$\begin{aligned}
\|X + Y\|^2 &= \|X\|^2 + \|Y\|^2 + 2X \cdot Y && (7.203a)\\
&\leqslant \|X\|^2 + \|Y\|^2 + 2|X \cdot Y| && (7.203b)\\
&\leqslant \|X\|^2 + \|Y\|^2 + 2\|X\|\|Y\| && (7.203c)\\
&= \big(\|X\| + \|Y\|\big)^2 && (7.203d)
\end{aligned}$$

Nous avons utilisé d'abord la majoration $|x| \geqslant x$ qui est évidente pour tout nombre x ; et ensuite l'inégalité de Cauchy-Schwarz 7.191. □

7.9.4 Théorème de Pythagore

Nous allons donner une preuve du théorème de Pythagore.

Théorème 7.203 (Pythagone[1])**.**
Soient $A, B, S \in \mathbb{R}^2$ un triangle rectangle en A, c'est-à-dire tel que

$$(B - A) \cdot (A - S) = 0. \tag{7.204}$$

Alors

$$\|S - B\|^2 = \|S - A\|^2 + \|B - A\|^2. \tag{7.205}$$

Démonstration. En développant l'hypothèse (7.204) nous avons :

$$B \cdot A - B \cdot S - \|A\|^2 + A \cdot S = 0. \tag{7.206}$$

Et de même,

$$\|S - B\|^2 = (S - B) \cdot (S - B) = \|S\|^2 - 2B \cdot S + \|B\|^2. \tag{7.207}$$

En substituant dans cette dernière $B \cdot S$ par $B \cdot S = B \cdot A - \|A\|^2 + A \cdot S$ tirée de (7.206), nous trouvons

$$\|S - B\|^2 = \|S\|^2 - 2B \cdot A + 2\|A\|^2 - 2A \cdot S + \|B\|^2 = \|S - A\|^2 + \|B - A\|^2. \tag{7.208}$$

□

Je profite de l'occasion pour montrer mon scepticisme quant aux preuves de Pythagore basées sur différents pliages et découpages des carrés construits sur les côtés du triangle. Pour autant que je le sache, la géométrie dans « le plan » (c'est-à-dire pas dans $\mathbb{R}^2$ muni de son produit scalaire) ne définit pas « longueur » et « aire ». Donc bon ... Il y a peut-être moyen de s'en sortir, mais je ne le connais pas.

7.9.5 Produit vectoriel

Définition 7.204.
Soient u et v, deux vecteurs de $\mathbb{R}^3$. Le ***produit vectoriel*** *de u et v est le vecteur $u \times v$ défini par*

$$u \times v = \det \begin{pmatrix} e_1 & e_2 & e_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{pmatrix} \tag{7.209}$$

où les vecteurs e_1, e_2 et e_3 sont les vecteurs de la base canonique de $\mathbb{R}^3$.

Lemme 7.205.
Le produit vectoriel $u \times v$ est également exprimé par

$$u \times v = (u_2v_3 - u_3v_2)e_1 + (u_3v_1 - u_1v_3)e_2 + (u_1v_2 - u_2v_1)e_3 \tag{7.210a}$$
$$= \sum_{i,j,k} \epsilon_{ijk} v_i w_j e_k \tag{7.210b}$$

où ϵ_{ijk} est défini par $\epsilon_{xyz} = 1$ et ensuite ϵ_{ijk} est 1 ou -1 suivant que la permutation des x, y et z est paire ou impaire. C'est-à-dire que ϵ_{ijk} est la signature de la permutation qui amène $(1, 2, 3)$ sur (i, j, k).

Démonstration. Il s'agit seulement de développer explicitement le déterminant (7.209). □

7.206.
Mettons que $a \times b = v$. En calculant le même produit vectoriel dans la base $f_i = -e_i$, les composantes de a et b changent de signe et la formule (7.210) dit que le produit vectoriel ne change pas. On serait tenter d'écrire, dans la base $\{f_i\}$

$$(-a) \times (-b) = v, \tag{7.211}$$

tout en pleurant parce que dans la base des f_i, le vecteur v devient $-v$.

Il a des personnes que cela tracasse tellement qu'on entend parler de « le produit vectoriel est une pseudo-vecteur sous SO(2) ».

Il suffit d'être clair. Le produit vectoriel n'est défini que sur $\mathbb{R}^3$, et est définit par sa formule dans la base canonique, point barre. Si vous avez des vecteurs a et b dont vous connaissez les composantes dans une autre base, vous devez calculer les composantes dans la base canonique, utiliser la formule pour trouver les composantes de $a \times b$ dans la base canonique. Ensuite, si ça vous chante, vous pouvez calculer à nouveau les composantes de $a \times b$ dans une autre base.

Tout cela pour dire que le produit vectoriel n'est pas une opération très généralisable. Il est possible, pour sembler plus intrinsèque, de tenter cette définition : le produit vectoriel $a \times b$ est le vecteur perpendiculaire à a et b, de longueur égale à l'aire du parallélogramme construit sur a et b.

Cette « définition » a plusieurs inconvénients.

— Elle demande quand même un produit scalaire et des aires ; bref, elle demande une structure métrique,

— Elle ne donne pas le sens. En effet, dans $\mathbb{R}^3$, il y a deux vecteurs de longueur donnée perpendiculaires à a et b. Il faut donc préciser le sens. Cela revient à donner une orientation et donc, fondamentalement, à choisir une base.

Bref, on retiendra que le produit vectoriel est une opération accrochée à $\mathbb{R}^3$ et a sa base canonique.

Une des principales utilités du produit vectoriel est donnée dans la proposition suivante.

Proposition 7.207.
Si u et v sont des vecteurs de $\mathbb{R}^3$ alors le vecteur $u \times v$ est perpendiculaire à u et à v.

La chose importante à retenir est que le produit vectoriel permet de construire un vecteur simultanément perpendiculaire à deux vecteurs donnés. Le vecteur $u \times v$ est donc linéairement indépendant de u et v. En pratique, si u et v sont déjà linéairement indépendants, alors le produit vectoriel permet de compléter une base de $\mathbb{R}^3$.

Lemme-définition 7.208.
Nous avons l'égalité suivante pour tout $u, v, w \in \mathbb{R}^3$:

$$(u \times v) \cdot w = \det \begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}. \tag{7.212}$$

*Le résultat est nommé le **produit mixte** de trois vecteurs de $\mathbb{R}^3$.*

7.209.
Nous avons donné un nom à la combinaison $(u \times v) \cdot w$. J'imagine que vous voyez pourquoi nous ne considérons pas la combinaison $(u \cdot v) \times w$.

Le lemme suivant donne un moyen compliqué et peu pratique de calculer la valeur absolue du produit mixte. La formule (7.213) ne sera utilisée que pour faire le lien entre un jacobien et un élément de volume en dimension trois lorsque nous verrons les intégrales sur des variétés. Voir l'équation (19.37).

Lemme 7.210 ([1])**.**
Le produit mixte peut également être exprimé par

$$|(u \times v) \cdot w|^2 = \det \begin{pmatrix} \|u\|^2 & u \cdot v & u \cdot w \\ v \cdot u & \|v\|^2 & v \cdot w \\ w \cdot u & w \cdot v & \|w\|^2 \end{pmatrix}. \tag{7.213}$$

Démonstration. Si nous notons

$$a = \begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}, \tag{7.214}$$

il faut simplement remarquer que

$$\begin{pmatrix} \|u\|^2 & u \cdot v & u \cdot w \\ v \cdot u & \|v\|^2 & v \cdot w \\ w \cdot u & w \cdot v & \|w\|^2 \end{pmatrix} = aa^t. \tag{7.215}$$

Donc au niveau des déterminants, en utilisant les propositions 4.99 et le lemme 4.76 nous avons

$$\det \begin{pmatrix} \|u\|^2 & u \cdot v & u \cdot w \\ v \cdot u & \|v\|^2 & v \cdot w \\ w \cdot u & w \cdot v & \|w\|^2 \end{pmatrix} = \det(aa^t) = \det(a)\det(a^t) = \det(a)^2. \tag{7.216}$$

Et maintenant, par définition, $\det(a) = (u \times w) \cdot w$. Donc le résultat annoncé. □

Proposition 7.211.
Les applications produit scalaire, vectoriel et mixte sont multilinéaires. Spécifiquement, nous avons les propriétés suivantes.

(1) Les applications produit scalaire et vectoriel sont bilinéaires. C'est-à-dire que pour tout vecteurs a, b, c et pour tout nombre α et β nous avons

$$\begin{aligned} a \times (\alpha b + \beta c) &= \alpha(a \times b) + \beta(a \times c) \\ (\alpha a + \beta b) \times c &= \alpha(a \times c) + \beta(b \times c). \end{aligned} \tag{7.217}$$

(2) Le produit mixte est trilinéaire.

(3) Le produit vectoriel est antisymétrique, c'est-à-dire $u \times v = -v \times u$.

(4) Nous avons $u \times v = 0$ si et seulement si u et v sont colinéaires, c'est-à-dire si et seulement si l'équation $\alpha u + \beta v = 0$ a une solution différente de la solution triviale $(\alpha, \beta) = (0, 0)$.

Proposition 7.212 (Identité de Lagrange[146])**.**
Si $x, y \in \mathbb{R}^n$, alors

$$\|x\|^2\|y\|^2 - (x \cdot y)^2 = \sum_j \sum_{i<j} (x_i y_j - x_j y_i)^2. \tag{7.218}$$

Et si $n = 3$ alors

$$\|x \times y\| = \|y\|^2\|y\|^2 - (x \cdot y)^2. \tag{7.219}$$

Démonstration. C'est un calcul. D'abord nous avons

$$\|x\|^2\|y\|^2 - (x \cdot y)^2 = \sum_i x_i^2 \sum_j y_j^2 - \big(\sum_k x_k y_k\big)^2 = \sum_{ij} x_i^2 y_j^2 - \sum_{kl} x_k y_k x_l y_l. \tag{7.220}$$

Ensuite nous coupons les sommes de la façon suivante

$$\sum_{ij} = \sum_j \sum_{i<j} + \sum_j (i = j) + \sum_j \sum_{i>j} \tag{7.221}$$

pour obtenir

$$\begin{aligned} \|x\|^2\|y\|^2 - (x \cdot y)^2 = & \sum_j \sum_{i<j} x_i^2 y_j^2 + \sum_j x_j^2 y_j^2 + \sum_j \sum_{i>j} x_i^2 y_j^2 \\ & - \sum_l \sum_{k<l} x_k y_k x_l y_l - \sum_k x_k^2 y_k^2 - \sum_l \sum_{k>l} x_k y_k x_l y_l. \end{aligned} \tag{7.222}$$

Il y a deux termes qui se simplifient. Notez que si A_{kl} est symétrique en kl nous avons

$$\sum_l \sum_{k<l} A_{kl} = \sum_k \sum_{l<k} A_{lk} = \sum_k \sum_{l<k} A_{kl}. \tag{7.223}$$

La première égalité était seulement un renommage des indices. Le coup des indices symétriques est justement ce qu'il se passe dans les deux termes en$x_k y_k x_l y_l$, donc nous les regroupons :

$$\|x\|^2\|y\|^2 - (x \cdot y)^2 = \sum_j \big(\sum_{i<j} x_i^2 x_j^2 + \sum_{i>j} x_i^2 y_j^2 - 2 \sum_{i>j} x_i y_i x_j y_j\big) \tag{7.224a}$$

$$= \sum_j \sum_{i<j} (x_i^2 y_j^2 + x_j^2 y_i^2 - 2 x_i y_i x_j y_j) \tag{7.224b}$$

$$= \sum_j \sum_{i<j} (x_i y_j - x_j y_i)^2. \tag{7.224c}$$

Voila qui prouve la première formule. Pour la seconde, il faut seulement poser $n = 3$ et écrire les sommes explicitement.

— Pour $j = 1$, la somme sur i est $\sum_{i<1}$, c'est-à-dire aucun termes.

— Pour $j = 2$, il y a seulement $i = 1$, donc le terme $(x_1 y_2 - x_2 y_1)^2$.

— Pour $j = 3$, il y a les termes $i = 1$ et $i = 2$, donc les termes $(x_1 y_3 - x_3 y_1)^2 + (x_2 y_3 - x_3 y_2)^2$.

Ces trois termes collectés sont justement les composants (au carré) de $x \times y$ données dans la formule (7.210a). □

Les trois vecteurs de base e_x, e_y et e_y ont des produits vectoriels faciles à retenir :

$$\begin{aligned} e_x \times e_y &= e_z \\ e_y \times e_z &= e_x \\ e_z \times e_x &= e_y \end{aligned} \tag{7.225}$$

Les deux formules suivantes, qui mêlent le produit scalaire et le produit vectoriel, sont souvent utiles en analyse vectorielle :

$$\begin{aligned}(u \times v) \cdot w &= u \cdot (v \times w)\\ (u \times v) \times w &= -(v \cdot w)u + (u \cdot w)v\end{aligned} \tag{7.226}$$

pour tout vecteurs u, v et w dans $\mathbb{R}^3$. Nous les admettons sans démonstration. La seconde formule est parfois appelée **formule d'expulsion**.

Exemple 7.213.
Calculons le produit vectoriel $v \times w$ avec

$$v = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} \quad w = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}. \tag{7.227}$$

Les vecteurs s'écrivent sous la forme $v = 3e_x - e_y + e_z$ et $w = e_x + 2e_y - e_z$. Le produit vectoriel s'écrit

$$\begin{aligned}(3e_x - e_y + e_z) \times (e_x + 2e_y - e_z) &= 6e_x \times e_y - 3e_x \times e_z\\ &\quad - e_y \times e_x + e_y \times e_z\\ &\quad + e_z \times e_x + 2e_z \times e_y\\ &= 6e_z + 3e_y + e_z + e_x + e_y - 2e_x\\ &= -e_x + 4e_y + 7e_z.\end{aligned} \tag{7.228}$$

△

7.9.6 Produit mixte

Si a, b et c sont trois vecteurs, leur **produit mixte** est le nombre $a \cdot (b \times c)$. En écrivant le produit vectoriel sous forme de somme de trois déterminants 2×2, nous avons

$$\begin{aligned}&a \cdot (b \times c)\\ &\quad = (a_1 e_x + a_2 e_y + a_3 e_z) \cdot \left(\begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} e_x - \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} e_y + \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \right)\\ &\quad = a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & b_3 \\ c_1 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}\\ &\quad = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.\end{aligned} \tag{7.229}$$

Le produit mixte s'écrit donc sous forme d'un déterminant. Nous retenons cette formule :

$$a \cdot (b \times c) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}. \tag{7.230}$$

Un grand intérêt du produit vectoriel est qu'il fournit un vecteur qui est simultanément perpendiculaire aux deux vecteurs donnés.

Proposition 7.214.
Le produit vectoriel[68] *$a \times b$ est un vecteur orthogonal à a et b.*

68. Définition 7.204.

Démonstration. Vérifions que $a \perp (a \times b)$. Pour cela, nous calculons $a \cdot (a \times b)$, c'est-à-dire le produit mixte

$$a \cdot (a \times b) = \begin{vmatrix} a_1 & a_2 & a_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = 0. \tag{7.231}$$

L'annulation de ce déterminant est due au fait que deux de ses lignes sont égales. □

Ces résultats admettent une intéressante généralisation.

Lemme 7.215.
Soit $X \in \mathbb{R}^n$ ainsi que $v_1, \dots, v_{n-1} \in \mathbb{R}^n$. Alors

(1) Nous avons

$$\det(X, v_1, \dots, v_{n-1}) = X \cdot \det \begin{pmatrix} e_1 & \dots & e_n \\ & v_1 & \\ & \vdots & \\ & v_{n-1} & \end{pmatrix} \tag{7.232}$$

(2) Le vecteur

$$\det \begin{pmatrix} e_1 & \dots & e_n \\ & v_1 & \\ & \vdots & \\ & v_{n-1} & \end{pmatrix} \tag{7.233}$$

est orthogonal à tous les v_i.

Démonstration. Vu que les deux côtés de (7.232) vus comme fonctions de X, sont des applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}$, il suffit de vérifier l'égalité sur une base.

Nous posons $\tau_i \colon \mathbb{R}^n \to \mathbb{R}^{n-1}$,

$$\tau_i(v)_k = \begin{cases} v_k & \text{si } k < i \\ v_{k+1} & \text{si } k \geqslant i. \end{cases} \tag{7.234}$$

et nous avons d'une part

$$e_k \cdot \det \begin{pmatrix} e_1 & \dots & e_n \\ & v_1 & \\ & \vdots & \\ & v_{n-1} & \end{pmatrix} = \det \begin{pmatrix} \tau_k v_1 \\ \vdots \\ \tau_k v_{n-1} \end{pmatrix} \tag{7.235}$$

et d'autre part,

$$\det(e_k, v_1, \dots, v_{n-1}) = \det \begin{pmatrix} 0 & & & \\ \vdots & & & \\ 1 & v_1 & \cdots & v_{n-1} \\ \vdots & & & \\ 0 & & & \end{pmatrix} = \det(\tau_k v_1, \dots, \tau_k v_{n-1}). \tag{7.236}$$

La première assertion est démontrée.

En ce qui concerne la seconde, il suffit d'appliquer la première et se souvenir qu'un déterminant est nul lorsque deux lignes sont égales[69]. En effet :

$$v_k \cdot \det \begin{pmatrix} e_1 & \dots & e_n \\ & v_1 & \\ & \vdots & \\ & v_{n-1} & \end{pmatrix} = \det(v_k, v_1, \dots, v_n) = 0. \tag{7.237}$$

□

69. Corolaire 4.79.

7.9.7 Procédé de Gram-Schmidt

Proposition 7.216 (Procédé de Gram-Schmidt)**.**
Un espace euclidien possède une base orthonormée.

Démonstration. Soit E un espace euclidien et $\{v_1, \dots, v_n\}$, une base quelconque de E. Nous posons d'abord

$$f_1 = v_1, \quad e_1 = \frac{f_1}{\|f_1\|}. \tag{7.238}$$

Ensuite

$$f_2 = v_2 - \langle v_2, e_1 \rangle e_1, \quad e_2 = \frac{f_2}{\|f_2\|}. \tag{7.239}$$

Notons que $\{e_1, e_2\}$ est une base de $\mathrm{Span}\{v_1, v_2\}$. De plus elle est orthogonale :

$$\langle e_1, f_2 \rangle = \langle e_1, v_2 \rangle - \langle v_2, e_1 \rangle \underbrace{\langle e_1, e_1 \rangle}_{=1} = 0. \tag{7.240}$$

Le fait que $\|e_1\| = \|e_2\| = 1$ est par construction. Nous avons donc donné une base orthonormée de $\mathrm{Span}\{v_1, v_2\}$.

Nous continuons par récurrence en posant

$$f_k = v_k - \sum_{i=1}^{k-1} \langle v_k, e_i \rangle e_i, \quad e_k = \frac{f_k}{\|f_k\|}. \tag{7.241}$$

Pour tout $j < k$ nous avons

$$\langle e_j, f_k \rangle = \langle e_j, v_k \rangle - \sum_{i=1}^{k-1} \langle v_k, e_i \rangle \underbrace{\langle e_i, e_j \rangle}_{=\delta_{ij}} = 0 \tag{7.242}$$

□

Cet algorithme de Gram-Schmidt nous donne non seulement l'existence de bases orthonormée pour tout espace euclidien, mais aussi le moyen d'en construire à partir de n'importe quelle base.

7.9.8 Approximation

Le lemme suivant est surtout intéressant en dimension infinie.

Lemme 7.217.
Soit un espace vectoriel normé V et un sous-espace vectoriel dense A. Soit $v \in V$; il existe une suite (v_n) dans A telle que $v_n \xrightarrow{V} v$ et $\|v_n\| \leqslant \|v\|$ pour tout n.

Démonstration. Vu que A est dense, il existe une suite a_n dans A telle que $a_n \to v$. Ensuite il suffit de poser

$$v_n = \frac{n}{n+1} \frac{\|v\|}{\|a_n\|} a_n. \tag{7.243}$$

Par construction nous avons toujours

$$\|v_n\| = \frac{n}{n+1} \|v\| \leqslant \|v\|. \tag{7.244}$$

Et de plus, la norme étant continue[70],

$$\lim_{n\to\infty} v_n = \lim_{n\to\infty} \frac{n}{n+1} \lim_{n\to\infty} \frac{\|v\|}{\|v_n\|} \lim_{n\to\infty} v_n = v. \tag{7.245}$$

Le fait que v_n soit dans A est dû au fait que A soit vectoriel. □

70. Où dans le calcul suivant nous utilisons la continuité de la norme ? Posez-vous la question.

Proposition 7.218.
Soit un espace vectoriel normé V et un sous-espace vectoriel dense A. Soit $v \in V$; pour tout $a \in \mathbb{R}$ nous avons

$$\sup\{|v \cdot a| \text{ tel que } a \in A \text{ et } \|a\| \leqslant \lambda\} = \lambda\|v\|. \tag{7.246}$$

Démonstration. D'abord pour tout $a \in A$ vérifiant $\|a\| \leqslant \lambda$ l'inégalité de Cauchy-Schwarz 7.191 donne

$$|v \cdot a| \leqslant \|v\|\|a\| \leqslant \lambda\|v\|. \tag{7.247}$$

Donc le supremum dont on parle est majoré par $\lambda\|v\|$.

Il nous faut l'inégalité dans l'autre sens. Par densité nous pouvons choisir une suite $v_n \in A$ tel que $v_n \to v$. Ensuite nous posons

$$a_n = \frac{\lambda}{\|v_n\|} v_n. \tag{7.248}$$

Nous avons $\|a_n\| = \lambda$ pour tout n et

$$|v \cdot a_n| = \frac{\lambda}{\|v_n\|} |v \cdot v_n|, \tag{7.249}$$

et en passant à la limite,

$$\lim_{n\to\infty} |v \cdot a_n| = \frac{\lambda}{\|v\|} \|v \cdot v\| = \lambda\|v\|. \tag{7.250}$$

Donc l'ensemble sur lequel nous prenons le supremum contient une suite convergente vers $\lambda\|v\|$. Le supremum est donc au moins aussi grand que cela. □

7.9.9 Quelques exemples de normes sur $\mathbb{R}^n$

Il est possible de définir de nombreuses normes sur $\mathbb{R}^n$. Citons-en quelques-unes.

Proposition-définition 7.219.
Les formules suivantes définissent des normes sur $\mathbb{R}^n$.

(1) Les normes $\|.\|_{L^p}$ ($p \in \mathbb{N}$) sont définies de la façon suivante :

$$\|x\|_{L^p} = \Big(\sum_{i=1}^{n} |x_i|^p\Big)^{1/p}, \tag{7.251}$$

pour tout $x = (x_1, \dots, x_n) \in \mathbb{R}^n$.

*(2) La norme L^2 est la **norme euclidienne**.*

*(3) Nous définissons également la **norme supremum** par*

$$\|x\|_\infty = \max_i |x_i|. \tag{7.252}$$

Démonstration. Point par point [71].

(1) Le cas $p = 1$ est déjà fait dans le lemme 7.194.

(2) Le fait que $x \mapsto \|x\|_{L^2}$ soit une norme provient de la propriété suivante :

$$\sqrt{(a+b)^2} \leqslant \sqrt{a^2} + \sqrt{b^2}, \tag{7.253}$$

laquelle se démontre en passant au carré :

$$(a+b)^2 = a^2 + b^2 + 2ab \leqslant a^2 + b^2 + 2|ab| = \big(\sqrt{a^2} + \sqrt{b^2}\big)^2. \tag{7.254}$$

(3)

□

71. Preuve non terminée

Parmi ces normes, celles qui seront le plus souvent utilisées dans ces notes sont

$$\begin{aligned} \|x\|_{L^1} &= \sum_{i=1}^{n} |x_i|, \\ \|x\|_{L^2} &= \Big(\sum_{i=1}^{n} |x_i|^2\Big)^{1/2}. \end{aligned} \tag{7.255}$$

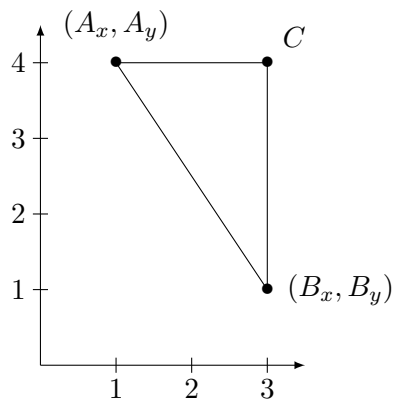


FIGURE 7.4 – La *norme* euclidienne induit la *distance* euclidienne. D'où son nom. Le point C est construit aux coordonnées (A_x, B_y).

Soient $A = (A_x, A_y)$ et $B = (B_x, B_y)$ deux éléments de $\mathbb{R}^2$. La distance [72] euclidienne entre A et B est donnée par $\|A - B\|_2$. En effet, sur la figure 7.4, la distance entre les points A et B est donnée par

$$|AB|^2 = |AC|^2 + |CB|^2 = |A_x - B_x|^2 + |A_y - B_y|^2, \tag{7.256}$$

par conséquent,

$$|AB| = \sqrt{|A_x - B_x|^2 + |A_y - B_y|^2} = \|A - B\|_2. \tag{7.257}$$

Remarque 7.220.
Si A, B et C sont trois points dans le plan $\mathbb{R}^2$, alors l'inégalité triangulaire $|AB| \leqslant |AC| + |CB|$ est précisément la propriété (3) de la norme (définition 7.113). En effet l'inégalité triangulaire s'exprime de la façon suivante en termes de la norme $\|.\|_2$:

$$\|A - B\|_2 \leqslant \|A - C\|_2 + \|C - B\|_2. \tag{7.258}$$

En notant $u = A - C$ et $v = C - B$, l'équation (7.258) devient exactement la propriété de définition de la norme :

$$\|u + v\|_2 \leqslant \|u\|_2 + \|v\|_2. \tag{7.259}$$

Ceci explique pourquoi cette propriété des normes est appelée « inégalité triangulaire ».

7.10 Équivalence des normes

Au premier coup d'œil, les notions dont nous parlons dans ce chapitre ont l'air très générales. Nous prenons en effet n'importe quel espace vectoriel V de dimension finie, et nous le munissons de n'importe quelle norme (rien que dans $\mathbb{R}^m$ nous en avons défini une infinité par l'équation (7.251)). À partir de ces données, nous définissons les boules, la topologie, l'adhérence, etc.

72. Ne pas confondre « distance » et « norme ».

7.10.1 En dimension finie

Dans $\mathbb{R}^n$, les normes $\|.\|_{L^1}$, $\|.\|_{L^2}$ et $\|.\|_\infty$ ne sont pas égales. Cependant elles ne sont pas complètement indépendantes au sens où l'on sent bien que si un vecteur sera grand pour une norme, il sera également grand pour les autres normes ; les normes « vont dans le même sens ». Cette notion est précisée par le concept de norme équivalente.

Définition 7.221.
Deux normes N_1 et N_2 sur $\mathbb{R}^m$ sont ***équivalentes*** *s'il existe deux nombres réels strictement positifs k_1 et k_2 tels que*

$$k_1 N_1(x) \leqslant N_2(x) \leqslant k_2 N_1(x), \tag{7.260}$$

pour tout x dans $\mathbb{R}^m$. Dans ce cas nous écrivons que $N_1 \sim N_2$.

Lemme 7.222.
La définition de norme équivalentes donne une relation d'équivalence (définition 1.20) sur l'ensemble des normes existantes sur $\mathbb{R}^m$.

Proposition 7.223.
Pour $\mathbb{R}^N$, nous avons les équivalences de normes $\|.\|_{L^1} \sim \|.\|_{L^2}$, $\|.\|_{L^1} \sim \|.\|_\infty$ et $\|.\|_{L^2} \sim \|.\|_\infty$. Plus précisément nous avons les inégalités

(1) $\|x\|_2 \leqslant \|x\|_1 \leqslant \sqrt{n}\|x\|_2$
(2) $\|x\|_\infty \leqslant \|x\|_1 \leqslant n\|x\|_\infty$
(3) $\|x\|_\infty \leqslant \|x\|_2 \leqslant \sqrt{n}\|x\|_\infty$

Démonstration. En mettant au carré la première inégalité nous voyons que nous devons vérifier l'inégalité

$$|x_1|^2 + \cdots + |x_n|^2 \leqslant \big(|x_1| + \cdots + |x_n|\big)^2 \tag{7.261}$$

qui est vraie parce que le membre de droite est égal au carré de chaque terme plus les double produits. La seconde inégalité provient de l'inégalité de Cauchy-Schwarz (théorème 7.191) sur les vecteurs

$$v = \begin{pmatrix} 1/n \\ \vdots \\ 1/n \end{pmatrix}, \quad w = \begin{pmatrix} |x_1| \\ \vdots \\ |x_n| \end{pmatrix}. \tag{7.262}$$

Nous trouvons

$$\frac{1}{n}\sum_i |x_i| \leqslant \sqrt{n \cdot \frac{1}{n^2}}\sqrt{\sum_i |x_i|^2}, \tag{7.263}$$

et par conséquent

$$\sum_i |x_i| \leqslant \sqrt{n}\|x\|_2. \tag{7.264}$$

La première inégalité de (3) se démontre en remarquant que si a et b sont positifs, $a \leqslant \sqrt{a^2+b}$. En appliquant cela à $a = \max_i |x_i|$, nous avons

$$\max_i |x_i| \leqslant \sqrt{|x_1|^2 + \cdots + |x_n|^2} \tag{7.265}$$

parce que $\max_i |x_i|$ est évidemment un des termes de la somme. Pour la seconde inégalité de (3), nous avons

$$\sqrt{\sum_k |x_k|^2} \leqslant \left(\sum_k \max_i |x_i|^2\right)^{1/2} = \sqrt{n}\|x\|_\infty. \tag{7.266}$$

Pour obtenir cette inégalité, nous avons remplacé tous les termes $|x_k|$ par le maximum. □

Pour les autres normes $\|.\|_p$, il y a des inégalités dans 11.364 et 16.96 ; voir aussi le thème 7.

En réalité, toutes les normes $\|.\|_{L^p}$ et $\|.\|_\infty$ sont équivalentes et, plus généralement, nous avons le résultat suivant, très étonnant à première vue, et en réalité assez difficile à prouver :

Théorème 7.224 ([147]).
Sur un espace vectoriel de dimension finie, toutes les normes sont équivalentes.

Corolaire 7.225.
Soit V un espace vectoriel de dimension finie et $\|.\|_1$, $\|.\|_2$ deux normes sur V. Alors l'identité $\mathrm{Id}\colon V \to V$ *est un isomorphisme d'espace topologique* $(V, \|.\|_1) \to (V, \|.\|_2)$.

De plus les ouverts sont les mêmes : une partie de V est ouverte dans $(V, \|.\|_1)$ si et seulement si elle est ouverte dans $(V, \|.\|_2)$.

7.226.
Le lemme 7.194 donnera une norme sur $\mathbb{R}^2$ qui ne dérive pas d'un produit scalaire. Vu que toutes les normes sur $\mathbb{R}^2$ produisent la même topologie (c'est le corolaire 7.225), il y a parfaitement moyen pour deux espaces vectoriels topologiques d'être isomorphes alors que l'un a une norme dérivant d'un produit scalaire et l'autre non.

7.227.
Le théorème d'équivalence de norme sera utilisé pour montrer que l'ensemble des formes quadratiques non dégénérées de signature (p, q) est ouvert dans l'ensemble des formes quadratiques, proposition 16.112. Plus généralement il est utilisé à chaque fois que l'on fait de la topologie sur les espaces de matrices en identifiant $\mathbb{M}(n, \mathbb{R})$ à $\mathbb{R}^{n^2}$, pour se rassurer en se disant que ce qu'on fait ne dépend pas de la norme choisie.

Voir aussi ce qu'on en fait en 11.240 pour démontrer la différentiabilité à partir des dérivées partielles.

Proposition 7.228 ([1]).
Let V be a finite dimensional complex vector space. For a basis $B = \{e_1, \dots, e_n\}$ of V we define
Soit un espace vectoriel V de dimension finie sur $\mathbb{C}$. Pour une base $B = \{e_i\}$ de V nous définissons

$$\|\sum_k v_k e_k\|_B = \sqrt{\sum_k |v_k|^2}. \tag{7.267}$$

(1) La formule (7.267) définit une norme sur V.

(2) Si B et B' sont des bases de V, alors les topologies induites par le norme $\|.\|_B$ et $\|.\|_{B'}$ sont égales.

Démonstration. Nous commençons par fixer une base $B = \{e_i\}_{i=1,\dots,n}$ de V. Cette base nous permet de définir

$$\begin{aligned} \varphi\colon V &\to \mathbb{C}^n \\ \sum_k v_k e_k &\mapsto (v_1, \dots, v_n). \end{aligned} \tag{7.268}$$

Cette application linéaire permet d'écrire

$$\|v\|_V = \|\varphi(v)\|_{\mathbb{C}^n}. \tag{7.269}$$

À partir de là, la vérification des propriétés de la définition 7.113 est immédiate. Par exemple :

$$\|v + w\| = \|\varphi(v + w)\| = \|\varphi(v) + \varphi(w)\| \leqslant \|\varphi(v)\| + \|\varphi(w)\| = \|v\| + \|w\|. \tag{7.270}$$

En ce qui concerne la seconde assertion, c'est le théorème 7.224. □

7.10.2 Contre-exemple en dimension infinie

Lorsque nous considérons des espaces vectoriels de dimension infinie, les choses ne sons plus aussi simples. Nous voyons ici sur l'exemple de l'espace des polynômes que le théorème 7.224 n'est plus valable si on enlève l'hypothèse de dimension finie.

On considère l'ensemble des fonctions polynomiales à coefficients réels sur l'intervalle $[0,1]$.

$$\mathcal{P}_{\mathbb{R}}([0,1]) = \{p : [0,1] \to \mathbb{R} \,|\, p : x \mapsto a_0 + a_1 x + a_2 x^2 + \ldots,\ a_i \in \mathbb{R},\ \forall i \in \mathbb{N}\}. \tag{7.271}$$

Cet ensemble, muni des opérations usuelles de somme entre polynômes et multiplications par les scalaires, est un espace vectoriel.

Sur $\mathcal{P}(\mathbb{R})$ on définit les normes suivantes

$$\begin{aligned} \|p\|_\infty &= \sup_{x\in[0,1]} \{p(x)\}, \\ \|p\|_1 &= \int_0^1 |p(x)|\, dx, \\ \|p\|_2 &= \left(\int_0^1 |p(x)|^2\, dx\right)^{1/2}. \end{aligned} \tag{7.272}$$

Les inégalités suivantes sont immédiates

$$\begin{aligned} \|p\|_1 &= \int_0^1 |p(x)|\, dx \leqslant \|p\|_\infty, \\ \|p\|_2 &= \left(\int_0^1 |p(x)|^2\, dx\right)^{1/2} \leqslant \|p\|_\infty, \end{aligned} \tag{7.273}$$

mais la norme $\|\cdot\|_\infty$ n'est équivalente ni à $\|\cdot\|_1$, ni à $\|\cdot\|_2$. Soit $p_k(x) = x^k$. Alors

$$\begin{aligned} \|p_k\|_\infty &= 1, \\ \|p_k\|_1 &= \int_0^1 x^k\, dx = \frac{1}{k+1}, \\ \|p_k\|_2 &= \left(\int_0^1 x^{2k}\, dx\right)^{1/2} = \sqrt{\frac{1}{2k+1}}. \end{aligned} \tag{7.274}$$

Pour $k \to \infty$ les normes $\|p_k\|_1$, $\|p_k\|_2$ tendent vers zéro, alors que la norme $\|p_k\|_\infty$ est constante, donc les normes ne sont pas équivalentes parce que il n'existe pas un nombre positif m tel que

$$\begin{aligned} m\|p_k\|_\infty &\leqslant \|p_k\|_1, \\ m\|p_k\|_\infty &\leqslant \|p_k\|_2, \end{aligned} \tag{7.275}$$

uniformément pour tout k dans $\mathbb{N}$.

7.11 Produit fini d'espaces vectoriels normés

Dans cette sections nous parlons de produits finis d'espaces. Cela ne signifie pas que chacun des espaces soient séparément de dimension finie.

7.11.1 Distance et norme produit

Proposition-définition 7.229 (Distance produit)**.**
Si (E_1, d_1),..., (E_n, d_n) sont des espaces métriques alors la formule

$$d(x,y) = \max_{i=1,\ldots,n} d_i(x_i, y_i) \tag{7.276}$$

*définit une distance sur le produit cartésien $E = E_1 \times \ldots \times E_n$. Elle est la **distance produit**.*

La définition de la norme sur un produit d'espaces vectoriels normés découle immédiatement de la définition de la distance 7.229 :

Lemme-définition 7.230 ([148]).
Soient V et W deux espaces vectoriels normés.

(1) L'ensemble

$$V \times W = \{(v, w) \,|\, v \in V,\, w \in W\} \tag{7.277}$$

est un espace vectoriel.

(2) L'opération

$$\|(v, w)\|_{V\times W} = \max\{\|v\|_V, \|w\|_W\}. \tag{7.278}$$

est une norme[73] *sur $V \times V$.*

(3) La topologie de cette norme est la même que la topologie produit sur $V \times W$.

*L'espace vectoriel $V \times W$ muni de cette norme est l'**espace produit** de V et W. La topologie ainsi définie de deux manières est la topologie qui sera toujours considérée dans le cas de produit d'espaces vectoriels normés.*

Démonstration. En plusieurs parties.

Espace vectoriel Il est presque immédiat de vérifier que le produit cartésien $V \times W$ est un espace vectoriel pour les opération de somme et multiplication par les scalaires définies composante par composante. C'est-à-dire, si (v_1, w_1), (v_2, w_2) sont dans $V \times W$ et a, b sont des scalaires, alors

$$a(v_1, w_1) + b(v_2, w_2) = (av_1, aw_1) + (bv_2, bw_2) = (av_1 + bv_2, aw_1 + bw_2). \tag{7.279}$$

Norme On doit vérifier les trois conditions de la définition 7.113.

— Soit (v, w) dans $V \times W$ tel que $\|(v, w)\|_{V\times W} = \max\{\|v\|_V, \|w\|_W\} = 0$. Alors $\|v\|_V = 0$ et $\|w\|_W = 0$, donc $v = 0_V$ et $w = 0_W$. Cela implique $(v, w) = (0_v, 0_w) = 0_{V\times W}$.

— Pour tout a dans $\mathbb{R}$ et (v, w) dans $V \times W$, la norme $\|a(v, w)\|_{V\times W}$ se calcule de la façon suivante :

$$\|a(v, w)\|_{V\times W} = \max\{\|av\|_V, \|aw\|_W\} = |a| \max\{\|v\|_V, \|w\|_W\} = |a|\|(v, w)\|_{V\times W}. \tag{7.280}$$

— Soient (v_1, w_1) et (v_2, w_2) dans $V \times W$.

$$\begin{aligned}\|(v_1, w_1) + (v_2, w_2)\|_{V\times W} &= \max\{\|v_1 + v_2\|_V, \|w_1 + w_2\|_W\}\\ &\leqslant \max\{\|v_1\|_V + \|v_2\|_V, \|w_1\|_W + \|w_2\|_W\}\\ &\leqslant \max\{\|v_1\|_V, \|w_1\|_W\} + \max\{\|v_2\|_V, \|w_2\|_W\}\\ &= \|(v_1, w_1)\|_{V\times W} + \|(v_2, w_2)\|_{V\times W}.\end{aligned} \tag{7.281}$$

Dans cette preuve, nous considérons la « topologie de $V \times W$ » comme étant la topologie produit et « la topologie métrique de $V \times W$ » la topologie de la norme produit.

Équivalence **Dans un sens** La proposition 7.13 dit qu'une prébase de $V \times W$ est donnée par

$$\big\{B(v, r) \times B(w, s) \text{ tel que } v \in V; w \in W; r, s > 0\big\}. \tag{7.282}$$

Nous prouvons maintenant que la partie $S = B(v_0, r) \times B(w_0, s)$ est un ouvert de l'espace $\big(V \times W, \|.\|_{V\times W}\big)$. Pour cela nous prouvons que tout élément de S contient un voisinage métrique contenu dans S.

Soit $(v_1, w_1) \in S$. Nous posons

$$d\big((v_1, w_1), (v_0, w_0)\big) = \delta < \max\{r, s\}. \tag{7.283}$$

73. Définition 7.113.

Nous considérons $\epsilon > 0$ et nous montrons que si ϵ est assez petit, $B\big((v_1, w_1), \epsilon\big) \subset S$. Pour cela nous considérons $(v, w) \in B\big((v_1, w_1), \epsilon\big)$ et nous calculons un tout petit peu :

$$d\big((v, w), (v_0, w_0)\big) \leqslant d\big((v, w), (v_1, w_1)\big) + d\big((v_1, w_1), (v_0, w_0)\big) \tag{7.284a}$$
$$< \epsilon + \delta. \tag{7.284b}$$

Si ϵ est assez petit, le tout reste plus petit que $\max\{r, s\}$.

Donc S est bien un ouvert métrique par le théorème 7.4. Vu que la topologie métrique contient une prébase de la topologie produit, tout ouvert de la topologie produit est un ouvert de la topologie métrique.

Dans l'autre sens Soient un ouvert métrique $\mathcal{O}$ ainsi que $(v_0, w_0) \in \mathcal{O}$; il existe $r > 0$ tel que

$$B\big((v_0, w_0), r\big) \subset \mathcal{O}. \tag{7.285}$$

Nous affirmons que $B(v_0, r) \times B(w_0, r)$ est contenu dans $\mathcal{O}$, de telle sorte que $\mathcal{O}$ soit un ouvert de la topologie produit. Pour $(v_1, w_1) \in B(v_0, r) \times B(w_0, r)$ nous avons

$$d\big((v_1, w_1), (v_0, w_0)\big) = \max\{\|v_1 - v_0\|, \|w_1 - w_0\|\} < r \tag{7.286}$$

parce que $v_1 \in B(v_0, r)$ et $w_1 \in B(w_0, r)$.

Donc tout élément de $\mathcal{O}$ admet un voisinage « produit » contenu dans $\mathcal{O}$; donc $\mathcal{O}$ est ouvert pour le topologie produit.

□

7.231.
En particulier, pour la topologie de la norme maximum, la convergence d'une suite implique la convergence « composante par composante » par la proposition 7.38.

7.12 Topologie réelle en dimension n

7.12.1 Ouverts et fermés

La définition suivante est là juste pour la facilité des notations ; nous ne disons pas qu'elle est liée à la topologie sur $\mathbb{R}^n$.

Définition 7.232.
*La **boule ouverte** de centre $x_0 \in \mathbb{R}^n$ et de rayon $r \in \mathbb{R}^+$ est définie par*

$$B(x_0, r) = \{x \in \mathbb{R}^n \text{ tel que } \|x - x_0\| < r\}, \tag{7.287}$$

*tandis que la **boule fermée** de centre x_0 et de rayon r est*

$$\overline{B(x_0, r)} = \{x \in \mathbb{R}^n \text{ tel que } \|x - x_0\| \leqslant r\}; \tag{7.288}$$

la différence est que l'inégalité dans la première est stricte.

Proposition-définition 7.233.
Sur $\mathbb{R}^n$, nous considérons les deux topologies suivantes :

(1) la topologie produit $\mathbb{R} \times \ldots \times \mathbb{R}$ des espaces topologiques $(\mathbb{R}, |.|)$,

(2) la topologie de la norme

$$\|(x_1, \ldots, x_n)\|_\infty = \max_i\{|x_i|\}, \tag{7.289}$$

(3) la topologie de la norme

$$\|(x_1, \ldots, x_n)\|_2 = \sqrt{\sum_{i=1}^n x_i^2}. \tag{7.290}$$

Ces deux topologies sont égales et sont la topologie que nous allons toujours considérer sur $\mathbb{R}^n$ (saut mention très explicite du contraire).

Démonstration. Les topologies (1) et (2) sont identiques par le lemme 7.230. Les topologies (2) et (3) sont identiques par la proposition 7.223 (ou plus généralement par le théorème 7.224 qui donne l'équivalence de toutes les normes). □

Proposition 7.234.
Une partie A de $\mathbb{R}^n$ est ouverte si et seulement si pour tout $a \in A$ il existe $r > 0$ tel que $B(a, r) \subset A$.

Cette proposition est évidemment à mettre en rapport avec le théorème 7.4.

Le lemme suivant justifie le vocabulaire des définitions 7.232.

Lemme 7.235.
Pour tout $x \in \mathbb{R}^n$ et tout $r > 0$ la boule $B(x, r)$ est ouverte.

Démonstration. Afin de prouver que la boule est ouverte, nous prenons un point $p \in B(x, r)$, et nous allons montrer qu'il existe une boule autour de p qui est contenue dans $B(x, r)$.

Étant donné que $p \in B(x, r)$, nous avons $d(p, x) < r$. Prouvons que la boule $B\big(p, r - d(p, x)\big)$ est contenue dans $B(x, r)$. Pour cela, nous prenons $p' \in B\big(p, r - d(p, x)\big)$, et nous essayons de prouver que $p' \in B(x, r)$. En effet, en utilisant l'inégalité triangulaire,

$$d(x, p') \leqslant d(x, p) + d(p, p') \leqslant d(x, p) + r - d(p, x) = r. \tag{7.291}$$

□

7.12.2 Intérieur, adhérence et frontière

Définition 7.236.
*Soient $A \subset \mathbb{R}^n$ et $x \in \mathbb{R}^n$. Le point x est **intérieur** à A s'il existe une boule autour de x complètement contenue dans A. L'ensemble des points intérieurs à A est noté $\operatorname{Int} A$ ou $\mathring{A}$, de sorte qu'on a précisément*

$$x \in \operatorname{Int} A \overset{def}{\iff} \exists \epsilon > 0 \text{ tel que } B(x, \epsilon) \subset A.$$

7.237.
La notion d'adhérence a déjà été définie en 7.22, et précisé par le lemme 7.23. Dans le cas de $\mathbb{R}^n$ dans lequel les boules forment une base de la topologie nous pouvons encore préciser de la façon suivante :

$$x \in \operatorname{Adh} A \overset{\text{def}}{\iff} \forall \epsilon > 0, B(x, \epsilon) \cap A \neq \varnothing \tag{7.292}$$

Proposition 7.238.
Pour $A \subset \mathbb{R}^n$, nous avons

$$\operatorname{Int} A \subseteq A \subseteq \operatorname{Adh} A$$

Définition 7.239.
*La **frontière** ou le **bord** de A est défini par $\partial A = \operatorname{Adh} A \backslash \operatorname{Int} A$. L'ensemble A est un **ouvert** si $A = \operatorname{Int} A$, et c'est un **fermé** si $A = \operatorname{Adh} A$.*

Lemme 7.240 (Caractérisation équivalente de la frontière)**.**
Soient X un espace topologique et $S \subset X$. Un point $x \in X$ est dans ∂S si et seulement si tout voisinage de x contient un point de S et un point de S^c.

Démonstration. Supposons que tout voisinage de x contienne un point de S et un point de S^c. Alors $x \in Adh(S)$ (définition 7.22), mais pas dans l'intérieur de S parce que x ne possède pas de voisinage contenu dans S. Donc $x \in \partial S$.

À l'inverse, si $x \in \partial S$ alors x est dans l'adhérence de S et tout voisinage de x contient un point de S. Mais x n'est pas dans l'intérieur de S et tout voisinage de x contient un point qui n'est pas dans S, aka un point de S^c. □

Corolaire 7.241.
Un ensemble et son complémentaire ont même frontière.

Démonstration. Conséquence du lemme 7.240. Les points de $\partial(S^c)$ sont caractérisés par le fait que tout voisinage contient un point de S^c et un point de $(S^c)^c = S$. □

Exemple 7.242.
Soit $X = [0, 1]$ muni de la topologie de la distance $|x - y|$ (définition 7.94). Les points 0 et 1 *ne sont pas* dans la frontière de X. En effet une boule ouverte autour de 1 est un ensemble de la forme

$$B(1, r) = \{x \in X \text{ tel que } |x - 1| < r\} =]1 - r, 1] \tag{7.293}$$

où nous avons supposé $r < 1$.

Les points 0 et 1 sont par contre sur la frontière de $[0, 1]$ lorsque cet ensemble est vu comme partie de l'espace métrique $\mathbb{R}$. △

Lemme 7.243 (Passage de douane[149, 150])**.**
Dans un espace topologique, toute partie connexe qui rencontre à la fois une partie A et son complémentaire rencontre nécessairement la frontière de A.

Démonstration. Nommons γ la partie connexe qui intersecte A et A^c. Les ouverts $\mathrm{Int}(A)$ et $X\backslash\bar{A}$ ne peuvent pas recouvrir γ parce que ce sont deux ouverts disjoints alors que γ est connexe (voir la définition 7.42 de la connexité). Donc γ doit contenir des points qui sont dans $\bar{A}$ mais pas dans $\mathrm{Int}(A)$. C'est-à-dire des points de ∂A. □

On vérifiera que les notations et les dénominations sont cohérentes en prouvant la proposition suivante.

Proposition 7.244.
Pour $\epsilon > 0$,

(1) l'adhérence de $B(x, \epsilon)$ est $\bar{B}(x, \epsilon)$,

(2) l'intérieur de $\bar{B}(x, \epsilon)$ est $B(x, \epsilon)$,

(3) la boule ouverte $B(x, \epsilon)$ est un ouvert,

(4) la boule fermée $\bar{B}(x, \epsilon)$ est un fermé.

Nous avons également les liens suivants entre intérieur, adhérence, ouvert, fermé et passage au complémentaire (noté ^{c}) :

Proposition 7.245.
Si $A \subset \mathbb{R}^n$ et $A^c = \mathbb{R}^n\backslash A$, nous avons

(1) $(\mathrm{Int}\, A)^c = \mathrm{Adh}(A^c)$ et $(\mathrm{Adh}\, A)^c = \mathrm{Int}(A^c)$,

(2) A est ouvert si et seulement si A^c est fermé,

(3) $\mathrm{Int}\, A$ est le plus grand ouvert contenu dans A,

(4) $\mathrm{Adh}\, A$ est le plus petit fermé contenant A,

Exemple 7.246.
Il n'est en général pas vrai que $\overline{A \cap B} = \bar{A} \cap \bar{B}$. Par exemple si $A = [0, 1[$ et $B =]1, 2]$. Dans ce cas, $A \cap B = \emptyset$ alors que $\bar{A} \cap \bar{B} = \{1\}$. △

7.12.3 Point d'accumulation, point isolé

Les définitions de point d'accumulation et de point isolé sont 7.27 et 7.28. Nous voyons maintenant ce que ces définitions donnent dans le cas de l'espace topologique $\mathbb{R}$.

Lemme 7.247.
Soit $D \subset \mathbb{R}$. Un point $a \in D$ est isolé dans D si et seulement si il existe $\varepsilon > 0$ tel que

$$[a - \varepsilon, a + \varepsilon] \cap D = \{a\}. \tag{7.294}$$

Autrement dit, il existe un intervalle autour de a dans lequel a est le seul élément de D.

Lemme 7.248.
Un point $a \in \mathbb{R}$ est un point d'accumulation de D si pour tout $\varepsilon > 0$,

$$\Big([a - \varepsilon, a + \varepsilon] \backslash \{a\}\Big) \cap D \neq \varnothing. \tag{7.295}$$

Autrement dit, quel que soit l'intervalle autour de a que l'on considère, le point a n'est pas tout seul dans D.

Exemple 7.249.
Prenons $D = [0, 1[\cup]2, 3]$. Cet ensemble n'a pas de point isolé, et l'ensemble de ses points d'accumulation est $[0, 1] \cup [2, 3]$.

Notez que les points 1 et 2 sont des points d'accumulation de D qui ne font pas partie de D. Il est possible d'être un point d'accumulation de D sans être dans D, mais pour être un point isolé dans D, il faut être dans D. △

Exemple 7.250.
Soit $D = \{\frac{1}{n}\}_{n \in \mathbb{N}}$. Tous les points de cet ensemble sont des points isolés (vérifier !). Aucun point de D n'est point d'accumulation. Cependant 0 est un point d'accumulation. △

Exemple 7.251.
Soit $D =]1, 2[\cup \{12\}$. Le point 12 est adhérence, mais pas d'accumulation parce que le voisinage $]9, 14[$ n'intersectionne pas $D \backslash \{12\}$. △

7.12.4 Limite de suite

Définition 7.252 (Limite d'une suite dans $\mathbb{R}^m$)**.**
Une suite de points (x_n) dans $\mathbb{R}^m$ est dite ***convergente*** *s'il existe un élément $\ell \in \mathbb{R}^m$ tel que*

$$\forall \varepsilon > 0, \exists N \in \mathbb{N} \text{ tel que } \forall n \geqslant N, \|x_n - \ell\| < \varepsilon. \tag{7.296}$$

Dans ce cas, nous disons que ℓ est la ***limite*** *de la suite (x_n) et nous écrivons $\lim x_n = \ell$ ou plus simplement $x_n \to \ell$.*

Notez aussi la similarité avec la définition 7.129.

Remarque 7.253.
Nous n'écrivons pas « $\lim_{n\to\infty} x_n$ » parce que, lorsqu'on parle de suites, la limite est *toujours* lorsque n tend vers l'infini. Il n'y a aucun intérêt à chercher par exemple $\lim_{n\to 4} x_n$ parce que cela vaudrait x_4 et rien d'autre.

Ceci est une différence importante avec les limites de fonctions.

Lemme 7.254 (Unicité de la limite)**.**
Il ne peut pas y avoir deux nombres différents qui satisfont à la condition (7.296). En d'autres termes, si ℓ et ℓ' sont deux limites de la suite (x_n), alors $\ell = \ell'$.

Démonstration. Soit $\varepsilon > 0$. Nous considérons N tel que

$$\|x_n - \ell\| < \varepsilon \tag{7.297}$$

pour tout $n \geqslant N$, et $N' > 0$ tel que

$$\|x_n - \ell'\| < \epsilon \tag{7.298}$$

pour tout $n > N'$. Maintenant, nous prenons n plus grand que N et N' de telle façon que les deux équations pour x_n soient vérifiées en même temps. Alors

$$\|\ell - \ell'\| = \|\ell - x_n + x_n - \ell'\| \leqslant \|\ell - x_n\| + \|x_n - \ell'\| < 2\varepsilon. \tag{7.299}$$

Cela prouve que $\|\ell - \ell'\| = 0$. □

Le théorème de Bolzano-Weierstrass 7.103 dit que dans le cas métrique, la compacité séquentielle est équivalente à la compacité.

7.13 Topologie et distance

Lemme 7.255.
Soient (X_1, d_1) et (X_2, d_2) des espaces métriques séparables. Alors $X_1 \times X_2$ admet une base dénombrable de topologie constituée de produits de boules de X_1 par des boules de X_2. Plus précisément si A_i est dénombrable et dense dans X_i alors l'ensemble des produits

$$\big\{B(y_1, r_1) \times B(y_2, r_2)\big\}_{\substack{y_i \in A_i \\ r_i \in \mathbb{Q}^+}} \tag{7.300}$$

est une base de topologie pour $X_1 \times X_2$.

Démonstration. Soit $\mathcal{O}$ un ouvert de $X_1 \times X_2$ et $(x_1, x_2) \in \mathcal{O}$. Par définition de la topologie produit[74], il existe $r_1, r_2 \in \mathbb{Q}^+$ tels que $B(x_1, r_1) \times B(x_2, r_2) \subset \mathcal{O}$. Les parties A_i étant denses, il existe $y_i \in B(x_i, r_i/2) \cap A_i$. Avec ces choix nous avons $x_i \in B(y_i, \frac{r_i}{2})$. Nous avons donc

$$(x_1, x_2) \in B(y_1, \frac{r_1}{2}) \times B(y_2, \frac{r_2}{2}). \tag{7.301}$$

Il est facile de voir que $B(y_i, r_i/2) \subset B(x_i, r_i)$. En effet si $z_i \in B(y_i, r_i/2)$ alors

$$d_i(z_i, x_i) \leqslant d(z_i, y_i) + d(y_i, x_i) \leqslant \frac{r_i}{2} + \frac{r_i}{2} = r_i. \tag{7.302}$$

Au final,

$$(x_1, x_2) \in B(y_1, \frac{r_1}{2}) \times B(y_2, \frac{r_2}{2}) \subset \mathcal{O}. \tag{7.303}$$

□

Définition 7.256.
Si (X, d_X) et (Y, d_Y) sont des espaces métriques, une ***isométrie*** *est une application bijective $f\colon X \to Y$ telle que pour tout $x, y \in X$ nous ayons*

$$d_Y\big(f(x), f(y)\big) = d_X(x, y). \tag{7.304}$$

Remarque 7.257.
Une application vérifiant (7.304) est automatiquement injective. En pratique, il ne faut donc vérifier que la surjectivité.

Exemple 7.258 (Manque de surjectivité)**.**
Si $X = [0, \infty[$ et $f(x) = x + 1$ alors f vérifie (7.304) pour la distance $d(x, y) = |x - y|$, mais n'est pas surjective. △

Proposition-définition 7.259 (Groupe des isométries)**.**
Si (X, d) est un espace métrique,

(1) l'ensemble des isométries de X, noté $\mathrm{Isom}(X)$ *est un groupe pour la composition.*

(2) Ce groupe agit fidèlement[75] *sur X.*

74. Définition 7.12.
75. Si vous ne savez pas ce que c'est, alors vous avez zappé la définition 2.49.

Proposition 7.260.
Une isométrie entre deux espaces métriques est continue.

Démonstration. Soient $f\colon X \to Y$ une application isométrique et $\mathcal{O}$ un ouvert de Y. Soit $a \in f^{-1}(\mathcal{O})$; si $d(a,b) < r$, alors $d\big(f(a), f(b)\big) < r$ et donc $b \in f^{-1}\big(B(f(a), r)\big)$. Donc autour de chaque point de $f^{-1}(\mathcal{O})$ nous pouvons trouver une boule ouverte contenue dans $f^{-1}(\mathcal{O})$, ce qui prouve que $f^{-1}(\mathcal{O})$ est ouvert. □

Exemple 7.261.
Si X est un ensemble, nous pouvons écrire la **distance discrète** :

$$d(x,y) = \begin{cases} 0 & \text{si } x = y \\ 1 & \text{si } x \neq y. \end{cases} \tag{7.305}$$

La topologie résultante est la topologie discrète, côtoyée dans l'exemple 7.6 [76].

Pour cette métrique, le groupe des isométries est le groupe symétrique de X, c'est-à-dire le groupe de toutes les bijections de X sur lui-même. △

7.13.0.1 Distance point-ensemble

Définition 7.262.
*Si A est une partie de l'espace métrique (X,d) et si $x \in X$, nous disons que la **distance** entre A et x est le nombre*

$$d(x,A) = \inf_{a \in A} d(x,a). \tag{7.306}$$

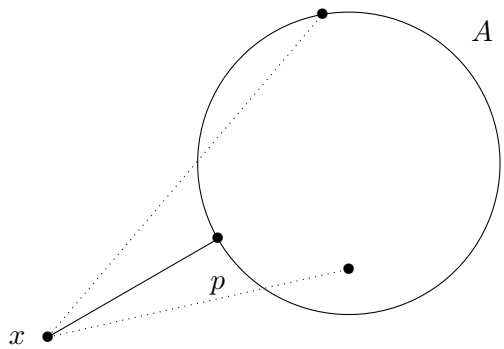


FIGURE 7.5 – La distance entre x et A est donnée par la distance entre x et p. Les distances entre x et les autres points de A sont plus grandes que $d(x,p)$.

7.13.1 Suites et espaces métriques

Proposition 7.263 (Caractérisation séquentielle de la limite[1])**.**
Soient deux espaces métriques X et Y ainsi qu'une fonction $f\colon X \to Y$. Soit $a \in X$ et $\ell \in Y$. On a

$$\lim_{x \to a} f(x) = \ell, \tag{7.307}$$

si et seulement si, pour toute suite (x_k) telle que $x_k \to a$, on a

$$\lim f(x_k) = \ell. \tag{7.308}$$

Par ailleurs, l'une des deux limites existe si et seulement si l'autre existe.

76. Vérifiez-le tout de même !

Démonstration. Le sens direct est la proposition 7.79. Pour la réciproque, nous passons par la contraposée. C'est-à-dire que nous supposons que ℓ n'est pas une limite de f pour $x \to a$. Il existe un ϵ tel que pour tout δ, il existe un x vérifiant $d_X(x;a) < \delta$ et $d_Y(f(x);\ell) > \epsilon$.

Nous construisons à présent une suite de la manière suivante. Pour $\delta = \frac{1}{n}$ nous considérons x_n tel que $d_X(x_n;a) < \delta$ et $d_Y(f(x_n);\ell) > \epsilon$. Cette suite converge vers a, mais la suite $f(x_n)$ ne converge manifestement pas vers ℓ : elle ne rentre jamais dans la boule $B(\ell,\epsilon)$. □

Une fonction continue est séquentiellement continue. Dans les espaces métriques la proposition suivante montre que la réciproque est également vraie et la continuité est équivalente à la continuité séquentielle. Cela n'est cependant pas vrai pour n'importe quel espace topologique.

Corolaire 7.264 (Caractérisation séquentielle de la continuité en un point[1])**.**
Si X et Y sont des espaces métriques, alors une fonction $f\colon X \to Y$ est continue en un point si et seulement si elle est séquentiellement continue en ce point.

Démonstration. Paraphrasons la preuve précédente. Nous supposons que X et Y sont métriques. Si f n'est pas continue en a, il existe $\epsilon > 0$ tel que pour tout $\delta > 0$, il existe x tel que $\|x - a\| \leqslant \delta$ et $\|f(x) - f(a)\| > \epsilon$. Nous considérons un tel ϵ et pour chaque $n \geqslant 1 \in \mathbb{N}$ nous considérons un x_n correspondant à $\delta = \frac{1}{n}$. Cela nous donne une suite $x_n \to a$ dans X mais $\|f(x_n) - f(a)\|$ reste plus grand que ϵ. Cela montre que f n'est pas non plus séquentiellement continue. □

Les espaces métriques ont une propriété importante que la fermeture séquentielle est équivalente à la fermeture.

Proposition 7.265 (Caractérisation séquentielle d'un fermé)**.**
Soient X un espace métrique et $F \subset X$. L'ensemble F est fermé si et seulement si toute suite contenue dans F et convergeant dans X converge vers un élément de F.

Démonstration. Une suite contenue dans un fermé ne peut converger que vers un élément de ce fermé : c'était la proposition 7.30. Le point le plus important est donc l'autre sens : si toute suite d'éléments de F converge dans F alors F est fermé.

Par contraposée, supposons que $X \backslash F$ ne soit pas ouvert. Alors il existe $x \in X \backslash F$ pour lequel tout voisinage intersecte F. En prenant $x_k \in B(x, \frac{1}{k})$, nous construisons une suite contenue dans F, convergeant vers x qui n'est pas dans F. □

Lemme 7.266.
Soit X un espace métrique, et soit (x_n) une suite convergente contenue dans un ensemble $A \subset X$. Alors la limite x_n appartient à $\bar{A}$.

Ce lemme est précisément la version « espace métrique » du corolaire 7.31 ; mais, donnons-en une preuve tout de même.

Démonstration. Supposons que nous ayons une partie A de X, et une suite (x_n) dont la limite ℓ se trouve hors de $\bar{A}$. Dans ce cas, il existe un $r > 0$ tel que[77] $B(\ell, r) \cap A = \varnothing$. Si tous les éléments x_n de la suite sont dans A, il n'y en a donc aucun tel que $d(x_n, \ell) < r$. Cela contredit la notion de convergence $x_n \to \ell$. □

Corolaire 7.267.
Soit X un espace métrique, $A \subset X$ et $a \in \bar{A}$. Alors il existe une suite d'éléments dans A qui converge vers a.

Démonstration. Si $a \in A$, alors nous pouvons prendre la suite constante $x_n = a$. Si a n'est pas dans A, alors a est dans ∂A, et pour tout n, il existe un point de A dans la boule $B(a, \frac{1}{n})$. Si nous nommons x_n ce point, la suite ainsi construite est une suite contenue dans A et qui converge vers a (ce dernier point est laissé à la sagacité du lecteur ou de la lectrice). □

77. Une autre manière de dire la même chose : si $\ell \notin \bar{A}$, alors $d(\ell, A) > 0$.

En termes savants, ce corolaire signifie que la fermeture $\bar{A}$ est composé de A plus de toutes les limites de toutes les suites contenues dans A.

Proposition 7.268 (Caractérisation séquentielle de la continuité[1])**.**
Soient X et Y deux espaces métriques. Une application $f\colon X \to Y$ est continue sur X si et seulement si elle est séquentiellement continue sur X.

Démonstration. Le sens direct est déjà prouvé dans la proposition 7.79. Nous nous concentrons donc sur la réciproque.

Soit $\mathcal{O}$ un ouvert de Y ; nous allons voir que le complémentaire de $f^{-1}(\mathcal{O})$ est fermé dans E. Pour cela nous considérons une suite convergente $x_k \xrightarrow{E} x$ avec $x_k \in \complement f^{-1}(\mathcal{O})$ pour tout k. Nous allons montrer que $x \in \complement f^{-1}(\mathcal{O})$ et la caractérisation séquentielle[78] de la fermeture conclura que $\complement f^{-1}(\mathcal{O})$ est fermé.

Pour tout k, nous avons $f(x_k) \in \complement\mathcal{O}$, mais $\mathcal{O}$ est ouvert et $f(x_k) \xrightarrow{Y} f(x)$ parce que f est séquentiellement continue. Par conséquent $f(x) \in \complement\mathcal{O}$ et $x \in \complement f^{-1}(\mathcal{O})$. □

Proposition 7.269.
Si X et Y sont deux espaces métriques et $f, g\colon X \to Y$ sont deux fonctions continues égales sur une partie dense de X alors $f = g$.

Démonstration. Les fonctions f et g sont séquentiellement continues (proposition 7.80, ou proposition 7.264). Soient A un ensemble dense dans X sur lequel f et g sont égales, et $x \notin A$. Vu que A est dense, il existe une suite a_n dans A telle que $a_n \to x$. La séquentielle continuité de f et g donnent

$$f(a_n) \to f(x) \tag{7.309a}$$
$$g(a_n) \to g(x), \tag{7.309b}$$

mais pour tout n, $f(a_n) = g(a_n)$. Par unicité de la limite[79] dans Y, $f(x) = g(x)$. □

Lemme 7.270.
La fonction

$$\begin{aligned} f\colon \mathbb{R} \times \mathbb{R} &\to \mathbb{R} \\ (x, y) &\mapsto x + y \end{aligned} \tag{7.310}$$

est continue.

Démonstration. Pour rappel, la topologie considérée sur $\mathbb{R}^n$ est celle de la définition 7.230. En vertu de la proposition 7.268, il est suffisant de prouver la continuité séquentielle. Soit donc une suite convergente

$$(x_n, y_n) \xrightarrow{\mathbb{R}\times\mathbb{R}} (x, y). \tag{7.311}$$

Nous devons prouver que

$$f(x_n, y_n) \xrightarrow{\mathbb{R}} f(x, y). \tag{7.312}$$

La proposition 7.38 nous permet de déduire la convergence composante par composante : $x_n \xrightarrow{\mathbb{R}} x$ et $y_n \xrightarrow{\mathbb{R}} x$. En permutant somme et limite (proposition 7.133) nous avons le calcul

$$f(x_n, y_n) = x_n + y_n \xrightarrow{\mathbb{R}} x + y = f(x, y). \tag{7.313}$$

D'où la convergence demandée. □

78. Proposition 7.265.
79. Proposition 7.70.

7.13.2 Espace métrisable

Définition 7.271 (Espace vectoriel topologique métrisable[151]).
*Un espace topologique est **métrisable** si il existe une distance compatible avec la topologie.*

Proposition 7.272 ([152]).
Soit un espace topologique métrisable X.

(1) Tout fermé de X est une intersection dénombrable d'ouverts.

(2) Tout ouvert de X est une union dénombrable de fermés.

Démonstration. Soit une métrique d compatible avec la topologie de X et un fermé A. Nous posons

$$V_n = \{x \in X \text{ tel que } d(x, A) < \frac{1}{n}\}. \tag{7.314}$$

Et juste pour faire simple nous notons $V_0 = X$.

Les parties V_n sont ouvertes Soit $x \in V_n$. Trouvons un voisinage de x contenu dans V_n afin de pouvoir encore invoquer le théorème 7.4. D'abord, vu que $x \in V_n$, il existe $a \in A$ tel que $d(x, a) < \frac{1}{n}$ (ici les inégalités strictes sont importantes).

Soient $\epsilon > 0$ que nous fixerons plus bas, et $y \in B(x, \epsilon)$. L'inégalité triangulaire donne

$$d(y, a) \leqslant d(y, x) + d(x, a) < \epsilon + \frac{1}{n}. \tag{7.315}$$

Nous pouvons donc choisir ϵ de telle sorte que $d(y, a) < 1/n$. Avec ce ϵ, nous avons, pour tout $y \in B(x, \epsilon)$:

$$d(y, A) \leqslant d(y, a) < \frac{1}{n} \tag{7.316}$$

et donc $y \in V_n$.

A est l'intersection des V_n Nous avons évidemment $A \subset V_n$ pour tout n. Et d'autre part, si $a \in \bigcap_{n\in\mathbb{N}} V_n$ alors $d(a, A) < \frac{1}{n}$ pour tout n. Cela implique $d(a, A) = 0$, et donc $a \in A$ par le lemme 7.108.

Ceci démontre le point (1).

En ce qui concerne la seconde partie, nous appliquons la première partie au complémentaire. Si $\mathcal{O}$ est ouvert, $\mathcal{O}^c$ est fermé et

$$\mathcal{O}^c = \bigcap_{n\in\mathbb{N}} V_n, \tag{7.317}$$

ce qui donne immédiatement

$$\mathcal{O} = \bigcup_{n\in\mathbb{N}} V_n^c \tag{7.318}$$

où les V_n^c sont fermés. □

Corolaire 7.273.
Si X est un espace topologique métrisable, alors X accepte une base dénombrable de topologie autour de chaque point.

Démonstration. Il s'agit seulement de remarquer que les singletons sont fermés et d'appliquer la proposition 7.272. □

7.14 Suites de Cauchy, métrique et espaces complets

7.14.1 Généralités

Définition 7.274 (Suite de τ-Cauchy, espace vectoriel topologique[153, 154]).
*Soit E un espace vectoriel topologique. Une suite (x_k) dans E est une **suite τ-Cauchy** si pour tout voisinage $\mathcal{U}$ de 0 il existe $N \in \mathbb{N}$ tel que $x_k - x_l \in \mathcal{U}$ pour tout $k, l \geqslant N$.*

Définition 7.275 (Espace τ-complet)**.**
Nous disons qu'une partie A d'un espace vectoriel topologique est ***τ-complet*** *si toute suite τ-Cauchy d'éléments de A converge*[80] *vers un élément de A.*

Définition 7.276 (Suite de Cauchy, espace métrique)**.**
Une suite (a_k) dans un espace métrique (V, d) est ***de Cauchy*** *si pour tout $\epsilon \in \mathbb{R}$, il existe N tel que si $n, m \geqslant N$ alors $d(a_n, a_m) < \epsilon$.*

Notons qu'ici, même si l'espace V n'a rien à voir avec $\mathbb{R}$, nous prenons ϵ dans $\mathbb{R}$ et la distance à valeurs dans $\mathbb{R}$. Cela semble une évidence, mais il faut se rendre compte que $\mathbb{R}$ commence à prendre une place centrale dans nos constructions. Ce n'était pas le cas du temps où nous parlions de suites de Cauchy et de complétude dans des corps totalement ordonnés (définitions 1.109). Dans ce contexte, le ϵ était pris dans le corps lui-même.

Définition 7.277 (Métrique complète)**.**
Soit (E, d) un espace métrique. Nous disons que la métrique d est ***complète*** *si toute suite de Cauchy dans (E, d) converge dans E.*

7.278.
Ces définitions méritent quelques remarques.

(1) Dans le cas des espaces vectoriels topologiques, nous définissons les notions de suite τ-Cauchy et d'espace topologique τ-complet. Nous ajoutons le préfixe τ pour indiquer que ce sont des notions topologiques.

(2) Dans le cas des espaces métriques, nous définissons la notion de *métrique* complète. C'est bien la métrique qui est complète, et non l'espace. En effet nous allons voir dans l'exemple 7.280 que le même espace topologique peut accepter plusieurs distances différentes (donnant la même topologie) donnant lieu à des suites de Cauchy différentes.

(3) Si un espace vectoriel a une topologie issue d'une distance, rien ne dit que ses suites τ-Cauchy et ses suites de Cauchy sont les mêmes. Ce sont deux notions à priori séparées. Si V est un espace vectoriel topologique que l'on peut munir de deux distances d_1, d_2 donnant toutes deux la topologie, dire que V est τ-complet, dire que d_1 est complète et dire que d_2 est complète sont trois choses différentes. Même si les trois topologies sont identiques.

(4) Nous allons bien entendu voir que dans de larges gammes d'exemples, les notions de suite de Cauchy et τ-Cauchy coincident.

Définition 7.279.
Un ***espace de Banach*** *est un espace vectoriel normé complet*[81] *pour la topologie de la norme.*

Exemple 7.280 (La complétude n'est pas une propriété topologique[155])**.**
Le fait pour un espace d'être complet n'est pas une propriété topologique, mais une propriété métrique. Plus exactement, il existe des espaces topologiques isomorphes, mais dont l'un est complet et l'autre non.

Nous considérons la distance suivante sur $\mathbb{N}$:

$$d_1(x, y) = \left|\frac{1}{x} - \frac{1}{y}\right|. \tag{7.319}$$

Pour vérifier que cette formule définit bien une distance (définition 7.93), le seul point non immédiat est l'inégalité triangulaire :

$$d_1(x, y) = \left|\frac{1}{x} - \frac{1}{y}\right| \leqslant \left|\frac{1}{x} - \frac{1}{z}\right| + \left|\frac{1}{z} - \frac{1}{y}\right| = d_1(x, z) + d_1(z, y). \tag{7.320}$$

80. Définition 7.10.
81. Définition 7.277.

Au niveau de la topologie induite par cette distance, c'est la topologie discrète. En effet, soit $x \in \mathbb{N}$ et $\epsilon > 0$; nous voulons déterminer la boule $B(x, \epsilon)$ en résolvant l'équation

$$\left|\frac{1}{x} - \frac{1}{y}\right| < \epsilon \tag{7.321}$$

pour $y \in \mathbb{N}$. Nous trouvons que $\frac{1}{y} > \frac{1}{x} - \epsilon$ et $\frac{1}{y} < \frac{1}{x} + \epsilon$, soit

$$\begin{cases} y > \dfrac{1}{\frac{1}{x} + \epsilon} & (7.322a) \\ y < \dfrac{1}{\frac{1}{x} - \epsilon}. & (7.322b) \end{cases}$$

Si ϵ est assez petit, la seule solution entière est $y = x$. Les ouverts sont donc toutes les parties parce que tous les singletons sont ouverts.

L'espace topologique associé à $(\mathbb{N}, d_1)$ est donc la topologie discrète[82].

Si nous considérons par contre la distance usuelle sur $\mathbb{N}$, à savoir $d(x, y) = |x-y|$, nous obtenons encore la topologie discrète. Nous avons donc un isomorphisme d'espaces topologiques

$$(\mathbb{N}, d) \simeq (\mathbb{N}, d_1). \tag{7.323}$$

Nous pouvons même donner un isomorphisme explicite : $f(n) = n$.

La suite $(x_n) = n$ est une suite de Cauchy dans $(\mathbb{N}, d_1)$ parce que si $\epsilon > 0$ est donné, il suffit de prendre N assez grand pour avoir $\frac{1}{N} < \epsilon$ (possible par le lemme 1.144) nous avons, pour $n, m > N$:

$$\left|\frac{1}{n} - \frac{1}{m}\right| < \frac{1}{n} < \frac{1}{N} < \epsilon. \tag{7.324}$$

Or cette suite ne converge pas. Soit en effet un candidat limite k. Calculons

$$d_1(x_n, k) = \left|\frac{1}{n} - \frac{1}{k}\right| \to \frac{1}{k} \neq 0. \tag{7.325}$$

L'espace $(\mathbb{N}, d_1)$ n'est pas complet.

Notons que cette suite n'est pas de Cauchy dans $(\mathbb{N}, d)$.

En résumé :

(1) Les espaces topologiques $(\mathbb{N}, d)$ et $(\mathbb{N}, d_1)$ sont isomorphes.

(2) Ils ont les mêmes notions de suites convergentes : une suite convergente pour l'un est convergente pour l'autre.

(3) Ils n'ont pas les mêmes notions de suites de Cauchy.

(4) Dans $(\mathbb{N}, d_1)$, il existe des suites de Cauchy qui ne convergent pas (pas complet).

(5) L'espace $(\mathbb{N}, d)$ est complet, mais $(\mathbb{N}, d_1)$ n'est pas complet.

(6) Le fait pour un espace topologique métrique d'être complet n'est pas intrinsèque à sa topologie : la complétude est une propriété de la distance. La complétude est une propriété de la métrique, et non de la topologie qui s'en suit.

△

7.14.2 Espace topologique métrique

Dans les espaces vectoriels topologiques métriques, il n'y a pas d'ambiguïté.

Proposition 7.281 (Caractérisations avec la distance d)**.**
Soit (E, d) un espace vectoriel topologique métrique.

82. Celle dont toutes les parties sont des ouverts.

(1) Une suite (x_n) dans E est convergente[83] *vers x si et seulement si pour tout $\epsilon \in \mathbb{R}$ il existe N_ϵ tel que pour tout $n \geqslant N_\epsilon$ nous avons $d(x_n, x) \leqslant \epsilon$.*

(2) Une suite (x_n) dans E est de Cauchy[84] *si pour tout $\epsilon \in \mathbb{R}$, il existe un N_ϵ tel que si $p, q \geqslant N_\epsilon$, nous avons $d(x_p, x_q) \leqslant \epsilon$.*

Démonstration. En ce qui concerne la convergence :

Sens direct Nous supposons que $x_k \to x$ dans E. Soit $\epsilon > 0$; vu que $B(x, \epsilon)$ est un ouvert contenant x, il existe un $N_\epsilon > 0$ tel que $k > N_\epsilon$ implique $x_k \in B(x, \epsilon)$. Cela signifie $d(x, x_k) \leqslant \epsilon$.

Réciproque Nous supposons que pour tout $\epsilon > 0$, il existe $N_\epsilon > 0$ tel que si $k > N_\epsilon$ alors $x_k \in B(x, \epsilon)$. Soit un ouvert $\mathcal{O}$ autour de x. Nous sommes dans un espace métrique ; ergo la topologie est donné par le théorème 7.94 et en particulier la liste des ouverts est donnée par (7.61). Il existe donc une boule $B(x, \epsilon)$ incluse à $\mathcal{O}$. Pour tout $k > N_\epsilon$ nous avons alors $x_k \in B(x, \epsilon) \subset \mathcal{O}$.

En ce qui concerne les suites de Cauchy :

Sens direct Si (x_n) est une suite de Cauchy et si $\epsilon > 0$ est donné, alors $B(0, \epsilon)$ est un voisinage de 0 et il existe N_ϵ tel que si $p, q \geqslant N_\epsilon$ alors $x_p - x_q \in B(0, \epsilon)$. Posons $u = x_p - x_q$; en utilisant l'invariance par translation (lemme 7.117(1)) nous avons

$$d(u, 0) = d(x_p - x_q, 0) = d(x_p, x_q). \tag{7.326}$$

Par conséquent $d(x_p, x_q) \leqslant \epsilon$.

Réciproque Soit $\mathcal{O}$ un voisinage de 0. Il existe ϵ tel que $B(0, \epsilon) \subset \mathcal{O}$. Par hypothèse il existe N_ϵ tel que $d(x_p, x_q) \leqslant \epsilon$ dès que $p, q \geqslant N_\epsilon$. En utilisant encore l'invariance par translation nous avons

$$d(x_p, x_q) = d(x_p - x_q, 0), \tag{7.327}$$

et comme cela est plus petit que ϵ, nous avons $x_p - x_q \in B(0, \epsilon) \subset \mathcal{O}$.

□

Proposition 7.282 ([156]).
Toute suite convergente dans un espace métrique est de Cauchy.

Démonstration. Nous utilisons les caractérisations de la proposition 7.281 des suites convergentes et de Cauchy.

Soit un espace métrique (X, d) et $x_n \to \ell$ une suite convergente. Si $\epsilon > 0$, la proposition 7.281(1), dit qu'il existe N tel que pour tout $n > N$ nous ayons $d(x_n, \ell) < \epsilon$. Par conséquent si $n, m > N$ alors

$$d(x_n, x_m) \leqslant d(x_m, \ell) + d(l, x_m) \leqslant 2\epsilon. \tag{7.328}$$

Cela prouve que (x_n) est une suite de Cauchy. □

7.15 Topologie et espace vectoriel

7.15.1 Espace vectoriel topologique

Définition 7.283.
Un espace vectoriel V sur le corps $\mathbb{K}$ muni d'une topologie est un ***espace vectoriel topologique*** *si*

(1) la somme de deux vecteurs est une application continue[85] *$V \times V \to V$; et*

83. Définition 7.10.
84. Définition 7.274.
85. Naturellement, l'espace $V \times V$ est muni de la topologie produit.

(2) la multiplication par un scalaire est une application continue[86] $\mathbb{K} \times V \to V$.

On le redit quand même : le corps[87] lui-même doit avoir sa topologie. Dans la grande majorité des cas, ce corps est $\mathbb{R}$ ou $\mathbb{C}$ muni de la topologie usuelle.

Mine de rien, le fait que les deux opérations usuelles soient continues a de belles conséquences sur la topologie de l'espace...

Proposition 7.284 ([154]).
Pour $x \in V$ et $\lambda \in \mathbb{K}$, $\lambda \neq 0$ fixés, les fonctions T_x et M_λ définies par :

$$T_x : V \to V \qquad \text{et} \qquad M_\lambda : V \to V \tag{7.329}$$

$$y \mapsto x + y \qquad\qquad y \mapsto \lambda y \tag{7.330}$$

sont des homéomorphismes de V dans V.

Démonstration. Ce sont des bijections continues, dont les inverses sont respectivement T_{-x} et $M_{1/\lambda}$. □

Corolaire 7.285 (Invariance de la topologie [154]).
Toute base de voisinage de 0 se transporte en tout point de l'espace vectoriel topologique.

Lemme 7.286 ([154]).
Soit V un espace vectoriel topologique, et W un voisinage de 0. Il existe U un voisinage de 0, symétrique[88]*, tel que $U + U = W$.*

Démonstration. Par continuité de l'addition et par la définition de la topologie produit, il existe U_1 et U_2 tels que $U_1 + U_2 \subset W$. En posant $U = U_1 \cap U_2 \cap (-U_1) \cap (-U_2)$, on a un sous-ensemble symétrique de U_1 et U_2, si bien que $U + U = W$. □

Définition 7.287.
Une distance d sur un espace vectoriel topologique V est dite ***compatible*** *avec la topologie si la topologie induite*[89] *de d est celle de V.*

Une distance d sur un espace vectoriel V est dite ***invariante*** *si pour tout $x, y, u \in V$ nous avons*

$$d(x + u, y + u) = d(x, y). \tag{7.331}$$

Notons que lorsque nous parlons d'une distance compatible avec un espace vectoriel topologique, nous parlons de compatibilité avec la topologie, pas avec la structure vectorielle.

Théorème 7.288 ([154]).
Si V est un espace vectoriel topologique possédant en tout point une base de topologie dénombrable, alors il existe une distance d sur V telle que

(1) d est compatible avec la topologie de V,

(2) d est invariante par translation.

Démonstration. Grâce à la proposition 7.285, on peut tout ramener en 0 puis faire les transports en tous les points de l'espace. Mieux : grâce à la proposition 7.286 (appliquée deux fois de suite), on peut créer une base de voisinage (U_n) de 0 telle que pour tout $n \in \mathbb{N}$,

$$U_{n+1} + U_{n+1} + U_{n+1} + U_{n+1} \subset U_n. \tag{7.332}$$

Pour tous entiers naturels n et k, on obtient alors

$$U_{n+1} + U_{n+2} + \cdots + U_{n+(k-1)} + U_{n+k} \subset U_{n+1} + U_{n+1} \subset U_n. \tag{7.333}$$

86. Naturellement, l'espace $\mathbb{K} \times V$ est muni (lui aussi) de la topologie produit.
87. Définition 1.97
88. C'est-à-dire que, pour tout $x \in V$, on a $x \in U$ si et seulement si $-x \in U$.
89. Définition 7.94.

On construit à présent, pour tout $n \in \mathbb{N}$, l'ensemble

$$D_n = \left\{ \sum_{i=1}^{n} \frac{c_i}{2^i} \text{ tel que } \forall i = 1, \cdots, n, c_i \in \{0, 1\} \right\}, \tag{7.334}$$

et $D = \cup_{n>0} D_n$. Ensuite, définissons ϕ sur $D \cup [1, +\infty[$ et à valeurs dans les parties de V :

$$\phi(r) = \begin{cases} V & \text{si } r \geqslant 1; \\ c_1 U_1 + \cdots + c_n U_n & \text{si } r \in D_n. \end{cases} \tag{7.335}$$

Quelques remarques sur cette fonction.

(1) $\phi(r) + \phi(s) \subset \phi(r+s)$ *:* Si déjà $r + s \geqslant 1$, c'est clair. Sinon, on se place dans D_n avec le n qui va bien – de telle sorte que r, s et $r + s$ soient dedans. Notons :

$$r = \sum_{i=1}^{n} \frac{r_i}{2^i}; \tag{7.336}$$

$$s = \sum_{i=1}^{n} \frac{s_i}{2^i}; \tag{7.337}$$

$$r + s = \sum_{i=1}^{n} \frac{t_i}{2^i}. \tag{7.338}$$

Deux cas se produisent. Si pour tout i, $t_i = r_i + s_i$, alors

$$\phi(r+s) = \sum_i t_i U_i = \sum_i r_i U_i + \sum_i s_i U_i = \phi(r) + \phi(s); \tag{7.339}$$

l'égalité a lieu car r_i et s_i ne peuvent jamais valoir 1 en même temps.

Sinon, posons k le plus petit entier tel que $t_k \neq r_k + s_k$. Alors, nécessairement, $r_k = 0$, $s_k = 0$ et $t_k = 1$. Il s'ensuit, grâce à (7.332) et (7.333), que

$$\phi(r) = \sum_{i=1}^{k-1} r_i V_i + \sum_{i=k+1}^{n} r_i V_i \subset \sum_{i=1}^{k-1} r_i V_i + V_{k+1} + V_{k+1}; \tag{7.340}$$

$$\phi(s) = \sum_{i=1}^{k-1} s_i V_i + \sum_{i=k+1}^{n} s_i V_i \subset \sum_{i=1}^{k-1} s_i V_i + V_{k+1} + V_{k+1}; \text{ d'où} \tag{7.341}$$

$$\phi(r) + \phi(s) = \sum_{i=1}^{k-1} r_i V_i + \sum_{i=1}^{k-1} s_i V_i + V_{k+1} + V_{k+1} + V_{k+1} + V_{k+1} = \sum_{i=1}^{k-1} t_i V_i + V_k \subset \phi(r+s). \tag{7.342}$$

(2) $0 \in \phi(r)$ *pour tout* r *:* en effet, $\phi(r)$ n'est jamais vide, c'est toujours un voisinage de 0.

(3) *si* $r < s$ *alors* $\phi(r) \subset \phi(s)$ *:* il suffit d'écrire

$$\phi(r) \subset \phi(r) + \phi(s - r) \subset \phi(s). \tag{7.343}$$

Enfin, on définit

$$d(x, y) = \inf\{r \in [0; 1] \text{ tel que } y - x \in \phi(r)\}. \tag{7.344}$$

Il suffit alors de voir que d convient. De par sa définition, il est clair qu'elle est invariante par translation ; reste à voir que c'est bien une distance, et qu'elle est compatible avec la topologie.

$\underline{d(x, x) = 0}$ Oui, car 0 est dans $\phi(r)$, pour tout r, puisque les U_i sont des voisinages de 0.

$\underline{d(x, y) = d(y, x)}$ Oui, car tous les voisinages considérés sont symétriques : pour tout i et tout $x \in V$, on a $x \in U_i$ si et seulement si $-x \in U_i$.

$d(x,z) \leqslant d(x,y) + d(y,z)$ Soit $\epsilon > 0$. Par définition des distances comme infimums, et grâce au corolaire 1.148, il existe r et s dans D tels que :

$$d(x,y) < r < d(x,y) + \frac{\epsilon}{2} \quad \text{et} \quad d(y,z) < s < d(y,z) + \frac{\epsilon}{2}. \tag{7.345}$$

Comme $d(x,y)+d(y,z) < r+s$, et par la remarque (3) sur ϕ, on a $y-x \in \phi(r)$ et $z-y \in \phi(s)$; donc

$$(y-x) + (z-y) = z - x \in \phi(r) + \phi(s) \subset \phi(r+s) \tag{7.346}$$

Ainsi, pour tout $\epsilon > 0$, on a

$$d(x,z) \leqslant r + s < d(x,y) + d(y,z) + \epsilon. \tag{7.347}$$

Compatibilité avec la topologie Si $d(0,y) < r$, alors $y \in \phi(r)$; en particulier pour $r = 1/2^k$, on a $y \in \phi(r) = V_k$. D'où, pour tout $n \in \mathbb{N}$, $B(0, 1/2^n) \subset V_n$.

□

Proposition 7.289.
Un espace vectoriel topologique[90] *est métrisable si et seulement si il possède en tout point une base dénombrable de topologie.*

Démonstration. Il s'agit seulement de mettre bout à bout les corolaires 7.273 et théorème 7.288. □

7.15.2 Équivalence entre Cauchy et $\tau-$Cauchy

Lemme 7.290.
Soit un espace vectoriel topologique[91] *V et une distance $d\colon V \times V \to \mathbb{R}^+$ compatible*[92] *avec la topologie de V. Si d est invariante*[93]*, alors les suites de Cauchy pour d et les suites τ-Cauchy sont les mêmes.*

Démonstration. Nous avons deux implications à prouver.

Cauchy pour d implique τ-Cauchy Soit (x_n), une suite de Cauchy dans V pour d, et un voisinage U de 0. Vu que d est compatible avec la topologie de V, il existe une boule ouverte $B(0,\epsilon)$ incluse à U. Soit $N > 0$ tel que $m, n > N$ implique $d(x_n, x_m) < \epsilon$. Par invariance de la métrique, nous avons aussi

$$d(0, x_m - x_n) < \epsilon, \tag{7.348}$$

c'est-à-dire $x_m - x_n \in B(0,\epsilon) \subset U$. La suite (x_n) est donc τ-Cauchy.

τ-Cauchy implique Cauchy pour d Soit (x_n), une suite τ-Cauchy dans V et $\epsilon > 0$. Vu que $B(0,\epsilon)$ est un voisinage de 0 dans V, il existe N tel que $m, n > N$ implique $x_n - x_m \in B(0,\epsilon)$. Cela signifie que $d(0, x_n - x_m) < \epsilon$ et toujours par invariance, que $d(x_n, x_m) < \epsilon$.

□

Tout ceci nous mène à donner une large classe d'espaces vectoriels topologiques sur lesquelles les notions de suites de Cauchy pour une distance et τ-Cauchy coïncident.

Théorème-définition 7.291.
Soit V un espace vectoriel topologique métrisable[94]*, alors il admet une métrique d compatible avec la topologie telle que une suite dans V est de Cauchy pour d si et seulement si elle est τ-Cauchy.*

*Une **suite de Cauchy** dans un espace vectoriel métrique (E,d) est une suite τ-Cauchy ou de Cauchy pour d.*

90. Définition 7.283.
91. Définition 7.283.
92. Définition 7.287.
93. Définition 7.287.
94. i.e. admet une base dénombrable de topologie, voir la proposition 7.289

Démonstration. Soit d une métrique sur V satisfaisant au théorème 7.288. Vu qu'elle est invariante par translation, les suites d-Cauchy sont exactement les suites τ-Cauchy par le lemme 7.290. □

Remarque 7.292.
Même si V est métrisable, si on choisit la métrique n'importe comment, on ne peut rien espérer.

7.293.
Sur les espaces vectoriels topologiques métrisables, nous pouvons donc parler de suite de Cauchy sans préciser si nous parlons de τ-Cauchy ou de d-Cauchy, parce que nous sous-entendons avoir choisi une métrique non seulement compatible avec la topologie, mais également invariante par translation.

Il reste cependant à traiter le cas d'un espace vectoriel topologique non métrisable. Dans ce cas, il n'y a pas de métrique, et la question de l'équivalence des définitions ne se pose pas.

Le théorème suivant donne la complétude de $\mathbb{R}$ et le critère de Cauchy pour les définitions métriques et topologiques usuelles. Lorsqu'on dit que $\mathbb{R}$ est complet, le plus souvent nous parlons de ce théorème, et non de 1.153 qui en est un lemme indispensable mais qui parle de notions différentes, bien que très liées.

Théorème 7.294 (Complétude de $\mathbb{R}$, critère de Cauchy[8])**.**
Nous avons :

(1) L'espace métrique $(\mathbb{R}, d)$ est complet (définition 7.277).

(2) Une suite dans $\mathbb{R}$ est convergente (définition 7.10) si et seulement si elle est de Cauchy (définition 7.291).

Démonstration. Tout ce théorème se base sur le fait que la définition de suite de Cauchy dans $(\mathbb{R}, d)$ et de suite convergente dans $(\mathbb{R}, d)$ coïncident avec les définitions correspondantes dans $\mathbb{R}$ vu comme simple corps ordonné (définitions 1.109).

Donc si (x_n) est de Cauchy dans $(\mathbb{R}, d)$, elle est de Cauchy dans le corps ordonné $(\mathbb{R}, \leqslant)$. Donc le théorème 1.153 nous dit que (x_n) est convergente dans $(\mathbb{R}, \leqslant)$. Et donc convergente dans $(\mathbb{R}, d)$.

Toutes les autres affirmations se prouvent de la même manière. □

Si vous n'êtes pas sûr ou si vous ne voulez pas étudier les notations de convergence et de suites de Cauchy dans les corps, vous pouvez simplement recopier la démonstration du théorème 1.153 en remplaçant partout $\mathbb{Q}$ par $\mathbb{R}$, et aussi en remplaçant les $|x - y|$ par $d(x, y)$.

7.295.
Nous pouvons également mettre une structure d'espace métrique sur $\mathbb{C}$ en posant

$$d(z, z') = |z - z'|. \tag{7.349}$$

Proposition 7.296.
L'espace métrique $(\mathbb{C}, d)$ est complet.

Démonstration. Commençons par nous rendre compte que pour tout $z \in \mathbb{C}$ nous avons $|\operatorname{Re}(z)| \leqslant |z|$. C'est bon ? Vous vous en êtes rendu compte ? Ok. Continuons.

Soit une suite de Cauchy (z_k) dans $\mathbb{C}$ et $\epsilon > 0$. Si $x_k = \operatorname{Re}(z_j)$, nous avons

$$|x_k - x_l| = |\operatorname{Re}(z_k - z_l)| \leqslant |z_k - z_l|. \tag{7.350}$$

Vu que (z_k) est de Cauchy, il existe un N tel que si $k, l \geqslant N$,

$$|x_k - x_l| \leqslant |z_k - z_l| \leqslant \epsilon. \tag{7.351}$$

Donc la suite des parties réelles converge par la complétude de $(\mathbb{R}, d)$ du théorème 7.294. Notez que le d ici n'est pas tout à fait le même, et que la démonstration fonctionne parce que la distance prise sur $\mathbb{R}$ est la restriction à $\mathbb{R}$ de la distance prise sur $\mathbb{C}$. Notons x la limite de (x_k).

De la même manière la suite des parties imaginaires $y_k = \mathrm{Im}(z_k)$ converge vers un réel que nous notons y. Avec tout cela, la suite z_k converge dans $\mathbb{C}$ vers $x + iy$. En effet pour ϵ donné et pour un k suffisament grand,

$$|z_k - (x + iy)| = \big|\mathrm{Re}(z_k) - x + i(\mathrm{Im}(z_k) - y)\big| \leqslant |x_k - x| + |y_k - y| \leqslant \epsilon. \tag{7.352}$$

□

7.16 Norme ; espace vectoriel normé

La valeur absolue est essentielle pour introduire les notions de limite et de continuité pour les fonctions d'une variable. Par exemple nous verrons dans la proposition 11.42 que la fonction $f\colon \mathbb{R} \to \mathbb{R}$ est continue en a si et seulement si pour tout $\varepsilon > 0$, il existe un $\delta > 0$ tel que

$$|x - a| \leqslant \delta \Rightarrow |f(x) - f(a)| \leqslant \varepsilon. \tag{7.353}$$

La quantité $|x - a|$ donne la « distance » entre x et a ; la définition de la continuité signifie que pour tout ε, il existe un δ tel que si a et x sont au plus à la distance δ l'un de l'autre, alors $f(x)$ et $f(a)$ ne seront éloignés au plus d'une distance ε.

La valeur absolue, dans $\mathbb{R}$, nous sert donc à mesurer des distances entre les nombres. Les principales propriétés de la valeur absolue sont :

(1) $|x| = 0$ implique $x = 0$,

(2) $|\lambda x| = |\lambda||x|$,

(3) $|x + y| \leqslant |x| + |y|$

pour tout $x, y \in \mathbb{R}$ et $\lambda \in \mathbb{R}$.

Afin de donner une notion de limite pour les fonctions de plusieurs variables, nous devons trouver un moyen de définir les notions de « taille » d'un vecteur et de distance entre deux points de $\mathbb{R}^n$, avec $n > 1$. La notion de « taille » doit satisfaire propriétés analogues à celles de la valeur absolue.

La première notion de « taille » pour un vecteur de $\mathbb{R}^2$ que nous vient à l'esprit est la longueur du segment entre l'origine et l'extrémité libre du vecteur. Cela peut être calculée à l'aide du théorème de Pythagore :

$$\text{taille de } (a, b) = \sqrt{a^2 + b^2}. \tag{7.354}$$

Nous pouvons introduire une notion de distance entre les éléments de $\mathbb{R}^2$ de façon similaire :

$$d\big((a_x, a_y), (b_x, b_y)\big) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2}. \tag{7.355}$$

Cette définition a l'air raisonnable ; est-elle mathématiquement correcte ? Peut-elle jouer le rôle de la valeur absolue dans $\mathbb{R}^2$? Est-elle la seule définition possibles de « taille » et distance en $\mathbb{R}^2$?

Nous voulons formaliser les notions de « taille » et de distance dans $\mathbb{R}^n$, et plus généralement dans un espace vectoriel V de dimension finie. Pour cela nous nous inspirons des propriétés de la valeur absolue.

7.16.0.1 Critère de Cauchy

Lemme 7.297.
Une suite de Cauchy[95] *dans un espace vectoriel normé admettant une sous-suite convergente est elle-même convergente vers la même limite.*

Démonstration. Soit (a_n) une suite de Cauchy dans un espace vectoriel normé E et ℓ la limite d'une sous-suite de (a_n). Soit $\epsilon > 0$ et $N \in \mathbb{N}$ tel que $\|a_m - a_p\| < \epsilon$ dès que $m, p \geqslant N$. Nous allons

95. Définition 7.276.

montrer que si $k > N$ alors $\|a_k - \ell\| < 2\epsilon$. Pour cela nous considérons un $n > N$ tel que $\|a_n - \ell\| \leqslant \epsilon$ et nous calculons

$$\|a_k - \ell\| \leqslant \|a_k - a_n\| + \|a_n - \ell\| \leqslant 2\epsilon. \tag{7.356}$$

□

Dans le cas des espaces de dimension finie, le fait d'être complet est automatique, comme le montre la proposition suivante.

Proposition 7.298.
Soit $\big(E, \|.\|\big)$ un espace vectoriel normé de dimension finie sur un corps $\mathbb{K}$ qui est complet[96]*. Alors E est complet*[97]*.*

Pour rappel, la complétude de l'espace métrique $\mathbb{R}$ est la proposition 1.115.

Démonstration. Nous considérons une suite de Cauchy (f_n) dans E et si $\{e_\alpha\}$ est une base orthonormée de E nous définissons les coefficients $f_n = \sum_\alpha a_{n\alpha} e_\alpha$. La somme sur α est finie par hypothèse sur la dimension de E.

Nous avons

$$\|f_n - f_m\| = \|\sum_\alpha (a_{n\alpha} - a_{m\alpha}) e_\alpha\| = \sum_\alpha |a_{n\alpha} - a_{m\alpha}|^2. \tag{7.357}$$

Pour tout ϵ, il existe N tel que si $m, n > N$ alors $|a_{n\alpha} - a_{m\alpha}| < \sqrt{\epsilon}$. Autrement dit, pour chaque α, la suite $(a_{n\alpha})_{\alpha \in \mathbb{N}}$ est de Cauchy dans $\mathbb{K}$ et converge donc dans $\mathbb{K}$. Soit a_α la limite et définissons $f = \sum_\alpha a_\alpha e_\alpha$. Nous avons alors

$$\|f_n - f\| = \|\sum_\alpha (a_{n\alpha} - a_\alpha) e_\alpha\|, \tag{7.358}$$

dont la limite $n \to \infty$ est bien zéro. Donc la suite (f_n) converge vers $f \in E$. L'espace E est alors complet. □

Proposition 7.299.
Soient V et W deux espaces vectoriels normés. Soient K une partie compacte de V et $f\colon K \to W$ une fonction continue. Alors l'image $f(K)$ est compacte dans W.

Ce résultat est démontré dans un cadre plus général par le théorème 7.92.

Démonstration. Nous allons prouver que $f(K)$ est fermée et bornée.

$f(K)$ est fermé Nous allons prouver que si (y_n) est une suite convergente contenue dans $f(K)$, alors la limite est également contenue dans $f(K)$. Dans ce cas, nous aurons que l'adhérence de $f(K)$ est contenue dans $f(K)$ et donc que $f(K)$ est fermé. Pour chaque $n \in \mathbb{N}$, le vecteur y_n appartient à $f(K)$ et donc il existe un $x_n \in K$ tel que $f(x_n) = y_n$. La suite (x_n) ainsi construite est une suite dans le fermé K et possède donc une sous-suite convergente (proposition 7.138). Notons (x'_n) cette sous-suite convergente, et a sa limite : $\lim(x'_n) = a \in K$. Le fait que la limite soit dans K provient du fait que K est fermé.

Nous pouvons considérer la suite $f(x'_n)$ dans W. Cela est une sous-suite de la suite (y_n), et nous avons $\lim f(x'_n) = a$ parce que f est continue. Par conséquent nous avons

$$f(a) = \lim f(x'_n) = \lim y_n. \tag{7.359}$$

Cela prouve que la limite de (y_n) est dans $f(K)$ et par conséquent que $f(K)$ est fermé.

96. La définition est 1.109, mais si vous n'avez pas envie de vous embarquer trop loin, dites juste « toutes les suites de Cauchy convergent ». Typiquement c'est $\mathbb{R}$ ou $\mathbb{C}$.
97. Définition 7.277.

$f(K)$ est borné Si $f(K)$ n'est pas borné, nous pouvons trouver une suite (x_n) dans K telle que

$$\|f(x_n)\|_W > n \tag{7.360}$$

Mais par ailleurs, l'ensemble K étant compact (et donc fermé), nous avons une sous-suite (x'_n) qui converge dans K. Disons $\lim(x'_n) = a \in K$.
Par la continuité de f nous avons alors $f(a) = \lim f(x'_n)$, et donc

$$|f(a)| = \lim |f(x'_n)|. \tag{7.361}$$

La suite $f(x'_n)$ est alors une suite bornée, ce qui n'est pas possible au vu de la condition (7.360) imposée à la suite de départ (x_n).

□

Corolaire 7.300.
Si $f\colon K \to \mathbb{R}$ est une application continue où K est une partie compacte d'un espace vectoriel normé, alors $f(K)$ est borné.

Démonstration. En effet, la proposition 7.299 montre que $f(K)$ est compact et donc borné. □

7.17 Espaces métriques

7.17.1 Espaces métrisables

Définition 7.301.
*Un espace topologique est **métrisable** s'il est homéomorphe à un espace métrique.*

Proposition 7.302.
Une fonction séquentiellement continue sur un espace métrisable et à valeurs dans un espace métrique est continue.

Démonstration. Soient E un espace métrique et $\phi\colon X \to (E, d)$ un homéomorphisme. Nous supposons que $f\colon X \to Y$ est séquentiellement continue. Nous considérons l'application $\tilde{f} = f \circ \phi^{-1}$, c'est-à-dire

$$\begin{aligned} \tilde{f}\colon E &\to Y \\ a &\mapsto f\big(\phi^{-1}(a)\big). \end{aligned} \tag{7.362}$$

L'application ϕ^{-1} est continue et donc séquentiellement continue. De plus $\tilde{f}$ est séquentiellement continue. En effet si $a_k \xrightarrow{E} a$, alors

$$\tilde{f}(a_k) = f\big(\phi^{-1}(a_k)\big), \tag{7.363}$$

mais ϕ^{-1} est séquentiellement continue, donc $\phi^{-1}(a_k) \xrightarrow{X} \phi^{-1}(a)$, ce qui signifie que $\phi^{-1}(a_k)$ est une suite convergente dans X et donc

$$\lim_{k\to\infty} \tilde{f}(a_k) = \lim_{k\to\infty} f\big(\phi^{-1}(a_k)\big) = f\big(\phi^{-1}(a)\big) = \tilde{f}(a). \tag{7.364}$$

L'application $\tilde{f}$ est donc séquentiellement continue. Mais étant donné que $\tilde{f}$ est définie sur un espace métrique (E) et à valeurs dans un métrique, elle est continue par la proposition 7.268. L'application $f = \tilde{f} \circ \phi$ est donc continue en tant que composée d'applications continues. □

7.17.2 Fonctions continues

La propriété suivante donne des caractérisations importantes de la continuité dans le cas des espaces métriques.

Proposition 7.303 (Continuité, ouverts et voisinages et limite[157]).
Soient $f\colon E \to F$ une application entre espaces métriques et $a \in E$. Alors nous avons équivalence entre les choses suivantes :

(1) f est continue en a,

(2) Pour tout voisinage ouvert W de $f(a)$, il existe un voisinage ouvert V de a tel que $f(V) \subset W$.

(3) Pour toute boule $W' = B\big(f(a), \epsilon\big)$, il existe une boule $V' = B(a, \delta)$ telle que $f(V) \subset W$.

(4) $\forall \epsilon > 0$, $\exists \delta > 0$ tel que $f\big(B(a, \delta)\big) \subset B\big(f(a), \epsilon\big)$.

(5) $\lim_{x\to a} f(x) = f(a)$ où la limite est donnée par la définition 7.67,

(6) Pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que $\|x - a\| < \delta$ implique $\|f(x) - f(a)\| < \epsilon$.

La proposition 7.330 nous montrera que ces équivalences tiennent encore lorsque l'espace a une topologie de semi-normes.

Démonstration. L'équivalence (1) $\Leftrightarrow$ (2) est la définition 7.71. L'équivalence (3) $\Leftrightarrow$ (4) est une simple paraphrase.

Montrons (2) $\Rightarrow$ (3). Si $W' = B\big(f(a), \delta\big)$, nous avons un voisinage V de a tel que $f(V) \subset W$. L'ensemble V contenant une boule autour de chacun de ses points[98], il en contient un autour de a : $V' = B(a, \delta) \subset V$. A fortiori nous avons $f(V') \subset W$.

Montrons (3) $\Rightarrow$ (2). Si W est un ouvert autour de $f(a)$, il contient une boule autour de $f(a)$: $B\big(f(a), \epsilon\big) \subset W$. Il existe donc une boule $V' = B(a, \delta)$ telle que $f(V') \subset B\big(f(a), \epsilon\big) \subset W$.

L'équivalence (1) $\Leftrightarrow$ (5) est la définition 7.71 de la continuité en un point couplée à l'unicité de la limite due à la proposition 7.70 parce qu'un espace métrique est séparé.

Prouvons (5) $\Rightarrow$ (6). Soient $\epsilon > 0$ et $V = B\big(f(a), \epsilon\big)$. Étant donné que $f(a)$ est une limite de f pour $x \to a$, il existe un voisinage W de a tel que $f(W) \subset V$. Soit $\delta > 0$ tel que $B(a, \delta) \subset W$; alors si $\|x - a\| < \delta$ nous avons $x \in B(x, \delta) \subset W$ et donc $f(x) \in B\big(f(a), \epsilon\big)$, c'est-à-dire $\|f(a) - f(x)\| < \epsilon$.

Enfin l'implication (2) $\Rightarrow$ (5) est une réécriture de la définition de la limite en un point. □

Voici un théorème qui parle de fermés emboîtés dans un espace métrique. Le corolaire 7.58 parle du cas $\cap_i A_i = \emptyset$ dans un compact.

Théorème 7.304 (Théorème des fermés emboîtés[158]).
Soit (E, d) un espace métrique. Il est complet si et seulement si toute suite décroissante de fermés non vides dont le diamètre tend vers zéro a une intersection qui se réduit à un seul point.

Démonstration. En deux parties.

Condition suffisante Soit $\{F_n\}_{n\in\mathbb{N}}$ une telle suite de fermés emboités. Si nous choisissons des points $x_n \in F_n$, nous obtenons une suite (x_n) de Cauchy et qui est par conséquent convergente vu que l'espace est par hypothèse complet. De plus, pour chaque $N \geqslant n$, la queue de suite $(x_n)_{n\geqslant N}$ est contenue dans F_N et donc converge vers un élément de F_N (parce que ce dernier est fermé). Donc la limite de (x_n) est dans $\bigcap_{n\in\mathbb{N}} F_n$.

De plus cette intersection a diamètre nul parce que le diamètre de $\bigcap_{n\in\mathbb{N}} F_n$ est majoré par tous les diamètres des F_n, lesquels sont arbitrairement petits par hypothèse. Donc l'intersection est réduite a un point.

Condition nécessaire Soit (x_n) un suite de Cauchy. Nous considérons les ensembles

$$F_n = \overline{\{x_i \text{ tel que } i \geqslant n\}}. \tag{7.365}$$

Le fait que la suite soit de Cauchy implique que $\operatorname{diam}(F_n) \to 0$. Par hypothèse, nous avons alors

$$\bigcap_{n\in\mathbb{N}} F_n = \{a\}. \tag{7.366}$$

98. Cela est le théorème-définition 7.94 des ouverts dans un espace métrique, à ne pas confondre avec le théorème 7.4.

Pour s'assurer que a est bien la limite de (x_n), il suffit de remarquer que

$$d(x_n, a) \leqslant \operatorname{diam} F_n \to 0. \tag{7.367}$$

□

Proposition 7.305.
Soient (X, d) un espace topologique métrique et F un fermé de X. Nous avons $d(x, F) = 0$ si et seulement si $x \in F$.

Démonstration. Si $x \in F$ alors $d(x, F) = 0$ parce que $d(x, x)$ fait partie de l'ensemble sur lequel nous prenons l'infimum.

Si réciproquement $d(x, F) = 0$, cela signifie que pour tout ϵ, il existe $x_\epsilon \in F$ tel que $d(x_\epsilon, x) \leqslant \psi$. En prenant $\epsilon = 1/k$ nous construisons une suite (x_k) d'éléments dans F vérifiant $d(x_k, x) = \frac{1}{k}$. Cela signifie que $\lim_{k\to\infty} x_k = x$ par la proposition 7.281(1).

Par la caractérisation séquentielle des fermés (un fermé contient les limites de toutes ses suites, proposition 7.265), la suite (x_k) étant dans F, la limite est dans F. Donc $x \in F$. □

Lemme 7.306.
Soit A_n une suite décroissante de fermés dans un espace métrique[99] *compact K. Alors*

$$C = \bigcap_{n\in\mathbb{N}} A_n \tag{7.368}$$

est non vide.

Démonstration. Soit (x_n) une suite dans K telle que $x_n \in A_n$. La suite étant contenue dans A_1, et A_1 étant compact (lemme 7.64), elle possède une sous-suite $(y_n = x_{\sigma_1(n)})$ convergente dont la limite est dans A_1 par le théorème de Bolzano-Weierstrass 7.103. Une queue de la suite y_n est dans A_2 et nous considérons donc une sous-suite convergente dans A_2 donnée par

$$z_n = y_{\sigma_2(n)} = x_{\sigma_1\sigma_2(n)}. \tag{7.369}$$

En continuant ainsi nous construisons une suite convergente dans A_k. Nous considérons enfin la suite

$$y_n = x_{\sigma_1\ldots\sigma_n(n)}. \tag{7.370}$$

Pour tout k, une queue de cette suite est une sous-suite de $x_{\sigma_1\ldots\sigma_k(n)}$ et par conséquent cette suite converge dans A_k. La limite de cette suite est donc dans l'intersection demandée. □

Remarque 7.307.
Cette propriété est fausse pour les ouverts. Par exemple

$$\bigcap_{n>1}]0, \frac{1}{n}[= \varnothing. \tag{7.371}$$

Lemme 7.308.
Si K est un compact dans un espace métrique et F un fermé disjoint de K, alors $d(K, F) > 0$.

Démonstration. La fonction

$$\begin{aligned} K &\to \mathbb{R} \\ x &\mapsto d(x, F) \end{aligned} \tag{7.372}$$

est une fonction continue sur K, et donc atteint son minimum par le théorème de Weierstrass 7.105. Soit $x_0 \in K$ un point de K qui réalise ce minimum. Si $d(x_0, F) = 0$, alors on aurait une suite (x_n) dans F qui convergerait vers x_0, mais F étant fermé cela signifierait que x_0 serait dans F, ce qui contredirait l'hypothèse que F et K sont disjoints. □

99. L'hypothèse métrique provient de l'utilisation de Bolzano-Weierstrass, lequel est vrai pour les espaces séquentiellement compacts, dont les espaces métriques.

Proposition 7.309 ([129]).
Une isométrie d'un espace métrique compact sur lui-même est une bijection.

Démonstration. Soient X un espace métrique compact et $f\colon X \to X$ une isométrie. Le fait que f soit injective est obligatoire (sinon il y a des images dont la distance est nulle). Il faut montrer que f est surjective.

Soit $x \in X$ hors de $f(X)$. Le lemme 7.308 appliqué au fermé $\{x\}$ et au compact $f(K)$ donne un $r > 0$ tel que

$$d\big(x, f(K)\big) > r. \tag{7.373}$$

Soit la suite $u_n = f^n(x)$; c'est une suite dans K et possède donc une sous-suite convergente (Bolzano-Weierstrass7.103) que l'on nomme (y_n). Vu que f est une isométrie,

$$d(y_n, y_{n+1}) = d(x, y_m) > r \tag{7.374}$$

pour un certain $m \leqslant n+1$. Cela signifie que pour tout n, nous avons $d(y_n, y_{n+1}) > r$, ce qui contredit le fait que la suite (y_n) converge. □

Proposition 7.310.
Soient (X, d) un espace métrique compact et (u_n) une suite de X telle que

$$\lim_{n\to\infty} d(u_n, u_{n+1}) = 0. \tag{7.375}$$

Alors l'ensemble des points d'accumulation[100] *de (u_n) est connexe.*

Démonstration. Nous notons Γ l'ensemble des points d'accumulation de la suite.

Γ est compact Nous notons $A_p = \{u_n \text{ tel que } n \geqslant p\}$ et nous avons

$$\Gamma = \bigcap_{p\in\mathbb{N}} \overline{A_p} \tag{7.376}$$

parce que si $x \in \Gamma$, alors pour tout n, il existe $m > n$ tel que $x_m \in B(x, \epsilon)$, et donc tel que $x \in B(x_m, \epsilon)$. Donc pour tout ϵ et pour tout p, l'intersection $B(x, \epsilon) \cap A_p$ est non vide.

En tant qu'intersection de fermés, Γ est fermé (lemme 7.3). En tant que fermé dans un compact, Γ est compact (lemme 7.64).

Recouvrement par deux compacts Supposons que Γ ne soit[101] pas connexe. Nous pouvons alors considérer S et O, deux ouverts disjoints recouvrant Γ et intersectant tout deux Γ. Nous posons alors

$$A = S \cap \Gamma \tag{7.377a}$$
$$B = O \cap \Gamma, \tag{7.377b}$$

et nous avons évidemment $\Gamma = A \cup B$. Montrons que A est fermé (B le sera aussi par le même raisonnement). Soit une suite d'éléments de $S \cap \Gamma$ convergent dans X. Alors la limite est dans $\bar{\Gamma} = \Gamma$ et donc elle est donc O ou S, mais elle est certainement dans $\bar{S}$. Cependant $\bar{S}$ n'intersecte pas O. En effet si $x \in \bar{S} \cap O$, alors tout voisinage de x intersecterait S, mais il y a des voisinages de x étant inclus dans O parce que O est ouvert ; cela donnerait une intersection entre O et S, ce qui est impossible. Donc la limite n'est pas dans O et donc elle est dans S. Au final la limite est dans $S \cap \Gamma$, ce qui prouve son caractère fermé.

Comme d'habitude, $\Gamma \cap S$ est compact parce que fermé dans un compact[102].

100. Définition 7.27.
101. est-ce qu'il faut vraiment un subjonctif ici ?
102. Lemme 7.64.

Décomposition en trois morceaux Vu que A et B sont des compacts disjoints, nous avons $d(A,B)=\alpha>0$ pour un certain α par le lemme 7.308. Nous notons

$$A'=\{x\in X \text{ tel que } d(x,A)<\frac{\alpha}{3}\} \tag{7.378a}$$

$$B'=\{x\in X \text{ tel que } d(x,B)<\frac{\alpha}{3}\} \tag{7.378b}$$

Nous avons $A'=\bigcup_{x\in A}B(x,\frac{\alpha}{3})$ et donc en tant qu'union d'ouverts, A' est ouvert (définition de la topologie). Même chose pour B'.

Enfin nous notons

$$K=X\backslash(A'\cup B') \tag{7.379}$$

qui est fermé en tant que complémentaire d'ouvert, et donc compact. Étant donné que $A\subset A'$ et $B\subset B'$, nous avons $K\cap\Gamma=\emptyset$.

L'idée est maintenant de montrer que K contient un point d'accumulation de (u_n).

Sous-suites de (u_n) L'hypothèse sur la suite (u_n) nous indique qu'il existe un N_0 tel que $\forall n\geqslant N_0$,

$$d(u_n,u_{n+1})<\frac{\alpha}{3}. \tag{7.380}$$

Soient $N>N_0$ et $x_0\in A$. Étant donné que x_0 est point d'accumulation de la suite, il existe $n_1>N$ tel que $d(x_0,u_{n_1})<\frac{\alpha}{3}$. Même chose dans B : nous prenons $y_0\in B$ et un naturel $n_2>n_1$ tel que $d(y_0,u_{n_2})<\frac{\alpha}{3}$. Nous avons $u_{n_1}\in A'$ et $u_{n_2}\in B'$.

Soit n_0 le plus petit naturel supérieur à n_1 tel que $u_{n_0}\notin A'$. Cela existe parce que $u_{n_2}\in B'$ et $B'\cap A'=\emptyset$, mais n_0 n'est pas n_2 lui-même parce que $d(A',B')\geqslant\frac{\alpha}{3}$ alors que nous considérons $n_0,n_1,n_2>N_0$ et donc pour tous les i entre n_1 et n_2 (compris), $d(u_i,u_{i+1})<\frac{\alpha}{3}$. Notons qu'ici le strict dans la condition (7.380) est important. Nous avons donc $N_0<n_1<n_0<n_2$.

Nous allons maintenant montrer que u_{n_0} est dans K. C'est fait pour : il est loin en même temps de A' et de B'. En utilisant l'inégalité triangulaire à l'envers, nous avons

$$\begin{aligned}d(u_{n_0},B)&\geqslant d(u_{n_0-1},B)-d(u_{n_0-1},u_{n0})\\&\geqslant d(A,B)-d(u_{n_0-1},A)-d(u_{n_0-1},u_{n_0})\\&\geqslant\alpha-\frac{\alpha}{3}-\frac{\alpha}{3}\\&=\frac{\alpha}{3}.\end{aligned} \tag{7.381}$$

Pour la dernière inégalité nous avons utilisé le fait que u_{n_0-1} n'est pas dans A'. Bref, nous avons montré que u_{n_0} n'est pas dans B' (dans la définition de ce dernier nous avons bien une inégalité stricte). Vu que par définition u_{n_0} n'est pas non plus dans A', nous avons $u_{n_0}\in K$.

Nous avons montré jusqu'à présent que pour tout $N\geqslant N_0$, il existe un $n_0\geqslant N$ tel que $u_{n_0}\in K$. Cela nous construit donc une sous-suite (v_n) de (u_n) contenue dans K. En tant que suite dans le compact K, la suite (v_n) admet un point d'accumulation dans K. Ce point est également point d'accumulation de la suite (u_n) complète, ce qui donne un point d'accumulation de (u_n) dans K et donc une contradiction.

Nous concluons que Γ est connexe. □

Encore une petite conséquence sans ambition du théorème de Bolzano-Weierstrass.

Proposition 7.311.
Si (x_n) est une suite dans un compact telle que toute sous-suite convergente ait le même point x comme limite. Alors la suite entière converge vers x.

Démonstration. Supposons que ce ne soit pas le cas. Alors il existe un ϵ tel que pour tout $N>0$, il existe $n>N$ avec $d(x_n,x)>\epsilon$. Cela nous donne une sous-suite de (x_n) composée d'éléments

tous à une distance de x supérieure à ϵ. Nous la nommons (y_n) ; c'est une suite dans un compact qui admet donc une sous-suite convergente (et une telle sous-suite est une sous-suite de (x_n)) dont la limite devrait être x, mais c'est impossible par construction. □

Lemme-définition 7.312.
Soi Ω un ouvert dans un espace métrique E. Il existe une suite (K_n) de compacts tels que

(1) $K_n \subset \Omega$
(2) $\bigcup_{n=0}^{\infty} K_n = \Omega$
(3) $K_n \subset \mathrm{Int}(K_{n+1})$.

Une telle suite de compacts vérifie alors

(1) Il existe δ_n tel que pour tout $z \in K_n$, $B(z, \delta_n) \subset K_{n+1}$.
(2) Tout compact de Ω est inclus dans $\mathrm{Int}(K_n)$ pour un certain n.

*Une telle suite de compacts est une **suite exhaustive** de compacts pour Ω.*

Démonstration. Nous considérons les ensembles

$$V_n = \{z \in E \text{ tel que } |z|\} \cup \bigcup_{a \notin \Omega} B(a, \frac{1}{n}), \tag{7.382}$$

et nous définissons $K_n = \complement V_n$. Vérifions que ces ensembles vérifient tout ce qu'il faut.

(1) Si $a \notin \Omega$ alors a est dans tous les V_n et donc dans aucun des K_n ; nous avons donc bien $K_n \subset \Omega$.
(2) Si $z \in \Omega$ alors nous prenons $n_1 > |z|$ puis n_2 tel que $B(z, \frac{1}{n_2}) \subset \Omega$. Alors $z \in K_n$ avec $n > \max(n_1, n_2)$.
(3) Une chose à comprendre est que si $z \in K_n$, alors $d(z, \complement\Omega) \geqslant \frac{1}{n}$. Du coup si nous prenons δ tel que
$$\frac{1}{n+1} < \delta < \frac{1}{n} \tag{7.383}$$
alors $B(z, \delta) \subset K_{n+1}$.
(4) Enfin, les K_n sont tous compacts. En effet ils sont bornés parce que $K_n \subset B(0, n)$ et ensuite K_n est fermé en tant que complémentaire d'un ouvert (V_n est ouvert en tant qu'union d'ouverts).

Nous passons maintenant aux propriétés, qui sont indépendantes de la façon dont nous avons construit les K_n vérifiant les conditions.

(1) Nous pouvons considérer la fonction $K_n \to \mathbb{R}$ donnée par $z \mapsto d(z, \complement K_{n+1})$. Vu que $K_n \subset \mathrm{Int}(K_{n+1})$, c'est une fonction (continue sur le compact K_n) prenant des valeurs strictement positives. Elle a donc un minimum strictement positif. Si δ_n est plus petit que ce minimum nous avons $B(z, \delta_n) \subset K_{n+1}$ pour tout $z \in K_n$.
(2) D'abord nous avons $\Omega = \bigcup_{n=0}^{\infty} \mathrm{Int}(K_n)$. En effet nous avons
$$\Omega = \bigcup_{n=0}^{\infty} K_n \subset \bigcup_{n=0}^{\infty} \mathrm{Int}(K_{n+1}) \subset \bigcup_{n=0}^{\infty} \mathrm{Int}(K_n). \tag{7.384}$$
L'inclusion dans l'autre sens est facile.
Soit K compact dans Ω. Vu que Ω est l'union des $\mathrm{Int}(K_n)$, nous avons
$$K \subset \bigcup_{n=0}^{\infty} \mathrm{Int}(K_n). \tag{7.385}$$
Cela donne à K un recouvrement par des ouverts dont nous pouvons extraire un sous-recouvrement fini par compacité. Les K_n étant croissants, du recouvrement fini, il suffit de prendre le plus grand (disons K_m) et nous avons $K \subset \mathrm{Int}(K_m)$.

□

Notons qu'avec la suite de K_n telle que construite, le dernier point est réglé en prenant

$$\frac{1}{n+1} < \delta_n < \frac{1}{n}. \tag{7.386}$$

Théorème 7.313 (Tykhonov)**.**
Un produit quelconque d'espaces métriques non vides est compact si et seulement si chacun de ses facteurs est compact.

Nous n'allons donner la preuve que dans le cas d'un produit fini dans le théorème 7.318.

7.17.3 Ensembles enchaînés

Soit (x, d) un espace métrique.

Définition 7.314.
Une ***ϵ-chaine*** *joignant les points a et b de X est une suite finie $(u_0, \dots, u_n)$ dans X telle que $u_0 = a$, $u_n = b$ et pour tout $0 \leqslant i \leqslant n-1$ nous avons $d(u_n, u_{n+1}) \leqslant \epsilon$.*

Une partie A de X est ***bien enchaînée*** *si pour tout $\epsilon > 0$ et pour tout $a, b \in A$, il existe une ϵ-chaine joignant a et b dans A.*

Les rationnels dans $\mathbb{R}$ sont bien enchaînés.

Proposition 7.315.
Un espace connexe est bien enchaîné.

Proposition 7.316.
La fermeture d'un ensemble bien enchaîné dans un espace métrique compact (X, d) est connexe.

Démonstration. Soit $A \subset X$ un ensemble bien enchaîné, et soient $a, b \in \bar{A}$. Nous construisons une suite (u_k) dans A de la façon suivante. Pour chaque $n > 0$ nous prenons $a' \in B(a, \frac{1}{n}) \cap A$ et $b' \in B(b, \frac{1}{n}) \cap A$. Ensuite nous considérons une $\frac{1}{n}$-chaine $\{v_i^{(n)}\}_{i \in I_n}$ dans A entre a' et b'. Ici l'ensemble I_n est fini. La suite (u_k) est simplement construite en mettant bout à bout les éléments $v_i^{(n)}$.

La suite ainsi construite est une suite dans A admettant a et b comme points d'accumulation (les autres points d'accumulation sont également dans $\bar{A}$) et telle que $\lim_{k \to \infty} d(u_k, u_{k+1}) = 0$. Par conséquent la proposition 7.310 nous dit que l'ensemble des points d'accumulation de (u_k) est connexe dans X. Nous le notons $C_{a,b}$.

Si nous fixons $a \in \bar{A}$, alors nous avons

$$\bigcup_{x \in \bar{A}} C_{a,x} = \bar{A}. \tag{7.387}$$

Vu que le membre de gauche est une union de connexes, c'est un connexe par la proposition 7.46. □

Corolaire 7.317.
Un espace métrique compact est connexe si et seulement s'il est bien enchaîné.

7.17.4 Produit fini d'espaces métriques

Pour rappel, la distance sur un espace produit est donnée par la définition 7.229.

Théorème 7.318 ([1])**.**
Un produit fini d'espaces métriques non vides est compact si et seulement si chacun de ses facteurs est compact.

Démonstration. Soient $K_1,\dots,K_n$ des compacts et $K = K_1 \times \dots \times K_n$ le produit muni de sa métrique usuelle de la définition (7.229) (attention : chacun des K_i peut être de dimension infinie) :

$$d(\alpha,\beta) = \max\{d_i(\alpha_i,\beta_i)\} \tag{7.388}$$

où d_i est la distance sur K_i. Si (α_n) est une suite dans K alors la suite $(\alpha_n)_1$ est une suite dans le compact K_1 dont nous pouvons extraire une sous-suite convergente (Bolzano-Weierstrass 7.103). De la sous-suite de α correspondante nous extrayons la sous-suite pour la seconde composante, etc.

En fin de compte nous avons une sous-suite (que nous nommons α également) donc chacune des composantes est convergente. Nous nommons ℓ_k les limites correspondantes. Soit $\epsilon > 0$ pour chaque $k = 1,\dots,n$, il existe $N_k > 0$ tel que si $p > N_k$ alors

$$d\big((\alpha_p)_k - \ell_p\big) \leqslant \epsilon. \tag{7.389}$$

Ici $\alpha_p \in K$ est le p^{e}élément de la suite α et $(\alpha_p)_i \in K_i$ est la i^{e}composante de α_p. En prenant $N = \max_k N_k$ et $n > N$ nous avons

$$d\big(\alpha_n,(\ell_1,\dots,\ell_n)\big) \leqslant \epsilon. \tag{7.390}$$

Par conséquent de la suite (α) nous avons extrait une sous-suite convergente et la partie « réciproque » de Bolzano-Weierstrass nous assure alors que K est compact.

À l'inverse si un des facteurs n'est pas compact (mettons K_1) alors nous prenons un recouvrement $\{\mathcal{O}_i\}_{i\in I}$ de K_1 par des ouverts duquel il est impossible d'extraire un sous-recouvrement fini. Ensuite nous posons

$$\mathcal{P}_i = \mathcal{O}_i \times K_2 \times \dots \times K_n, \tag{7.391}$$

qui est un recouvrement de K par des ouverts (de K) d'où aucun sous-recouvrement fini ne peut être extrait. □

Pour la culture générale, il y a bien entendu moyen de faire des produits dénombrables et pire d'espaces métriques.

Définition 7.319 ([159]).
Soient (E_n,d_n) des espaces métriques. Sur l'ensemble produit $E = \prod_{i=1}^{\infty} E_i$ nous définissons la métrique

$$d(x,y) = \sum_{k=1}^{\infty} \frac{1}{2^k} d'_k(x_i,y_i) \tag{7.392}$$

où $d'_i = \min(d_i,1)$.

On peut montrer que ce d est bien une distance et que (E,d) devient un espace métrique.

Théorème 7.320 (Tykhonov dénombrable[159]).
Un produit dénombrable d'espaces métriques non vides est compact si et seulement si chacun de ses facteurs est compact.

Note : ce résultat est encore valable pour un produit quelconque, c'est le théorème de Tykhonov 7.313.

7.17.5 Équicontinuité

Définition 7.321 ([160]).
*Soit une famille de fonctions $f_i\colon X \to E$ indexée par un ensemble I où X est un espace topologique et E un espace métrique. Cette famille est **équicontinue** en $x \in X$ si pour tout $\epsilon > 0$, il existe un voisinage V de x tel que*

$$\|f_i(x) - f_i(y)\| < \epsilon \tag{7.393}$$

pour tout i dès que $x,y \in V$.

Nous disons qu'une famille est équicontinue sans préciser en quel point si elle est équicontinue en tout pout.

La proposition suivante permet de montrer que certaines fonctions définies par une limite sont continues. Ce sera par exemple le cas de la fonction puissance, proposition 11.349.

Proposition 7.322 ([1, 160]).
Soit une suite équicontinue (f_i) de fonctions qui converge simplement vers f, alors f est continue.

Démonstration. Soit une suite équicontinue $f_i\colon X \to E$ convergeant simplement vers f. Soit $a \in X$. Nous prouvons que f est continue en a.. Pour cela nous considérons $\epsilon > 0$ et, conformément à l'hypothèse équicontinuité un voisinage V de a tel que $|f_i(a) - f_i(x)| < \epsilon$ pour tout $x \in V$.

Nous avons la majoration

$$|f(x) - f(a)| \leqslant |f(x) - f_i(x)| + |f_i(x) - f_i(a)| + |f_i(a) - f(a)|. \tag{7.394a}$$

Plusieurs majorations.

— Vu que $f_i \to f$, il existe N_1 tel que $|f(x) - f_i(x)| < \epsilon$ pour tout $i > N_1$.

— De plus, par définition de V, nous avons aussi $|f_i(x) - f_i(a)| \leqslant \epsilon$.

— Vu que $f_i \to f$, il existe N_2 tel que $|f_i(a) - f(a)| < \epsilon$ pour tout $i > N_2$.

Donc en prenant $x \in V$ et $i > \max\{N_1, N_2\}$ nous avons

$$|f(x) - f(a)| \leqslant 3\epsilon. \tag{7.395}$$

□

7.17.6 Continuité uniforme

Définition 7.323 ([161]).
*Soient deux espaces métriques (E, d) et (E', d'). Une application $f\colon E \to E'$ est **uniformément continue** si pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que $d(x, y) \leqslant \delta$ implique $d'\big(f(x), f(y)\big) \leqslant \epsilon$.*

Dans l'uniforme continuité, le α qui fait fonctionner ϵ doit le faire fonctionner pour tous les $x, y \in E$. C'est la différence avec la continuité simple dans laquelle nous pouvons choisir, pour un même ϵ, un δ différent en chaque point.

7.18 Ensembles nulle part denses

Nous allons nous limiter au cas de $\mathbb{R}$, mais je crois que ça se généralise sans trop de peine aux espaces métriques, voire plus. Voir aussi la section 7.20 sur les espaces de Baire.

Définition 7.324.
*Un ensemble est dit **nulle part dense** s'il n'est dense dans aucun intervalle.*

*Un ensemble dans $\mathbb{R}$ est de **première catégorie** ou **maigre** s'il est une union dénombrable d'ensembles nulle part dense (c'est-à-dire d'ensembles denses sur aucun intervalle).*

Théorème 7.325 (Baire[162]).
Une réunion dénombrable d'ensembles nulle part denses est d'intérieur vide.

Démonstration. Soient $a \in S$ et $\epsilon > 0$. Nous allons trouver un élément dans $B(a, \epsilon)$ qui n'est pas dans S. Nous commençons par choisir $x_1 \in B(a, \epsilon)$ et $r_1 < \frac{\epsilon}{2}$ tel que

$$B(x_1, r_1) \cap A_1 = \varnothing. \tag{7.396}$$

Ensuite nous choisissons $x_2 \in B(x_1, r_1)$ et $r_2 < \epsilon/4$ tel que $B(x_2, r_2) \subset B(x_1, r_1)$ et $B(x_2, r_2) \cap A_2 = \varnothing$. Notons que $B(x_2, r_2) \cap A_1 = \varnothing$ aussi, par construction.

Par récurrence nous construisons une suite d'éléments x_n et de rayons $r_n < \epsilon/2^n$ tels que

(1) $B(x_n, r_n) \cap A_j = \varnothing$ pour tout $j \leqslant n$,

(2) $\overline{B(x_n, r_n)} \subset B(x_{n-1}, r_{r-1})$.

Cette suite étant de Cauchy (parce que contenue dans des intervalles emboîtés de rayon décroissant vers zéro), elle converge [103] donc vers un point qui en particulier appartient à $B(a, \epsilon)$. Mais la limite n'est dans aucun des A_n et donc pas dans S. □

7.19 Topologie des semi-normes

Les principaux espaces topologiques construit avec des semi-normes seront les espaces de fonctions de la définition 29.10. Nous verrons également la topologie $*$-faible sur $\mathscr{D}'(\Omega)$ en la définition 29.17.

Définition 7.326.
*Si E est un espace vectoriel, une **semi-norme** sur E est une application $p\colon E \to \mathbb{R}$ telle que*

(1) $p(x) \geqslant 0$,

(2) $p(\lambda x) = |\lambda| p(x)$

(3) $p(x + y) \leqslant p(x) + p(y)$.

La seule différence avec une norme est simplement qu'une semi-norme peut s'annuler en des éléments non-nuls de l'espace.

Lemme 7.327 ([163])**.**
Si p est une semi-norme nous avons

$$|p(x) - p(y)| \leqslant p(x - y). \tag{7.397}$$

Démonstration. Nous avons d'une part $p(x + h) \leqslant p(x) + p(h)$ et d'autre part $p(x) \leqslant p(x + h) + p(-h) = p(x + h) + p(h)$. En isolant $p(x + h) - p(x)$ dans chacune ce des deux inégalités,

$$-p(h) \leqslant p(x + h) - p(x) \leqslant p(h) \tag{7.398}$$

ou encore

$$|p(x + h) - p(x)| \leqslant p(h) \tag{7.399}$$

qui donne le résultat demandé en posant $h = y - x$. □

Soit $(p_i)_{i \in I}$ une famille de semi-normes sur E. Nous construisons alors une topologie sur E de la façon suivante.

Définition 7.328 (Topologie et semi-normes[164, 165])**.**
*Pour tout J fini dans I nous définissons les **boules ouvertes***

$$B_J(x, r) = \{y \in E \text{ tel que } p_j(y - x) < r \, \forall j \in J\}. \tag{7.400}$$

*La **topologie** sur E donnée par la famille de semi-norme est définie en disant que $\mathcal{O} \subset E$ est ouvert si et seulement si chaque point de $\mathcal{O}$ est dans une boule contenue dans $\mathcal{O}$.*

Proposition 7.329.
Une suite (x_n) dans E converge vers x au sens de la topologie des semi-normes si et seulement si pour tout $i \in I$,

$$p_i(x - x_n) \to 0. \tag{7.401}$$

Démonstration. Si la suite (x_n) converge [104] vers x, alors pour tout ouvert $\mathcal{O}$ autour de x, il existe un N tel que si $n \geqslant N$, alors $x_n \in \mathcal{O}$. En particulier pour tout j et pour tout $\epsilon > 0$, il doit exister un $n \geqslant N_j$ tel que $x_n \in B_j(x, \epsilon)$.

103. Par la proposition 1.115
104. Définition 7.10.

Voyons l'implication inverse. Soit $\epsilon > 0$. Pour tout $i \in I$, il existe un N_i tel que $n \geqslant N_i$ implique $p_i(x - x_n) \leqslant \epsilon$. Si $\mathcal{O}$ est un ouvert, il doit contenir une boule du type $B_J(x, r)$ pour un certain ensemble fini $J \subset I$.

En prenant $N = \max\{N_j \text{ tel que } j \in J\}$, nous avons $p_j(x - x_n) \leqslant \epsilon$ pour tout j et donc $x_n \in B_J(x, r)$. □

La proposition suivante est un vulgaire plagiat de la proposition 7.303.

Proposition 7.330.
Soit $f\colon \mathbb{R} \to (E, p_i)_{i\in I}$ une application. Nous avons équivalence entre

(1) la fonction f est continue en $t_0 \in \mathbb{R}$,

(2) si W est un voisinage ouvert de $f(t_0)$ il existe un voisinage ouvert V de t_0 (dans $\mathbb{R}$) tel que $f(V) \subset W$,

(3) pour tout $i \in I$ et $\epsilon > 0$ il existe $\delta > 0$ tel que

$$f\big(B(t_0, \delta)\big) \subset B_i\big(f(t_0), \epsilon\big). \tag{7.402}$$

Démonstration. L'équivalence (1) ⇔ (2) est la définition 7.71.

Prouvons (2) ⇒ (3). Soient $i \in I$ et $\epsilon > 0$. Considérons la boule $B_i\big(f(t_0), \epsilon\big)$, qui est un ouvert de E contenant $f(t_0)$. Il existe donc un ouvert V autour de t_0 tel que $f(V) \subset B_i\big(f(t_0), \epsilon\big)$. En particulier V contient une boule $B(t_0, \delta)$ et nous avons

$$f\big(B(t_0, \delta)\big) \subset f(V) \subset B_i\big(f(t_0), \epsilon\big). \tag{7.403}$$

Prouvons (3) ⇒ (2). Soit W un ouvert autour de $f(t_0)$. Il existe un $i \in I$ et $\epsilon > 0$ tel que $B_i\big(f(t_0), \epsilon\big) \subset W$. Nous avons alors un $\delta > 0$ tel que

$$f\big(B(t_0, \delta)\big) \subset B_i\big(f(t_0), \epsilon\big) \subset W. \tag{7.404}$$

□

Lorsqu'on a un espace E muni d'une quantité dénombrable de semi-normes $\{p_k\}_{k\in I}$ nous définissons l'écart [105]

$$d(x, y) = \sup_{k\geqslant 1} \min\Big\{\frac{1}{k}, p_k(x - y)\Big\}. \tag{7.405}$$

Notons que cet écart est invariant par translation au sens où pour tout x, y, h dans E nous avons

$$d(x + h, y + h) = \sup_{k\geqslant 1} \min\Big\{\frac{1}{k}, p_k(x - y)\Big\} = d(x, y). \tag{7.406}$$

Proposition 7.331.
Si X est un espace topologique dont la topologie est donnée par une famille dénombrable de semi-normes, alors il est métrisable.

Proposition 7.332 ([163])**.**
La topologie donnée par les boules

$$B_k(a, r) = \{x \in E \text{ tel que } \forall\, k \leqslant \frac{1}{r}, p_k(x - a) < r\} \tag{7.407}$$

est la même que celle « usuelle » donnée par les semi-normes. En disant « la même » nous entendons le fait que les ouverts sont les mêmes : A est ouvert pour une des deux topologies si et seulement s'il est ouvert pour l'autre.

105. Dans le cas de $E = \mathscr{D}(K)$, la première semi-norme est numérotée à zéro, donc il faudra poser $d(\varphi_1, \varphi_2)$ avec p_{k-1} au lieu de p_k.

Démonstration. Pour cette démonstration nous allons préfixer par d les notions topologiques issues des boules (7.407) et par P celle des semi-normes : P-continue, d-ouvert, etc.

D'abord nous avons

$$B(a,r) = \bigcap_{k\leqslant\frac{1}{k}} B_k(a,r). \tag{7.408}$$

Si $\mathcal{O}$ est un d-ouvert, il contient une d-boule autour de chacun de ses points. Or d'après la formule (7.408), une d-boule est une intersection *finie* de P-ouverts et donc est un P-ouvert par définition. Donc $\mathcal{O}$ contient un P-ouvert autour de tous ses points et est donc P-ouvert.

Inversement nous supposons que $\mathcal{O}$ est un P-ouvert. Commençons par prouver que les semi-normes p_k sont d-continues. En effet soient $k\in\mathbb{N}$, $\epsilon\leqslant\frac{1}{k}$ et $x,y\in E$ tels que $d(x,y)\leqslant\epsilon$; nous avons

$$\begin{aligned}
|p_k(y)-p_k(x)| &\leqslant p_k(x-y) && (7.409a)\\
&= \min\{\frac{1}{k},p_k(x-y)\} && (7.409b)\\
&\leqslant d(x,y) && (7.409c)\\
&\leqslant \epsilon. && (7.409d)
\end{aligned}$$

Montrons à présent que $\mathcal{O}$ est d-ouverte. Si $a\in\mathcal{O}$, il existe k et r tels que $B_k(a,r)\subset\mathcal{O}$. Soit $x\in B_k(a,r)$. Montrons que si ϵ est suffisamment petit, la d-boule $B(x,\epsilon)$ est inclue à $B_k(a,r)$. Pour cela prenons $y\in B(x,\epsilon)$; nous avons

$$\big|p_k(a-x)-p_k(a-y)\big|\leqslant d(x,y)\leqslant\epsilon. \tag{7.410}$$

Par conséquent le nombre $p_k(a-y)$ est dans l'intervalle

$$p_k(a-x)\pm\epsilon \tag{7.411}$$

et il suffit de prendre $\epsilon<\frac{r-p_k(a-x)}{2}$. □

7.19.1 Espace dual

Nous parlerons plus en détail d'espace dual d'un espace normé en la section 10.9.

Définition 7.333.
Soient F un espace métrique et E un espace topologique vectoriel. Une topologie possible[106] *sur l'espace des applications linéaires $\mathcal{L}(E,F)$ est la* ***topologie $*$-faible*** *qui est la topologie des semi-normes*

$$p_v(T)=\|T(v)\|_F. \tag{7.412}$$

C'est une famille de semi-normes indicées par les éléments de E. Si E est un espace métrique, c'est cette topologie qui sera considérée sur son dual topologique E' des applications continues $E\to\mathbb{R}$.

La proposition suivante indique qu'elle est un peu la topologie de la convergence ponctuelle.

Proposition 7.334.
Soient E un espace muni de la topologie des semi-normes $\{p_i\}_{i\in I}$ et F un espace métrique. Soient une suite (T_n) dans $\mathcal{L}(E,F)$ et $T\in\mathcal{L}(E,F)$. Nous avons $T_n\xrightarrow{*}T$ si et seulement si $T_n(v)\xrightarrow{F}T(v)$ pour tout $v\in E$.

Démonstration. Nous avons équivalence entre les lignes suivantes :

$$\begin{aligned}
T_n &\xrightarrow{*} T && && (7.413a)\\
p_v(T_n-T)&\to 0\,\forall v\in E && \text{proposition 7.329} && (7.413b)\\
\|T_n(v)-T(v)\|_E&\to 0\,\forall v\in E && && (7.413c)\\
T_n(v)&\xrightarrow{E} T(v). && && (7.413d)
\end{aligned}$$

106. C'est, dans l'idée, celle qui sera choisie pour les espaces de distributions, voir la définition 29.17.

□

7.19.2 Espace $C^k(\mathbb{R}, E')$

Nous revenons à nos histoires de limites de la définition 7.10.

Proposition 7.335 (Unicité de la limite dans un dual topologique)**.**
Soient E un espace métrique et E' son dual topologique muni de sa topologie de la définition 7.333. Il y a unicité de l'élément de E' vers lequel une fonction $u\colon \mathbb{R} \to E'$ peut converger.

Démonstration. Soit T un élément vers lequel u_t converge lorsque $t \to t_0$. Soient $\epsilon > 0$ et $x \in E$. La boule $B_x(T, \epsilon)$ de E' subordonnée à la norme p_x et centrée en T est un ouvert de E'. Étant donné que u converge vers T il existe $\delta > 0$ tel que $u_t \in B_x(T, \epsilon)$ dès que $|t - t_0| \leqslant \delta$. Nous avons donc, pour tout $x \in E$, la limite (dans $\mathbb{R}$) :

$$\lim_{t \to t_0} u_t(x) = T(x). \tag{7.414}$$

Cela prouve que la convergence de u vers T implique l'existence pour tout x de la limite de $u_t(x)$ dans $\mathbb{R}$. Si T' est un autre élément vers lequel u_t converge, nous avons par le même raisonnement que

$$\lim_{t \to t_0} u_t(x) = T'(x). \tag{7.415}$$

Par unicité de la limite dans $\mathbb{R}$ nous devons alors avoir $T(x) = T'(x)$ pour tout x, c'est-à-dire $T = T'$. □

Proposition 7.336.
Soit $u\colon \mathbb{R} \to E'$ une fonction continue. Alors

(1) pour tout $x \in E$ la fonction $t \mapsto u_t(x)$ est continue,

(2) pour tout $x \in E$ nous avons la limite dans $\mathbb{R}$

$$\lim_{t \to t_0} u_t(x) = u_{t_0}(x), \tag{7.416}$$

(3) nous avons la limite dans E'

$$\lim_{t \to t_0} u_t = u_{t_0}. \tag{7.417}$$

Démonstration. Soient $x \in E$ et $\epsilon > 0$. Par la proposition 7.330 la continuité de u donne un $\delta > 0$ tel que

$$u_{B(t_0,\delta)} \subset B_x(u_{t_0}, \epsilon). \tag{7.418}$$

C'est-à-dire que si $|t - t_0| \leqslant \delta$ nous avons

$$\big|u_{t_0}(x) - u_t(x)\big| < \epsilon, \tag{7.419}$$

ce qui signifie bien que la fonction $t \mapsto u_t(x)$ est continue en tant que fonction $\mathbb{R} \to \mathbb{R}$. Cela est le point (1). Le théorème de limite et continuité dans $\mathbb{R}$ nous donne immédiatement la limite (7.416).

Nous passons à la preuve du point (3). Soit $\mathcal{O}$ un ouvert de E' contenant u_{t_0}. Il existe donc un $i \in I$ et $\epsilon > 0$ tel que $B_i(u_{t_0}, \epsilon) \subset \mathcal{O}$. Étant donné que u est continue, il existe $\delta > 0$ tel que

$$u_{B(t_0,\delta)} \subset B_i(u_{t_0}, \epsilon) \subset \mathcal{O}. \tag{7.420}$$

Cela signifie bien que

$$|t - t_0| \leqslant \delta \Rightarrow u_t \in \mathcal{O}, \tag{7.421}$$

c'est-à-dire que nous avons la limite $\lim_{t \to t_0} u_t = u_{t_0}$ dans E'. Pour dire cela nous avons utilisé la définition 7.67 de la limite et le résultat d'unicité 7.335. □

Définition 7.337.
*Si nous avons une application $u\colon \mathbb{R} \to E'$ nous considérons sa **dérivée** donnée par la limite*

$$u'_{t_0} = \lim_{t\to t_0} \frac{u_t - u_{t_0}}{t - t_0}. \tag{7.422}$$

Cela est un nouvel élément de E' (pour peu que la limite existe). La fonction $u'\colon \mathbb{R} \to E'$ ainsi définie peut être continue ou non. Cela nous permet de définir les espaces $C^k(\mathbb{R}, E')$ et $C^\infty(\mathbb{R}, E')$.

Une des principales utilisations que nous ferons de ces espaces seront les espaces de fonctions à valeurs dans les distributions tempérées dont nous parlerons dans la section 29.5.

7.20 Espaces de Baire

Définition 7.338.
*Un **espace de Baire** est un espace topologique dans lequel toute intersection dénombrable d'ouverts denses est dense.*

Théorème 7.339 (Théorème de Baire[166])**.**
Les espaces suivants sont de Baire :

(1) les espaces topologiques localement compacts,

(2) les espaces métriques complets (donc ceux de Banach en particulier),

(3) tout ouvert d'un espace de Baire.

Démonstration. **Espaces topologiques localement compacts**

Espaces métriques complets Soit (E, d) un espace métrique complet. Soient V un ouvert quelconque de E et U_n une suite d'ouverts denses. Le but est de prouver que l'ensemble $\bigcap_{n\in\mathbb{N}} U_n$ intersecte V. Vu que V est ouvert dans un espace métrique, il contient une boule ouverte et donc une boule fermée B_0 de rayon strictement positif. L'ensemble U_1 est dense et intersecte donc un ouvert contenu dans B_0. L'intersection est un ouvert qui contient alors une boule fermée B_1 de rayon strictement positif. Continuant ainsi nous construisons une suite de fermés emboités B_n telle que

$$\bigcap_{n\in\mathbb{N}} U_n \cap V \tag{7.423}$$

contient l'intersection des B_n. Par le théorème 7.304 des fermés emboîtés (que nous utilisons parce que E est métrique et complet), cette intersection est non vide.

Ouvert d'un espace de Baire

□

Une des applications du théorème de Baire est le théorème de Banach-Steinhaus 10.84.

Chapitre 8

Espaces affines

8.1 Vecteurs agissant sur un espace

Définition 8.1.
Soit E, un espace vectoriel. Un ***espace affine modelé sur*** *E est un ensemble $\mathcal{E}$ sur lequel le groupe $(E, +)$ agit à droite transitivement et librement*[1].

Étant donné que E est un groupe commutatif, l'action peut être vue indifféremment à gauche ou à droite. Si $M \in \mathcal{E}$ et si $x \in E$ nous notons $M + x$ au lieu de $x \cdot M$ le résultat de l'action de x sur M.

8.2.
Lorsque nous écrivons « $M + x$ », le symbole plus n'est pas une loi de composition interne de $\mathcal{E}$, mais une action.

Proposition-définition 8.3.
Soient $N, M \in \mathcal{E}$. Il existe un unique $x \in E$ tel que $M + x = N$.
Nous noterons $\overrightarrow{MN}$ ce vecteur.

Démonstration. La transitivité de l'action assure l'existence et la liberté assure l'unicité. □

Proposition 8.4.
Soit un espace affine $\mathcal{E}$ modelé sur l'espace vectoriel E. Soient $A, B, C \in \mathcal{E}$. Nous avons les égalités suivantes dans E :
(1) $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$ (relations de Chasles),
(2) $\overrightarrow{AA} = 0$,
(3) $\overrightarrow{BA} = -\overrightarrow{AB}$.

Démonstration. Point par point.

Pour (1) Nous avons, par définition 8.3 les égalités

$$\begin{cases} C = A + \overrightarrow{AC} & (8.1a) \\ B = A + \overrightarrow{AB} & (8.1b) \\ C = B + \overrightarrow{BC} & (8.1c) \end{cases}$$

En substituant les deux premières dans la troisième, nous trouvons $A + \overrightarrow{AB} + \overrightarrow{BC} = A + \overrightarrow{AC}$. Par liberté de l'action, nous pouvons « simplifier » par A et trouver la relation de Chasles.

Pour (2) Nous avons $A + \overrightarrow{AA} = A$, mais aussi $A + 0 = A$. Par unicité nous avons $\overrightarrow{AA} = 0$.

Pour (3) Nous avons $B + \overrightarrow{BA} = A$ et $A + \overrightarrow{AB} = B$. En mettant bout à bout,

$$B + \overrightarrow{BA} + \overrightarrow{AB} = B. \tag{8.2}$$

1. Définition 2.62.

Donc $\overrightarrow{BA} + \overrightarrow{AB} = 0$.

□

8.5.
Si E est un espace vectoriel, le groupe $(E, +)$ agit sur E par l'action $t_y(x) = y + x$. Utilisant cette action nous construisons l'**espace affine canonique** de E. En particulier nous notons $\mathcal{E}_n(\mathbb{K})$ l'espace affine canonique de $\mathbb{K}^n$ vu comme espace vectoriel sur $\mathbb{K}$.

— En tant qu'ensembles, $\mathcal{E}_n(\mathbb{K}) = \mathbb{K}^n$.

— Sur cet espace en particulier, si $M, N \in \mathcal{E}_n(\mathbb{K})$, nous avons $\overrightarrow{MN} = N - M$ où à droite, la différence est la différence vectorielle dans $\mathbb{K}^n$.

Ces deux points se généralisent immédiatement à un espace vectoriel E au lieu de $\mathbb{K}^n$.

8.2 Repères cartésiens affines

Soit E un $\mathbb{K}$-espace vectoriel de dimension n et $\mathcal{E}$ un espace affine construit sur E.

Définition 8.6.
*Un multiplet $(A, e_1, \dots, e_n)$ où A est un point de $\mathcal{E}$ et $\{e_i\}$ est une base de E est un **repère cartésien** de $\mathcal{E}$.*

*Nous disons que $\{e_i\}$ est la **base associée** au repère.*

Proposition 8.7.
Si $\mathcal{E}$ est un espace affine modelé sur l'espace vectoriel E de dimension n sur le corps $\mathbb{K}$, et si $\big(A, \{e_i\}_{i=1,\dots,n}\big)$ est un repère cartésien, alors

$$\begin{aligned} \phi\colon \mathbb{K}^n &\to \mathcal{E} \\ (x_1, \dots, x_n) &\mapsto A + \sum_i x_i e_i. \end{aligned} \tag{8.3}$$

est une bijection.

*Ces nombres x_i sont les **coordonnées** du point $A + \sum_i x_i e_i$ dans le repère (A, e_i).*

Démonstration. L'application φ est surjective parce que l'action de E sur $\mathcal{E}$ est transitive et injective parce que l'action est libre. □

8.3 Classification affine des conique

Soit une conique $f(x, y) = 0$ avec

$$f(x, y) = ax^2 + 2bxy + cy^2 + 2dx + 2ey + f \tag{8.4}$$

dans le repère $R = (A, e_i)$. La signature de la quadratique

$$q(x, y) = ax^2 + 2bx + cy^2 \tag{8.5}$$

ne dépend pas de la base choisie et un changement de variables

$$\begin{cases} \tilde{x} = \alpha x + \beta y & \text{(8.6a)} \\ \tilde{y} = \gamma x + \delta y & \text{(8.6b)} \end{cases}$$

peut nous amener dans trois cas :

$$q(x, y) = \begin{cases} \tilde{x}^2 + \tilde{y}^2 & \text{genre ellipse} \\ \tilde{x}^2 - \tilde{y}^2 & \text{genre hyperbole} \\ \tilde{x}^2 & \text{genre parabole.} \end{cases} \tag{8.7}$$

Dans le troisième cas, la matrice de q est de rang 1.

Nous cherchons maintenant à savoir si un point $I = (x_0, y_0)$ est un centre de symétrie de $f(x, y) = 0$. Pour cela nous choisissons le repère centré en I, c'est-à-dire que nous posons

$$\begin{cases} x = x_0 + \tilde{x} & (8.8a) \\ y = y_0 + \tilde{y}. & (8.8b) \end{cases}$$

Un peu de calcul montre qu'alors la conique s'écrit

$$f(x_0, y_0) + q(\tilde{x}, \tilde{y}) + (2ax_0 + 2by_0 + 2d)\tilde{x} + (2bx_0 + 2cy_0 + 2e)\tilde{y} = 0. \tag{8.9}$$

Le point I sera un centre de symétrique si les termes linéaires en $\tilde{x}$ et $\tilde{y}$ s'annulent, c'est-à-dire si

$$\begin{cases} ax_0 + by_0 + d = 0 & (8.10a) \\ bx_0 + cy_0 + e = 0. & (8.10b) \end{cases}$$

Nous supposons que $(d, e) \neq (0, 0)$, sinon la conique de départ serait déjà centrée. Le déterminant du système (8.10) est

$$\delta = ac - b^2. \tag{8.11}$$

Si ce dernier est différent de zéro, le système possède une unique solution et la conique aura alors un unique centre de symétrie.

Si le déterminant du système est nul, il y a soit pas de centre de symétrie, soit une infinité. Dans le premier cas nous sommes en présence d'une parabole, et dans le second cas de deux droites parallèles.

Exemple 8.8.
Soit

$$f(x, y) = x^2 + 2xy - y^2 - 6x + 2y - 1 = 0 \tag{8.12}$$

donnée dans le repère affine $R = (A, \{e_i\})$. Nous commençons par étudier la signature de $q(x, y) = x^2 + 2xy - y^2$ dont la matrice symétrique est

$$Q = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{8.13}$$

Son polynôme caractéristique est $\lambda^2 - 2$ sont les racines sont $\pm\sqrt{2}$. La signature est donc $(1, -1)$ et nous sommes en présence d'une conique de genre hyperbole. Nous cherchons le centre en posant $x = \tilde{x} + x_0$, $y = \tilde{y} + y_0$. Le système à résoudre est

$$\begin{cases} x_0 + y_0 - 3 = 0 & (8.14a) \\ x_0 - y_0 + 1 = 0, & (8.14b) \end{cases}$$

dont l'unique solution est $(x_0, y_0) = (1, 2)$. Nous considérons le repère centré en (x_0, y_0), c'est-à-dire le repère

$$R' = (I, \{e_i\}) \tag{8.15}$$

avec $I = A + x_0e_1 + y_0e_2$ où A est l'origine du repère dans lequel l'équation (8.12) était donnée.

Par construction dans ce repère nous avons la conique

$$f(x_0, y_0) + q(\tilde{x}, \tilde{y}) = 0, \tag{8.16}$$

c'est-à-dire

$$\tilde{x}^2 + 2\tilde{x}\tilde{y} - \tilde{y}^2 = 0. \tag{8.17}$$

Maintenant la nous avons une quadrique centrée nous voulons la mettre sous une forme plus canonique :

$$\left(\frac{1}{\sqrt{2}}(\tilde{x} + \tilde{y})\right)^2 - \tilde{y}^2 - 1 = 0. \tag{8.18}$$

Nous posons donc

$$\begin{cases} X = \frac{1}{\sqrt{2}}(\tilde{x} + \tilde{y}) & (8.19a) \\ Y = \tilde{y}, & (8.19b) \end{cases}$$

et nous trouvons l'hyperbole

$$X^2 - Y^2 - 1 = 0. \tag{8.20}$$

Cela revient à faire le changement de base

$$\begin{cases} e_1' = \sqrt{2}e_1 & (8.21a) \\ e_2' = -e_1 + e_2. & (8.21b) \end{cases}$$

Pour rappel, les vecteurs de bases se transforment avec la matrice inverse des coefficients. Étant donné que

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix}, \tag{8.22}$$

nous avons

$$\begin{pmatrix} e_1' \\ e_2' \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} e_1 \\ e_2 \end{pmatrix}. \tag{8.23}$$

C'est de là que provient le changement (8.21). △

8.4 Applications affines

La définition 4.48 donnait déjà la définition d'une application affine entre deux espaces vectoriels. Voici maintenant le pendant entre deux espaces affines.

Définition 8.9.
*Soient $\mathcal{E}$ et $\mathcal{E}'$ deux espaces affines sur les espaces vectoriels E et E' (sur le même corps $\mathbb{K}$). Une application $f\colon \mathcal{E} \to \mathcal{E}'$ est dite **affine** si pour tout $M \in \mathcal{E}$, il existe une application linéaire $u_M\colon E \to E'$ telle que*

$$f(M + x) = f(M) + u_M(x) \tag{8.24}$$

pour tout $x \in E$.

La définition suivante permet de décomposer une application affine en une partie linéaire et une translation. À partir de là, la proposition 8.57 nous donnera une structure de groupe sur $\mathrm{Aff}(\mathbb{R}^n)$.

Lemme-définition 8.10 ([1])**.**
Soient $\mathcal{E}$ et $\mathcal{E}'$ deux espaces affines sur les espaces vectoriels E et E' (sur le même corps $\mathbb{K}$). Nous considérons une application affine $f\colon \mathcal{E} \to \mathcal{E}'$.

Il existe une unique application linéaire $u\colon E \to E'$ telle que

$$f(M + x) = f(M) + u(x) \tag{8.25}$$

pour tout $x \in E$ et pour tout $M \in \mathcal{E}$.

*Cette application linéaire est appelée **partie linéaire** de f. Pour varier les notations, nous noterons souvent $f = \alpha \circ \tau_v$ pour une application linéaire α et la translation τ_v de vecteur v.*

Démonstration. En plusieurs coups.

Unicité Supposons que u_1 et u_2 vérifient la propriété, alors pour tout $x \in E$ et tout $M \in \mathcal{E}$ nous avons $f(M + x) = f(M) + u_1(x)$ et $f(M + x) = f(M) + u_2(x)$. Cela suffit à nous convaincre que $u_1 = u_2$.

$\underline{u_M = u_N}$ Avant de prouver l'existence, nous considérons $M, N \in \mathcal{E}$ et les applications linéaires u_M et u_N vérifiant l'équation (8.24) pour M et N respectivement. Prouvons que $u_M = u_N$. Posons

$$f(M + x) = f(M) + u_M(x) \tag{8.26a}$$
$$f(N + y) = f(N) + u_N(y). \tag{8.26b}$$

Définissons $a \in E$ par $N = M + a$; nous avons d'une part que

$$f(N + y) = f(M + y + a) = f(M) + u_M(y + a), \tag{8.27}$$

et d'autre part

$$f(N + y) = f(M + a) + u_N(y) = f(M) + u_M(a) + u_N(y). \tag{8.28}$$

Par conséquent $u_M(y + a) = u_M(a) + u_N(y)$. Par linéarité $u_N = u_M$.

Existence Soit $M \in \mathcal{E}$. Nous affirmons que u_M fait l'affaire. En effet, soient $N \in \mathcal{E}$ et $x \in E$. Vu que $u_M = u_N$ nous avons

$$f(N + x) = f(M) + u_N(x) = f(M) + u_M(x). \tag{8.29}$$

Donc effectivement u_M peut être utilisé en tout point de $\mathcal{E}$.

□

Ce lemme est important car il permet de démonter qu'une application est affine en prouvant la linéarité des u_M séparément sans devoir prouver qu'elles sont égales.

8.4.1 Autres propriétés

Lemme 8.11 ([1]).
Soient $M \in \mathcal{E}$ et $A, B \in \mathcal{E}$ deux points donnés par $A = M + x_a$, $B = M + x_b$. Soit encore une application affine f sur $\mathcal{E}$. Alors

$$\overrightarrow{AB} = u_f(x_b - x_a). \tag{8.30}$$

Démonstration. En appliquant f à $A = M + x_a$ et $B = M + x_b$,

$$f(A) = f(M) + u_f(x_a) \tag{8.31a}$$
$$f(B) = f(M) + u_f(x_b). \tag{8.31b}$$

Donc $f(B) = f(A) - u_f(x_a) + u_f(x_b)$ ou encore

$$f(B) = f(A) + u_f(x_b - x_a). \tag{8.32}$$

□

Remarque 8.12.
La condition (8.24) pour tout $M \in \mathcal{E}$ est équivalente à demander

$$f \circ t_x = t_{u(x)} \circ f \tag{8.33}$$

pour tout $x \in E$.

Proposition 8.13.
Soit f une application affine.

(1) Il existe une unique application linéaire u_f telle que $f(M + x) = f(M) + u_f(x)$ pour tout $M \in \mathcal{E}$ et tout $x \in E$.

(2) L'application u_f est injective si et seulement si f est injective.

(3) L'application u_f est surjective si et seulement si f est surjective.

Si de plus les espaces $\mathcal{E}$ et $\mathcal{E}'$ ont même dimension finie, alors f est injective si et seulement si f est surjective.

Démonstration. La partie (1) est le lemme 8.10. □

Exemple 8.14.
L'espace $\mathbb{R}^n$ est très particulier parce qu'il agit sur lui-même ; il est donc un espace affine à lui tout seul : $\mathcal{E} = E = \mathbb{R}^n$.

Dans le cas de $\mathbb{R}^n$, en posant $M = 0$ dans la condition (8.24), si f est une application affine il existe une application linéaire α et un vecteur v tel que $f = \tau_v \circ \alpha$.

Notons que ça n'a pas de sens de poser $M = 0$, et la décomposition $f = \tau_v \circ \alpha$ n'a aucun sens en général. En particulier, nous ne pouvons pas appliquer une application linéaire à un élément d'un espace affine général. △

Proposition 8.15.
Si $f\colon \mathcal{E} \to \mathcal{E}'$ et $g\colon \mathcal{E} \to \mathcal{E}''$ sont des applications affines, alors $g \circ f\colon \mathcal{E} \to \mathcal{E}''$ est affine et $u_{g\circ f} = u_g \circ u_f$.

Démonstration. Si $M \in \mathcal{E}$ et $x \in E$ nous avons

$$\begin{aligned}(g \circ f)(M + x) &= g\big(f(M) + u_f(x)\big)\\ &= f\big(f(M)\big) + u_g\big(u_f(x)\big)\\ &= (g \circ f)(M) + (u_g \circ u_f)(x).\end{aligned} \tag{8.34}$$

□

Théorème 8.16.
Soient $\mathcal{E}$ et $\mathcal{E}'$ deux espaces affines de dimensions finies p et q sur $\mathbb{K}$. Soient les repères cartésiens $R = (O, \{e_i\})$ et $R' = (O', \{e'_i\})$. Une application $f\colon \mathcal{E} \to \mathcal{E}'$ est affine si et seulement s'il existe une matrice $a \in \mathbb{M}_{p,q}(\mathbb{K})$ et $b \in \mathbb{K}^q$ tels que

$$f(x) = b + ax. \tag{8.35}$$

Remarque 8.17.
L'équation (8.35) est écrite en utilisant un abus de notation entre le vecteur $x \in \mathbb{K}^p$ et le point de $\mathcal{E}$ qui est représenté par x dans le repère $(A, \{e_i\})$.

8.5 Isomorphismes

Définition 8.18.
*Un **isomorphisme** entre les espaces affines $\mathcal{E}$ $\mathcal{E}'$ est une application affine $f\colon \mathcal{E} \to \mathcal{E}'$ inversible dont l'inverse est affine.*

Proposition 8.19.
Une application affine bijective est un isomorphisme. Si f est un isomorphisme d'espaces affines, alors $u_{f^{-1}} = (u_f)^{-1}$.

Proposition 8.20.
Un espace affine de dimension finie n sur un corps $\mathbb{K}$ est isomorphe à l'espace affine canonique $\mathcal{E}_n(\mathbb{K})$.

Démonstration. Si nous considérons le repère $R = (A, \{e_i\})$ de l'espace affine $\mathcal{E}$ alors l'application

$$\begin{aligned}\varphi\colon \mathbb{K}^n &\to \mathcal{E}\\ (x_1, \dots, x_n) &\mapsto A + \sum_i x_i e_i\end{aligned} \tag{8.36}$$

est un isomorphisme. □

8.6 Sous espaces affines

Définition 8.21.
Soit $\mathcal{E}$ un espace affine sur l'espace vectoriel E. Un ***sous-espace affine*** *de $\mathcal{E}$ est une orbite de l'action d'un sous-espace vectoriel de E.*

Si $\mathcal{F}$ est un sous-ensemble de $\mathcal{E}$, il sera un sous-espace affine de $\mathcal{E}$ si et seulement si l'ensemble

$$F = \{AB \text{ tel que } A, B \in \mathcal{F}\} \tag{8.37}$$

est un sous-espace vectoriel de E. Dans ce cas nous disons que F est la **direction** de $\mathcal{F}$. Si $A \in \mathcal{F}$, alors l'orbite de A sous F est $\mathcal{F}$. La **dimension** de $\mathcal{F}$ est la dimension de sa direction.

Si $\mathcal{F}$ et $\mathcal{G}$ sont des sous-espaces affines de $\mathcal{E}$ de directions F et G, nous disons que $\mathcal{F}$ est **parallèle** à $\mathcal{G}$ si $F \subset G$.

Proposition 8.22.
Soit $\mathcal{F}$ un sous-espace affine de dimension k dans l'espace affine $\mathcal{E}$ de dimension n. Alors il existe une application affine $f\colon \mathcal{E} \to \mathbb{K}^{n-k}$ telle que $\mathcal{F} = f^{-1}(0)$.

Démonstration. Soient F la direction de $\mathcal{F}$ et $A \in \mathcal{F}$. Nous considérons une base $\{e_i\}$ adaptée à F au sens $\{e_1, \dots, e_k\}$ est une base de F. Nous considérons maintenant le repère cartésien $(A, \{e_i\})$ avec $A \in \mathcal{F}$ et nous construisons l'application affine

$$\begin{aligned} f\colon \mathcal{E} &\to \mathbb{K}^{n-k} \\ A + \sum_{i=1}^{n} x_i e_i &\mapsto \begin{pmatrix} x_{k+1} \\ \vdots \\ x_n \end{pmatrix}. \end{aligned} \tag{8.38}$$

Par construction nous avons $f(M) = 0$ si et seulement si $M \in \mathcal{F}$. □

Proposition 8.23 ([35])**.**
Soit σ une partie de l'espace affine $\mathcal{E}$.

(1) L'intersection de tous les sous-espaces affines contenant σ est un sous-espace affine, noté $\mathcal{F}$.

(2) Si $A \in \sigma$, alors la direction de $\mathcal{F}$ est le sous-espace vectoriel

$$F = \operatorname{Span}\{\overrightarrow{AM} \text{ tel que } M \in \sigma\}. \tag{8.39}$$

Le sous-espace affine donné par la proposition 8.23 est le sous-espace affine **engendré** par la partie σ, et il est noté $\operatorname{eae}(\sigma)$.

Proposition 8.24.
Soit $\mathcal{E}$ un espace affine de dimension n sur $\mathbb{K}$, soit $f\colon \mathcal{E} \to \mathbb{K}^r$ une fonction affine. Pour tout $a = (a_1, \dots, a_r) \in \mathbb{K}^r$, l'ensemble $f^{-1}(a)$ est un sous-espace affine de dimension $\dim \ker(u_f)$.

Démonstration. Nous considérons le repère $(A, \{e_i\})$ de $\mathcal{E}$. Étant donné que f est affine nous avons

$$f\big(A + \sum_i x_i e_i\big) = f(A) + u_f\big(\sum_i x_i e_i\big). \tag{8.40}$$

Nous avons donc $f\big(A + \sum_i x_i e_i\big) = a$ lorsque

$$u_f(\sum_i x_i e_i) = a - f(A). \tag{8.41}$$

Nous avons donc

$$f^{-1}(a) = A + (u_f)^{-1}\big(a - f(A)\big), \tag{8.42}$$

dont la dimension est le rang de $(u_f)^{-1} = u_{f^{-1}}$ (proposition 8.19). Le rang de $(u_f)^{-1}$ est le dimension du noyau de u_f. □

Définition 8.25 (Partie convexe)**.**
Une partie A d'un espace vectoriel est ***convexe*** *si pour tout $a, b \in A$ et pour tout $t \in [0,1]$, le point $ta + (1-t)b$ est dans A.*

Autrement dit, une partie est convexe lorsqu'elle contient tous les segments joignant ses points.

Proposition 8.26.
Soit un espace vectoriel normé[2] *$(V, \|.\|)$. Pour tout $a \in V$ et $r > 0$, la boule $B(a,r)$ est convexe. La boule fermée $\overline{B(a,r)}$ également.*

Démonstration. En deux parties.

La boule centrée en zéro Soient $x, y \in B(0,r)$ et $\lambda \in]0,1[$. Alors

$$\|\lambda x + (1-\lambda)y\| \leqslant |\lambda|\|x\| + |1-\lambda|\|y\| < (|\lambda| + |1-\lambda|)r \leqslant r \tag{8.43}$$

où nous avons utilisé le fait que $|\lambda| = \lambda$ et $|1-\lambda| = 1-\lambda$.

Cela prouve que $\lambda x + (1-\lambda)y \in B(0,r)$. Notez l'inégalité stricte due au fait que $\|x\| < r$ et $\|y\| < r$. Dans le cas de la boule fermée, nous avons une inégalité large.

La boule centrée autre part Soient $x, y \in B(a,r)$. Alors $x - a$ et $y - a$ sont dans $B(0,r)$, de telle sorte que

$$\lambda(x-a) + (1-\lambda)(y-a) \in B(0,r) \tag{8.44}$$

par la première partie. En développant et simplifiant,

$$\lambda x + (1-\lambda)y - a \in B(0,r), \tag{8.45}$$

ce qui signifie que $\lambda x + (1-\lambda)y \in B(a,r)$.

□

Proposition 8.27.
Soit A un ensemble convexe[3] *dans un espace vectoriel et $v_1, \dots, v_n$ des éléments de A. Alors toute combinaison*

$$a_1 v_1 + \cdots + a_n v_n \tag{8.46}$$

telle que $a_1 + \cdots + a_n = 1$ et $a_i \in [0,1]$ appartient à A.

Démonstration. Nous prouvons la proposition pour $n = 3$. Nous devons trouver des nombres $t_1, t_2 \in [0,1]$ tels que

$$t_2\big(t_1 v_1 + (1-t_1)v_2\big) + (1-t_2)v_3 = av_1 + bv_2 + cv_3. \tag{8.47}$$

La réponse est immédiatement donnée par

$$t_2 a = 1 - c \tag{8.48a}$$
$$t_1 = a/t_2. \tag{8.48b}$$

Étant donné que $c \in [0,1]$ nous avons $t_2 \in [0,1]$. En ce qui concerne t_1 nous avons

$$t_1 = \frac{a}{t_2} \leqslant \frac{1-c}{1-c} = 1. \tag{8.49}$$

□

2. Définition 7.113.
3. Définition 8.25.

8.7 Barycentre

Soit $\mathcal{E}$ un espace affine sur le $\mathbb{K}$-espace vectoriel E. Un couple (A, λ) avec $A \in \mathcal{E}$ et $\lambda \in \mathbb{K}$ est un **point pondéré**.

Lemme-définition 8.28 ([167]).
Soit une famille de points pondérés $\{(A_i, \lambda_i)\}_{i=1\ldots r}$. Si $\sum_i \lambda_i \neq 0$, alors il existe un unique $G \in \mathcal{E}$ tel que

$$\sum_{i=1}^{r} \lambda_i \overrightarrow{GA_i} = 0. \tag{8.50}$$

*Le point G donné par le lemme 8.28 est le **barycentre** des points pondérés (A_i, λ_i).*

Notons que l'on peut toujours supposer que $\sum_i \lambda_i = 1$ parce que le barycentre ne change pas lorsque tous les λ_i sont multipliés par un même nombre.

Définition 8.29 (Combinaison convexe).
*Des nombres $\lambda_1, \ldots, \lambda_n$ vérifiant $\sum_i \lambda_i = 1$ forment une **combinaison convexe**.*

Le théorème suivant donné quelques caractérisations équivalentes du barycentre.

Théorème 8.30 ([167]).
Soient $\{(A_i, \lambda_i)\}_{i=1,\ldots,r}$ une famille de points pondérés. Les conditions suivantes sur le point $G \in \mathcal{E}$ sont équivalentes.

(1) Le point G est barycentre de la famille.
(2) Pour tout $\alpha \in \mathbb{R}^*$, $\sum_i (\alpha\lambda_i)\overrightarrow{GA_i} = 0$.
(3) Il existe $A \in \mathcal{E}$ tel que $\left(\sum_i \lambda_i\right)\overrightarrow{AG} = \sum_i \lambda_i \overrightarrow{AA_i}$.
(4) Pour tout $B \in \mathcal{E}$, nous avons $\left(\sum_i \lambda_i\right)\overrightarrow{BG} = \sum_i \lambda_i \overrightarrow{BA_i}$.

Définition 8.31.
*Si $A, B \in \mathcal{E}$, le **segment** $[AB]$ est l'ensemble des barycentres de A et B pondérés par des poids positifs (ouvert ou fermé suivant que l'on accepte que l'un ou l'autre des poids soit nul).*

Lorsque tous les λ_i sont égaux, nous parlons d'**isobarycentre**. Autrement dit, l'isobarycentre des points A_i est le barycentre des points pondérés $(A_i, 1)$.

8.7.1 Sous-espaces affines

Proposition 8.32.
Une partie $\mathcal{F}$ des $\mathcal{E}$ est un sous-espace affine si et seulement si elle est stable par barycentrisation.

Démonstration. Soit $\mathcal{F}$ une sous-espace affine de direction F et $A_1, \ldots, A_n$ des points de $\mathcal{F}$. Nous devons voir que le barycentre des points A_i pondérés de n'importe quelles masses appartient à $\mathcal{F}$. Pour ce faire nous faisons appel à la caractérisation (4) du théorème 8.30 : pour tout $B \in \mathcal{F}$,

$$\overrightarrow{BG} = \sum_i \lambda_i \overrightarrow{BA_i}. \tag{8.51}$$

Vu que B et A_i sont dans $\mathcal{F}$, nous avons $\overrightarrow{BA_i \in F}$ et donc $\overrightarrow{BG} \in F$. Mais comme $B \in \mathcal{F}$, le point G est à son tour dans $\mathcal{F}$.

Réciproquement, nous supposons que $\mathcal{F}$ est stable par barycentrisme. Nous voudrions montrer que l'ensemble

$$F = \{\overrightarrow{AB} \text{ tel que } A, B \in \mathcal{F}\} \tag{8.52}$$

est un sous-espace vectoriel. Soit $A \in \mathcal{F}$. Nous commençons par prouver que les vecteurs de la forme $\overrightarrow{AX}$ $(X \in \mathcal{F})$ forment un espace vectoriel. Considérons $\overrightarrow{AX} + \overrightarrow{AY}$ qui est un élément de E ; il existe donc $V \in \mathcal{E}$ tel que

$$\overrightarrow{AV} = \overrightarrow{AX} + \overrightarrow{AY}. \tag{8.53}$$

Par les relations de Chasles,

$$\overrightarrow{AV} = \overrightarrow{AV} + \overrightarrow{VX} + \overrightarrow{AV} + \overrightarrow{VY}, \tag{8.54}$$

donc

$$0 = \overrightarrow{VX} - \overrightarrow{VA} + \overrightarrow{VY}, \tag{8.55}$$

ce qui prouve que V est un barycentre de X, A, Y, et donc que $V \in \mathcal{F}$. De la même manière si $W \in \mathcal{E}$ est défini par $\overrightarrow{AW} = \mu\overrightarrow{AX}$, alors

$$\overrightarrow{AW} = \mu\overrightarrow{AX} = \mu(\overrightarrow{AW} + \overrightarrow{WX}), \tag{8.56}$$

ce qui signifie que

$$(1-\mu)\overrightarrow{AW} + \mu\overrightarrow{XW} = 0 \tag{8.57}$$

et que W est un barycentre.

Afin de montrer que (8.52) est bien un espace vectoriel, nous devons considérer $A, B, X, Y \in \mathcal{F}$ et prouver que $\overrightarrow{AX} + \overrightarrow{BY} \in F$. Nous avons

$$\begin{aligned}
\overrightarrow{AX} + \overrightarrow{BY} &= \overrightarrow{AX} + \overrightarrow{BA} + \overrightarrow{AY} && (8.58a)\\
&= \overrightarrow{AV} + \overrightarrow{BA} \qquad V \text{ est celui donné plus haut} && (8.58b)\\
&= \overrightarrow{AV} - \overrightarrow{AB} && (8.58c)\\
&= \overrightarrow{AV} + \overrightarrow{AW} \qquad W \text{ est donné par } \mu = -1. && (8.58d)\\
&= \overrightarrow{AV'}. && (8.58e)
\end{aligned}$$

□

Proposition 8.33 ([167])**.**
Soient $A_0, \dots, A_r$ des points de $\mathcal{E}$. L'ensemble des barycentres de ces points (avec des masses de somme 1) est le sous-espace affine engendré par les A_i que nous nommons $\mathcal{F}$.

Démonstration. Soit G le barycentre associé aux poids λ_i. Nous avons

$$G = A_0 + \overrightarrow{A_0G} = A_0 + \sum_{i=1}^{r} \lambda_i \overrightarrow{A_0A_i}. \tag{8.59}$$

Notons que les vecteurs $\overrightarrow{A_0A_i}$ sont dans la direction du sous-espace affine engendré par les A_i par (8.39). Donc G est bien dans $\mathcal{F}$.

Inversement si X est dans $\mathcal{F}$, on a

$$X = A_0 + \sum_i \lambda_i \overrightarrow{A_0A_i} \tag{8.60}$$

parce que $\sum_i \lambda_i \overrightarrow{A_0A_i}$ est un élément général de la direction de $\mathcal{F}$. Du coup

$$\overrightarrow{A_0X} = \sum_i \lambda_i \overrightarrow{A_0A_i}, \tag{8.61}$$

et en utilisant la relation de Chasles sur chacun des $\overrightarrow{A_0A_i}$,

$$\overrightarrow{A_0X} = \sum_i \lambda_i \big(\overrightarrow{A_0X} + \overrightarrow{XA_i}\big). \tag{8.62}$$

De là nous concluons que

$$\big(1 - \sum_i \lambda_i\big)\overrightarrow{A_0X} + \sum_i \lambda_i \overrightarrow{A_iX} = 0, \tag{8.63}$$

ce qui signifie précisément que X est un barycentre des A_i. □

Proposition 8.34.
Soient $r+1$ point $A_0, \dots, A_r$ dans $\mathcal{E}$. Le sous-espace affine engendré par les A_i est au plus de dimension r.

Démonstration. La direction de l'espace engendré $\mathrm{Aff}\{A_i\}$ est l'espace

$$\mathrm{Span}\{\overrightarrow{A_0A_i}_{i=1,\dots,r}\} \tag{8.64}$$

qui est engendré par r vecteurs et donc est au plus de dimension r. □

En deux mots, la proposition suivante signifie que le barycentre des barycentres est le barycentre.

Proposition 8.35 (Associativité des barycentres[168])**.**
Soit $I = \{0, 1, \dots, n\}$ et une partition $I = J_0 \cup \dots \cup J_r$. Soient des points $a_0, \dots, a_n \in \mathcal{E}$ et $\lambda_0, \dots, \lambda_m$ des nombres tels que $\sum_i \lambda_i \neq 0$. Nous supposons que $\mu_k = \sum_{i\in J_k} \lambda_i \neq 0$ pour tout k, et enfin nous nommons b_k le barycentre de la famille $\{(a_i, \lambda_i), i \in J_k\}$.

Alors le barycentre de la famille $\{(b_k, \mu_k)\}_{k=1,\dots,n}$ est le barycentre de la famille $\{(a_i, \lambda_i)\}_{i\in I}$.

Démonstration. Nous nommons b le barycentre des b_k pondérés par les μ_k, donc par définition

$$\begin{aligned}
0 &= \sum_{k=0}^{r} \mu_k \overrightarrow{bb_k} && (8.65a)\\
&= \sum_k \sum_{i\in J_k} \lambda_i \overrightarrow{bb_k} && (8.65b)\\
&= \sum_{k=0}^{r} \sum_{i\in J_k} \lambda_i (\overrightarrow{ba_i} + \overrightarrow{a_ib_k}) && (8.65c)\\
&= \sum_{k=0}^{r} \sum_{i\in J_k} \lambda_i \overrightarrow{ba_i} + \sum_{k=0}^{r} \underbrace{\sum_{i\in J_k} \lambda_i \overrightarrow{a_ib_k}}_{=0} && (8.65d)\\
&= \sum_{i\in I} \lambda_i \overrightarrow{ba_i}. && (8.65e)
\end{aligned}$$

Donc b est bien barycentre des a_i avec les poids λ_i. □

8.7.2 Enveloppe convexe

Définition 8.36.
*Soit A une partie d'un espace vectoriel E. L'**enveloppe convexe** de A, notée $\mathcal{C}\mathrm{onv}(A)$ est l'intersection de tous les convexes contenant A.*

L'enveloppe convexe est un convexe. En effet soit C un convexe contenant A et $x, y \in \mathcal{C}\mathrm{onv}(A)$; alors x et y sont dans C et par conséquent le segment $[x, y]$ est inclus dans C. Ce segment étant inclus dans tout convexe contenant A, il est inclus dans $\mathcal{C}\mathrm{onv}(A)$.

Proposition 8.37 ([169])**.**
Soit C un convexe dans l'espace affine $\mathcal{E}$ et une famille de points pondérés $\{(a_i, \lambda_i)\}_{i=1,\dots,r}$ dont tous les poids sont positifs (et non tous nuls). Alors le barycentre est aussi dans C.

En d'autre termes, un convexe est stable par barycentrage à poids positifs[4].

Démonstration. Nous prouvons par récurrence. D'abord pour $r = 2$. Le barycentre des points pondérés (a_1, λ_1), (a_2, λ_2) est le point b tel que

$$\lambda_1 \overrightarrow{ba_1} + \lambda_2 \overrightarrow{ba_2} = 0. \tag{8.66}$$

4. Sauf si on prend tous les poids nuls ; mais contre ce genre d'idées, on ne peut rien faire.

Par définition, ce qui est noté $\overrightarrow{ab}$ n'est rien d'autre que $b - a$; en déballant (8.66), nous trouvons

$$\lambda_1(a_1 - b) + \lambda_2(a_2 - b) = 0 \tag{8.67}$$

et donc

$$b = \frac{\lambda_1}{\lambda_1 + \lambda_2} a_1 + \frac{\lambda_2}{\lambda_1 + \lambda_2} a_2, \tag{8.68}$$

qui est bien un point du segment $[a_1, a_2]$ parce que c'est une combinaison à coefficients positifs de somme 1.

Nous passons maintenant à la vraie récurrence avec un ensemble de points pondérés

$$A_r = \{(a_1, \lambda_1), \dots, (a_r, \lambda_r)\} \tag{8.69}$$

de masse totale non nulle ; et en vous laissant deviner ce que va désigner A_{r-1}. Si une des masses est nulle (disons λ_r), alors le barycentre de A_r est le même que celui de A_{r-1} et l'hypothèse de récurrence nous enseigne que ledit barycentre est dans C. Nous supposons donc que $\lambda_i \neq 0$ pour tout i. Dans ce cas le théorème d'associativité des barycentres 8.35 dit que le barycentre de A_r est le barycentre entre le barycentre de A_{r-1} et (a_r, λ_r), qui sont deux points de C par hypothèse de récurrence. □

Si E est un espace vectoriel et si $x_i \in E$ et $\lambda_i \in \mathbb{R}$, alors le barycentre des couples (x_i, λ_i) est le point g tel que $\sum_i \lambda_i \overrightarrow{gx_i}$, c'est-à-dire $\sum_i \lambda_i (x_i - g) = 0$ ou encore

$$\sum_i \lambda_i x_i = \sum_i \lambda_i g. \tag{8.70}$$

Donc quitte à diviser tous les λ_i par la somme, nous pouvons supposer que la somme des poids est 1. C'est pourquoi lorsque nous parlerons de barycentre dans un espace vectoriel sans contexte affin, nous allons toujours supposer $\sum_i \lambda_i = 0$ et avoir le barycentre

$$g = \sum_i \lambda_i x_i. \tag{8.71}$$

Proposition 8.38.
Soit E, un espace vectoriel et $A \subset E$. L'enveloppe convexe $\mathcal{C}\mathrm{onv}(A)$ est l'ensemble des barycentres de familles finies de points affublés de masses positives.

Démonstration. Nous notons $\mathcal{B}$ l'ensemble des dits barycentres. Par la proposition 8.37, ces barycentres sont dans l'enveloppe convexe et donc $\mathcal{B} \subset \mathcal{C}\mathrm{onv}(A)$. A contrario, si nous prouvons que $\mathcal{B}$ était convexe, alors nous aurions $\mathcal{C}\mathrm{onv}(A) \subset \mathcal{B}$ parce que l'enveloppe convexe est l'intersection des convexes contenant A.

Soient $a, b \in \mathcal{B}$, c'est-à-dire que l'on a $a_0, \dots, a_n$ et $b_0, \dots, b_m$ dans A ainsi que les nombres strictement positifs $\lambda_0, \dots, \lambda_n$ et $\mu_0, \dots, \mu_m$ tels que

$$a = \sum_i \lambda_i a_i \qquad \sum_{i=1}^n \lambda_i = 1 \tag{8.72a}$$

$$b = \sum_j \mu_j b_j \qquad \sum_{j=1}^n \mu_j = 1 \tag{8.72b}$$

Un point du segment $[a, b]$ est de la forme $p = ta + (1 - t)b$ avec $t \in [0, 1]$. En développant,

$$p = \sum_{i=0}^n (t\lambda_i) a_i + \sum_{j=0}^m (1 - t)\mu_j b_j. \tag{8.73}$$

Cela est le barycentre de la famille $\{(a_i, \lambda_i), (b_j, \mu_j)\}$, parce que la somme des coefficients est bien 1 :

$$\sum_i (t\lambda_i) + \sum_j (1 - t)\mu_j = t + (1 - t) = 1. \tag{8.74}$$

□

Théorème 8.39 (Carathéodory[54]).
Dans un espace affine de dimension n, l'enveloppe convexe[5] *de A est l'ensemble des barycentres à coefficients positifs ou nuls de familles de $n+1$ points.*

Démonstration. Soit $x \in \mathcal{C}\mathrm{onv}(A)$; on sait par la proposition 8.38 que x est barycentre de points de A avec des coefficients positifs :

$$x = \sum_{k=1}^{p} \lambda_k x_k \tag{8.75}$$

avec $\sum_k \lambda_k = 1$. Nous supposons que $p > n+1$ (sinon le théorème est réglé), et nous allons faire une récurrence à l'envers en montrant qu'on peut aussi écrire x sous forme d'un barycentre de strictement moins de p points.

Étant donné que $p-1 > n$, la famille $\{x+i-x_1\}_{i=2,\dots,p}$ est liée et il existe donc $\alpha_1, \dots, \alpha_p \in \mathbb{R}$ tels que $\sum_{i=2}^{p} \alpha_i(x_i - x_1) = 0$, c'est-à-dire telle que

$$\sum_{i=2}^{p} \alpha_i x_i = \sum_{i=2}^{p} \alpha_i x_1. \tag{8.76}$$

Nous posons $\alpha_1 = -\sum_{i=2}^{p} \alpha_1$. Remarquons qu'alors $\sum_{i=1}^{p} \alpha_i x_i = 0$ parce que

$$\sum_{i=1}^{p} \alpha_i x_i = \alpha_1 x_1 + \sum_{i=2}^{p} \alpha_i x_i = \alpha_1 x_1 + \sum_{i=2}^{p} \alpha_i x_1 = \sum_{i=1}^{p} \alpha_i x_1 = 0. \tag{8.77}$$

Par conséquent ça ne coûte rien de récrire (8.75) sous la forme

$$x = \sum_{i=1}^{p} (\lambda_i + t\alpha_i) x_i. \tag{8.78}$$

Les α_i ne sont pas tous nuls, mais leur somme est nulle, donc il y en a au moins un négatif. Nous notons

$$\tau = \min\{-\frac{\lambda_i}{\alpha_i} \text{ tel que } \alpha_i < 0\}, \tag{8.79}$$

et J l'ensemble de i pour lesquels ce minimum est atteint. Nous considérons aussi le nombres $\mu_i = \lambda_i + \tau\alpha_i$. Plusieurs remarques.

(1) Si $j \in J$, alors $\mu_j = 0$

(2) Si $\alpha_i > 0$ alors $\mu_i \geqslant 0$, mais si $\alpha_i < 0$ alors

$$\lambda_i + \tau\alpha_i \geqslant \lambda_i + (-\frac{\lambda_i}{\alpha_i})\alpha_i = 0m \tag{8.80}$$

donc $\mu_i \geqslant 0$ quand même.

(3) $\sum_{i=1}^{p} \mu_i = 1$, toujours parce que $\sum_{i=1}^{p} \alpha_i = 0$.

Avec tout ça, nous avons

$$\sum_{i \notin J} \mu_i x_i = \sum_{i=1}^{p} \mu_i x_i = x. \tag{8.81}$$

Et voilà, nous avons écrit x comme un barycentre à coefficients positifs de moins de p éléments parce que J n'est pas vide. □

Corolaire 8.40.
Dans un espace affine de dimension finie, l'enveloppe convexe d'un compact est compacte.

5. Définition 8.36.

Démonstration. Soit A une partie compacte de l'espace vectoriel E, et $\mathcal{C}\mathrm{onv}(A)$ son enveloppe convexe. Nous allons montrer que toute suite dans $\mathcal{C}\mathrm{onv}(A)$ admet une sous-suite convergente en écrivant un point de $\mathcal{C}\mathrm{onv}(A)$ comme le théorème de Carathéodory 8.39 nous le suggère. Pour cela nous considérons le simplexe

$$\Lambda = \left\{ \lambda \in \mathbb{R}^{n+1} \text{ tel que } \sum_{k=1}^{n+1} \lambda_k = 1 \text{ et } \lambda_k \geqslant 0 \forall k \right\}. \tag{8.82}$$

Montrons en passant que Λ est compact. Si $\lambda_k \in \Lambda$ est une suite, alors chacun des λ_k est un $(n+1)$-uple de nombres dans $[0,1]$:

$$k \mapsto (\lambda_k)_i \tag{8.83}$$

est une suite qui possède une sous-suite convergente. En passant $n+1$ fois à une sous-suite, nous tombons sur une suite convergente vers $\lambda \in \Lambda$, grâce à la convergence composante par composante. De plus pour chaque k nous avons $\sum_{i=1}^{n+1} (\lambda_k)_i = 1$, et en passant à la limite, la somme étant une application continue, $\sum_i \lambda_i = 1$.

Considérons l'application

$$\begin{aligned} f \colon \Lambda \times A^{n+1} &\to \mathcal{C}\mathrm{onv}(A) \\ (\lambda, x) &\mapsto \sum_{k=1}^{n+1} \lambda_k x_k. \end{aligned} \tag{8.84}$$

C'est une application continue parce qu'elle est bilinéaire en dimension finie ; son image est contenue dans $\mathcal{C}\mathrm{onv}(A)$ par la proposition 8.37, et elle est surjective par le théorème de Carathéodory 8.39. Bref, $\mathcal{C}\mathrm{onv}(A) = f(\Lambda \times A^{n+1})$ est donc l'image d'un compact par une application continue ; elle est donc compacte par le théorème 7.92. □

Notons que sans le théorème de Carathéodory, peut être que le nombre de points utiles pour décomposer les différents a_k n'était pas borné ; dans ce cas nous aurions du prendre une infinité de sous-suites et rien n'aurait été sûr.

8.7.3 Applications affines et barycentre

Proposition 8.41 ([170]).
Une application $f \colon \mathcal{E} \to \mathcal{E}'$ entre deux espaces affines est affine si et seulement si pour tout système $\{(A_i, \lambda_i)\}_{i=1,\dots,k}$ de barycentre G et de poids total non nul, le point $f(G)$ est barycentre du système $\{\big(f(A_i), \lambda_i\big)\}$.

Démonstration. En deux parties.

Si f est affine Par définition d'un barycentre,

$$\sum_i \lambda_i \overrightarrow{GA_i} = 0. \tag{8.85}$$

Nous considérons un point arbitraire $O \in \mathcal{E}$ et nous écrivons $A_i = O + x_i$, $G = O + x_g$. Ensuite nous utilisons le lemme 8.11 pour le calcul suivant :

$$\begin{aligned} \sum_i \lambda_i \overrightarrow{g(G)f(A_i)} &= \sum_i \lambda_i u_f(x_i - x_g) && \text{(8.86a)} \\ &= u_f\big(\sum_i \lambda_i (x_i - x_g)\big) && \text{(8.86b)} \\ &= u_f\big(\sum_i \lambda_i \overrightarrow{GA_i}\big) && \text{(8.86c)} \\ &= u_f(0) = 0. && \text{(8.86d)} \end{aligned}$$

Donc $f(G)$ est bien le barycentre du nouveau système.

Si f conserve les barycentres Nous définissons u par $f(O+x) = f(O) + u(x)$. À priori, ce u dépend de O et n'est pas linéaire.

u est linéaire Soient $M, N \in \mathcal{E}$ et les éléments $x_m, x_n \in E$ tels que $\overrightarrow{OM} = x_m$ et $\overrightarrow{ON} = x_n$. Nous définissons enfin P par

$$\overrightarrow{OP} = \alpha\overrightarrow{OM} + \beta\overrightarrow{ON}, \tag{8.87}$$

et $P = O + x_p$. En décomposant $\overrightarrow{MO}$ et $\overrightarrow{NO}$ par les relations de Chasles de la proposition 8.4(1) nous avons

$$(\alpha+\beta-1)\overrightarrow{PO} - \alpha\overrightarrow{PM} - \beta\overrightarrow{PN} \tag{8.88}$$

et donc P est barycentre du système

$$\big\{(\alpha+\beta-1, O), (\alpha, M), (\beta, N)\big\}. \tag{8.89}$$

Le point $f(P)$ sera barycentre du système

$$\Big\{\big(\alpha+\beta-A, f(O)\big), \big(\alpha, f(M)\big), \big(\beta, f(N)\big)\Big\}. \tag{8.90}$$

Cela signifie que

$$(\alpha+\beta-1)\overrightarrow{f(P)f(O)} - \alpha\overrightarrow{f(P)f(M)} - \beta\overrightarrow{f(P)f(N)} = 0. \tag{8.91}$$

En y substituant $\overrightarrow{f(P)f(O)} = u(-x_p)$, $\overrightarrow{f(P)f(M)} = u(x_m - x_p)$ et $\overrightarrow{f(P)f(N)} = u(x_m m - x_p)$ ainsi que $x_p = \alpha x_m + \beta x_b$ nous trouvons

$$u(\alpha x_m + \beta x_n) = \alpha u(x_m) + \beta u(x_n). \tag{8.92}$$

Donc u est linéaire.

u ne dépend pas du point O Il n'est pas besoin de démonter cela parce que la définition 8.9 ne le demande pas. Note : c'est le lemme 8.10 qui dit que c'est par ailleurs vrai.

□

8.8 Repères, coordonnées cartésiennes et barycentriques

Définition 8.42.
*On dit que les points $A_0, \dots, A_r \in \mathcal{E}$ sont **affinement indépendants** si le sous-espace affine engendré est de dimension r.*

Proposition 8.43 ([167]).
Pour $r+1$ points $A_0, \dots, A_r$ dans $\mathcal{E}$, les propriétés suivantes sont équivalentes.

(1) Les A_i sont affinement indépendants.

(2) Pour tout $i = 0, \dots, r$, le point A_i n'est pas dans $\mathrm{Aff}\{A_0, \dots, \hat{A}_i, \dots, A_r\}$.

(3) Les points $A_0, \dots, A_{r-1}$ sont affinement indépendants et $A_r \notin \mathrm{Aff}\{A_0, \dots, A_{r+1}\}$.

(4) Il existe i tel que les vecteurs $\overrightarrow{A_kA_i}$ $(k \in i)$ sont linéairement indépendants.

(5) Pour tout $i \in \{1, \dots, r\}$, les vecteurs $\overrightarrow{A_kA_i}$ $(k \neq i)$ sont linéairement indépendants.

Notons à propos de la condition (3) que l'existence d'un i tel que $A_i \notin \mathrm{Aff}\{A_0, \dots, \hat{A}_i, \dots, A_r\}$ n'implique pas l'indépendance des $r+1$ points. En effet dans $\mathbb{R}^2$ nous considérons les 4 points $A_0 = (0,0)$, $A_1 = (1,0)$, $A_2 = (2,0)$ et $A_3 = (0,1)$. Évidemment le point A_3 n'est pas dans l'espace engendré par les trois autres ; il n'empêche que ces points ne sont pas affinement indépendants parce que la direction est de dimension 2 au lieu de 3.

Définition 8.44.
Soit $\mathcal{E}$ un espace affine de dimension n et $\mathcal{F}$ un sous-espace affine de dimension k. Un ***repère affine*** *de $\mathcal{F}$ est la donnée de $k+1$ points affinement indépendants de $\mathcal{F}$.*

Si $\{A_0, \dots, A_n\}$ est un repère affine, le point A_0 est l'**origine**. C'est un choix complètement arbitraire ; et c'est bien cet arbitraire qui nous amènera à considérer les coordonnées barycentriques au lieu des coordonnées cartésiennes.

Soit $M \in \mathcal{E}$; par définition nous avons

$$M = A_0 + \overrightarrow{A_0M}. \tag{8.93}$$

Mais nous savons que les vecteurs $\overrightarrow{A_0A_i}$ forment une base de E, nous avons donc des nombres λ_i tels que

$$\overrightarrow{A_0M} = \sum_{i=1}^{n} \lambda_i \overrightarrow{A_0A_i}. \tag{8.94}$$

Les nombres λ_i ainsi construits sont les **coordonnées cartésiennes** du point M dans le repère $\{A_0, \dots, A_n\}$ d'origine A_0.

À partir de ces coordonnées, le point $M \in \mathcal{E}$ se retrouve par la formule

$$M = A_0 + \sum_{i=1}^{n} \lambda_i \overrightarrow{A_0A_i}. \tag{8.95}$$

Proposition 8.45 ([1])**.**
La paire $\big(O, \{e_1, \dots, e_n\}\big)$ est un repère cartésien de $\mathcal{E}$ si et seulement si $\{O, O+e_1, \dots, O+e_n\}$ est un repère affine.

Démonstration. En deux parties.

Sens direct Vue la proposition 8.43, il suffit de prouver que les vecteurs $\overrightarrow{O(O+e_i)}$ sont linéairement indépendants. Mais $\overrightarrow{O(O+e_i)} = e_i$, donc oui, ils sont linéairement indépendants.

Sens inverse Il s'agit d'utiliser la même proposition 8.43 qui est encore fonctionnelle parce qu'elle est une équivalence.

□

Soient (A, e_i) et (A', e'_i) deux repères cartésiens pour l'espace affine $\mathcal{E}$. Soit (a_{ij}) la matrice de changement de base entre $\{e_i\}$ et $\{e'_i\}$ dans E. Nous voudrions trouver les x_i en termes des x'_i.

Pour cela nous considérons un point M dans $\mathcal{E}$ et nous l'écrivons dans les deux bases. Cela fournit l'égalité

$$A + \sum_i x_i e_i = A' + \sum_i x'_i e'_i. \tag{8.96}$$

Nous considérons les coordonnées (a_i) de A' dans le repère (A, e_i), c'est-à-dire

$$A' = A + \sum_i a_i e_i. \tag{8.97}$$

En substituant $e'_i = \sum_k a_{jk} e_k$ et (8.97) dans (8.96) nous trouvons

$$\sum_k x_k e_k = \sum_k a_k e_k + \sum_{jk} a_{jk} x'_j e_k, \tag{8.98}$$

et par conséquent

$$x_k = a_k + \sum_j a_{jk} x'_j. \tag{8.99}$$

Les coordonnées barycentriques sont données par la proposition suivante.

Proposition 8.46 ([167]).
Soient $A_0, \dots, A_r$ des points affinement indépendants dans $\mathcal{E}$ et $\mathcal{F} = \mathrm{Aff}\{A_0, \dots, A_r\}$. Tout point $M \in \mathcal{F}$ s'écrit de façon unique comme barycentre[6] *des A_i affectés de poids λ_i tels que $\sum_{i=0}^{r} \lambda_i = 1$.*

Démonstration. Nous avons vu plus haut (définition 8.44) que l'affine indépendance des points A_i assurait que $(A_0, \dots, A_r)$ était un repère de $\mathcal{F}$.

En ce qui concerne l'existence de l'écriture de M comme barycentre, nous savons que les sous-espace affine sont exactement les ensembles de barycentres (proposition 8.33), c'est-à-dire que si on a des points dans un sous-espace affine, alors les barycentres de ces points est encore dans le sous-espace affine.

L'unicité est comme suit. Si M est barycentre des A_i avec poids λ_i, nous écrivons la caractérisation (4) du théorème 8.30 avec $B = A_0$:

$$\overrightarrow{A_0M} = \sum_{i=1}^{r} \lambda_i \overrightarrow{A_0A_i} \tag{8.100}$$

où la somme à droite s'étend à priori de 0 à r, mais comme $\overrightarrow{A_0A_0} = 0$, nous l'avons limitée à 1. Si M s'écrit comme barycentre de deux façons différentes, nous aurions

$$\overrightarrow{A_0M} = \sum_{i=1}^{r} \lambda_i \overrightarrow{A_0A_i} = \sum_{i=1}^{r} \mu_i \overrightarrow{A_0A_i} \tag{8.101}$$

avec $\sum_i \lambda_i = \sum_i \mu_i = 1$. Étant donné que les points $A_0, \dots, A_r$ forment un repère, les vecteurs $\overrightarrow{A_0A_i}$ sont linéairement indépendants (point (5) de la proposition 8.43) et donc $\lambda_i = \mu_i$ pour $i = 1, \dots, r$. La condition de somme des points égale à 1 impose alors immédiatement $\lambda_0 = \mu_0$. □

Définition 8.47.
Soit un espace affine $\mathcal{E}$ de dimension n. Soient des points affinement indépendants $A_1, \dots, A_n$. Pour $M \in \mathcal{E}$, la proposition 8.46 indique qu'il existe un unique choix de λ_i tel que

$$\begin{cases} \sum_i \lambda_i = 1 & (8.102a) \\ \sum_i \lambda_i \overrightarrow{MA_i} = 0. & (8.102b) \end{cases}$$

Ces λ_i sont les ***coordonnées barycentriques*** *de M dans le repère $\{A_i\}_{i=1,\dots,n}$.*

8.48.
Soit $\mathbb{R}^2$ et les points non alignés A, B, C. Les coordonnées barycentriques (α, β, γ) dans ce système correspondent à l'unique $X \in \mathbb{R}^2$ tel que

$$\alpha\overrightarrow{XA} + \beta\overrightarrow{XB} + \gamma\overrightarrow{XC} = 0. \tag{8.103}$$

Exemple 8.49.
Soient les points $A = (3, 1)$, $B = (-1, 2)$ et $C = (0, -1)$ dans $\mathbb{R}^2$. Nous allons montrer qu'il forment un repère affine de $\mathbb{R}^2$. L'espace engendré par ces trois points est l'espace des

$$A + \alpha\overrightarrow{AB} + \beta\overrightarrow{AC}, \tag{8.104}$$

et la direction correspondante est l'espace vectoriel donné par $\alpha\overrightarrow{AB} + \beta\overrightarrow{AC}$ qui est de dimension deux. Donc l'espace affine engendré par A, B et C est de dimension 2. △

Exemple 8.50.
Dans le repère (A, B, C), quel est le point de coordonnées barycentriques $(\frac{1}{6}, \frac{1}{3}, \frac{1}{2})$? D'abord nous vérifions que

$$\frac{1}{6} + \frac{1}{3} + \frac{1}{2} = 1. \tag{8.105}$$

6. Définition 8.28.

Ensuite nous cherchons $X \in \mathbb{R}^2$ tel que

$$\frac{1}{6}\overrightarrow{AX} + \frac{1}{3}\overrightarrow{BX} + \frac{1}{2}\overrightarrow{CX} = 0, \tag{8.106}$$

c'est-à-dire

$$\frac{1}{6}\begin{pmatrix} x-3 \\ y-1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} x+1 \\ y-2 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} x \\ y+1 \end{pmatrix} = 0. \tag{8.107}$$

Nous trouvons immédiatement $x = 1/6$ et $y = 1/3$. Le point cherché est donc le point $\begin{pmatrix} 1/6 \\ 1/3 \end{pmatrix}$. △

Lemme 8.51 ([1])**.**
Une application affine $f\colon \mathcal{E} \to \mathcal{E}$ qui préserve les points d'une base affine de $\mathcal{E}$ est l'identité.

Démonstration. Une base affine de $\mathcal{E}$ consiste en $n+1$ points $\{A_0, \dots, A_n\}$ affinement indépendants. Nous utilisons la proposition 8.45 pour dire que $\big(A_0, \{\overrightarrow{A_0A_i}\}_{i=1,\dots,n}\big)$ est un repère cartésien.

En utilisant la formule du lemme 8.10,

$$f(A_i) = f(A_0 + \overrightarrow{A_0A_i}) = f(A_0) + u(\overrightarrow{A_0A_i}). \tag{8.108}$$

Donc $A_i = A_0 + u(\overrightarrow{A_0A_i})$, ce qui signifie que

$$u(\overrightarrow{A_0A_i}) = \overrightarrow{A_0A_i} \tag{8.109}$$

Par ailleurs, tout point M [7] de $\mathcal{E}$ peut être écrit sous la forme

$$M = A_0 + \sum_i \lambda_i \overrightarrow{A_0A_i}. \tag{8.110}$$

En appliquant M, et en utilisant (8.109),

$$f(M) = f(A_0) + \sum_i \lambda_i u(\overrightarrow{A_0A_i}) = A_0 + \sum_i \lambda_i \overrightarrow{A_0A_i} = M. \tag{8.111}$$

Donc tout point de $\mathcal{E}$ est fixé par f, ce qui signifie que f est l'identité. □

8.8.1 Équation de droite

Soit $\mathcal{E}$ un espace affine de dimension trois muni d'un repère barycentrique. Une droite est donnée par trois nombres : $D = D(a, b, c)$ est l'ensemble des points dont les coordonnées barycentriques (normalisées) (x, y, z) vérifient $ax + by + cz = 0$. C'est un espace de dimension un parce qu'il y a aussi la condition $x + y + z = 1$.

La droite $D(1, 1, 1,)$ n'existe pas parce que ce serait $x + y + z = 0$, qui est incompatible avec $x + y + z = 1$.

Les droites $D(a, b, c)$ et $D(a', b', c')$ s'intersectent selon les solutions du système

$$\begin{cases} x + y + z = 1 & (8.112a) \\ ax + by + cz = 0 & (8.112b) \\ a'x + b'y + c'z = 0 & (8.112c) \end{cases}$$

Donc deux droites affines ont un unique point d'intersection si et seulement si

$$d = \begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a' & b' & c' \end{vmatrix} \neq 0. \tag{8.113}$$

Elles seront parallèles ou confondues si et seulement si $d = 0$.

7. Même les points qui ne s'appellent pas « M » en fait.

8.8.2 Associativité, coordonnées barycentriques dans un triangle

Lemme 8.52 ([171]).
Soient trois points non alignés A, B, C ainsi que des nombres α, β, γ tels que $\alpha+\beta\neq 0$ et $\alpha+\beta+\gamma\neq 0$.

Soit H le barycentre du système $\{(A,\alpha),(B,\beta)\}$ et G le barycentre de $\{(A,\alpha),(B,\beta),(C,\gamma)\}$. Alors G est barycentre de $\{(H,\alpha+\beta),(C,\gamma)\}$.

Démonstration. Vues les définitions de H et G nous avons

$$\alpha\overrightarrow{HA}+\beta\overrightarrow{HB}=0 \tag{8.114a}$$
$$\alpha\overrightarrow{GA}+\beta\overrightarrow{GB}+\gamma\overrightarrow{GC}=0. \tag{8.114b}$$

En utilisant les relations de Chasles nous introduisons H dans la seconde relation :

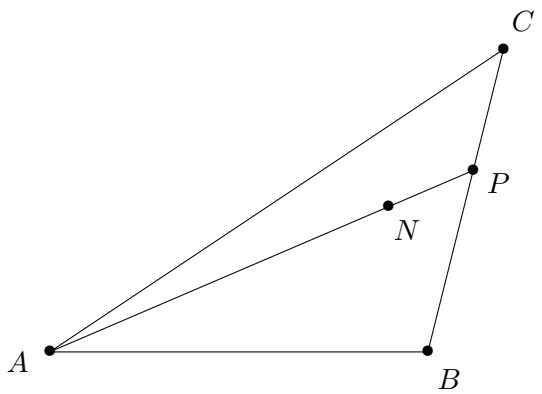
$$\alpha(\overrightarrow{GH}+\overrightarrow{HA})+\beta(\overrightarrow{GH}+\overrightarrow{HB})+\gamma\overrightarrow{GC}=0 \tag{8.115a}$$
$$(\alpha+\beta)\overrightarrow{GH}+\underbrace{\alpha\overrightarrow{HA}+\beta\overrightarrow{HB}}_{=0}+\gamma\overrightarrow{GC}=0 \tag{8.115b}$$
$$(\alpha+\beta)\overrightarrow{GH}+\gamma\overrightarrow{GC}=0. \tag{8.115c}$$

□

Les coordonnées barycentriques dans un triangle (et plus généralement en fait) permettent de faire des projections.

Proposition 8.53.
Soient trois points non alignés A, B, C ainsi qu'un point N de coordonnées barycentriques (α,β,γ) dans le système (A,B,C). Si P est l'intersection $(AN)\cap(BC)$ alors les coordonnées de P sont $(0,\beta,\gamma)$.

Démonstration. Un dessin de la situation :



Dire que les coordonnées de N sont (α,β,γ) signifie que

$$\alpha\overrightarrow{NA}+\beta\overrightarrow{NB}+\gamma\overrightarrow{NC}=0. \tag{8.116}$$

Nous voudrions montrer que le point P est bien le point de coordonnées $(0,\beta,\gamma)$. Soit donc le point P tel que

$$\beta\overrightarrow{PB}+\gamma\overrightarrow{PC}=0 \tag{8.117}$$

et montrons que ce point est l'intersection $(BC)\cap(NA)$.

D'abord la relation (8.117) nous dit immédiatement que P est sur la droite (BC). Ensuite, en utilisant les relations de Chasles pour introduire N :

$$\beta(\overrightarrow{PN}+\overrightarrow{NB})+\gamma(\overrightarrow{PN}+\overrightarrow{NC})=0. \tag{8.118}$$

Nous remplaçons $\beta\overrightarrow{NB} + \gamma\overrightarrow{NC}$ par $-\alpha\overrightarrow{NA}$ pour obtenir :

$$(\beta + \gamma)\overrightarrow{PN} - \alpha\overrightarrow{NA} = 0. \tag{8.119}$$

Cela montre que les vecteurs $\overrightarrow{PN}$ et $\overrightarrow{NA}$ sont colinéaires, et donc que P, N et A sont alignés. □

8.9 Applications affines sur $\mathbb{R}^n$

Soit $v \in \mathbb{R}^n$; nous notons $\tau_v \colon \mathbb{R}^n \to \mathbb{R}^n$ la translation donnée par $\tau_v(x) = x + v$. Le groupe de toutes les translations de $\mathbb{R}^n$ est noté $T(n)$ et est isomorphe au groupe abélien $(\mathbb{R}^n, +)$.

Nous avons déjà discuté de la structure d'un espace vectoriel (en particulier $\mathbb{R}^n$) comme espace affine en 8.5.

Lemme 8.54.
Décomposition d'une application affine.

(1) Une application $f \colon E \to E$ est affine si et seulement si il existe $v \in E$ et une application linéaire α sur E telle que $f = \tau_v \circ \alpha$.

(2) Dans ce cas, le choix de (v, α) est unique.

(3) Si f est bijective, alors α est bijective.

Démonstration. Nous supposons d'abord que f est affine. Alors il existe une application linéaire u_f sur E telle que

$$f(M + x) = f(M) + u_f(x) = (\tau_{f(M)} + u_f)(x) \tag{8.120}$$

pour tout x et M. De plus l'application u_f ne dépend ni de M ni de x (c'est la proposition 8.13(1)). En posant $M = 0$ nous avons :

$$f(x) = (\tau_{f(0)} \circ u_f)(x). \tag{8.121}$$

Dans l'autre sens nous supposons avoir $v \in E$ et α linéaire sur E telles que

$$f(M) = (\tau_v \circ \alpha)(M). \tag{8.122}$$

Notons qu'il y a un abus de notation entre α qui est linéaire sur l'espace *vectoriel* E et l'application α qui est une application sur l'espace *affine* E. Cet abus est légitime parce que les deux espaces sont identiques en tant qu'ensembles. Ce qui est vraiment abuser par contre, c'est de se poser ce genre de questions.

Nous avons :

$$\begin{aligned} f(M + x) = \tau_v\big(\alpha(M + x)\big) = \alpha(M + x) + v &= \alpha(M) + v + \alpha(x) \\ &= (\tau_v \circ \alpha)(M) + \alpha(x) = f(M) + \alpha(x). \end{aligned} \tag{8.123}$$

Donc la fonction f vérifie la définition 8.9. La partie (1) est prouvée.

Pour prouver l'unicité de la partie (2), nous supposons que $\tau_v \circ \alpha = \tau_w \circ \alpha$. En appliquant cela à 0 nous trouvons $v = w$. Nous avons donc $\tau_v \circ \alpha = \tau_v \circ \beta$. Comme τ_v est inversible, nous en déduisons $\alpha = \beta$.

Enfin le point (3) est relativement évident du fait que τ_v, elle, est surement bijective. □

Corolaire 8.55.
Une application affine qui conserve l'origine est linéaire.

Démonstration. Conserver l'origine demande de poser $v = 0$ dans l'expression du lemme 8.54. □

Proposition 8.56.
Soit une application affine $f \colon \mathbb{R}^n \to \mathbb{R}^n$. L'ensemble des points fixes

$$\mathrm{Fix}(f) = \{x \in \mathbb{R}^n \text{ tel que } f(x) = x\} \tag{8.124}$$

est soit vide soit un sous-espaces affine de $\mathbb{R}^n$.

Démonstration. Soit $f = \tau_v \circ \alpha$; nous avons $x \in \mathrm{Fix}(f)$ si et seulement si

$$x = \tau_v\big(\alpha(x)\big) = \alpha(x) + v, \tag{8.125}$$

autrement dit, en considérant l'application linéaire $\beta = \mathrm{Id} - \alpha$, si et seulement si $\beta(x) = v$. Nous écrivons $\mathrm{Fix}(f) = \beta^{-1}(v)$. Supposons que ce soit non vide et considérons $x_0 \in \beta^{-1}(v)$. Nous avons

$$\begin{aligned}
\beta^{-1}(v) &= \{x \in \mathbb{R}^n \text{ tel que } \beta(x) = \beta(x_0)\} && (8.126a)\\
&= \{x \text{ tel que } \beta(x - x_0) = 0\} && (8.126b)\\
&= \{x \text{ tel que } x - x_0 \in \ker(\beta)\} && (8.126c)\\
&= \ker(\beta) + x_0 && (8.126d)\\
&= \tau_{x_0}\big(\ker(\beta)\big). && (8.126e)
\end{aligned}$$

Mais comme $\ker(\beta)$ est un sous-espace vectoriel, $\beta^{-1}(v)$ est le translaté d'un sous-espace vectoriel, c'est-à-dire un sous-espace affine. □

8.9.1 Structure de groupe pour les applications affines

Proposition-définition 8.57 ([1]).
L'ensemble des applications affines bijectives de $\mathbb{R}^n$ forment un groupe pour la composition. Les lois de groupe sont données par les formules suivantes :

(1) Le neutre est l'identité.

(2) Le produit est donné par

$$(\tau_v \circ \alpha)(\tau_w \circ \beta) = \tau_{\alpha(w)+v} \circ \alpha\beta. \tag{8.127}$$

(3) L'inverse est donné par

$$(\tau_v \circ \alpha)^{-1} = \tau_{-\alpha^{-1}(v)} \circ \alpha^{-1}. \tag{8.128}$$

Ce groupe est noté $\mathrm{Aff}(\mathbb{R}^n)$.

Démonstration. Pour l'identité, oui, composer par l'identité est neutre.

Le fait que la formule (8.127) soit vraie est un simple calcul :

$$(\tau_v \circ \alpha) \circ (\tau_w \circ \beta)(x) = (\alpha\beta)(x) + \alpha(w) + v = \big(\tau_{\alpha(w)+v} \circ \alpha\beta\big)x. \tag{8.129}$$

Le fait que la formule (8.127) donne bien un produit pour tous les éléments de $\mathrm{Aff}(\mathbb{R}^n)$ est le lemme 8.54.

En ce qui concerne l'inverse, c'est un calcul :

$$\begin{aligned}
(\tau_{-\alpha^{-1}(v)}\alpha^{-1})(\tau_v\alpha)(x) &= (\tau_{-\alpha^{-1}(v)\alpha^{-1}})\big(\alpha(x) + v\big) && (8.130a)\\
&= \tau_{-\alpha^{-1}(v)}\big(x + \alpha^{-1}(v)\big) && (8.130b)\\
&= x. && (8.130c)
\end{aligned}$$

□

Si $f\colon \mathbb{R}^n \to \mathbb{R}^n$ est une application affine, la proposition 8.54 affirme qu'il existe une application linéaire u telle que

$$f(x + y) = f(x) + u(y). \tag{8.131}$$

En écrivant cela pour $x = 0$,

$$f(y) = f(0) + u(y), \tag{8.132}$$

ou encore $f = \tau_{f(0)} \circ u$.

Proposition 8.58.
L'ensemble $\mathrm{Aff}(\mathbb{R}^n)$ *est isomorphe au produit semi-direct*[8]

$$\mathrm{Aff}(\mathbb{R}^n) \simeq T(n) \times_{\boldsymbol{Ad}} \mathrm{GL}(n,\mathbb{R}) \tag{8.133}$$

où $\boldsymbol{Ad}$ *est l'action adjointe, c'est-à-dire*

$$\begin{aligned} \boldsymbol{Ad}\colon \mathrm{GL}(n,\mathbb{R}) &\to \mathrm{Aut}\big(T(n)\big) \\ \alpha &\mapsto \big(\tau_v \mapsto \alpha\circ\tau_v\circ\alpha^{-1}\big). \end{aligned} \tag{8.134}$$

Démonstration. L'application que nous allons montrer être un isomorphisme est ψ qui à $f=\tau_v\circ\alpha$ fait correspondre le couple $(\tau_v,\alpha)\in T(n)\times\mathrm{GL}(n,\mathbb{R})$.

Égalité d'ensembles Il faut que $\mathrm{Aff}(\mathbb{R}^n)$ soit en bijection avec $T(n)\times\mathrm{GL}(n,\mathbb{R})$. En effet si $f\in\mathrm{Aff}(\mathbb{R}^n)$, la décomposition $f=\tau_v\circ\alpha$ est unique. D'abord en appliquant à 0, $f(0)=\tau_v\big(\alpha(v)\big)=v$. Donc v est fixé par la valeur de $f(0)$. Ensuite $\alpha=f\circ\tau_v^{-1}$, donc α fixé.

L'action adjointe fonctionne Il faut vérifier que $\alpha\circ\tau_v\circ\alpha^{-1}$ est bien dans $T(n)$. Pour cela, en agissant sur $x\in\mathbb{R}^n$ nous trouvons

$$\alpha\tau_v\alpha^{-1}(x)=\alpha\big(\alpha^{-1}(x)+v\big)=x+\alpha(v)=\tau_{\alpha(v)}(x). \tag{8.135}$$

Le fait que $\mathbf{Ad}(\alpha)$ soit un automorphisme est toujours correct.

Morphisme Il faut vérifier que l'application ψ est un morphisme de groupe. D'abord la loi de groupe sur $\mathrm{Aff}(\mathbb{R}^n)$ est donnée par

$$(\tau_v\circ\alpha)\circ(\tau_w\circ\beta)=\tau_{v+\alpha(w)}\circ(\alpha\circ\beta). \tag{8.136}$$

Ensuite le loi de groupe de le produit semi-direct est donnée par

$$(\tau_v,\alpha)\cdot(\tau_w,\beta)=\big(\tau_v\,\mathbf{Ad}(\alpha)\tau_w,\alpha\beta\big)=\big(\tau_v\tau_{\alpha(w)},\alpha\beta\big)=\big(\tau_{\alpha(w)+v},\alpha\beta\big). \tag{8.137}$$

Nous avons donc bien

$$\psi\big((\tau_v,\beta)\cdot(\tau_w,\beta)\big)=\psi(\tau_v,\beta)\circ\psi(\tau_w,\beta). \tag{8.138}$$

□

8.10 Isométries

Définition 8.59 (Isométrie d'espace affine)**.**
Si $\mathcal{E}$ *est un espace affine muni d'une distance* d*, une isométrie de* $\mathcal{E}$ *est une application* $f\colon\mathcal{E}\to\mathcal{E}$ *préservant* d*.*

Notons que toutes les applications affines ne sont pas des isométries : par exemple les homothéties.

Proposition 8.60.
Si $\mathcal{E}$ *est modelé sur un espace euclidien* $(E,\|.\|)$ *alors la formule*

$$d(A,B)=\|\overrightarrow{AB}\| \tag{8.139}$$

définit une distance sur $\mathcal{E}$*.*

Démonstration. Étant donné la proposition 8.3, la formule a un sens parce qu'à A et B donnés dans $\mathcal{E}$, il est associé un unique vecteur $\overrightarrow{AB}\in E$. □

Nous parlons d'isométries affines ou linéaires dans le thème 66.

8. Définition 2.79.

Chapitre 9

Espaces vectoriels (encore)

9.1 Déterminants

9.1.1 Formes multilinéaires alternées

Définition 9.1.
Soit E, un $\mathbb{K}$-espace vectoriel. Une forme linéaire ***alternée*** *sur E est une application linéaire $f\colon E \to \mathbb{K}$ telle que $f(v_1, \dots, v_k) = 0$ dès que $v_i = v_j$ pour certains $i \neq j$.*

Lemme 9.2.
Une forme linéaire alternée est antisymétrique. Si $\mathbb{K}$ est de caractéristique différente de 2, alors une forme antisymétrique est alternée.

Démonstration. Soit f une forme alternée ; quitte à fixer toutes les autres variables, nous pouvons travailler avec une 2-forme et simplement montrer que $f(x, y) = -f(y, x)$. Pour ce faire nous écrivons

$$0 = f(x + y, x + y) = f(x, x) + f(x, y) + f(y, x) + f(y, y) = f(x, y) + f(y, x). \tag{9.1}$$

Pour la réciproque, si f est antisymétrique, alors $f(x, x) = -f(x, x)$. Cela montre que $f(x, x) = 0$ lorsque $\mathbb{K}$ est de caractéristique différente de deux. □

Proposition 9.3 ([172])**.**
Soit E, un $\mathbb{K}$-espace vectoriel de dimension n, où la caractéristique de $\mathbb{K}$ n'est pas deux. L'espace des n-formes multilinéaires alternées sur E est de $\mathbb{K}$-dimension 1.

Démonstration. Soient $\{e_i\}$, une base de E, une n-forme linéaire alternée $f\colon E \to \mathbb{K}$ ainsi que des vecteurs $(v_1, \dots, v_n)$ de E. Nous pouvons les écrire dans la base

$$v_j = \sum_{i=1}^{n} \alpha_{ij} e_i \tag{9.2}$$

et alors exprimer f par

$$f(v_1, \dots, v_n) = f\Big(\sum_{i_1=1}^{n} \alpha_{1i_1} e_{i_1}, \dots, \sum_{i_n=1}^{n} \alpha_{ni_n} e_{i_n}\Big) \tag{9.3a}$$

$$= \sum_{i,j} \alpha_{1i_1} \dots \alpha_{ni_n} f(e_{i_1}, \dots, e_{i_n}). \tag{9.3b}$$

Étant donné que f est alternée, les seuls termes de la somme sont ceux dont les i_k sont tous différents, c'est-à-dire ceux où $\{i_1, \dots, i_n\} = \{1, \dots, n\}$. Il y a donc un terme par élément du groupe des permutations S_n et

$$f(v_1, \dots, v_n) = \sum_{\sigma \in S_n} \alpha_{\sigma(1)1} \dots \alpha_{\sigma(n)n} f(e_{\sigma(1)}, \dots, e_{\sigma(n)}). \tag{9.4}$$

En utilisant encore une fois le fait que la forme f soit alternée, $f = f(e_1, \dots, e_n)\Pi$ où

$$\Pi(v_1, \dots, v_n) = \sum_{\sigma \in S_n} \epsilon(\sigma)\alpha_{\sigma(1)1} \dots \alpha_{\sigma(n)n}. \tag{9.5}$$

Pour rappel, la donnée des v_i est dans les nombres α_{ij}.

L'espace des n-formes alternées est donc *au plus* de dimension 1. Pour montrer qu'il est exactement de dimension 1, il faut et suffit de prouver que Π est alternée. Par le lemme 9.2, il suffit de prouver que cette forme est antisymétrique[1].

Soient donc $v_1, \dots, v_n$ tels que $v_i = v_j$. En posant $\tau = (1i)$ et $\tau' = (2j)$ et en sommant sur $\sigma\tau\tau'$ au lieu de σ, nous pouvons supposer que $i = 1$ et $j = 2$. Montrons que $\Pi(v, v, v_3, \dots, v_n) = 0$ en tenant compte que $\alpha_{i1} = \alpha_{i2}$:

$$\begin{aligned}
\Pi(v, v, v_3, \dots, v_n) &= \sum_{\sigma \in S_n} \epsilon(\sigma)\alpha_{\sigma(1)1}\alpha_{\sigma(2)2}\alpha_{\sigma(3)3} \dots \alpha_{\sigma(n)n} && (9.6a)\\
&= \sum_{\sigma \in S_n} \epsilon(\sigma\tau)\alpha_{\sigma\tau(1)1}\alpha_{\sigma\tau(2)2}\alpha_{\sigma\tau(3)3} \dots \alpha_{\sigma\tau(n)n} \qquad \text{où } \tau = (12) && (9.6b)\\
&= -\sum_{\sigma \in S_n} \epsilon(\sigma)\alpha_{\sigma(1)1}\alpha_{\sigma(2)2}\alpha_{\sigma(3)3} \dots \alpha_{\sigma(n)n} && (9.6c)\\
&= -\Pi(v, v, v_3, \dots, v_n). && (9.6d)
\end{aligned}$$

□

9.1.2 Déterminant d'une famille de vecteurs

Nous considérons un corps $\mathbb{K}$ et l'espace vectoriel E de dimension n sur $\mathbb{K}$.

Définition 9.4 (Déterminant d'une famille de vecteurs[6])**.**
*Le **déterminant** de la famille de vecteurs $(v_1, \dots, v_n)$ dans la base B est l'élément de $\mathbb{K}$*

$$\det_{(e_1,\dots,e_n)}(v_1, \dots, v_n) = \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^{n} e^*_{\sigma(i)}(v_i) \tag{9.7}$$

où

- *la somme porte sur le groupe symétrique,*
- *le nombre $\epsilon(\sigma)$ est la signature de la permutation σ,*
- *les éléments $\{e_i\}$ forment la base canonique de $\mathbb{K}^n$.*
- *les éléments $\{e_i^*\}$ sont la base duale de $\{e_i\}$.*

Nous le notons $\det_{(e_1,\dots,e_n)}(v_1, \dots, v_n)$.

9.5.
La base $\{e_i\}$ est la base canonique de $\mathbb{K}^n$, et l'élément e_k^* est la forme linéaire définie par

$$\begin{aligned}
e_k^* \colon \mathbb{K}^n &\to \mathbb{K}\\
\sum_i x_i e_i &\mapsto x_k.
\end{aligned} \tag{9.8}$$

Il n'est pas sous-entendu que $\mathbb{K}^n$ ait un produit scalaire. Il n'est donc pas autorisé de dire que $\{e_i\}$ est une base orthonormée et que $e_k^*(x) = \langle e_k, x\rangle$. Ce genre d'égalités sont vraies dans le cas $\mathbb{K} = \mathbb{R}$, mais n'ont pas de sens en général.

Le lemme 9.9 va un peu parler du cas où $\mathbb{K}^n$ est muni d'une base orthonormée.

Lemme 9.6 ([6])**.**
Les propriétés du déterminant. Soit B une base de E.

1. C'est ici que joue l'hypothèse sur la caractéristique de $\mathbb{K}$.

(1) *L'application* $\det_B \colon E^n \to \mathbb{K}$ *est* n*-linéaire.*

(2) *L'application* $\det_B \colon E^n \to \mathbb{K}$ *est* n*-linéaire est antisymétrique et alternée*[2].

(3) *Pour toute base,* $\det_B(B) = 1$.

(4) *Le déterminant ne change pas si on remplace un vecteur par une combinaison linéaire des autres :*

$$\det_B(v_1, \dots, v_n) = \det_B\big(v_1 + \sum_{s=2}^{n} a_s v_s, v_2, \dots, v_n\big). \tag{9.9}$$

(5) *Si on permute les vecteurs,*

$$\det_B(v_1, \dots, v_n) = \epsilon(\sigma) \det_B(v_{\sigma(1)}, \dots, v_{\sigma(n)}). \tag{9.10}$$

(6) *Si* B' *est une autre base :*

$$\det_B = \det_B(B') \det_{B'} \tag{9.11}$$

(7) *Nous avons aussi la formule* $\det_B(B') \det_{B'}(B) = 1$.

(8) *Les vecteurs* $\{v_1, \dots, v_n\}$ *forment une base si et seulement si* $\det_B(v_1, \dots, v_n) \neq 0$.

Démonstration. Point par point.

(1) En posant $v_1 = x_1 + \lambda x_2$ nous avons

$$\det_B(x_1 + \lambda x_2, v_2, \dots, v_n) = \sum_\sigma \epsilon(\sigma) \prod_{i=1}^{n} e^*_{\sigma(i)}(v_i) \tag{9.12a}$$

$$= \sum_\sigma \epsilon(\sigma) \Big(e^*_{\sigma(1)}(x_1 + \lambda x_2)\Big) \prod_{i=2}^{n} e^*_{\sigma(i)}(v_i). \tag{9.12b}$$

À partir de là, la linéarité de $e^*_{\sigma(1)}$ montre que $\det_B$ est linéaire en son premier argument. Pour les autres arguments, le même calcul tient.

(2) Nous prouvons à présent que det est alternée. Si votre corps est de caractéristique différente de deux, vous pouvez lire la proposition 9.7.

Supposons $v_k = v_l$, et considérons la permutation $\beta = (k, l)$. Nous savons par la proposition 5.32 que $S_n = A_n \cup A_n\beta$. Cela nous permet de décomposer la somme sur S_n en deux parties :

$$\sigma_{\sigma \in S_n} (-1)^\sigma \prod_i \epsilon^*_{\sigma(i)}(v_i) = \sum_{\sigma \in A_n} (-1)^\sigma \prod_i \epsilon^*_{\sigma(i)}(v_i) + \sum_{\sigma \in A_n} (-1)^{\sigma\beta} \prod_i \epsilon^*_{(\sigma\beta)(i)}(v_i). \tag{9.13}$$

D'abord $(-1)^\sigma = 1$ et $(-1)^{\sigma\beta} = -1$. Ensuite, pour un $\sigma \in A_n$ donné, nous avons

$$\prod_i \epsilon^*_{(\sigma\beta)(i)}(v_i) = \epsilon^*_{(\sigma\beta)(k)}(v_k)\epsilon^*_{(\sigma\beta)(l)}(v_l) \prod i \overset{i \neq k}{\neq} l \; \epsilon^*_{(\sigma\beta)(i)}(v_i) \tag{9.14a}$$

$$= \epsilon^*_{\sigma(l)}(v_k)\epsilon^*_{\sigma(k)}(v_l) \prod_{\substack{i \neq k \\ i \neq l}} \epsilon^*_{\sigma(i)}(v_i) \tag{9.14b}$$

$$= \epsilon^*_{\sigma(l)}(v_l)\epsilon^*_{\sigma(k)}(v_k) \prod_{\substack{i \neq k \\ i \neq l}} \epsilon^*_{\sigma(i)}(v_i) \tag{9.14c}$$

$$= \prod_i \epsilon^*_{\sigma(i)}(v_i). \tag{9.14d}$$

Donc les deux termes de la somme (9.13) ne diffèrent que par un signe. Elle est donc nulle, et la forme déterminant est alternée.

La fonction det est antisymétrique parce que alternée, voir le lemme 9.2.

2. Alternée, définition 9.1. En caractéristique 2, alternée n'est pas équivalent à symétrique.

(3) Nous avons

$$\det_B(B) = \sum_{\sigma\in S_n} \epsilon(\sigma) \prod_{i=1}^{n} \underbrace{e^*_{\sigma(i)}(e_i)}_{=\delta_{\sigma(i),i}}. \tag{9.15}$$

Si σ n'est pas l'identité, le produit contient forcément un facteur nul. Il ne reste de la somme que $\sigma = \mathrm{Id}$ et le résultat est 1.

(4) Vu que $\det_B$ est linéaire en tous ses arguments,

$$\det_B\left(v_1 + \sum_{s=2}^{n} a_s v_s, v_2, \dots, v_n\right) = \det_B(v_1, \dots, v_n) + \sum_{s=2}^{n} a_s \det_B(v_s, v_2, \dots, v_n). \tag{9.16}$$

Chacun des termes de la somme est nul parce qu'il y a répétition de v_s parmi les arguments alors que la forme est alternée.

(5) Nous devons calculer $\det_B(v_{\sigma(1)}, \dots, v_{\sigma(n)})$, et pour y voir plus clair nous posons $w_i = v_{\sigma(i)}$. Alors :

$$\begin{aligned}
\det_B(v_{\sigma(1)}, \dots, v_{\sigma(n)}) &= \sum_{\sigma'} \epsilon(\sigma') \prod_{i=1}^{n} e^*_{\sigma'(i)}(w_i) && (9.17a)\\
&= \sum_{\sigma'} \epsilon(\sigma') \prod_{i=1}^{n} e^*_{\sigma'(i)}(v_{\sigma(i)}) && (9.17b)\\
&= \sum_{\sigma'} \epsilon(\sigma') \prod_{i=1}^{n} e^*_{\sigma^{-1}\sigma'(i)}(v_i) && (9.17c)\\
&= \sum_{\sigma'} \epsilon(\sigma\sigma') \prod_{i=1}^{n} e^*_{\sigma'(i)}(v_i) && (9.17d)\\
&= \epsilon(\sigma) \det_B(v_1, \dots, v_n). && (9.17e)
\end{aligned}$$

Justifications : nous avons d'abord modifié l'ordre des éléments du produit et ensuite l'ordre des éléments de la somme. Nous avons ensuite utilisé le fait que $\epsilon\colon S_n \to \{0, 1\}$ était un morphisme de groupe (proposition 2.76).

(6) Étant donné que l'espace des formes multilinéaires alternées est de dimension 1, il existe un $\lambda \in \mathbb{K}$ tel que $\det_B = \lambda \det_{B'}$. Appliquons cela à B' :

$$\det_B(B') = \lambda \det_{B'}(B'), \tag{9.18}$$

donc $\lambda = \det_B(B')$.

(7) Il suffit d'appliquer l'égalité précédente à B en nous souvenant que $\det_B(B) = 1$.

(8) Si $B' = \{v_1, \dots, v_n\}$ est une base alors $\det_B(B') \neq 0$, sinon il n'est pas possible d'avoir $\det_B(B') \det_{B'}(B) = 1$.

À l'inverse, si B' n'est pas une base, c'est que $\{v_1, \dots, v_n\}$ est liée par la proposition 4.16. Il y a donc moyen de remplacer un des vecteurs par une combinaison linéaire des autres. Le déterminant s'annule alors.

□

Proposition 9.7.
Si la caractéristique du corps de base n'est pas deux, le déterminant est antisymétrique et alterné.

Démonstration. Si la caractéristique du corps de base n'est pas deux, une forme antisymétrique est alternée (lemme 9.2).

Pour prouver que le déterminant est antisymétrique, remarquez que permuter v_k et v_l revient à calculer le nombre $\det_B(v_{\sigma_{kl}(1)}, \dots, v_{\sigma_{kl}(n)})$ au lieu de $\det_B(v_1, \dots, v_n)$. Cela revient à changer la somme $\sum_\sigma$ en $\sum_{\sigma\circ\sigma_{kl}}$. Cela ajoute 1 à $\epsilon(\sigma)$ vu que l'on ajoute une permutation.

Donc le déterminant est antisymétrique. Nous en déduisons qu'il est alterné parce que, en permutant trivialement v_1 et v_1, nous obtenons $\det_B(v_1, v_1) = -\det_B(v_1, v_1)$. Si le corps est de caractéristique différente de deux, cela implique que $\det_B(v_1, v_1) = 0$. □

D'après la proposition 9.3, il existe une unique forme n-linéaire alternée égale à 1 sur B, et c'est $\det_B \colon E^n \to \mathbb{K}$.

9.1.3 Déterminant d'un endomorphisme

L'interprétation géométrique du déterminant en termes d'aires et de volumes est donnée après la théorème 13.254.

Lemme-définition 9.8.
Si $f \colon E \to E$ est un endomorphisme, et si les parties B et B' sont deux bases, alors

$$\det_B\big(f(B)\big) = \det_{B'}\big(f(B')\big). \tag{9.19}$$

*Ce nombre, indépendant de la base choisie est nommé le **déterminant** de f et est noté $\det(f)$.*

Démonstration. L'application

$$\begin{aligned} \varphi \colon E^n &\to \mathbb{K} \\ v_1, \dots, v_n &\mapsto \det_B\big(f(v_1), \dots, f(v_n)\big) \end{aligned} \tag{9.20}$$

est n-linéaire et alternée ; il existe donc $\lambda \in \mathbb{K}$ tel que $\varphi = \lambda \det_B$. En appliquant cela à B :

$$\det_B\big(f(B)\big) = \lambda \det_B(B) = \lambda. \tag{9.21}$$

Nous avons donc déjà prouvé que $\lambda = \det_B\big(f(B)\big)$, c'est-à-dire

$$\det_B\big(f(v)\big) = \det_B\big(f(B)\big) \det_B(v). \tag{9.22}$$

Nous allons maintenant introduire B' là où il y a du v en utilisant les formules (9.11) :

$$\det_B\big(f(v)\big) = \det_B(B') \det_{B'}\big(f(v)\big) \tag{9.23a}$$

$$\det_B(v) = \det_B(B') \det_{B'}(v). \tag{9.23b}$$

Nous obtenons

$$\det_{B'}\big(f(v)\big) = \det_B\big(f(B)\big) \det_{B'}(v). \tag{9.24}$$

Et on applique cela à $v = B'$:

$$\det_{B'}\big(f(B')\big) = \det_B\big(f(B)\big) \underbrace{\det_{B'}(B')}_{=1}. \tag{9.25}$$

□

Couplé à la formule (9.7), nous pouvons écrire la formule pratique à utiliser le plus souvent.

Lemme 9.9.
Soit un espace vectoriel euclidien[3] *E sur le corps $\mathbb{K}$. Si $\{e_i\}_{i=1,\dots,n}$ est une base orthonormée de E et si $f \colon E \to E$ est un endomorphisme, alors*

$$\det(f) = \sum_{\sigma \in S_n} \epsilon(\sigma) \prod_{i=1}^{n} \langle e_{\sigma(i)}, f(e_i) \rangle. \tag{9.26}$$

Démonstration. Nous utilisons la définition 9.8 du déterminant d'un endomorphisme $\det(f) = \det_B\big(f(B)\big)$ en prenant la liste des vecteurs $\{e_i\}$ comme B. En l'occurrence, le i^{e} vecteur de la famille B est $f(e_i)$.

Vu que la base est orthonormée, nous avons $e_k^*(v) = \langle e_k, v \rangle$ et donc aussi

$$e_{\sigma(i)}^*(v_i) = \langle e_{\sigma(i)}^*, f(e_i) \rangle. \tag{9.27}$$

□

3. C'est-à-dire qu'il possède un produit scalaire, voir la définition 7.188.

Et si vous avez tout suivi, vous aurez remarqué que les produits scalaires impliqués dans la formule (9.26) sont les éléments de la matrice de f dans la base $\{e_i\}$ parce que $\langle e_i, f(e_j)\rangle$ est la composante i de l'image de e_j par f. Si la matrice est composée en mettant en colonne les images des vecteurs de base, le compte est bon.

Proposition 9.10.
Principales propriétés géométriques du déterminant d'un endomorphisme.

(1) Si f et g sont des endomorphismes, alors $\det(f\circ g)=\det(f)\det(g)$.

(2) L'endomorphisme f est un automorphisme[4] *si et seulement si* $\det(f)\neq 0$.

(3) Si $\det(f)\neq 0$ *alors* $\det(f^{-1})=\det(f)^{-1}$.

(4) L'application $\det\colon \mathrm{GL}(E)\to\mathbb{K}\backslash\{0\}$ *est un morphisme de groupe.*

Démonstration. Point par point.

(1) Nous considérons l'application

$$\begin{aligned}\varphi\colon E^n&\to\mathbb{K}\\ v&\mapsto\det_B\big(f(v)\big).\end{aligned}\tag{9.28}$$

Comme d'habitude nous avons $\varphi(v)=\lambda\det_B(v)$. En appliquant à B et en nous souvenant que $\det_B(B)=1$ nous avons $\det_B\big(f(B)\big)=\lambda$. Autrement dit :

$$\lambda=\det(f).\tag{9.29}$$

Calculons à présent $\varphi\big(g(B)\big)$: d'une part,

$$\varphi\big(g(B)\big)=\det_B\big((f\circ g)(B)\big)\tag{9.30}$$

et d'autre part,

$$\varphi\big(g(B)\big)=\lambda\det_B\big(g(B)\big)=\lambda\det(g)\tag{9.31}$$

En égalisant et en reprenant la la valeur déjà trouvée de λ,

$$\det\big(f\circ g)(B)\big)=\det(f)\det(g),\tag{9.32}$$

ce qu'il fallait.

(2) Supposons que f soit un automorphisme. Alors si B est une base, $f(B)$ est une base. Par conséquent $\det(f)=\det_B\big(f(B)\big)\neq 0$ parce que $f(B)$ est une base (lemme 9.6(8)).

Réciproquement, supposons que $\det(f)\neq 0$. Alors si B est une base quelconque nous avons $\det_B\big(f(B)\big)\neq 0$, ce qui est uniquement possible lorsque $f(B)$ est une base. L'application f transforme donc toute base en une base et est alors un automorphisme d'espace vectoriel.

(3) Vu que le déterminant de l'identité est 1 et que f est inversible, $1=\det(f\circ f^{-1})=\det(f)\det(f^{-1})$.

□

Proposition 9.11.
Soient deux espaces vectoriels E et F de dimension finies n et m sur le corps $\mathbb{K}$ munis de bases $\{e_i\}$ et $\{f_\alpha\}$. À une matrice $A\in$
$eM(m\times n,\mathbb{K})$ nous associons l'application linéaire[5]

$$f_A(x)=\sum_{i\alpha}A_{\alpha i}x_if_\alpha.\tag{9.33}$$

Alors, en ce qui concerne les déterminants[6]*, nous avons*

4. Endomorphisme inversible, définition 4.29.
5. Dont nous avons déjà beaucoup parlé entre autres dans la proposition 4.70.
6. Définition 9.8 pour les applications linéaires et 4.74 pour les matrices.

(1) $\det(f_A) = \det(A)$

(2) $\det(f_{AB}) = \det(f_A)\det(f_B)$

Démonstration. Nous devons étudier la formule

$$\det(f_A) = \sum_{\sigma\in S_n} \epsilon(\sigma) \prod_{i=1}^{n} e^*_{\sigma(i)}\big(f_A(e_i)\big). \tag{9.34}$$

En premier lieu nous avons

$$f_A(e_i) = \sum_{jk} A_{jk}(e_i)_k e_j = \sum_j A_{ji} e_j. \tag{9.35}$$

Nous avons alors

$$e^*_{\sigma(i)}\big(f_A(e_i)\big) = \sum_j A_{ji} \underbrace{e^*_{\sigma(i)}(e_j)}_{\delta_{j\sigma(i)}} = A_{\sigma(i)i}. \tag{9.36}$$

Au final,

$$\det(f_A) = \sum_\sigma \epsilon(\sigma) \prod_{i=1}^{n} A_{\sigma(i)i} = \det(A^t) = \det(A) \tag{9.37}$$

où la dernière égalité est autorisée par le lemme 4.76.

Cela prouve la formule $\det(f_A) = \det(A)$.

En ce qui concerne la seconde formule, il s'agit de se souvenir de la proposition 4.70 qui donne $f_{AB} = f_A \circ f_B$, et ensuite de la proposition 9.10(1) qui donne $\det(f_A \circ f_B) = \det(f_A)\det(f_B)$. □

9.1.4 Déterminant de Vandermonde

Proposition 9.12 ([71]).
*Le **déterminant de Vandermonde** est le polynôme en n variables donné par*

$$V(T_1,\dots,T_n) = \det\begin{pmatrix} 1 & 1 & \dots & 1 \\ T_1 & T_2 & \dots & T_n \\ \vdots & \ddots & \ddots & \vdots \\ T_1^{n-1} & T_2^{n-1} & \dots & T_n^{n-1} \end{pmatrix} = \prod_{1\leqslant i<j\leqslant n} (T_j - T_i). \tag{9.38}$$

Notez que l'inégalité du milieu est stricte (sinon d'ailleurs l'expression serait nulle).

Démonstration. Nous considérons le polynôme

$$f(X) = V(T_1,\dots,T_{n-1},X) \in \big(\mathbb{K}[T_1,\dots,T_{n-1}]\big)[X]. \tag{9.39}$$

C'est un polynôme de degré au plus $n-1$ en X et il s'annule aux points $T_1,\dots,T_{n-1}$. Par conséquent il existe $\alpha \in \mathbb{K}[T_1,\dots,T_{n-1}]$ tel que

$$f = \alpha(X - T_{n-1})\dots(X - T_1). \tag{9.40}$$

Nous trouvons α en écrivant $f(0)$. D'une part la formule (9.40) nous donne

$$f(0) = \alpha(-1)^{n-1} T_1 \dots T_{n-1}. \tag{9.41}$$

D'autre par la définition donne

$$f(0) = \det \begin{pmatrix} 1 & \cdots & 1 & 1 \\ T_1 & & T_{n-1} & 0 \\ \vdots & & \vdots & \vdots \\ T_1^{n-1} & \cdots & T_{n-1}^{n-1} & 0 \end{pmatrix} \tag{9.42a}$$

$$= (-1)^{n-1} \det \begin{pmatrix} T_1 & \dots & T_{n-1} \\ \vdots & \ddots & \vdots \\ T_1^{n-1} & \dots & T_{n-1}^{n-1} \end{pmatrix} \tag{9.42b}$$

$$= (-1)^{n-1} T_1 \dots T_{n-1} \det \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ T_1^{n-1} & \dots & T_{n-1}^{n-1} \end{pmatrix} \tag{9.42c}$$

$$= (-1)^{n-1} T_1 \dots T_{n-1} V(T_1, \dots, T_{n-1}) \tag{9.42d}$$

En égalisant avec (9.41), nous trouvons $\alpha = V(T_1, \dots, T_{n-1})$, et donc

$$f = V(T_1, \dots, T_{n-1}) \prod_{j \leqslant n-1} (X - T_j) \tag{9.43}$$

Enfin, une récurrence montre que

$$V(T_1, \dots, T_n) = f(T_n) \tag{9.44a}$$

$$= V(T_1, \dots, T_{n-1}) \prod_{j \leqslant n-1} (T_n - T_j) \tag{9.44b}$$

$$= \prod_{k \leqslant n} \prod_{j \leqslant k-1} (T_k - T_j) \tag{9.44c}$$

$$= \prod_{1 \leqslant j < k \leqslant n} (T_i - T_j). \tag{9.44d}$$

□

Exemple 9.13.
Le déterminant de Vandermonde (proposition 9.12) est alterné, semi-symétrique et non symétrique. Le fait qu'il soit alterné est le fait qu'il soit un déterminant. Étant donné qu'il est alterné, il est semi-symétrique parce que sur A_n, nous avons $\epsilon = 1$. Étant donné qu'il est alterné, il change de signe sous l'action des éléments impairs de S_n et n'est donc pas symétrique. △

Proposition 9.14.
Un polynôme semi-symétrique $f \in \mathbb{K}[T_1, \dots, T_n]$ se décompose de façon unique en

$$f = P + VQ \tag{9.45}$$

où P et Q sont deux polynômes symétriques.

Démonstration. Nous commençons par prouver l'unicité en montrant que si $f = PVQ$ avec P et Q symétrique, alors P et Q sont donnés par des formules explicites en termes de f.

Si σ_1 et σ_2 sont deux permutations impaires de $\{1, \dots, n\}$, alors $\sigma_1 \cdot f = \sigma_2 \cdot f$ parce que l'élément $\sigma_2^{-1}\sigma_1$ est pair (proposition 2.76), de telle sorte que $\sigma_2^{-1}\sigma_1 \cdot f = f$. Nous posons donc $g = \tau \cdot f$ où τ est une permutation impaire quelconque – par exemple une transposition.

Vu que V est alternée et que τ est une transposition nous avons

$$g = \tau \cdot f = P - VQ. \tag{9.46}$$

Donc $f + g = 2P$ et $f - g = 2VQ$. Cela donne P et Q en termes de f et g, et donc l'unicité.

Attention : cela ne donne pas un moyen de prouver l'existence parce que rien ne prouve pour l'instant que $f-g$ peut effectivement être écrit sous la forme VQ, c'est-à-dire que $f-g$ soit divisible par V. C'est cela que nous allons nous atteler à démontrer maintenant.

Nous commençons par prouver que $f+g$ est symétrique et $f-g$ alterné. Si σ est une transposition,

$$\sigma \cdot (f+g) = \sigma \cdot f + \sigma\tau \cdot f = g + f \tag{9.47}$$

parce que $\sigma\tau$ est pair. De la même façon,

$$\sigma \cdot (f-g) = g - f = \epsilon(\sigma)(f-g). \tag{9.48}$$

Dans les deux cas nous concluons en utilisant le fait que toute permutation est un produit de transpositions (proposition 2.73) et que ϵ est un homomorphisme.

Soient maintenant deux entiers $h<k$ dans $\{1,\dots,n\}$ et l'anneau

$$\big(\mathbb{K}[T_1,\dots,\hat{T}_k,\dots,T_n]\big)[T_k]. \tag{9.49}$$

Cet anneau contient le polynôme $T_k - T_h$ où T_k est la variable et T_h est un coefficient. Nous faisons la division euclidienne de $f-g$ par $T_k - T_h$ parce que nous avons dans l'idée de faire arriver le déterminant de Vandermonde et donc le produit de toutes les différences $T_k - T_h$:

$$f - g = (T_k - T_h)q + r \tag{9.50}$$

où $\deg_{T_k} r < 1$, c'est-à-dire que r ne dépends pas de T_k. Nous revoyons maintenant l'égalité (9.50) dans $\mathbb{K}[T_1,\dots,T_n]$ et nous y appliquons la transposition τ_{kh}. Nous savons que $\tau_{kh}(f-g) = -(f-g)$ et $\tau_{kh}(T_k - T_h) = -(T_k - T_h)$, et donc

$$-(f-g) = -(T_k - T_h)\tau_{kh} \cdot q + \tau_{kh} \cdot r \tag{9.51}$$

où $\tau_{kh} \cdot r$ ne dépend pas de T_h. Nous appliquons à (9.51) l'application

$$\begin{aligned} t\alpha\colon \mathbb{K}[T_1,\dots,T_n] &\to \mathbb{K}[T_1,\dots,\hat{T}_k,\dots,T_n] \\ \alpha(PT_1,\dots,\hat{T}_k,\dots,T_n) &= P(T_1,\dots,T_h,\dots,T_n). \end{aligned} \tag{9.52}$$

Cette application vérifie $\alpha\big(\tau_{kh} \cdot r\big) = \alpha(r)$ et nous avons

$$-\alpha(f-g) = \alpha(r). \tag{9.53}$$

Puis en appliquant α à la relation $f - g = (T_k - T_h)q + r$, nous trouvons

$$\alpha(f-g) = \alpha(r), \tag{9.54}$$

et par conséquent $\alpha(r) = 0$. Ici nous utilisons l'hypothèse de caractéristique différente de deux. Dire que $\alpha(r) = 0$, c'est dire que r est divisible par $T_k - T_h$, mais r étant de degré zéro en T_k, nous avons $r = 0$. Par conséquent $T_k - T_h$ divise $f-g$ pour tout $h<k$, et nous pouvons définir un polynôme Q par

$$f - g = 2Q \prod_{h<k} \prod_{k\leqslant n} (T_k - T_h) = 2Q(T_1,\dots,T_n)V(T_1,\dots,T_n), \tag{9.55}$$

où nous avons utilisé la formule du déterminant de Vandermonde de la proposition 9.12.

Étant donné que $f+g$ est un polynôme symétrique, nous allons aussi poser $f+g = 2P$ avec P symétrique.

Montrons à présent que Q est un polynôme symétrique. Soit $\sigma \in S_n$; vu que nous savons déjà que $f-g$ est alternée, nous avons

$$\sigma \cdot (f-g) = \epsilon(\sigma)(f-g) = \epsilon(\sigma)2QV, \tag{9.56}$$

Mais en appliquant σ à l'équation (9.55),

$$\sigma \cdot (f - g) = 2(\sigma \cdot V)(T_1, \dots, , T_n)(\sigma \cdot Q)(T_1, \dots, T_n) \tag{9.57a}$$
$$= 2\epsilon(\sigma)V(T_1, \dots, T_n)(\sigma \cdot Q)(T_1, \dots, T_n). \tag{9.57b}$$

Nous égalisons cela avec (9.56) et nous souvenant que l'anneau $\mathbb{K}[T_1, \dots, T_n]$ est intègre par le théorème 3.172. Ensuite nous simplifions par $2\epsilon(\sigma)V$ pour obtenir

$$Q = \sigma \cdot Q, \tag{9.58}$$

c'est-à-dire que Q est symétrique.

Au final nous avons $f + q = 2P$ et $f - g = 2VQ$ avec P et Q symétriques. En faisant la somme,

$$f = P + VQ. \tag{9.59}$$

□

9.1.5 Déterminant de Gram

Si $x_1, \dots, x_r$ sont des vecteurs d'un espace vectoriel, alors le **déterminant de Gram** est le déterminant

$$G(x_1, \dots, x_r) = \det\big(\langle x_i, x_j \rangle\big). \tag{9.60}$$

Notons que la matrice est une matrice symétrique.

Proposition 9.15.
Si F est un sous-espace vectoriel de base $\{x_1, \dots, x_n\}$ et si x est un vecteur, alors le déterminant de Gram est un moyen de calculer la distance entre x et F par

$$d(x, F)^2 = \frac{G(x, x_1, \dots, x_n)}{G(x_1, \dots, x_n)}. \tag{9.61}$$

9.1.6 Déterminant de Cauchy

Soient des nombres a_i et b_i $(i = 1, \dots, n)$ tels que $a_i + b_j \neq 0$ pour tout couple (i, j). Le **déterminant de Cauchy** est

$$D_n = \det\left(\frac{1}{a_i + b_j}\right). \tag{9.62}$$

Proposition 9.16 ([173])**.**
Le déterminant de Cauchy est donné par la formule

$$D_n = \frac{\prod_{i<j}(a_j - a_i)\prod_{i<j}(b_j - b_i)}{\prod_{ij}(a_i + b_j)}. \tag{9.63}$$

9.1.7 Matrice de Sylvester

La définition est pompée de wikipédia. Soient P et Q deux polynômes non nuls, de degrés respectifs m et n :

$$P(x) = p_0 + p_1 x + \dots + p_n x^n \tag{9.64a}$$
$$Q(x) = q_0 + q_1 x + \dots + q_m x^m. \tag{9.64b}$$

La **matrice de Sylvester** associée à P et Q est la matrice carrée $m + n \times m + n$ définie ainsi :

(1) la première ligne est formée des coefficients de P, suivis de 0 :

$$\begin{pmatrix} p_n & p_{n-1} & \cdots & p_1 & p_0 & 0 & \cdots & 0 \end{pmatrix}; \tag{9.65}$$

(2) la seconde ligne s'obtient à partir de la première par permutation circulaire vers la droite ;
(3) les $(m-2)$ lignes suivantes s'obtiennent en répétant la même opération ;
(4) la ligne $(m+1)$ est formée des coefficients de Q, suivis de 0 :

$$\begin{pmatrix} q_m & q_{m-1} & \cdots & q_1 & q_0 & 0 & \cdots & 0 \end{pmatrix}; \tag{9.66}$$

(5) les $(m-1)$ lignes suivantes sont formées par des permutations circulaires.

Ainsi dans le cas $n = 4$ et $m = 3$, la matrice obtenue est

$$S_{p,q} = \begin{pmatrix} p_4 & p_3 & p_2 & p_1 & p_0 & 0 & 0 \\ 0 & p_4 & p_3 & p_2 & p_1 & p_0 & 0 \\ 0 & 0 & p_4 & p_3 & p_2 & p_1 & p_0 \\ q_3 & q_2 & q_1 & q_0 & 0 & 0 & 0 \\ 0 & q_3 & q_2 & q_1 & q_0 & 0 & 0 \\ 0 & 0 & q_3 & q_2 & q_1 & q_0 & 0 \\ 0 & 0 & 0 & q_3 & q_2 & q_1 & q_0 \end{pmatrix}. \tag{9.67}$$

Le déterminant de la matrice de Sylvester associée à P et Q est appelé le **résultant** de P et Q et noté $\operatorname{res}(P, Q)$.

Attention : si P est de degré n et Q de degré m, il y a m lignes pour P et n pour Q dans le déterminant du résultant (et non le contraire).

Lemme 9.17 ([174]).
Si P et Q sont deux polynômes de degrés n et m à coefficients dans l'anneau $\mathbb{A}$, alors pour tout $\lambda \in \mathbb{A}$,

$$\operatorname{res}(\lambda P, Q) = \lambda^m \operatorname{res}(P, Q) \tag{9.68a}$$
$$\operatorname{res}(P, \lambda Q) = \lambda^n \operatorname{res}(P, Q). \tag{9.68b}$$

Démonstration. Cela est simplement un comptage du nombre de lignes. Il y a m lignes contenant les coefficients de P ; donc prendre λP revient à multiplier m lignes dans un déterminant et donc le multiplier par λ^m. □

L'équation de Bézout (6.71) peut être traitée avec une matrice de Sylvester. Soient P et Q, deux polynômes donnés et à résoudre l'équation

$$xP + yQ = 0 \tag{9.69}$$

par rapport aux polynômes inconnus x et y dont les degrés sont $\deg(x) < \deg(Q)$ et $\deg(y) < \deg(P)$. Si nous notons $\tilde{x}$ et $\tilde{y}$ la liste des coefficients de x et y (dans l'ordre décroissant de degré), nous pouvons récrire l'équation (9.69) sous la forme

$$S_{PQ}^t \begin{pmatrix} \tilde{x} \\ \tilde{y} \end{pmatrix} = 0. \tag{9.70}$$

Pour s'en convaincre, écrivons pour les polynômes de l'exemple (9.67) :

$$\begin{pmatrix} p_4 & 0 & 0 & q_3 & 0 & 0 & 0 \\ p_3 & p_4 & 0 & q_2 & q_3 & 0 & 0 \\ p_2 & p_3 & p_4 & q_1 & q_2 & q_3 & 0 \\ p_1 & p_2 & p_3 & q_0 & q_1 & q_2 & q_3 \\ p_0 & p_1 & p_2 & 0 & q_0 & q_1 & q_2 \\ 0 & p_0 & p_1 & 0 & 0 & q_0 & q_1 \\ 0 & 0 & p_0 & 0 & 0 & 0 & q_0 \end{pmatrix} \begin{pmatrix} x_2 \\ x_1 \\ x_0 \\ y_3 \\ y_2 \\ y_1 \\ y_0 \end{pmatrix} = \begin{pmatrix} x_2 p_4 + y_2 q_3 \\ p_3 x_2 + p_4 x_1 + q_2 y_3 + q_3 y_2 \\ \\ \vdots \\ \\ \end{pmatrix} \tag{9.71}$$

Nous voyons que sur la ligne numéro k (en partant du bas et en numérotant de à partir de zéro) nous avons les produits $p_i x_j$ et $q_i y_j$ avec $i + j = k$. La colonne de droite représente donc bien les coefficients du polynôme $xP + yQ$.

Proposition 9.18.
Le résultant de deux polynômes est non nul si et seulement si les deux polynômes sont premiers entre eux.

Un polynôme P a une racine double en a si et seulement si P et P' ont a comme racine commune, ce qui revient à dire que P et P' ne sont pas premiers entre eux.

Une application importante de ces résultats sera le théorème de Rothstein-Trager 19.93 sur l'intégration de fractions rationnelles.

Exemple 9.19.
Si nous prenons $P = aX^2 + bX + c$ et $P' = 2aX + b$ alors la taille de la matrice de Sylvester sera $2 + 1 = 3$ et

$$S_{P,P'} = \begin{pmatrix} a & b & c \\ 2a & b & 0 \\ 0 & 2a & b \end{pmatrix}. \tag{9.72}$$

Le résultant est alors

$$\operatorname{res}(P, P') = -a(b^2 - 4ac). \tag{9.73}$$

Donc un polynôme du second degré a une racine double si et seulement si $b^2 - 4ac = 0$. Cela est un résultat connu depuis longtemps mais qui fait toujours plaisir à revoir. △

La matrice de Sylvester permet aussi de récrire l'équation de Bézout pour les polynômes ; voir le théorème 6.41 et la discussion qui s'ensuit.

Une proposition importante du résultant est qu'il peut s'exprimer à l'aide des racines des polynômes.

Proposition 9.20.
Si

$$P(X) = a_p \prod_{i=1}^{p} (X - \alpha_i) \tag{9.74a}$$

$$Q(X) = b_q \prod_{j=1}^{q} (X - \beta_i) \tag{9.74b}$$

alors nous avons les expressions suivantes pour le résultant :

$$\operatorname{res}(P, Q) = a_p^q b_q^p \prod_{i=1}^{p} \prod_{j=1}^{q} (\beta_j - \alpha_i) = b_q^p \prod_{j=1}^{q} P(\beta_j) = (-1)^{pq} a_p^q \prod_{i=1}^{p} Q(\alpha_i). \tag{9.75}$$

Démonstration. Si P et Q ne sont pas premiers entre eux, d'une part la proposition 9.18 nous dit que $\operatorname{res}(P, Q) = 0$ et d'autre part, P et Q ont un facteur irréductible en commun, ce qui signifie que nous devons avoir un des $X - \alpha_i$ égal à un des $X - \beta_j$. Autrement dit, nous avons $\alpha_i = \beta_j$ pour un couple (i, j). Par conséquent tous les membres de l'équation (9.75) sont nuls.

Nous supposons donc que P et Q sont premiers entre eux. Nous commençons par supposer que les polynômes P et Q sont unitaires, c'est-à-dire que $a_p = b_q = 1$. Nous considérons alors l'anneau

$$\mathbb{A} = \mathbb{Z}[\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q]. \tag{9.76}$$

Dans cet anneau, l'élément $\beta_j - \alpha_i$ est irréductible (tout comme $X - Y$ est irréductible dans $\mathbb{Z}[X, Y]$). Le résultant $R = \operatorname{res}(P, Q)$ est un élément de $\mathbb{A}$ parce que tous leurs coefficients peuvent être exprimés à l'aide des α_i et des β_j. Dans $\mathbb{A}$, l'élément $\beta_j - \alpha_i$ divise R. En effet lorsque $\beta_j = \alpha_i$, le déterminant définissant le résultant est nul, ce qui signifie que $\beta_j - \alpha_i$ est un facteur irréductible de R.

Par conséquent il existe un polynôme $T \in \mathbb{A}$ tel que

$$R = \lambda(\alpha_1, \ldots, \beta_q) \prod_{i=1}^{p} \prod_{j=1}^{r} (\beta_j - \alpha_i). \tag{9.77}$$

Comptons les degrés. Pour donner une idée de ce calcul de degré, voici comment se présente, au niveau des dimensions, le déterminant :

$$\begin{array}{ccccccc}
\multicolumn{4}{c}{\overleftrightarrow{\qquad\qquad p+1 \qquad\qquad}} & & \overleftrightarrow{q-1} & \\
a_p & a_{p-1} & \cdots & a_0 & 0 & \cdots & 0 \\
 & & \ddots & & & \ddots & \\
0 & \cdots & 0 & a_p & \cdots & a_1 & a_0 \\
\multicolumn{7}{c}{\overleftrightarrow{\qquad\qquad\qquad\qquad p+q \qquad\qquad\qquad\qquad}}
\end{array} \quad \updownarrow q \tag{9.78}$$

si les a_i sont les coefficients de P. Mais chacun des a_i est de degré 1 en les α_i, donc le déterminant dans son ensemble est de degré q en les α_i, parce que R contient q lignes telles que (9.78). Le même raisonnement montre que R est de degré p en les β_j. Par ailleurs le polynôme $\prod_{i=1}^{p}\prod_{j=1}^{r}(\beta_j-\alpha_i)$ est de degré p en les β_j et q en les α_i. Nous en déduisons que T doit être un polynôme ne dépendant pas de α_i ou de β_j.

Nous pouvons donc calculer la valeur de T en choisissant un cas particulier. Avec $P(X)=X^p$ et $Q(X)=X^q+1$, il est vite vu que $R(P,Q)=1$ et donc que $T=1$.

Si les polynômes P et Q ne sont pas unitaires, le lemme 9.17 nous permet de conclure.

□

9.1.8 Théorème de Kronecker

Nous considérons K_n l'ensemble des polynômes de $\mathbb{Z}[X]$

(1) unitaires de degré n,

(2) dont les racines dans $\mathbb{C}$ sont de modules plus petits ou égaux à 1,

(3) et qui ne sont pas divisés par X.

Un tel polynôme s'écrit sous la forme

$$P = X^n + \sum_{k=0}^{n-1} a_k X^k. \tag{9.79}$$

Théorème 9.21 (Kronecker[54]).
Les racines des éléments de K_n sont des racines de l'unité.

Démonstration. Vu que $\mathbb{C}$ est algébriquement clos nous pouvons considérer les racines $\alpha_1,\dots,\alpha_n$ de P dans $\mathbb{C}$. Nous les considérons avec leurs multiplicités.

Soit $R=X^n+\sum_{k=0}^{n-1} b_k X^k$ un élément de K_n dont nous notons $\beta_1,\dots,\beta_n$ les racines dans $\mathbb{C}$. Les relations coefficients-racines stipulent que

$$b_k = \sum_{1\leqslant i_1<\dots<i_{n-k}\leqslant n} \prod_{j=1}^{n-k} \beta_{i_j}. \tag{9.80}$$

En prenant le module et en se souvenant que $|\beta_l|\leqslant 1$ pour tout l, nous trouvons que

$$|b_k| \leqslant \binom{n}{n-k}. \tag{9.81}$$

Mais comme $b_k\in\mathbb{Z}$, nous avons

$$b_k \in \Big\{ -\binom{n}{n-k}, -\binom{n}{n-k}+1, \dots, 0, \cdots, \binom{n}{n-k}\Big\} \tag{9.82}$$

qui est de cardinal $\binom{n}{n-k}+1$. Nous avons donc

$$\mathrm{Card}(K_n) \leqslant \prod_{k=0}^{n-1}\Big(1+\binom{n}{n-k}\Big) < \infty. \tag{9.83}$$

La conclusion jusqu'ici est que K_n est un ensemble fini.

Pour chaque $k \in \mathbb{N}^*$ nous considérons les polynômes

$$P_k = \prod_{i=1}^{n}(X-\alpha_i^k) \tag{9.84a}$$

$$Q_k = X^k - Y \in \mathbb{Z}[X,Y], \tag{9.84b}$$

et puis nous considérons le résultant $R_k = \mathrm{res}_X(P,Q_k) \in \mathbb{Z}[Y]$:

$$R_k = \mathrm{res}_X(P,Q_k) = \begin{pmatrix} 1 & a_{n-1} & \cdots & a_0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & a_{n-1} & \cdots & a_0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & & & \\ 0 & \cdots & 0 & 1 & a_{n-1} & \cdots & a_0 & 0 & 0 \\ 0 & \cdots & 0 & 0 & 1 & a_{n-1} & \cdots & a_0 & 0 \\ 0 & \cdots & 0 & 0 & 0 & 1 & a_{n-1} & \cdots & a_0 \\ & & & & & & & & \\ 1 & 0 & \dots & 0 & -Y & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & -Y & 0 & \dots & 0 \\ & & \ddots & & & \ddots & \ddots & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & -Y & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & -Y \end{pmatrix} \tag{9.85}$$

Cela est un polynôme en Y dont le terme de plus haut degré est $(-1)^nY^n$. Les petites formules de la proposition 9.20 nous permettent d'exprimer $R_k(Y)$ en termes des racines de P :

$$R_k(Y) = \prod_{i=1}^{n} Q_k(\alpha_i) = \prod_{i=1}^{n}(\alpha_i^k - Y) = (-1)^n\prod_{i=1}^{n}(Y-\alpha_i^k) = (-1)^nP_k(Y). \tag{9.86}$$

Vu que $P \in K_n$ nous savons que les α_i ne sont pas tous nuls ; donc $P_k \in K_n$. Cependant nous avons vu que K_n est un ensemble fini ; donc parmi les P_k, il y a des doublons (et pas un peu) [7]. Nous regardons même l'ensemble des P_{2^n} dans lequel nous pouvons en trouver deux les mêmes. Soit $l > k$ tels que $P_{2^k} = P_{2^l}$. Si α est racine de P_{2^k}, alors il est de la forme $\alpha = \beta^{2^k}$ pour une certaine racines β de P. Par conséquent

$$\alpha^{2^l/2^k} = \alpha^{2^{l-k}} \tag{9.87}$$

est racine de P_{2^l}. Notons que dans cette expression il n'y a pas de problèmes de définition d'exposant fractionnaire dans $\mathbb{C}$ parce que $l > k$. Vu que (9.87) est racine de P_{2^l}, il est aussi racine de P_{2^k}. Donc

$$\big(\alpha^{2^{l-k}}\big)^{2^{l-k}} = \alpha^{2^{2(l-k)}} \tag{9.88}$$

est racine de P_{2^l} et donc de P_{2^k}. Au final nous savons que tous les nombres de la forme $\alpha^{2^{n(l-k)}}$ sont racines de P_{2^k}. Mais comme P_{2^k} a un nombre fini de racines, nous pouvons en trouver deux égales. Si nous avons

$$\alpha^{2^{n(l-k)}} = \alpha^{2^{m(l-k)}} \tag{9.89}$$

pour certains entiers $m > n$, alors

$$\alpha^{2^{n(l-k)}-2^{m(l-k)}} = 1, \tag{9.90}$$

ce qui prouver que α est une racine de l'unité. Nous avons donc prouvé que toutes les racines de P_{2^k} sont des racines de l'unité et donc que les racines de P sont racines de l'unité. □

7. Ici dans [54], il déduit qu'on a un k tel que $P_k = P_1 = P$. Mois je vois pourquoi on a un k et un l tels que $P_k = P_l$, mais pourquoi on peut en trouver un spécialement égal au premier ? Une réponse à cette question permettrait de solidement réduire la lourdeur de la suite de la preuve.

9.2 Orientation

9.2.1 Cas vectoriel

Proposition-définition 9.22 ([175]).
Soient deux bases $\mathcal{B}$ et $\mathcal{B}'$ d'un espace vectoriel réel E. Nous définissons la relation $\mathcal{B} \sim \mathcal{B}'$ si et seulement si $\det_{\mathcal{B}}(\mathcal{B}') > 0$ [8].

Cela est une relation d'équivalence[9] *sur l'ensemble des bases de E, et les classes sont les* ***orientations*** *de E.*

Démonstration. Tout est dans le lemme 9.6. D'abord quand $\mathcal{B}$ et $\mathcal{B}'$ sont des bases, $\det_{\mathcal{B}}(\mathcal{B}') \neq 0$ ensuite, nous passons en revue les points qu'il faut pour être une relation d'équivalence.

(1) $\mathcal{B} \sim \mathcal{B}$ parce que $\det_{\mathcal{B}}(\mathcal{B}) = 1 > 0$.

(2) Vu que $\det_{\mathcal{B}}(\mathcal{B}') = \frac{1}{\det_{\mathcal{B}'}(\mathcal{B})}$, les deux sont positifs en même temps ou pas du tout.

(3) Si $\mathcal{B} \sim \mathcal{B}'$ et $\mathcal{B}' \sim \mathcal{B}''$, alors en utilisant la formule

$$\det_{\mathcal{B}}(\mathcal{B}'') = \det_{\mathcal{B}}(\mathcal{B}') \det_{\mathcal{B}'}(\mathcal{B}''), \tag{9.91}$$

nous voyons que $\det_{\mathcal{B}}(\mathcal{B}'') > 0$.

□

Lemme 9.23.
Soit un espace vectoriel réel E. L'ensemble des bases de E possède exactement deux orientations[10]

Démonstration. Nous considérons une base $\mathcal{B} = (e_1, \dots, e_n$ [11] à partir de laquelle nous définissons une autre base : $\mathcal{B}' = (-e_1, e_2, \dots, e_n)$. Nous allons prouver que ces deux bases ne sont pas équivalentes, et que toute base de E est équivalente soit à $\mathcal{B}$ soit à $\mathcal{B}'$.

Au moins deux classes Le fait que $\det_{\mathcal{B}}(\mathcal{B}') = -1$ vient du fait que $\det_{\mathcal{B}}(\mathcal{B}') = 1$ et que l'application $\det_{\mathcal{B}}$ est n-linéaire ; en multipliant par -1 le premier argument, la valeur du déterminant est multipliée par -1.

Donc les bases $\mathcal{B}$ et $\mathcal{B}'$ ne sont pas équivalentes et il existe au moins deux classes.

Au plus deux classes Nous montrons à présent que toute base est équivalente soit à $\mathcal{B}$ soit à $\mathcal{B}'$. Supposons que $\mathcal{B}''$ ne soit pas équivalente à $\mathcal{B}$, c'est-à-dire que $\det_{\mathcal{B}}(\mathcal{B}'') < 0$. Nous utilisons encore la formule (9.11),

$$\underbrace{\det_{\mathcal{B}}(\mathcal{B}'')}_{<0} = \underbrace{\det_{\mathcal{B}}(\mathcal{B}')}_{<0} \det_{\mathcal{B}'}(\mathcal{B}''), \tag{9.92}$$

et nous déduisons que $\det_{\mathcal{B}'}(\mathcal{B}'') > 0$.

□

9.24.
Vu qu'il n'y a que deux classes d'équivalence parmi les bases, nous pouvons utiliser le vocable « avoir la même orientation que » ou « avoir l'orientation contraire de ». Ce n'est pas ambigu.

Proposition 9.25 ([175]).
Si $\mathcal{B}$ est une base de l'espace vectoriel E de dimension n, et si τ est une transposition[12] *de S_n, alors la base $\tau(\mathcal{B})$ est de sens contraire.*

8. Définition 9.4.
9. Définition 1.20.
10. Définition 9.22.
11. Nous notons (e_1, e_2) et non $\{e_1, e_2\}$ parce que l'ordre est important.
12. Définition 2.64.

Démonstration. Le lemme 9.6(2) dit que $\det_{\mathcal{B}}$ est une forme anti-symétrique ; donc

$$\det_{\mathcal{B}}(\mathcal{B}') = -\det_{\mathcal{B}}\big(\tau(\mathcal{B})'\big). \tag{9.93}$$

Si l'un est positif, l'autre est négatif. Elles ont donc des orientations contraires. □

Corolaire 9.26.
Si $\mathcal{B}$ est une base de l'espace vectoriel E de dimension n, et si $\sigma \in S_n$, la base $\sigma(\mathcal{B})$ a même orientation que $\mathcal{B}$ si et seulement si $\sigma \in A_n$.

Démonstration. Notons c_1 la classe d'orientation de $\mathcal{B}$ et c_2 l'autre classe. La permutation σ se décompose en produit de transpositions dont la parité est fixée (proposition 2.74). Posons $\sigma = \tau_k \dots \tau_1$.

En posant $\mathcal{B}_0 = \mathcal{B}$ et $\mathcal{B}_{l+1} = \tau_{l+1}(\mathcal{B}_l)$, pour tout l, la base $\mathcal{B}_l$ est d'orientation contraire à celle de la base $\mathcal{B}_{l-1}$. Une base sur deux a l'orientation de $\mathcal{B}$ et l'autre sur deux a l'orientation contraire.

Donc $\sigma(\mathcal{B})$ a la même orientation que $\mathcal{B}$ si et seulement si k est pair. Mais $\sigma \in A_n$ si et seulement si k est pair. C'est bon. □

Proposition-définition 9.27 ([175]).
Soit un espace vectoriel réel, et un endomorphisme f de E. Deux définitions.

*(1) L'endomorphisme f est **direct** si sont déterminant est strictement positif.*

*(2) L'endomorphisme **préserve l'orientation** si il transforme toutes base de E en une base de même orientation.*

Un endomorphisme est direct si et seulement si il préserve l'orientation.

Démonstration. En deux sens.

Direct implique préserve l'orientation Soit une base $\mathcal{B}$ de E et un endomorphisme direct u. D'abord, u est inversible du fait que son déterminant est non nul par la proposition 9.10(2). Donc u transforme une base en une base par le lemme 4.7.

La définition 9.8 du déterminant de u est que

$$\det(u) = \det_{\mathcal{B}}\big(u(\mathcal{B})\big) > 0. \tag{9.94}$$

Donc $\mathcal{B}$ et $u(\mathcal{B})$ ont même orientation.

Préserve l'orientation implique direct Le fait que u préserve l'orientation signifie en particulier qu'il transforme une base en une base et qu'il est inversible par le lemme 4.7.

Donc si $\mathcal{B}$ est une base, $u(\mathcal{B})$ est encore une base et nous avons, parce que $\mathcal{B}$ et $u(\mathcal{B})$ ont même orientation,

$$0 < \det_{\mathcal{B}}\big(u(\mathcal{B})\big) = \det(u). \tag{9.95}$$

□

9.2.2 Cas affine

Définition 9.28.
Soit un espace affine $\mathcal{E}$ modelé sur E. Les repères cartésiens[13] *$(O, \mathcal{B})$ et $(O', \mathcal{B}')$ ont **même orientation** si les bases $\mathcal{B}$ et $\mathcal{B}'$ ont même orientation.*

Les classes d'équivalence (il y en a deux) sont les orientations de $\mathcal{E}$.

*Une application affine $f\colon \mathcal{E} \to \mathcal{E}$ **préserve l'orientation** si sa partie linéaire*[14] *préserve l'orientation.*

13. Définition 8.6.
14. Définition 8.10.

9.3 Hermitien, orthogonal, adjoint

9.29.
Une des choses à retenir de la définition de l'opérateur adjoint est que la notion de A^* dépend du produit scalaire considéré.

Il se fait que le plus souvent, sur $\mathbb{R}^n$, nous considérons le produit scalaire usuel et la base canonique. De ce fait, les notions d'opérateur adjoint et d'opérateur transposés se confondent avec la notion de matrice transposée. Ce sont pourtant, en général, trois notions distinctes.

Proposition-définition 9.30 (Définition de la transposée[1])**.**
Soient deux espaces vectoriels euclidiens ou hermitiens E et F et une application linéaire $A\colon E \to F$.

(1) Il existe une unique application linéaire $B\colon F \to E$ telle que

$$\langle Ax, y\rangle_F = \langle x, By\rangle_E \tag{9.96}$$

pour tout $x \in E$ et $y \in F$.

(2) Si $\{e_i\}$ est une base orthonormée de E et $\{f_\alpha\}$ est une base orthonormée de F, alors la matrice de A et B pour ces bases sont liées par

$$B_{i\alpha} = A_{\alpha i}. \tag{9.97}$$

*L'application B ainsi définie set nommée **adjoint** de A et sera notée $B = A^*$.*

Démonstration. Pour l'unicité, nous écrivons la condition avec $x = e_j$ pour obtenir :

$$\langle Ae_j, y\rangle = \langle e_j, By\rangle = (By)_j \tag{9.98}$$

c'est-à-dire que les coefficients $B(y)_j$ de $B(y)$ dans la base canonique sont fixés par la condition.

Pour l'existence, il suffit de vérifier que poser

$$B(y) = \sum_j \langle Ae_j, y\rangle e_j \tag{9.99}$$

fonctionne. Pour cela il faut utiliser la bilinéarité du produit scalaire et le fait que $\langle x, e_j\rangle = x_j$. Nous avons :

$$\langle x, B(y)\rangle = \langle x, \sum_j \langle A(e_j), y\rangle e_j\rangle \tag{9.100a}$$

$$= \sum_j \langle A(e_j), y\rangle\langle x, e_j\rangle \tag{9.100b}$$

$$= \sum_j \langle A(x_j e_j), y\rangle \tag{9.100c}$$

$$= \langle A(x), y\rangle. \tag{9.100d}$$

En ce qui concerne la matrice de l'application B ainsi définie, nous écrivons la condition (9.96) avec $y = e'_\alpha$ et $x = e_i$, de telle sorte que

$$A(x) = A(e_i) = \sum_\beta A_{\beta i} e'_\beta \tag{9.101}$$

et

$$B(y) = B(e'_\alpha) = \sum_j B_{j\alpha} e_j. \tag{9.102}$$

Alors nous avons :

$$\sum_\beta A_{\beta i}\langle e'_\beta, e'_\alpha\rangle = \sum_j B_{j\alpha}\langle e_i, e_j\rangle, \tag{9.103}$$

donc

$$A_{\alpha i} = B_{i\alpha}. \tag{9.104}$$

□

9.31.
À cause de l'expression (9.97) pour la matrice de A^*, cette application est souvent appelé **transposé** de A et noté A^t. Nous savons, nous, que la transposée de A est une application $A^t\colon F^* \to E^*$ donnée par la définition 4.139. Il nous arrivera donc d'écrire des égalités comme $\langle Ax, y\rangle = \langle x, A^t y\rangle$.

Proposition 9.32.
En ce qui concerne le déterminant,

$$\det(A^*) = \det(A)^* \tag{9.105}$$

où l'étoile à droite dénote la conjugaison complexe dans $\mathbb{C}$.

Démonstration. Écrivons l'expression explicite (9.7) du déterminant. Le tout avec la base canonique :

$$\det(A) = \det_{(e_1,\dots,e_n)}(Ae_1,\dots,Ae_n) = \sum_{\sigma\in S_n} \epsilon(\sigma)\prod_{i=1}^n e^*_{\sigma(i)}(Ae_i). \tag{9.106}$$

Mais nous pouvons développer :

$$e^*_{\sigma(i)}(Ae_i) = \langle e_{\sigma(i)}, Ae_i\rangle = \langle A^*e_{\sigma(i)}, e_i\rangle = \langle e_i, A^*e_{\sigma(i)}\rangle^* = e_i^*(A^*e_{\sigma(i)})^*. \tag{9.107}$$

Notez que dans la dernière expression, les trois * ont trois significations différentes. Par conséquent,

$$\det(A) = \sum_{\sigma\in S_n} \epsilon(\sigma)\prod_{i=1}^n e_i^*(A^*e_{\sigma(i)})^*. \tag{9.108}$$

Mais $e_i^*(A^*e_{\sigma(i)}) = e^*_{\sigma(j)}(A^*e_j)$ pour $j = \sigma(i)$, donc le produit ne change pas si on déplace le σ :

$$\det(A) = \sum_{\sigma\in S_n} \epsilon(\sigma)\prod_{i=1}^n e^*_{\sigma(i)}(A^*e_i)^* = \det(A^*)^*. \tag{9.109}$$

Nous avons donc $\det(A) = \det(A^*)^*$, c'est-à-dire $\det(A)^* = \det(A^*)$. Pour information, la dernière étoile est la conjugaison complexe. □

Proposition 9.33 ([1]).
Si $A\colon E_2 \to E_3$ et $B\colon E_1 \to E_2$ sont des applications linéaires, alors

$$(AB)^* = B^*A^* \tag{9.110}$$

où la « multiplication » est la composition.

Démonstration. L'existence de $(AB)^*$, de A^* et de B^* ne donne pas lieu à débat parce que la proposition 9.30 ne souffre pas de discussions. La propriété que $(AB)^*$ est unique a avoir est que

$$\langle ABx, y\rangle = \langle x, (AB)^*y\rangle \tag{9.111}$$

pour tout $x \in E_1$ et $y \in E_3$. Or l'application B^*A^* possède également cette propriété parce que

$$\langle x, B^*A^*y\rangle = \langle Bx, A^*y\rangle = \langle ABx, y\rangle. \tag{9.112}$$

La partie unicité de la proposition 9.30 nous impose donc d'accepter que les applications $(AB)^*$ et B^*A^* sont en réalité les mêmes [15]. □

9.34.
Un grand moment d'utilisation de la notion d'adjoint pour un opérateur non carré sera la définition d'une intégrale sur une variété ; en particulier dans la proposition 19.9.

15. Et ce même si vous croyez les avoir déjà vu ensemble dans la même pièce.

Définition 9.35.
Un opérateur A est ***hermitien*** *si $A^* = A$. On dit aussi* ***autoadjoint****.*

9.36.
Le mot « hermitien » est réservé aux opérateurs sur des espaces hermitiens, c'est-à-dire des espaces vectoriels sur $\mathbb{C}$. Le mot « autoadjoint » par contre est plutôt utilisé dans le cadre d'opérateurs sur les espaces réels. En conséquence de quoi, ces deux mots sont synonymes, mais il est préférable d'utiliser « hermitien » lorsque l'espace vectoriel est sur $\mathbb{C}$ et « autoadjoint » lorsqu'il est sur $\mathbb{R}$.

L'ensemble des opérateurs autoadjoints de E est noté $\mathrm{S}(E)$. Cette notation provient du fait que dans $\mathbb{R}^n$ muni du produit scalaire usuel, les opérateurs autoadjoints sont les matrices symétriques.

Remarque 9.37.
Le fait d'être hermitien n'implique en rien le fait d'être inversible.

Lemme 9.38.
Si E est un espace euclidien, un endomorphisme $f\colon E \to E$ est autoadjoint si et seulement si pour tout $x, y \in E$ nous avons $\langle x, f(y)\rangle = \langle f(x), y\rangle$.

Démonstration. Dans le sens direct, nous avons

$$\langle f(x), y\rangle = \langle x, f^*(y)\rangle = \langle x, f(y)\rangle. \tag{9.113}$$

La première égalité est la définition de f^* et la seconde est l'hypothèse $f = f^*$.

Dans l'autre sens, l'hypothèse est que l'endomorphisme f vérifie $\langle x, f(y)\rangle = \langle f(x), y\rangle$. Mais la proposition 9.30(1) spécifie que f^* est l'unique endomorphisme à satisfaire cette égalité. Donc $f = f^*$. □

9.3.1 Opérateur orthogonal, matrice orthogonale

Définition 9.39.
Un opérateur est ***orthogonal*** *lorsque $A^* = A^{-1}$ où A^* est l'adjoint de A définit en 9.30.*

Définition 9.40.
Une matrice U est ***orthogonale*** *si $U^t = U^{-1}$. Le* ***groupe orthogonal*** *noté $\mathrm{O}(n)$ est l'ensemble des matrices orthogonales $n \times n$.*

Lemme 9.41.
Soit un opérateur $A\colon \mathbb{R}^n \to \mathbb{R}^n$ muni du produit scalaire usuel. Il est orthogonal si et seulement si sa matrice dans la base canonique est orthogonale[16].

Démonstration. Soit la base canonique $\{e_i\}_{i=1,\dots,n}$ de $\mathbb{R}^n$. Nous avons

$$\langle AA^* e_i, e_j\rangle = \langle e_i, e_j\rangle = \delta_{ij}, \tag{9.114}$$

donc $\big((AA^*)e_i\big)_j = \delta_{ij}$, ou encore $(AA^*)_{ij} = \delta_{ij}$, ce qui signifie que la matrice AA^* est l'identité. □

Proposition 9.42 (Thème 66)**.**
À propos de matrices orthogonales.

(1) L'ensemble des matrices réelles orthogonales forme un groupe noté $\mathrm{O}(n, \mathbb{R})$.

(2) Si A est une matrice orthogonale, alors $\det(A) = \pm 1$.

(3) Le groupe $\mathrm{O}(n)$ est le groupe des isométries linéaires[17] *de $\mathbb{R}^n$.*

16. Définition 9.40.

17. Au sens où, parmi les applications linéaires, les isométries sont les éléments de $\mathrm{O}(n)$. À part ça, il y a aussi les translations, mais c'est une autre histoire qui vous sera contée une autre fois.

Démonstration. Si A et B sont orthogonales, alors

$$(AB)(AB)^t = ABB^tA^t = A\mathbb{1}A^t = \mathbb{1}. \tag{9.115}$$

Vu que $\mathbb{1}$ est orthogonale, nous avons bien un groupe.

En ce qui concerne le déterminant, $AA^t = \mathbb{1}$ donne $\det(A)\det(A^t) = 1$, mais la proposition 9.32 dit que $\det(A) = \det(A^t)$, donc $\det(A)^2 = 1$. D'où le fait que $\det(A) = \pm 1$.

D'autre part si A est une isométrie de $\mathbb{R}^n$ alors pour tout $x, y \in \mathbb{R}^n$ nous avons $\langle Ax, Ay\rangle = \langle x, y\rangle$. En particulier,

$$\langle A^tAx, y\rangle = \langle x, y\rangle \tag{9.116}$$

pour tout $x, y \in \mathbb{R}^n$. En prenant $y = e_i$ nous trouvons

$$(A^tAx)_i = x_i, \tag{9.117}$$

ce qui signifie que pour tout x, $A^tAx = x$, ou encore que A^tA est l'identité.

Réciproquement si A^tA est l'identité nous avons

$$\langle x, y\rangle = \langle A^tAx, y\rangle = \langle Ax, Ay\rangle, \tag{9.118}$$

ce qui prouve que A est une isométrie. □

En ce qui concerne les valeurs propres des matrices de $\mathrm{O}(n)$ ainsi que leurs formes canoniques (avec des fonctions trigonométriques) pour $\mathrm{O}(3)$ et $\mathrm{SO}(3)$, ce sera pour la proposition 17.189 et ce qui s'ensuit.

Définition 9.43.
Le sous-groupe des matrices orthogonales de déterminant 1 *est le groupe* ***spécial orthogonal*** *noté* $\mathrm{SO}(n)$.

9.4 Topologie

9.4.1 Boules et sphères

Définition 9.44.
Soit $(V, \|.\|)$*, un espace vectoriel normé,* $a \in V$ *et* $r > 0$*. Nous allons abondamment nous servir des ensembles suivants :*

(1) la ***boule ouverte*** $B(a, r) = \{x \in V \text{ tel que } \|x - a\| < r\}$ *;*
(2) la ***boule fermée*** $\bar{B}(a, r) = \{x \in V \text{ tel que } \|x - a\| \leqslant r\}$ *;*
(3) la ***sphère*** $S(a, r) = \{x \in V \text{ tel que } \|x - a\| = r\}$*.*

Les différences entre ces trois ensembles sont très importantes. D'abord, les *boules* sont pleines tandis que la *sphère* est creuse. En comparant à une pomme, la boule ouverte serait la pomme « sans la peau », la boule fermée serait « avec la peau » tandis que la sphère serait seulement la peau. Nous avons

$$\bar{B}(a, r) = B(a, r) \cup S(a, r). \tag{9.119}$$

Définition 9.45.
Une partie A *de* V *est dite* ***bornée*** *s'il existe un réel* R *tel que* $A \subset B(0_V, R)$*.*

Une partie est donc bornée si elle est contenue dans une boule de rayon fini.

Exemple 9.46.
Dans $\mathbb{R}$, les boules sont les intervalles ouverts et fermés tandis que la sphère est donnée par les points extrêmes des intervalles :

$$\begin{aligned} B(a, r) &=]a - r, a + r[, \\ \bar{B}(a, r) &= [a - r, a + b], \\ S(a, r) &= \{a - r, a + r\}. \end{aligned} \tag{9.120}$$

△

Exemple 9.47.
Si nous considérons $\mathbb{R}^2$, la situation est plus riche parce que nous avons plus de normes. Essayons de voir les sphères de centre $(0,0) \in \mathbb{R}^2$ et de rayon r pour les normes $\|.\|_1$, $\|.\|_2$ et $\|.\|_\infty$.

Pour la norme $\|.\|_1$, la sphère de rayon r est donnée par l'équation

$$|x| + |y| = r. \tag{9.121}$$

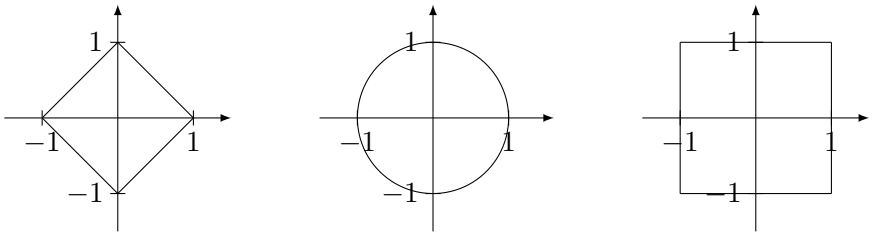
Pour la norme $\|.\|_2$, l'équation de la sphère de rayon r est

$$\sqrt{x^2 + y^2} = r, \tag{9.122}$$

et pour la norme supremum, la sphère de rayon r a pour équation

$$\max\{|x|, |y|\} = r. \tag{9.123}$$

Elles sont dessinées sur la figure 9.1



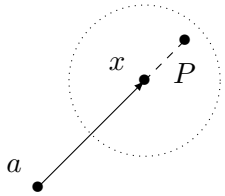
(a) La sphère unité pour la norme $\|.\|_1$ (b) La sphère unité pour la norme $\|.\|_2$ (c) La sphère unité pour la norme $\|.\|_\infty$

FIGURE 9.1 – Les sphères de rayon 1 pour les trois normes classiques.

△

Proposition 9.48.
Soient V un espace vectoriel normé, a dans V et x tel que $d(a,x) = r$, c'est-à-dire $x \in S(a,r)$. Dans ce cas, toute boule centrée en x contient un point P tel que $d(P,a) > r$ et un point Q tel que $d(Q,a) < r$.

Démonstration. Soit une boule de rayon δ autour de x. Le but est de trouver un point P tel que $d(P,a) > r$ et $d(P,x) < \delta$. Pour cela, nous prenons P sur la même droite que x (en partant de a), mais juste « un peu plus loin », comme sur la figure suivante :



Plus précisément, nous considérons le point

$$P = x + \frac{v}{N} \tag{9.124}$$

où $v = x - a$ et N est suffisamment grand pour que $d(x,P)$ soit plus petit que δ. Cela est toujours possible parce que

$$d(P,x) = \|P - x\| = \frac{\|v\|}{N} \tag{9.125}$$

peut être rendu aussi petit que l'on veut par un choix approprié de N. Montrons maintenant que $d(a,P) > d(a,x)$:

$$\begin{aligned} d(a,P) &= \|a - x - \frac{v}{N}\| \\ &= \|a - x + \frac{a}{N} - \frac{x}{N}\| \\ &= \|\big(1 + \frac{1}{N}(a-x)\big)\| \\ &> \|a - x\| = d(a,x). \end{aligned} \tag{9.126}$$

Nous laissons en exercice le soin de trouver un point Q tel que $d(Q,a) < r$ et $d(Q,x) < \delta$. □

9.4.2 Ouverts, fermés, intérieur et adhérence

Définition 9.49.
*Soit $(V, \|.\|)$ un espace vectoriel normé et A, une partie de V. Un point a est dit **intérieur** à A s'il existe une boule ouverte centrée en a et contenue dans A.*

*On appelle **l'intérieur** de A l'ensemble des points qui sont intérieurs à A. Nous notons* $\mathrm{Int}(A)$ *l'intérieur de A.*

Notons que $\mathrm{Int}(A) \subset A$ parce que si $a \in \mathrm{Int}(A)$, nous avons $B(a,r) \subset A$ pour un certain r et en particulier $a \in A$.

Exemple 9.50.
Trouver l'intérieur d'un intervalle dans $\mathbb{R}$ consiste à « ouvrir là où c'est fermé ».

(1) $\mathrm{Int}\big([0,1[\big) =]0,1[$.
Prouvons d'abord que $]0,1[\subset \mathrm{Int}([0,1[)$. Si $a \in]0,1[$, alors a est strictement supérieur à 0 et strictement inférieur à 1. Dans ce cas, la boule de centre a et de rayon $\frac{\min\{a,1-a\}}{2}$ est contenue dans $]0,1[$ (voir figure 9.2). Cela prouve que a est dans l'intérieur de $[0,1[$.

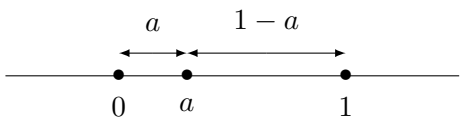


FIGURE 9.2 – Trouver le rayon d'une boule autour de a. Une boule qui serait centrée en a avec un rayon strictement plus petit à la fois de a et de $1-a$ est entièrement contenue dans le segment $]0,1[$.

Prouvons maintenant que $\mathrm{Int}\big([0,1[\big) \subset]0,1[$. Vu que l'intérieur d'un ensemble est inclus dans l'ensemble, nous savons déjà que $\mathrm{Int}\big([0,1[\big) \subset [0,1[$. Nous devons donc seulement montrer que 0 n'est pas dans l'intérieur de $[0,1[$. C'est le cas parce que toute boule du type $B(0,r)$ contient le point $-r/2$ qui n'est pas dans $[0,1[$.

(2) $\mathrm{Int}\Big([0,\infty[\Big) =]0,\infty[$.

(3) $\mathrm{Int}\big(]2,3[\big) =]2,3[$.

△

Exemple 9.51.
Les intérieurs des boules et sphères sont importantes à savoir.

(1) $\mathrm{Int}\big(B(a,r)\big) = B(a,r)$. Si $x \in B(a,r)$, nous avons $d(a,x) < r$. Alors la boule $B\big(x, r - d(x,a)\big)$ est incluse à $B(a,r)$, et donc x est dans l'intérieur de $B(a,r)$. Conseil : faire un dessin.

(2) $\mathrm{Int}\big(\bar{B}(a,r)\big) = B(a,r)$. Par le point précédent, la boule $B(a,r)$ est certainement dans l'intérieur de la boule fermée. Il reste à montrer que les points de $\bar{B}(a,r)$ qui ne sont pas dans $B(a,r)$ ne sont pas dans l'intérieur. Ces points sont ceux dont la distance à a est *égale* à r. Le résultat découle alors de la proposition 9.48.

(3) $\mathrm{Int}\big(S(a,r)\big) = \varnothing$. Si $x \in S(a,r)$, toute boule centrée en a contient des points qui ne sont pas à distance r de a.

Notez que la sphère est un exemple d'ensemble non vide mais d'intérieur vide.

△

Définition 9.52.
Une partie A de l'espace vectoriel normé $(V, \|.\|)$ est dite ***ouverte*** *si chacun de ses points est intérieur. La partie A est donc ouverte si $A \subset \mathrm{Int}(A)$. Par convention, nous disons que l'ensemble vide $\varnothing$ est ouvert.*

Une partie est dite ***fermée*** *si son complémentaire est ouvert. La partie A est donc fermée si $V \backslash A$ est ouverte.*

Remarque : un ensemble A est ouvert si et seulement si $\mathrm{Int}(A) = A$.

Définition 9.53.
Une partie A de l'espace vectoriel normé V est dite ***compacte*** *si elle est fermée et bornée.*

Nous verrons tout au long de ce cours que les ensembles compacts, et les fonctions définies sur ces ensembles ont de nombreuses propriétés agréables.

Exemple 9.54.
En ce qui concerne les intervalles de $\mathbb{R}$,

— $]1,2[$ est ouvert ;

— $[3,4]$ est fermé ;

— $[5,6[$ n'est ni ouvert ni fermé ;

Les intervalles fermés de $\mathbb{R}$ sont toujours compacts. △

Proposition 9.55.
Soit V un espace vectoriel normé.

(1) L'ensemble V lui-même et le vide sont à la fois fermés et ouverts.

(2) Toute union d'ouverts est ouverte.

(3) Toute intersection finie *d'ouverts est ouverte.*

(4) Le vide et V sont les seules parties de V à être à la fois fermées et ouvertes.

Démonstration. L'ingrédient principal de cette démonstration est que si a est un point d'un ouvert $\mathcal{O}$, alors il existe une boule autour de a contenue dans $\mathcal{O}$ parce que a doit être dans l'intérieur de $\mathcal{O}$.

(1) Nous avons déjà dit que, par définition, l'ensemble vide est ouvert. Cela implique que V lui-même est fermé (parce que son complémentaire est le vide). De plus, V est ouvert parce que toutes les boules sont inclues à V. Le vide est alors fermé (parce que son complémentaire est V).

(2) Soit une famille $(\mathcal{O}_i)_{i\in I}$ d'ouverts[18], et l'union

$$\mathcal{O} = \bigcup_{i\in I} \mathcal{O}_i. \tag{9.127}$$

Soit maintenant $a \in \mathcal{O}$. Nous devons prouver qu'il existe une boule centrée en a entièrement contenue dans $\mathcal{O}$. Étant donné que $a \in \mathcal{O}$, il existe $i \in I$ tel que $a \in \mathcal{O}_i$ (c'est-à-dire que a est au moins dans un des $\mathcal{O}_i$). Par hypothèse l'ensemble $\mathcal{O}_i$ est ouvert et donc tous ses points (en particulier a) sont intérieurs ; il existe donc une boule $B(a,r)$ centrée en a telle que $B(a,r) \subset \mathcal{O}_i \subset \mathcal{O}$.

18. L'ensemble I avec lequel nous « numérotons » les ouverts $\mathcal{O}_i$ est *quelconque*, c'est-à-dire qu'il peut être $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}^n$ ou n'importe quel autre ensemble, fini ou infini.

(3) Soit une famille finie d'ouverts $(\mathcal{O}_k)_{k\in\{1,\dots,n\}}$, et $a \in \mathcal{O}$ où

$$\mathcal{O} = \bigcap_{k=1}^{n} \mathcal{O}_k. \tag{9.128}$$

Vu que a appartient à chaque ouvert $\mathcal{O}_k$, nous pouvons trouver, pour chacun de ces ouverts, une boule $B(a, r_k)$ contenue dans $\mathcal{O}_k$. Chacun des r_k est strictement positif, et nous n'en avons qu'un nombre fini, donc le nombre $r = \min\{r_1, \dots, r_n\}$ est strictement positif. La boule $B(a, r)$ est inclue dans toutes les autres (parce que $B(a, r) \subset B(a, r')$ lorsque $r \leqslant r'$), par conséquent

$$B(a, r) \subset \bigcap_{k=1}^{n} B(a, r_k) \subset \bigcap_{k=1}^{n} \mathcal{O}_k = \mathcal{O}, \tag{9.129}$$

c'est-à-dire que la boule de rayon r est une boule centrée en a contenue dans $\mathcal{O}$, ce qui fait que a est intérieur à $\mathcal{O}$.

(4) Nous acceptons ce point sans démonstration.

□

La proposition dit que toute intersection *finie* d'ouvert est ouverte. Il est faux de croire que cela se généralise aux intersections infinies, comme le montre l'exemple suivant :

$$\bigcap_{i=1}^{\infty}]-\frac{1}{n}, \frac{1}{n}[= \{0\}. \tag{9.130}$$

Chacun des ensembles $]-\frac{1}{n}, \frac{1}{n}[$ est ouvert, mais le singleton $\{0\}$ est fermé (pourquoi ?).

Nous reportons à la proposition 1.157 la preuve du fait que tout ensemble borné de $\mathbb{R}$ possède un infimum et un supremum.

Définition 9.56.
*L'ensemble des ouverts de V est la **topologie** de V. La topologie dont nous parlons ici est dite **induite** par la norme $\|.\|$ de V (parce que cette norme définit la notion de boule et qu'à son tour la notion de boule définit la notion d'ouverts). Un **voisinage** de a dans V est un ensemble contenant un ouvert contenant a.*

Il existe de nombreuses topologies sur un espace vectoriel donné, mais certaines sont plus fameuses que d'autres. Dans le cas de $V = \mathbb{R}^n$, la topologie **usuelle** est celle induite par la norme euclidienne. Lorsque nous parlons de boules, de fermés, de voisinages ou d'autres notions topologiques (y compris de convergence, voir plus bas) dans $\mathbb{R}^n$, nous sous-entendons toujours la topologie de la norme euclidienne.

Exemple 9.57.
Les ensembles suivants sont des voisinages de 3 dans $\mathbb{R}$:

— $]1, 5[$;
— $[0, 10]$;
— $\mathbb{R}$.

Les ensembles suivants ne sont pas des voisinages de 3 dans $\mathbb{R}$:

— $]1, 3[$;
— $]1, 3]$;
— $[0, 5[\setminus\{3\}$.

△

Proposition 9.58.
Dans un espace vectoriel normé,

(1) toute intersection de fermés est fermée ;

(2) toute union finie *de fermés est fermée.*

Encore une fois, l'hypothèse de finitude de l'intersection est indispensable comme le montre l'exemple suivant :

$$\bigcup_{n=1}^{\infty}[-1+\frac{1}{n},1-\frac{1}{n}]=]-1,1[. \tag{9.131}$$

Chacun des intervalles dont on prend l'union est fermé tandis que l'union est ouverte.

Définition 9.59.
Soit A, une partie de l'espace vectoriel normé V. Un point $a \in V$ est dit ***adhérent*** *à A dans V si pour tout $\varepsilon > 0$,*

$$B(a,\varepsilon) \cap A \neq \emptyset. \tag{9.132}$$

Nous notons $\bar{A}$ l'ensemble des points adhérents à a et nous disons que $\bar{A}$ est l'adhérence de A. L'ensemble $\bar{A}$ sera aussi souvent nommé ***fermeture*** *de l'ensemble A.*

Un point peut être adhérent à A sans faire partie de A, et nous avons toujours $A \subset \bar{A}$.

Exemple 9.60.
La terminologie « fermeture » de A pour désigner $\bar{A}$ provient de deux origines.

(1) L'ensemble $\bar{A}$ est le plus petit fermé contenant A. Cela signifie que si B est un fermé qui contient A, alors $\bar{A} \subset A$. Cela est fondamentalement le sens de la définition 7.22.

(2) Pour les intervalles dans $\mathbb{R}$, trouver $\bar{A}$ revient à fermer les extrémités qui sont ouvertes, comme on en a parlé dans l'exemple 9.54.

△

Exemple 9.61.
Dans $\mathbb{R}$, l'infimum et le supremum d'un ensemble sont des points adhérents. En effet si M est le supremum de $A \subset \mathbb{R}$, pour tout ε, il existe un $a \in A$ tel que $a > M - \varepsilon$, tandis que $M > a$. Cela fait que $a \in B(M,\varepsilon)$, et en particulier que pour tout rayon ε, nous avons $B(M,\varepsilon) \cap A \neq \emptyset$.

Le même raisonnement montre que l'infimum est également dans l'adhérence de A. △

Exemple 9.62.
Il ne faut pas conclure de l'exemple précédent qu'un point limite ou adhérent est automatiquement un minimum ou un maximum. En effet, si nous regardons l'ensemble formé par les points de la suite $x_n = (-1)^n/n$, le nombre zéro est un point adhérent et une limite, mais pas un infimum ni un maximum. △

Lemme 9.63.
Si B est une partie fermée de V, alors $B = \bar{B}$.

Démonstration. Supposons qu'il existe $a \in \bar{B}$ tel que $a \notin B$. Alors il n'y a pas d'ouverts autour de a qui soit contenu dans $\complement B$. Cela prouve que $\complement B$ n'est pas ouvert, et par conséquent que B n'est pas fermé. Cela est une contradiction qui montre que tout point de $\bar{B}$ doit appartenir à B lorsque B est fermé. □

Exemple 9.64.
Au niveau des intervalles dans $\mathbb{R}$, prendre l'adhérence consiste à « fermer là où c'est ouvert ». Attention cependant à ne pas fermer l'intervalle en l'infini.

(1) $\overline{[0,2[} = [0,2]$.

(2) $\overline{]3,\infty]} = [3,\infty[$.

△

Proposition 9.65.
Soit V un espace vectoriel normé et $a \in V$. Les trois conditions suivantes sont équivalentes :

(1) $a \in \bar{A}$ *;*

(2) il existe une suite d'éléments x_n *dans* A *qui converge vers* a *;*

(3) $d(a, A) = 0$.

Notez que dans cette proposition, nous ne supposons pas que a soit dans A.

Proposition 9.66.
Pour toute partie A *d'un espace vectoriel normé nous avons*

(1) $V \backslash \bar{A} = \operatorname{Int}(V \backslash A)$,

(2) $V \backslash \operatorname{Int}(A) = \overline{V \backslash A}$.

En utilisant les notations du complémentaire (1.1.4), les deux points de la proposition se récrivent

(1) $\complement \bar{A} = \operatorname{Int}(\complement A)$,

(2) $\complement \operatorname{Int}(A) = \overline{\complement A}$.

Démonstration. Nous avons $a \in V \backslash \bar{A}$ si et seulement si $a \notin \bar{A}$. Or ne pas être dans $\bar{A}$ signifie qu'il existe un rayon ε tel que la boule $B(a, \varepsilon)$ n'intersecte pas A. Le fait que la boule $B(a, \varepsilon)$ n'intersecte pas A est équivalent à dire que $B(a, \varepsilon) \subset V \backslash A$. Or cela est exactement la définition du fait que a est à l'intérieur de $V \backslash A$. Nous avons donc montré que $a \in V \backslash \bar{A}$ si et seulement si $a \in \operatorname{Int}(V \backslash A)$. Cela prouve la première affirmation.

Pour prouver la seconde affirmation, nous appliquons la première au complémentaire de A : $\complement(\overline{\complement A}) = \operatorname{Int}(\complement\complement A)$. En prenant le complémentaire des deux membres nous trouvons successivement

$$\begin{aligned} \complement\complement(\overline{\complement A}) &= \complement \operatorname{Int}(\complement\complement A), \\ \overline{\complement A} &= \complement \operatorname{Int}(A), \end{aligned} \tag{9.133}$$

ce qu'il fallait démontrer. □

Attention à ne pas confondre $\complement \bar{A}$ et $\overline{\complement A}$. Ces deux ensembles ne sont pas égaux. En effet, en tant que complément d'un fermé, l'ensemble $\complement \bar{A}$ est certainement ouvert, tandis que, en tant que fermeture, l'ensemble $\overline{\complement A}$ est fermé. Pouvez-vous trouver des exemples d'ensembles A tels que $\complement \bar{A} = \overline{\complement A}$?

Proposition 9.67.
Soient A *et* B *deux parties de l'espace vectoriel normé* V.

(1) Pour les inclusions, si $A \subset B$*, alors* $\operatorname{Int}(A) \subset \operatorname{Int}(B)$ *et* $\bar{A} \subset \bar{B}$.

(2) Pour les unions, $\overline{A \cup B} = \bar{A} \cup \bar{B}$ *et* $\overline{A \cap B} \subset \bar{A} \cap \bar{B}$.

(3) Pour les intersections, $\operatorname{Int}(A) \cap \operatorname{Int}(B) = \operatorname{Int}(A \cap B)$ *et* $\operatorname{Int}(A) \cup \operatorname{Int}(B) \subset \operatorname{Int}(A \cup B)$.

Démonstration. (1) Si a est dans l'intérieur de A, il existe une boule autour de a contenue dans A. Cette boule est alors contenue dans B et donc est une boule autour de a contenue dans B, ce qui fait que a est dans l'intérieur de B. Si maintenant a est dans l'adhérence de A, toute boule centrée en a contient un élément de A et donc un élément de B, ce qui prouve que a est dans l'adhérence de B.

(2) Nous avons $A \subset A \cup B$ et donc, en utilisant le premier point, $\bar{A} \subset \overline{A \cup B}$. De la même manière, $\bar{B} \subset \overline{A \cup B}$. En prenant l'union, $\bar{A} \cup \bar{B} \subset \overline{A \cup B}$.
Réciproquement, soit $a \in \overline{A \cup B}$ et montrons que $a \in \bar{A} \cup \bar{B}$. Supposons par l'absurde que a ne soit ni dans $\bar{A}$ ni dans $\bar{B}$. Il existe donc des rayons ε_1 et ε_2 tels que

$$\begin{aligned} B(a, \varepsilon_1) \cap A &= \varnothing, \\ B(a, \varepsilon_2) \cap B &= \varnothing. \end{aligned} \tag{9.134}$$

En prenant $r = \min\{\varepsilon_1, \varepsilon_2\}$, la boule $B(a, r)$ est inclue aux deux boules citées et donc n'intersecte ni A ni B. Donc $a \notin \overline{A \cup B}$, d'où la contradiction.

(3) Si nous appliquons le second point à $\complement A$ et $\complement B$, nous trouvons

$$\overline{\complement A \cup \complement B} = \overline{\complement A} \cup \overline{\complement B}. \tag{9.135}$$

En utilisant les propriétés du lemme 1.17, le membre de gauche devient

$$\overline{\complement A \cup \complement B} = \overline{\complement(A \cap B)} = \complement \operatorname{Int}(A \cap B), \tag{9.136}$$

tandis que le membre de droite devient

$$\overline{\complement A} \cup \overline{\complement B} = \complement \operatorname{Int}(A) \cup \complement \operatorname{Int}(A) = \complement\Big(\operatorname{Int}(A) \cap \operatorname{Int}(B)\Big). \tag{9.137}$$

En égalisant le membre de droite de (9.136) avec celui de (9.137) et en passant au complémentaire nous trouvons

$$\operatorname{Int}(A \cap B) = \operatorname{Int}(A) \cap \operatorname{Int}(B), \tag{9.138}$$

comme annoncé.

La dernière affirmation provient du fait que $\operatorname{Int}(A) \subset \operatorname{Int}(A \cup B)$ et de la propriété équivalente pour B.

□

Remarque 9.68.
Nous avons prouvé que $\overline{A \cap B} \subset \bar{A} \cap \bar{B}$. Il arrive que l'inclusion soit stricte, comme dans l'exemple suivant. Si nous prenons $A = [0, 1]$ et $B =]1, 2]$, nous avons $A \cap B = \emptyset$ et donc $\overline{A \cap B} = \emptyset$. Par contre nous avons $\bar{A} \cap \bar{B} = \{1\}$.

Définition 9.69.
*La **frontière** d'un sous-ensemble A de l'espace vectoriel normé V est l'ensemble des points $a \in V$ tels que*

$$\begin{aligned} B(a, r) \cap A &\neq \emptyset, \\ B(a, r) \cap \complement A &\neq \emptyset, \end{aligned} \tag{9.139}$$

pour tout rayon r. En d'autres termes, toute boule autour de a contient des points de A et des points de $\complement A$. La frontière de A se note ∂A.

Proposition 9.70.
La frontière d'une partie A d'un espace vectoriel normé V s'exprime sous la forme

$$\partial A = \bar{A} \backslash \operatorname{Int}(A). \tag{9.140}$$

Démonstration. Le fait pour un point a de V d'appartenir à $\bar{A}$ signifie que toute boule centrée en a intersecte A. De la même façon, le fait de ne pas appartenir à $\operatorname{Int}(A)$ signifie que toute boule centrée en a intersecte $\complement A$. □

La description de la frontière donnée par la proposition 9.70 est celle qu'en pratique nous utilisons le plus souvent. Dans certains textes, elle est prise comme définition de la frontière.

Lemme 9.71.
La frontière de A peut également s'exprimer des façons suivantes :

$$\partial A = \bar{A} \cap \complement \operatorname{Int}(A) = \bar{A} \cap \overline{\complement A}, \tag{9.141}$$

Démonstration. En partant de $\partial A = \bar{A} \backslash \operatorname{Int}(A)$, la première égalité est une application de la propriété (4) du lemme 1.17. La seconde égalité est alors la proposition 9.66. □

Exemple 9.72.
Dans $\mathbb{R}$, la frontière d'un intervalle est la paire constituée des points extrêmes. En effet

$$\partial[a, b[= \overline{[a, b[} \backslash \operatorname{Int}\big([a, b[\big) = [a, b] \backslash]a, b[= \{a, b\}. \tag{9.142}$$

Toujours dans $\mathbb{R}$ nous avons

$$\partial\mathbb{R} = \bar{\mathbb{R}}\backslash \operatorname{Int}(\mathbb{R}) = \mathbb{R}\backslash\mathbb{R} = \varnothing, \tag{9.143}$$

et

$$\partial\mathbb{Q} = \bar{\mathbb{Q}}\backslash \operatorname{Int}(\mathbb{Q}) = \mathbb{R}\backslash\varnothing = \mathbb{R}. \tag{9.144}$$

△

Exemple 9.73.
Dans $\mathbb{R}^n$, nous avons

$$\partial B(a,r) = \partial\bar{B}(a,r) = S(a,r). \tag{9.145}$$

Cela est un boulot pour la proposition 9.48. Si $x \in S(a,r)$ alors tout boule autour de x contient des points à distance strictement plus grande et plus petite que $d(a,x)$, c'est-à-dire des points dans $B(a,r)$ et hors de $B(a,r)$. Cela prouve que les points de $S(a,r)$ font partie de $\partial B(a,r)$, c'est-à-dire que $S(a,r) \subset \partial B(a,r)$; et idem pour $\bar{B}(a,r)$.

Pour prouver l'inclusion inverse, soit $x \in \partial B(a,r)$. Vu que toute boule autour de x contient des points intérieurs à $B(a,r)$, pour tout $\epsilon > 0$, $d(a,x) - \epsilon < r$, c'est-à-dire que $d(a,x) \leqslant r$. De la même manière toute boule autour de x contient des points hors de $B(a,r)$ signifie que pour tout ϵ, $d(a,x) + \epsilon > r$ ou encore que $d(a,x) \geqslant r$. Nous avons donc $d(a,x) = r$. △

Remarque 9.74.
Il serait toutefois faux de croire que $\partial A = \partial\bar{A}$ pour toute partie A de $\mathbb{R}^n$. En effet si $A = \mathbb{R}\backslash\{0\}$ nous avons $\partial A = \{0\}$ et $\bar{A} = \mathbb{R}$, donc $\partial\bar{A} = \varnothing$.

9.4.3 Point isolé, point d'accumulation

Définition 9.75.
Soit D, une partie de V.

*(1) Un point $a \in D$ est dit **isolé** dans D relativement à V s'il existe un $\varepsilon > 0$ tel que*

$$B(a,\varepsilon) \cap D = \{a\}. \tag{9.146}$$

*(2) Un point $a \in V$ est un **point d'accumulation** de D si pour tout $\varepsilon > 0$,*

$$\Big(B(a,\varepsilon)\backslash\{a\}\Big) \cap D \neq \varnothing. \tag{9.147}$$

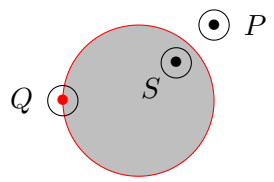


FIGURE 9.3 – L'ensemble décrit par l'équation (9.148). Le point P est un point isolé de D, tandis que les points S et Q sont des points d'accumulation.

Exemple 9.76.
Considérons la partie suivante de $\mathbb{R}^2$:

$$D = \{(x,y) \text{ tel que } x^2 + y^2 < 1\} \cup \{(1,1)\}. \tag{9.148}$$

Comme on peut le voir sur la figure 9.3, le point $P = (1,1)$ est un point isolé de D parce qu'on peut tracer une boule autour de P sans inclure d'autres points de D que P lui-même. Le point $Q = (-1,0)$ est un point d'accumulation de D parce que toute boule autour de Q contient des points de D.

Le point S, étant un point intérieur, est un point d'accumulation : toute boule autour de S intersecte D.

Notez cependant que le point Q lui-même n'est pas dans D parce que l'inégalité qui définit D est stricte. △

Remarque 9.77.
À propos de la position des points d'accumulation et des points isolés.

(1) Les points intérieurs sont tous des points d'accumulation.
(2) Les points isolés ne sont jamais intérieurs.
(3) Certains points d'accumulation ne font pas partie de l'ensemble. Par exemple le point 1 est un point d'accumulation de $E =]0, 1[$.
(4) Les points de la frontière sont soit d'accumulation soit isolés.

Exemple 9.78.
Tous les points de $\mathbb{R}$ sont des points d'accumulation de $\mathbb{Q}$ parce que dans toute boule autour d'un réel, on peut trouver un nombre rationnel. △

Remarque 9.79.
L'ensemble des points d'accumulation d'un ensemble n'est pas exactement son adhérence. En effet, un point isolé dans A est dans l'adhérence de A, mais n'est pas un point d'accumulation de A.

9.4.4 Des exemples

Exemple 9.80.
Nous considérons l'ensemble.

$$A_1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 - 5x + 6 < y \leqslant 2\}. \tag{9.149}$$

Si un point $(x, y) \in \mathbb{R}^2$ est tel que $x^2 - 5x + 6 < y$, alors dans une boule centrée en (x, y) (de rayon r_1), l'inégalité reste vraie (parce que la fonction $x^2 - 5x + 6 - y$ est une fonction continue). De la même manière, si nous avons $y < 2$ en (x, y), alors nous avons encore l'inégalité dans une boule de rayon r_2. En prenant $r = \min\{r_1, r_2\}$, les deux inégalités restent vraies dans la boule de rayon r.

Donc les points (x, y) tels que $x^2 - 5x + 6 < y < 2$ sont dans l'intérieur de A_1.

Pour les mêmes raisons, autour d'un point (x, y) tel que $x^2 - 5x + 6 > y$, nous pouvons trouver une boule dans laquelle l'inégalité reste stricte. Ces points ne sont donc pas dans l'adhérence de A_1. Un point qui vérifie $x^2 - 5x + 6 = y = 2$ est par contre dans l'adhérence parce que dans toute boule, on pourra trouver un x tel que $x^2 - 5x + 6 < y$, et un y. L'adhérence est donc donnée par les inéquations

$$\bar{A}_1 \equiv x^2 - 5x + 6 \leqslant y \leqslant 2. \tag{9.150}$$

La frontière est donnée par les points de l'adhérence qui ne sont pas dans l'intérieur de A_1. Attention : **ne pas dire** que la frontière est alors donnée simplement en remplaçant les inégalités par des égalités : $\partial A_1 \equiv x^2 - 5x + 6 = y = 2$. Quel est cet ensemble ?

Trouver la frontière demande un peu plus de travail. Le point marqué sur la figure 9.4 est sur la frontière parce que toute boule intersecte l'intérieur et l'extérieur. Cela est dû au fait que, sur ce point, nous ayons $x^2 - 5x + 6 = y$ en même temps que $y < 2$. Donc si on prend une boule assez petite, on conserve $y < 2$, mais on obtient des points tels que $x^2 - 5x + 6 < y$.

En voyant le dessin, la chose à faire pour écrire la frontière est de trouver les deux points d'intersection entre la parabole et la droite horizontale. Ces points sont les points (x, y) qui satisfont au système

$$\begin{cases} x^2 - 5x + 6 = y & \text{(9.151a)} \\ y = 2. & \text{(9.151b)} \end{cases}$$

En substituant la seconde équation dans la première, il vient $x^2 - 5x + 6 = 2$, ce qui nous donne à résoudre le polynôme du second degré $x^2 - 5x + 4 = 0$. Les solutions sont $x = 1$ et $x = 4$, et les

deux points d'intersection sont les points $P = (1, 2)$ et $Q = (4, 2)$. Les points de la frontière de A_1 sont donc donnés par

$$\begin{aligned}\partial A_1 = \{(x, y) \in \mathbb{R}^2 \text{ tels que } x^2 - 5x + 6 = y \text{ et } 1 \leqslant x \leqslant 4\} \\ \cup \{(x, y) \in \mathbb{R}^2 \text{ tels que } y = 2 \text{ et } 1 \leqslant x \leqslant 4\}.\end{aligned} \tag{9.152}$$

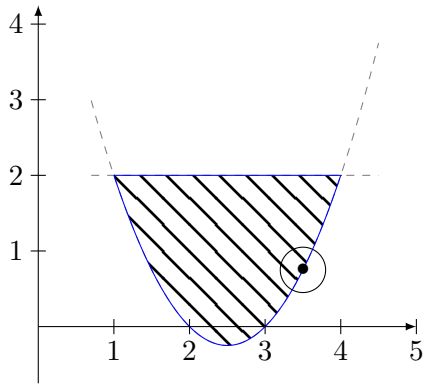


FIGURE 9.4 – En hachuré : l'intérieur ; en trait plein : la frontière. L'adhérence est l'union des deux. Exemple9.80.

Notez que les points de la parabole qui sont sur la frontière ne font pas partie de l'ensemble A_1 lui-même, tandis que ceux de la frontière qui sont sur la droite horizontale en font partie sauf $(4, 2)$ et $(1, 2)$.

L'intérieur de A_1 n'étant pas égal à A_1, cet ensemble n'est pas ouvert ; de la même manière, vu que $\bar{A}_1 \neq A_1$, l'ensemble n'est pas fermé. L'ensemble A_1 est par contre borné parce qu'il est contenu par exemple dans la boule de centre $(0, 0)$ et de rayon 5. Les points d'accumulation de A_1 sont les points de sa fermeture. △

Exemple 9.81.
Nous étudions

$$A_2 = \{(x, y) \in \mathbb{R}^2 \mid x + 1 < y < 2x\}. \tag{9.153}$$

Pour les mêmes raisons que dans l'exemple 9.80 l'intérieur est donné par

$$\mathrm{Int}(A_2) \equiv x + 1 < y < 2x; \tag{9.154}$$

L'adhérence est donnée par

$$\overline{A_2} \equiv x + 1 \leqslant y \leqslant 2x, \tag{9.155}$$

Pour la frontière, les deux droites dont il est question dans la définition de A_2 (les droites $y = x+1$ et $y = 2x$) se coupent en $x = 1$ (refaire soi-même le dessin de la figure 9.5). Lorsque $x < 1$, les conditions $x + 1 < y$ et $y < 2x$ sont incompatibles : aucun point de A_2 n'est dans la partie $x < 1$ du plan. Lorsque $x > 1$, alors les points situés *entre* les deux droites font partie de A_2. La frontière est donc donnée par ces deux droites pour $x \geqslant 1$.

Étant donné que $\mathrm{Int}(A_2) = A_2$, cet ensemble est ouvert (et donc pas fermé par la proposition 9.55(4)). Il n'est par contre pas borné parce qu'il contient des point (x, y) avec des x arbitrairement grands. △

9.5 Valeur propre et vecteur propre

9.5.1 Généralités

Nous savons qu'une application *linéaire* $A\colon \mathbb{R}^3 \to \mathbb{R}^3$ est complètement définie par la donnée de son action sur les trois vecteurs de base, c'est-à-dire par la donnée de

$$Ae_1, \quad Ae_2 \quad \text{et} \quad Ae_3. \tag{9.156}$$

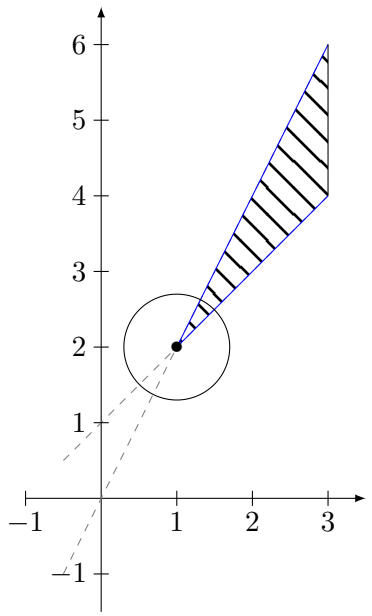


FIGURE 9.5 – Notez que le point d'angle fait partie de la frontière, mais pas de l'ensemble. Exemple 9.81.

Nous allons former la matrice de A en mettant simplement les vecteurs Ae_1, Ae_2 et Ae_3 en colonne. Donc la matrice

$$A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \tag{9.157}$$

signifie que l'application linéaire A envoie le vecteur e_1 sur $\begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}$, le vecteur e_2 sur $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ et le vecteur e_3 sur $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Pour savoir comment A agit sur n'importe quel vecteur, on applique la règle de produit vecteur$\times$matrice :

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + 2y + 3z \\ 4x + 5y + 6z \\ 7x + 8y + 9z \end{pmatrix}. \tag{9.158}$$

Une chose intéressante est de savoir quelles sont les directions invariantes de la transformation linéaire. Par exemple, on peut lire sur la matrice (9.157) que la direction $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ est invariante : elle est simplement multipliée par 3. Dans cette direction, la transformation est juste une dilatation. Afin de savoir si v est un vecteur d'une direction conservée, il faut voir s'il existe un nombre λ tel que $Av = \lambda v$, c'est-à-dire voir si v est simplement dilaté.

L'équation $Av = \lambda v$ se récrit $(A - \lambda \mathbb{1})v = 0$, c'est-à-dire qu'il faut résoudre l'équation

$$(A - \lambda \mathbb{1}) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{9.159}$$

Nous savons qu'une telle équation ne peut avoir de solutions que si $\det(A - \lambda \mathbb{1}) = 0$. La première étape est donc de trouver les λ qui vérifient cette condition.

9.5.2 Dans le vif du sujet

Définition 9.82.
*Soit un $\mathbb{K}$-espace vectoriel E et un endomorphisme $A\colon V \to V$. Un **vecteur propre** de A est un vecteur $v \neq 0$ tel que $Av = \lambda v$ pour un certain $\lambda \in \mathbb{K}$. Dans ce cas, λ est la **valeur propre** de v.*

*L'**espace propre** de A pour la valeur λ*[19] *est l'ensemble des vecteurs propres de A pour la valeur propre λ et zéro.*

Définition 9.83.
*L'ensemble de valeurs propres de l'endomorphisme u est son **spectre** et est noté* $\operatorname{Spec}(u)$.

Remarque 9.84.
Le nombre zéro peut être une valeur propre ; c'est le vecteur zéro qui ne peut pas être vecteur propre. La matrice nulle est une matrice diagonalisable.

Lemme 9.85.
Soit u un endomorphisme et $E_\lambda(u)$ ses espaces propres. La somme des V_λ est directe.

Démonstration. Soit $v_i \in V_{\lambda_i}$ un choix de vecteurs propres de u. Si la somme n'est pas directe, nous pouvons considérer une combinaison linéaire des v_i qui soit nulle :

$$v_1 + \cdots + v_p = 0. \tag{9.160}$$

Appliquons $(A - \lambda_1 \mathbb{1})$ à cette égalité :

$$(\lambda_2 - \lambda_1)v_1 + \cdots + (\lambda_p - \lambda_1)v_p = 0. \tag{9.161}$$

En appliquant encore successivement les opérateurs $(A - \lambda_i \mathbb{1})$ nous réduisons le nombre de termes jusqu'à obtenir $v_p = 0$. □

Proposition 9.86 ([176]).
Soit E, un espace vectoriel sur un corps infini et $(F_k)_{k=1,\dots,r}$, des sous-espaces vectoriels propres[20] *de E tels que $\bigcup_{i=1}^r F_i = E$. Alors $E = F_k$ pour un certain k.*

Autrement dit, l'union finie de sous-espaces propres ne peut être égal à l'espace complet.

9.6 Polynômes d'endomorphismes

Soit A un anneau commutatif et $\mathbb{K}$, un corps commutatif. L'injection canonique $A \to A[X]$ se prolonge en une injection

$$\mathbb{M}(A) \to \mathbb{M}\big(A[X]\big). \tag{9.162}$$

9.6.1 Polynômes d'endomorphismes

Soit $u \in \operatorname{End}(E)$ où E est un $\mathbb{K}$-espace vectoriel. Nous considérons l'application

$$\begin{aligned} \varphi_u\colon \mathbb{K}[X] &\to \operatorname{End}(E) \\ P &\mapsto P(u). \end{aligned} \tag{9.163}$$

L'image de φ_u est un sous-espace vectoriel. En effet si $A = \varphi_u(P)$ et $B = \varphi_u(Q)$, alors $A + B = \varphi_u(P + Q)$ et $\lambda A = (\lambda P)(u)$. En particulier c'est un espace fermé.

Soit u un endomorphisme d'un $\mathbb{K}$-espace vectoriel E et P, un polynôme. Nous disons que P est un polynôme **annulateur** de u si $P(u) = 0$ en tant que endomorphisme de E.

Lemme 9.87.
Si P et Q sont des polynômes dans $\mathbb{K}[X]$ et si u est un endomorphisme d'un $\mathbb{K}$-espace vectoriel E, nous avons

$$(PQ)(u) = P(u) \circ Q(u). \tag{9.164}$$

19. Nous laissons au lecteur le soin de vérifier que c'est bien un sous-espace vectoriel de E.
20. Définition 9.82.

Démonstration. Si $P = \sum_i a_i X^i$ et $Q = \sum_j b_j X^j$, alors le coefficient de X^k dans PQ est

$$\sum_l a_l b_{k-l}. \tag{9.165}$$

Par conséquent $(PQ)(u)$ contient $\sum_l a_l b_{k-l} u^k$. Par ailleurs $P(u) \circ Q(u)$ est donné par

$$\sum_i a_i u^i \left(\sum_j b_j u^j \right)(x) = \sum_{ij} a_i b_j u^{i+j}(x). \tag{9.166}$$

Le coefficient du terme en u^k est bien le même que celui donné par (9.165). □

Théorème 9.88 (Décomposition des noyaux ou lemme des noyaux)**.**
Soit u un endomorphisme du $\mathbb{K}$-espace vectoriel E. Soit $P \in \mathbb{K}[X]$ un polynôme tel que $P(u) = 0$. Nous supposons que P s'écrive comme le produit $P = P_1 \dots P_n$ de polynômes deux à deux étrangers[21]*. Alors*

$$E = \ker P_1(u) \oplus \dots \oplus \ker P_n(u). \tag{9.167}$$

De plus les projecteurs associés à cette décomposition sont des polynômes en u.

Ce résultat est utilisé pour prouver que toute représentation est décomposable en représentations irréductibles, proposition 15.9 ainsi que pour le théorème 9.128 qui dit que si le polynôme minimal d'un endomorphisme est scindé à racine simple alors il est diagonalisable.

Démonstration. Nous posons

$$Q_i = \prod_{j \neq i} P_i. \tag{9.168}$$

Par le lemme 6.42 ces polynômes sont étrangers entre eux et le théorème de Bézout (théorème 6.41) donne l'existence de polynômes R_i tels que

$$R_1 Q_1 + \dots + R_n Q_n = 1. \tag{9.169}$$

Si nous appliquons cette égalité à u et ensuite à $x \in E$ nous trouvons

$$\sum_{i=1}^n (R_i Q_i)(u)(x) = x, \tag{9.170}$$

et en particulier si nous posons $E_i = \text{Image}\big(P_i Q_i(u)\big)$ nous avons

$$E = \sum_{i=1}^n E_i. \tag{9.171}$$

Cette dernière somme n'est éventuellement pas une somme directe. Si $i \neq j$, alors $Q_i Q_j$ est multiple de P et nous avons, en utilisant le lemme 9.87,

$$(R_i Q_i)(u) \circ (R_j Q_j)(u) = \big(R_i Q_i R_j Q_j\big)(u) = S_{ij}(u) \circ P(u) = 0 \tag{9.172}$$

où S_{ij} est un polynôme.

Nous pouvons voir E comme un $\mathbb{K}$-module et appliquer le théorème 3.80. Les opérateurs $R_i Q_i(u)$ ont l'identité comme somme et sont orthogonaux, et nous avons donc la décomposition en somme directe :

$$E = \bigoplus_{i=1}^n R_i Q_i(u) E. \tag{9.173}$$

Afin de terminer la preuve, nous devons montrer que $R_i Q_i(u) E = \ker P_i(u)$. D'abord nous avons

$$P_i R_i Q_i(u) = (R_i P)(u) = R_i(u) \circ P(u) = 0, \tag{9.174}$$

21. Définition 3.179.

par conséquent $\text{Image}(R_iQ_i(u)) \subset \ker P_i(u)$. Pour obtenir l'inclusion inverse, nous reprenons l'équation (9.170) avec $x \in \ker P_i(u)$. Elle se réduit à

$$(R_iQ_i)(u)x = x. \tag{9.175}$$

Par conséquent $x \in \text{Image}\big(R_iQ_i(u)\big)$. □

Corolaire 9.89.
Soit E, un $\mathbb{K}$-espace vectoriel de dimension finie et f, un endomorphisme semi-simple dont la décomposition du polynôme minimal μ_f en facteurs irréductibles sur $\mathbb{K}[X]$ est $\mu_f = M_1^{\alpha_1} \cdots M_r^{\alpha_r}$. Si F est un sous-espace stable par f, alors

$$F = \bigoplus_{i=1}^{r} \ker M_i^{\alpha_i}(f) \cap F \tag{9.176}$$

Démonstration. Nous posons $E_i = \ker M_i^{\alpha_i}(f)$ et $F_i = E_i \cap F$. Les polynômes $M_i^{\alpha_i}$ sont deux à deux étrangers et $\mu_f(f) = 0$, donc le lemme des noyaux (9.88) s'applique et

$$E = E_1 \oplus \ldots \oplus E_r. \tag{9.177}$$

Nous pouvons décomposer $x \in F$ en termes de cette somme :

$$x = x_1 + \cdots + x_r \tag{9.178}$$

avec $x_i \in E_i$. Toujours selon le lemme des noyaux, les projections sur les espaces E_i sont des polynômes en f. Par conséquent F est stable sous toutes ces projections $\texttt{proj}_i \colon E \to E_i$, et en appliquant $\texttt{proj}_i$ à (9.178), $\texttt{proj}_i(x) = x_i$. Vu que $x \in F$, le membre de gauche est encore dans F et $x_i \in E_i \cap F$. Nous avons donc

$$F \subset \bigoplus_{i=1}^{r} F_i. \tag{9.179}$$

L'inclusion inverse est immédiate parce que $F_i \subset F$ pour chaque i. □

Lemme 9.90.
Si x est un vecteur propre de valeur propre λ pour l'endomorphisme u et si P est un polynôme, alors x est vecteur propre de u pour la valeur propre $P(\lambda)$.

Démonstration. C'est un simple calcul de $P(u)x$ en ayant noté $P(X) = \sum_{k=0}^{n} c_k X^n$:

$$P(u)x = \sum_{k=0}^{n} c_k u^k(x) = \sum_{k=0}^{n} c_k \lambda^k u = P(\lambda)x. \tag{9.180}$$

□

9.6.2 Polynôme minimal et minimal ponctuel

Lemme-définition 9.91.
Soit un endomorphisme $f \colon E \to E$ d'un $\mathbb{K}$-espace vectoriel de dimension finie. Il existe un unique polynôme annulateur normalisé de degré minimum.

*Il est nommé le **polynôme minimal** de f et il est noté μ_f ou simplement μ lorsque la dépendance en f est claire.*

Démonstration. Pour l'unicité, soient P et Q deux polynômes annulateur de f de même degré N et ayant tous deux 1 comme coefficient de x^N. Alors $P - Q$ est de degré $N - 1$ tout en étant encore annulateur.

Pour l'existence, les endomorphismes Id, f, f^2, ... ne peuvent pas être tous linéairement indépendants parce que la dimension de $\text{End}(E)$ est finie. Il existe donc un nombre N et des coefficients a_k tels que $\sum_{k=0}^{N} a_k f^k = 0$. Le polynôme $P(X) = \sum_{k=0}^{N} a_k X^k$ est donc annulateur de f.

Une autre façon de le dire est que l'application linéaire $\varphi \colon \mathbb{K}[X] \to \text{End}(E)$ donnée par $\varphi(P) = P(f)$ est un endomorphisme d'un espace vectoriel de dimension infinie vers un espace vectoriel de dimension finie. Il ne peut donc pas être injectif et possède donc un noyau non réduit à zéro. □

Remarque 9.92.
La preuve donnée ci-dessus montre que $\deg(\mu) \leqslant \dim(E)^2$. Comme conséquence du théorème de Caley-Hamilton 9.115 nous verrons qu'en réalité le degré du polynôme minimal est majoré par la dimension de l'espace.

Exemple 9.93 (Pas en dimension infinie)**.**
L'endomorphisme de dérivation △

Dans la suite, l'endomorphisme f du $\mathbb{K}$-espace vectoriel E de dimension n est fixé. Pour $x \in E$ nous notons

$$E_x = \{P(f)x \text{ tel que } P \in \mathbb{K}[X]\}. \tag{9.181}$$

Nous considérons le morphisme d'algèbres

$$\begin{aligned} \varphi\colon \mathbb{K}[X] &\to \operatorname{End}(E) \\ P &\mapsto P(f) \end{aligned} \tag{9.182}$$

et si $x \in E$ est donné nous considérons le morphisme de $\mathbb{K}$-espaces vectoriels

$$\begin{aligned} \varphi_x\colon \mathbb{K}[X] &\to E \\ P &\mapsto P(f)x. \end{aligned} \tag{9.183}$$

Les noyaux de ces applications sont des idéaux, entre autres par le lemme 9.87. Ils ont donc un unique générateur unitaire (chacun) par le théorème 6.37. En termes de vocabulaire, l'ensemble

$$\ker(\phi) = \{Q \in \mathbb{K}[X] \text{ tel que } Q(f) = 0\} \tag{9.184}$$

est l'**idéal annulateur** de f et un polynôme Q tel que $Q(f) = 0$ est une polynôme annulateur de f.

Définition 9.94.
Le générateur unitaire de $\ker(\varphi_x)$ *est le* ***polynôme minimal ponctuel*** *de* f *en* x*. Il sera noté* $\mu_{f,x}$ *ou* μ_x *lorsque la dépendance en* f *est claire dans le contexte.*

Nous notons μ le générateur unitaire du noyau de φ et μ_x celui de φ_x. Vu que $\mu \in \ker(\varphi_x)$ pour tout x nous avons $\mu_x \mid \mu$ pour tout x.

Exemple 9.95 (Pas en dimension infinie)**.**
En dimension infinie, il n'y a pas toujours de polynôme annulateur. Si E est un espace vectoriel de dimension infine ayant une base dénombrable $\{e_i\}_{i\in\mathbb{N}}$ alors l'opérateur donné par $f(e_i) = e_{i+1}$ n'a pas de polynôme annulateur. Même pas ponctuel en quel que point que ce soir.

De même l'opérateur donné par $g(e_1) = 0$ et $g(e_i) = e_{i-1}$ si $i \neq 1$ n'a pas de polynôme annulateur, mais il a un polynôme annulateur ponctuel évident en $x = e_1$. L'exemple 14.63 donnera un habillage à peine subtil à cet exemple. △

Proposition 9.96.
Si P *est un polynôme tel que* $P(f) = 0$*, alors le polynôme minimal* μ_f *divise* P*. Autrement dit, le polynôme minimal engendre l'idéal des polynômes annulateurs.*

Démonstration. L'ensemble $\ker(\varphi) = \{Q \in \mathbb{K}[X] \text{ tel que } Q(u) = 0\}$ est un idéal par le lemme 9.87. Le polynôme minimal de u est un élément de degré plus bas dans I et par conséquent $I = (\mu_u)$ par le théorème 6.37. Nous concluons que μ_u divise tous les éléments de I. □

La proposition suivante permet de caractériser le polynôme minimal.

Proposition 9.97 ([80])**.**
Soit une application linéaire f *sur un* $\mathbb{K}$*-espace vectoriel. Il existe un unique polynôme unitaire*[22] $P \in \mathbb{K}[X]$ *tel que*

22. À mon avis, « unitaire » manque dans [80].

(1) $P(f) = 0$;

(2) l'application

$$\begin{aligned} \varphi\colon \frac{\mathbb{K}[X]}{(P)} &\to \operatorname{End}(E) \\ \bar{Q} &\mapsto Q(f) \end{aligned} \tag{9.185}$$

est injective.

Démonstration. En ce qui concerne l'existence, il existe le polynôme minimal de f qui satisfait les conditions. Pour l'unicité nous travaillons maintenant.

Supposons que l'application (9.185) soit injective. Alors pour tout $Q \in \mathbb{K}[X]$ tel que $Q(f) = 0$ nous avons $\bar{Q} = 0$, c'est-à-dire $Q = PR$ pour un certain $R \in \mathbb{K}[X]$. Autrement dit : P est un générateur unitaire de l'idéal annulateur de f. Le théorème 6.37(3) nous dit alors que $P = \mu$ parce que μ est également générateur unitaire. □

Lemme 9.98 ([177]).
Soit $f\colon E \to E$ un endomorphisme de l'espace vectoriel E. Il existe un élément $x \in E$ tel que $\mu_{f,x} = \mu_f$.

Démonstration. Soit une décomposition en irréductibles du polynôme minimal $\mu = P_1^{\alpha_1} \dots P_r^{\alpha_r}$. Nous notons $E_i = \ker\big(P_i^{\alpha_i}(f)\big)$. Les polynômes P_i sont étrangers deux à deux (un diviseur commun aurait a fortiori été un diviseur et aurait contredit l'irréductibilité). Le lemme des noyaux 9.88 nous donne la somme directe

$$E = \bigoplus_{i=1}^{r} \ker\big(P_i^{\alpha_i}(f)\big). \tag{9.186}$$

Si $x_i \in E_i$ alors μ_{x_i} est une puissance de P_i. En effet $\mu_{x_i} \mid \mu$ et est donc un produit des puissances des P_j. Or si $(QP_j)(f)x_i = 0$ alors $(P_jQ)(f)x_i = 0$, ce qui donne $Q(f)x_i \in E_j \cap E_i = \{0\}$. Donc μ_{x_i} n'est pas de la forme QP_j pour $j \neq i$. Nous en déduisons que μ_{x_i} est une puissance de P_i dès que $x_i \in E_i$. Nous choisissons $x_i \in E_i$ tel que $\mu_{x_i} = P_i^{\alpha_i}$.

Nous posons enfin $a = x_1 + \dots + x_r$; par définition du polynôme annulateur μ_a, nous avons

$$0 = \mu_a(f)a = \mu_a(f)x_1 + \dots + \mu_a(f)x_r. \tag{9.187}$$

Mais $m_a(f)x_j \in E_i$, et la somme des E_j est directe, donc l'annulation de la somme (9.187) implique l'annulation de chacun des termes : $\mu_a(f)x_i = 0$ pour tout i. Cela prouve que $\mu_{x_i} \mid \mu_a$. Mais comme les μ_{x_i} sont premiers deux à deux (parce que ce sont les $P_i^{\alpha_i}$), nous avons que le produit divise encore μ_a :

$$\prod_{i=1}^{r} \mu_{x_i} \mid \mu_a, \tag{9.188}$$

c'est-à-dire $\mu \mid \mu_a$. Comme nous avons aussi $\mu_a \mid \mu$, nous déduisons $\mu_a = \mu$. □

Définition 9.99 (Matrices, endomorphismes et vecteurs cycliques).
*Une matrice est **cyclique** si elle est semblable à une matrice compagnon. Un endomorphisme $f\colon E \to E$ est **cyclique** s'il existe un vecteur $x \in E$ tel que $\{f^k(x)\}_{k=0,\dots,n-1}$ est une base de E. Un vecteur ayant cette propriété est un **vecteur cyclique** pour f.*

Lemme 9.100.
Soit E un espace vectoriel de dimension finie et un endomorphisme cyclique[23] *f de E. Soit un vecteur cyclique v de f, alors le polynôme minimal de f est égal au polynôme minimal de f au point v : $\mu_f = \mu_{f,v}$.*

Démonstration. Montrons que $\mu_{f,v}$ est un polynôme annulateur de f, ce qui prouvera que μ_f divise $\mu_{f,v}$ par la proposition 9.96. Étant donné que v est cyclique, tout élément de E s'écrit sous la forme

23. Voir la définition 9.99.

$x = Q(f)v$. Prenons un polynôme P annulateur de f en v : $P(f)v = 0$. Nous montrons que P est alors un polynôme annulateur de f. En effet, nous avons

$$P(f)x = \big(P(f) \circ Q(f)\big)v = \big(Q(f) \circ P(f)\big)v = 0 \tag{9.189}$$

où nous avons utilisé le lemme 9.87. □

Lemme 9.101 ([177]).
Soit $a \in E$ tel que $\mu_a = \mu$. Alors E_a est un sous-espace stable pour f pour lequel il existe un supplémentaire stable.

Démonstration. Soit $l = \deg(\mu) = \deg(\mu_a)$. L'espace E_a étant engendré par les $f^k(a)$ nous savons que $e_1 = a$, $e_2 = f(a)$,..., $e_l = k^{l-1}(a)$ forment une base de E_a. Nous pouvons la compléter en une base $\{e_1, \dots, e_n\}$ de E. Et nous posons [24]

$$G = \{x \in E \text{ tel que } e_l^*\big(f^k(x)\big) = 0\, \forall k \geqslant 0\} \tag{9.190a}$$
$$= \bigcap_{k\geqslant 0} \ker\{e_l^* \circ f^k\} \tag{9.190b}$$
$$= \bigcap_{k=0}^{l-1} \ker(e_l^* \circ f^k). \tag{9.190c}$$

La dernière égalité est due au fait que l soit le degré de μ. Du coup f^l est une combinaison linéaire des f^i avec $i \leqslant l-1$.

Nous avons $f(G) \subset G$ et de plus $E_a \cap G = \{0\}$ parce qu'un élément de E_a est une combinaison linéaire d'éléments de la forme $f^j(a)$ $(j \leqslant l)$. Après application de f^{l-j}, ces éléments obtiennent une composante $f^l(a) = e_l$. De plus G est un sous-espace vectoriel du fait que $e_l^* \circ f^i$ est une application linéaire.

Montrons enfin que $\dim(G) = n - l$. Pour cela nous remarquons que G est une intersection d'hyperplans, et nous montrons que les équations définissant ces hyperplans sont linéairement indépendantes. Soit donc

$$\sum_{j=0}^{l-1} \lambda_j \big(e_l^* \circ f^j\big) = 0 \tag{9.191}$$

et montrons que $\lambda_j = 0$ pour tout j est l'unique solution. Soit $x \in E$ et appliquons l'opération (9.191) au vecteur $f^i(x)$; le résultat est zéro :

$$0 = \sum_{j=0}^{l-1} \lambda_j (e_l^* \circ f^i \circ f^j) = (e_l^* \circ f^i)P(u) \tag{9.192}$$

où nous avons posé $P(X) = \sum_{j=0}^{l-1} \lambda_j X^j$. Appliquons cela à a : pour tout i nous avons

$$(e_l^* \circ f^i)\big(P(f)a\big) = 0. \tag{9.193}$$

Mais par définition de E_a, l'élément $P(f)a$ est dans E_a. Nous en déduisons que

$$P(f)a \in G \cap E_a = \{0\}, \tag{9.194}$$

c'est-à-dire que P est un polynôme annulateur de a. Mais P est de degré $l-1$ alors que le polynôme minimal de a est de degré l. Par conséquent $P = 0$ et $\lambda_j = 0$ pour tout j. □

Définition 9.102.
*Un endomorphisme d'un espace vectoriel est **semi-simple** si tout sous-espace stable par u possède un supplémentaire stable.*

24. ici, comme presque partout, e_l^* est le dual de e_l, c'est-à-dire l'application linéaire sur E donnée par $e_l^*(e_i) = \delta_{li}$.

Lemme 9.103.
Si le polynôme minimal d'un endomorphisme est irréductible, alors il est semi-simple[25].

Démonstration. Soit f, un endomorphisme dont le polynôme minimal est irréductible et F, un sous-espace stable par f. Nous devons en trouver un supplémentaire stable. Si $F = E$, il n'y a pas de problèmes. Sinon nous considérons $u_1 \in E\backslash F$ et

$$E_{u_1} = \{P(f)u_1 \text{ tel que } P \in \mathbb{K}[X]\}, \tag{9.195}$$

qui est un espace stable par f.

Montrons que $E_{u_1} \cap F = \{0\}$. Pour cela nous regardons l'idéal

$$I_{u_1} = \{P \in \mathbb{K}[X] \text{ tel que } P(f)u_1 = 0\}. \tag{9.196}$$

Cela est un idéal non réduit à $\{0\}$ parce que le polynôme minimal de f par exemple est dans I_{u_1}. Soit P_{u_1} un générateur unitaire de I_{u_1}. Étant donné que $\mu_f \in I_{u_1}$, nous avons que P_{u_1} divise μ_f et donc $P_{u_1} = \mu_f$ parce que μ_f est irréductible par hypothèse.

Soit $y \in E_{u_1} \cap F$. Par définition il existe $P \in \mathbb{K}[X]$ tel que $y = P(f)u_1$ et si $y \neq 0$, ce la signifie que $P \notin I_{u_1}$, c'est-à-dire que P_{u_1} ne divise pas P. Étant donné que P_{u_1} est irréductible cela implique que P_{u_1} et P sont premiers entre eux (ils n'ont pas d'autre pgcd que 1).

Nous utilisons maintenant Bézout (théorème 6.41) qui nous donne $A, B \in \mathbb{K}[X]$ tels que

$$AP + BP_{u_1} = 1. \tag{9.197}$$

Nous appliquons cette égalité à f et puis à u_1 :

$$u_1 = A(f) \circ \underbrace{P(f)u_1}_{=y} + B(f) \circ \underbrace{P_{u_1}(u_1)}_{=0} = A(f)y. \tag{9.198}$$

Mais $y \in F$, donc $A(f)y \in F$. Nous aurions donc $u_1 \in F$, ce qui est impossible par choix. Nous avons maintenant que l'espace $E_{u_1} \oplus F$ est stable sous f. Si cet espace est E alors nous arrêtons. Sinon nous reprenons le raisonnement avec $E_{u_1} \oplus F$ en guise de F et en prenant $u_2 \in E\backslash(E_{u_1} \oplus F)$. Étant donné que E est de dimension finie, ce procédé s'arrête à un certain moment et nous aurons

$$E = F \oplus E_{u_1} \oplus \ldots \oplus E_{u_k} \tag{9.199}$$

où chacun des E_{u_i} sont stables. □

Théorème 9.104.
Un endomorphisme est semi-simple si et seulement si son polynôme minimal est produit de polynômes irréductibles distincts deux à deux.

Démonstration. Supposons que f soit semi-simple et que son polynôme minimal soit donné par $\mu_f = M_1^{\alpha_1} \ldots M_r^{\alpha_r}$ où les M_i sont des polynômes irréductibles deux à deux distincts. Nous devons montrer que $\alpha_i = 1$ pour tout i. Soit i tel que $\alpha_i \geqslant 1$ et $N \in \mathbb{K}[X]$ tel que $\mu_f = M^2 N$ où l'on a noté $M = M_i$. Nous étudions l'espace

$$F = \ker M(f) \tag{9.200}$$

qui est stable par f, et qui possède donc un supplémentaire S également stable par f. Nous allons montrer que MN est un polynôme annulateur de f.

D'abord nous prenons $x \in S$. Étant donné que F est le noyau de $M(f)$,

$$M(f)\big(MN(f)x\big) = \mu_f(f)x = 0, \tag{9.201}$$

25. Définition 9.102.

ce qui signifie que $MN(f)x \in F$. Mais vu que S est stable par f nous avons aussi que $MN(f)x \in S$. Finalement $MN(f)x \in F \cap S = \{0\}$. Autrement dit, $MN(f)$ s'annule sur S.

Prenons maintenant $y \in F$. Nous avons

$$MN(f) = N(f)\big(M(f)y\big) = 0 \tag{9.202}$$

parce que $y \in F = \ker M(f)$.

Nous avons prouvé que $MN(f)$ s'annule partout et donc que $MN(f)$ est un polynôme annulateur de f, ce qui contredit la minimalité de $\mu_f = M^2N$.

Nous passons au sens inverse. Soit $m_f = M_1 \dots M_r$ une décomposition du polynôme minimal de l'endomorphisme f en irréductibles distincts deux à deux. Soit F un sous-espace vectoriel stable par f. Nous notons

$$E_i = \ker(M_i(f)) \tag{9.203}$$

et $f_i = f|_{E_i}$. Par le lemme 9.89 nous avons

$$F = \bigoplus_{i=1}^{r}(F \cap E_i). \tag{9.204}$$

Les espaces E_i sont stables par f et étant donné que M_i est irréductible, il est le polynôme minimal de f_i. En effet, M_i est annulateur de f_i, ce qui montre que le minimal de f_i divise M_i. Mais M_i étant irréductible, M_i est le polynôme minimal. Étant donné que $\mu_{f_i} = M_i$, l'endomorphisme f_i est semi-simple par le lemme 9.103.

L'espace $F \cap E_i$ étant stable par l'endomorphisme semi-simple f_i, il possède un supplémentaire stable que nous notons S_i :

$$E_i = S_i \oplus (F \cap E_i). \tag{9.205}$$

Étant donné que sur chaque S_i nous avons $f|_{S_i} = f_i$, l'espace $S = S_1 \oplus \dots \oplus S_r$ est stable par f. Du coup nous avons

$$\begin{aligned}
E &= E_1 \oplus \dots \oplus E_r && (9.206a)\\
&= \big(S_1 \oplus (F \cap E_1)\big) \oplus \dots \oplus \big(S_r \oplus (F \cap E_r)\big) && (9.206b)\\
&= \big(\bigoplus_{i=1}^{r} S_i\big) \oplus \big(\bigoplus_{i=1}^{r} F \cap E_i\big) && (9.206c)\\
&= S \oplus F, && (9.206d)
\end{aligned}$$

ce qui montre que F a bien un supplémentaire stable par f et donc que f est semi-simple. □

Exemple 9.105 (L'espace engendré par $\mathbb{1}$, A, A^2,...).
Soit A une matrice, et

$$V = \mathrm{Span}\{A^k \text{ tel que } k \in \mathbb{N}\}. \tag{9.207}$$

Nous montrons que $\dim(V)$ est le degré du polynôme minimal de A.

D'abord l'idéal annulateur de A est engendré par le polynôme minimal[26] que nous notons $\mu = \sum_{k=0}^{p} a_k X^k$. La partie $\{\mathbb{1}, \dots, A^{p-1}\}$ est libre parce qu'une combinaison linéaire nulle de cela serait un polynôme annulateur en A de degré plus petit que p. Donc $\dim(V) \geqslant p$.

La partie $\{\mathbb{1}, A, \dots, A^p\}$ est liée à cause du polynôme minimal. Isoler A^p dans $\mu(A) = 0$ donne un polynôme f de degré $p-1$ tel que $A^p = f(A)$.

Nous allons montrer à présent que la famille $\{\mathbb{1}, A, \dots, A^{p-1}\}$ est génératrice (alors $\dim(V) \leqslant p$). Soit un entier $q \geqslant p$et de division euclidienne[27] $np + r = q$ avec $r < p$. Nous avons $A^q = A^{np}A^r$. D'une part

$$A^{np} = (A^p)^n = f(A)^n \tag{9.208}$$

est de degré $n(p-1)$. Par conséquent

$$A^q = f(A)^n A^r \tag{9.209}$$

26. Proposition 9.96.
27. Théorème 3.6.

qui est de degré $n(p-1)+r=q-n$. Autrement dit il existe un polynôme g_1 de degré $q-n$ tel que $A^q=g_1(A)$. Si $q-n>p-1$ alors nous pouvons recommencer et obtenir un polynôme g_2 de degré strictement inférieur à celui de g_1 tel que $A^q=g_2(A)$. Au bout du compte, il existe un polynôme g de degré au maximum $p-1$ tel que $A^q=g(A)$. Cela prouve que la partie $\{\mathbb{1},A,\dots,A^{p-1}\}$ est génératrice de V.

La dimension de V est donc p, le degré du polynôme minimal. △

Proposition 9.106.
Soit f un endomorphisme d'un espace vectoriel de dimension finie. Nous avons l'isomorphisme d'espace vectoriel

$$\mathbb{K}[f]\simeq\frac{\mathbb{K}[X]}{(\mu_f)} \tag{9.210}$$

La dimension en est $\deg(\mu_f)$.

Démonstration. Notons avant de commencer que (μ) est l'idéal engendré par μ. Les classes dont il est question dans le quotient $\mathbb{K}[X]/(\mu)$ sont

$$\bar{P}=\{P+S\mu\}_{S\in\mathbb{K}[X]}. \tag{9.211}$$

Nous allons montrer que l'application suivante fournit l'isomorphisme :

$$\begin{aligned}\psi\colon\frac{\mathbb{K}[X]}{(\mu)}&\to\mathbb{K}[f]\\ \bar{P}&\mapsto P(f).\end{aligned} \tag{9.212}$$

<u>ψ **est bien définie**</u> Si $Q\in\bar{P}$ alors $Q=P+S\mu$ pour un certain $S\in\mathbb{K}[X]$. Du coup nous avons

$$\psi(\bar{Q})=P(f)+(S\mu)(f). \tag{9.213}$$

Mais $\mu(f)=0$ donc le deuxième terme est nul. Donc $\psi(\bar{P})$ est bien définit.

<u>**Injectif**</u> Si $\psi(\bar{P})=0$ nous avons $P(f)=0$, ce qui signifie que $P=S\mu$ pour un polynôme S. Par conséquent $P\in(\mu)$ et donc $\bar{P}=0$.

<u>**Surjectif**</u> Soit $P\in\mathbb{K}[X]$. L'élément $P(f)$ de $\mathbb{K}[f]$ est dans l'image de ψ parce que c'est $\psi(\bar{P})$.

En ce qui concerne la dimension, le corolaire 6.38 en parle déjà : une base est donné par les projections de $1,X,\dots,X^{\deg(\mu_a)-1}$. □

9.6.3 Polynôme caractéristique

Définition 9.107.
Soit un anneau commutatif A. Si $u\in\mathbb{M}(n,A)$, nous définissons le ***polynôme caractéristique de*** *u :*

$$\chi_u(X)=\det(u-X\mathbb{1}_n). \tag{9.214}$$

Nous définissons de même le polynôme caractéristique d'un endomorphisme $u\colon E\to E$.

Remarque 9.108.
Quelques remarques à propos du signe[28].

- Certains auteurs définissent le polynôme caractéristique par $\det(X-u)$ au lieu de $\det(u-X)$.
- Wikipédia francophone prend la définition $\det(X-u)$ (donc inverse de la notre). Allez lire la page de discussion.
- Sur les wikipédias d'autre langues, ça varie.
- Un avantage de $\det(u-X)$ est que $\det(u)=\chi_u(0)$.
- Un avantage de $\det(X-u)$ est qu'il est unitaire.

28. Attention : je crois qu'il y a des incohérences dans le Frido à propos de ce choix

Lemme 9.109.
Le polynôme caractéristique χ_u est unitaire en dimension paire et a pour degré la dimension de l'espace vectoriel E..

Théorème 9.110.
Soit E un $\mathbb{K}$-espace vectoriel de dimension finie n et un endomorphisme $u \in \mathrm{End}(E)$. Alors

(1) Le polynôme caractéristique divise $(\mu_u)^n$ dans $\mathbb{K}[X]$.
(2) Les polynômes caractéristiques et minimaux ont mêmes facteurs irréductibles dans $\mathbb{K}[X]$.
(3) Les polynômes caractéristiques et minimaux ont mêmes racines dans $\mathbb{K}[X]$.
(4) Le polynôme caractéristique est scindé si et seulement si le polynôme minimal est scindé.

Théorème 9.111.
Soit $u \in \mathrm{End}(E)$ et $\lambda \in \mathbb{K}$. Les conditions suivantes sont équivalentes

(1) $\lambda \in \mathrm{Spec}(u)$
(2) $\chi_u(\lambda) = 0$
(3) $\mu_u(\lambda) = 0$.

Démonstration. (1) $\Leftrightarrow$ (2). Dire que λ est dans le spectre de u signifie que l'opérateur $u - \lambda\mathbb{1}$ n'est pas inversible, ce qui est équivalent à dire que $\det(u - \lambda\mathbb{1})$ est nul par la proposition 9.10(1) ou encore que λ est une racine du polynôme caractéristique de u.

(2) $\Leftrightarrow$ (3). Cela est une application directe du théorème 9.110 qui précise que le polynôme caractéristique a les mêmes racines dans $\mathbb{K}$ que le polynôme minimal. □

Exemple 9.112.
Sur $\mathbb{R}^2$, nous considérons la matrice $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ qui a pour polynôme caractéristique[29] le polynôme $\chi_A = (X-1)^2$. Le nombre $\lambda = 1$ est une racine double de ce polynôme, et pourtant il n'y a qu'une seule dimension d'espace propre :

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \tag{9.215}$$

entraine $x = 0$.

Ici la multiplicité algébrique est différente de la multiplicité géométrique. △

La proposition suivante donne une utilisation amusante de la notion de polynôme caractéristique[30].

Proposition 9.113 ([178])**.**
Soit un espace vectoriel V de dimension finie pour lequel il existe un endomorphisme $f\colon V \to V$ tel que $(f \circ f)(v) = -v$ pour tout $v \in V$. Alors la dimension de V est paire.

Démonstration. Cherchons les valeurs propres de f en résolvant l'équation $f(v) = \lambda v$. Nous appliquons f à cette égalité :

$$-v = \lambda f(v) = \lambda^2 v. \tag{9.216}$$

Donc λ ne peut pas être réel. Nous avons montré que f n'a pas de valeurs propres réelles. Or le polynôme caractéristique de f est de degré égal à la dimension. Si la dimension est impaire, le polynôme caractéristique est de degré impair, et possède donc une racine réelle. Autrement dit, l'absence de racines réelles au polynôme caractéristique indique une dimension paire. □

Une autre preuve possible est d'utiliser le déterminant : si la dimension de V est n nous avons :

$$\det(f^2) = \det(-\mathrm{Id}) = (-1)^n. \tag{9.217}$$

Donc $(-1)^n$ est positif, ce qui montre que n est pair.

29. Définition 9.107.
30. Définition 9.107.

Proposition 9.114 ([176]).
Soit f, un endomorphisme de E et $x \in E$. Alors

(1) L'espace $E_{f,x}$ est stable par f.

(2) L'espace $E_{f,x}$ est de dimension

$$p_{f,x} = \dim E_{f,x} = \deg(\mu_{f,x}) \tag{9.218}$$

où $\mu_{f,x}$ est le générateur unitaire de $I_{f,x}$.

(3) Le polynôme caractéristique de $f|_{E_{f,x}}$ est $\mu_{f,x}$.

(4) Nous avons

$$\chi_{f|_{E_{f,x}}}(f)x = \mu_{f,x}(f)x = 0. \tag{9.219}$$

Démonstration. Le fait que $E_{f,x}$ soit stable par f est classique. Le point (4) est un une application du point (3). Les deux gros morceaux sont donc les points (2) et (3).

Étant donné que $\mu_{f,x}$ est de degré minimal dans $I_{f,x}$, l'ensemble

$$B = \{f^k(x) \text{ tel que } 0 \leqslant k \leqslant p_{f,x} - 1\} \tag{9.220}$$

est libre. En effet une combinaison nulle des vecteurs de B donnerait un polynôme en f de degré inférieur à $p_{f,x}$ annulant x. Nous écrivons

$$\mu_{f,x}(X) = X^{p_{f,x}} - \sum_{i=0}^{p_{f,x}-1} a_i X^k. \tag{9.221}$$

Étant donné que $\mu_{f,x}(f)x = 0$ et que la somme du membre de droite est dans $\mathrm{Span}(B)$, nous avons $f^{p_{f,x}}(x) \in \mathrm{Span}(B)$. Nous prouvons par récurrence que $f^{p_{f,x}+k}(x) \in \mathrm{Span}(B)$. En effet en appliquant f^k à l'égalité

$$0 = f^{p_{f,x}}(x) - \sum_{i=0}^{p_{f,x}-1} a_i f^i(x) \tag{9.222}$$

nous trouvons

$$f^{p_{f,x}+k}(x) = \sum_{i=0}^{p_{f,x}-1} a_i f^{i+k}(x), \tag{9.223}$$

alors que par hypothèse de récurrence le membre de droite est dans $\mathrm{Span}(B)$. L'ensemble B est alors générateur de $E_{f,x}$ et donc une base d'icelui. Nous avons donc bien $\dim(E_{f,x}) = p_{f,x}$.

Nous montrons maintenant que $\mu_{f,x}$ est annulateur de f au point x. Nous savons que

$$\mu_{f,x}(f)x = 0. \tag{9.224}$$

En y appliquant f^k et en profitant de la commutativité des polynômes sur les endomorphismes (proposition 9.87), nous avons

$$0 = f^k\big(\mu_{f,x}(f)x\big) = \mu_{f,x}(f)f^k(x), \tag{9.225}$$

de telle sorte que $\mu_{f,x}(f)$ est nul sur B et donc est nul sur $E_{f,x}$. Autrement dit,

$$\mu_{f,x}\big(f|_{E_{f,x}}\big) = 0. \tag{9.226}$$

Montrons que $\mu_{f,x}$ est même minimal pour $f|_{E_{f,x}}$. Sot Q, un polynôme non nul de degré $p_{f,x} - 1$ annulant $f|_{E_{f,x}}$. En particulier $Q(f)x = 0$, alors qu'une telle relation signifierait que B est un système lié, alors que nous avons montré que c'était un système libre. Nous concluons que $\mu_{f,x}$ est le polynôme minimal de $f|_{E_{f,x}}$. □

Cette histoire de densité permet de donner une démonstration alternative du théorème de Cayley-Hamilton.

Théorème 9.115 (Cayley-Hamlilton)**.**
Le polynôme caractéristique est un polynôme annulateur.

Une démonstration plus simple via la densité des diagonalisables est donnée en théorème 12.23.

Démonstration. Nous devons prouver que $\chi_f(f)x = 0$ pour tout $x \in E$. Pour cela nous nous fixons un $x \in E$, nous considérons l'espace $E_{f,x}$ et $\chi_{f,x}$, le polynôme caractéristique de $f|_{E_{f,x}}$. Étant donné que $E_{f,x}$ est stable par f, le polynôme caractéristique de $f|_{E_{j,x}}$ divise χ_f, c'est-à-dire qu'il existe un polynôme Q_x tel que

$$\chi_f = Q_x \chi_{f,x}, \tag{9.227}$$

et donc aussi

$$\chi_f(f)x = Q_x(f)\big(\chi_{f,x}(f)x\big) = 0 \tag{9.228}$$

parce que la proposition 9.114 nous indique que $\chi_{f,x}$ est un polynôme annulateur de $f|_{E_{f,x}}$. □

Corolaire 9.116.
Le degré du polynôme minimal est majoré par la dimension de l'espace.

Démonstration. Le polynôme minimal divise le polynôme caractéristique parce qu'il engendre l'idéal des polynômes annulateurs par la proposition 9.96. Or le degré du polynôme caractéristique est la dimension de l'espace par le lemme 9.109. □

Exemple 9.117 (Calcul de l'inverse d'un endomorphisme)**.**
Le polynôme de Cayley-Hamilton donne un moyen de calculer l'inverse d'un endomorphisme inversible pourvu que l'on sache son polynôme caractéristique. En effet, supposons que

$$\chi_f(X) = \sum_{k=0}^{n} a_k X^k. \tag{9.229}$$

Nous aurons alors

$$0 = \chi_f(f) = \sum_{k=0}^{n} a_k f^k. \tag{9.230}$$

Nous appliquons f^{-1} à cette dernière égalité en sachant que $f^{-1}(0) = 0$:

$$0 = a_0 f^{-1} + \sum_{k=1}^{n} a_k f^{k-1}, \tag{9.231}$$

et donc

$$u^{-1} = -\frac{1}{\det(f)} \sum_{k=1}^{n} a_k f^{k-1} \tag{9.232}$$

où nous avons utilisé le fait que $a_0 = \chi_f(0) = \det(f)$. △

Proposition 9.118.
Si $(X-z)^l$ $(l \geqslant 1)$ est la plus grande puissance de $(X-z)$ dans le polynôme caractéristique d'un endomorphisme u alors

$$1 \leqslant \dim(E_e) \leqslant l. \tag{9.233}$$

C'est-à-dire que nous avons au moins un vecteur propre pour chaque racine du polynôme caractéristique.

Démonstration. Si $(X-z)$ divise χ_u alors en posant $\chi_u = (X-z)P(X)$ nous avons

$$\det(u - X\mathbb{1}) = (X-z)P(X), \tag{9.234}$$

ce qui, évalué en $X = z$, donne $\det(u - z\mathbb{1}) = 0$. L'annulation du déterminant étant équivalente à l'existence d'un noyau non trivial, nous avons $v \neq 0$ dans E tel que $(u - z\mathbb{1})v = 0$. Cela donne $u(v) = zv$ et donc que v est vecteur propre de u pour la valeur propre z. Donc aussi $\dim(E_z) \geqslant 1$.

Si $\dim(E_z) = k$ alors le théorème de la base incomplète 4.11 nous permet d'écrire une base de E dont les k premiers vecteurs forment une base de E_z. Dans cette base, la matrice de u est de la forme

$$\begin{pmatrix} z & & & * \\ & \ddots & & \vdots \\ & & z & * \\ & & & * \end{pmatrix} \tag{9.235}$$

où les étoiles représentent des blocs à priori non nuls. En tout cas il est vu sous cette forme que $(X - z\mathbb{1})^k$ divise χ_u. □

9.7 Diagonalisation et trigonalisation

Ici encore $\mathbb{K}$ est un corps commutatif.

9.7.1 Matrices semblables

Définition 9.119 (matrices semblables)**.**
*Sur l'ensemble $\mathrm{M}_n(\mathbb{K})$ des matrices $n \times n$ à coefficients dans $\mathbb{K}$ nous introduisons la relation d'équivalence $A \sim B$ si et seulement s'il existe une matrice $P \in \mathrm{GL}(n, \mathbb{K})$ telle que $B = P^{-1}AP$. Deux matrices équivalentes en ce sens sont dites **semblables**.*

Le polynôme caractéristique [31] est un invariant sous les similitudes. En effet si P est une matrice inversible,

$$\chi_{PAP^{-1}} = \det(PAP^{-1} - \lambda X) \tag{9.236a}$$
$$= \det\left(P^{-1}(PAP^{-1} - \lambda X)P^{-1}\right) \tag{9.236b}$$
$$= \det(A - \lambda X). \tag{9.236c}$$

La permutation de lignes ou de colonnes ne sont pas de similitudes, comme le montrent les exemples suivants :

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}. \tag{9.237}$$

Nous avons $\chi_A = x^2 - 5x - 2$ tandis que $\chi_B = x^2 - 5x + 2$ alors que le polynôme caractéristique est un invariant de similitude.

9.7.2 Endomorphismes nilpotents

La **trace** d'une matrice $A \in \mathrm{M}(n, \mathbb{K})$ est la somme de ses éléments diagonaux :

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} A_{ii}. \tag{9.238}$$

Une propriété importante est son invariance cyclique.

Lemme 9.120.
Quelques propriétés de la trace.

(1) Si A et B sont des matrices carrées, alors $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

(2) La trace est un invariant de similitude.

Démonstration. C'est un simple calcul :

$$\mathrm{Tr}(AB) = \sum_{ik} A_{ik}B_{ki} = \sum_{ik} A_{ki}B_{ik} = \sum_{ik} B_{ik}A_{ki} = \sum_{i} (BA)_{ii} = \mathrm{Tr}(BA) \tag{9.239}$$

31. Définition 9.107.

où nous avons simplement renommé les indices $i \leftrightarrow k$.

En particulier, la trace est un invariant de similitude parce que $\mathrm{Tr}(ABA^{-1}) = \mathrm{Tr}(A^{-1}AB) = \mathrm{Tr}(B)$ par l'invariance cyclique démontrée en 4.62(2). □

La trace étant un invariant de similitude, nous pouvons donc définir la **trace** comme étant la trace de sa matrice dans une base quelconque. Si la matrice est diagonalisable, alors la trace est la somme des valeurs propres.

Lemme 9.121 ([71]).
L'endomorphisme $u \in \mathrm{End}(\mathbb{C}^n)$ est nilpotent si et seulement si $\mathrm{Tr}(u^p) = 0$ pour tout p.

Démonstration. Supposons que u est nilpotent. Alors ses valeurs propres sont toutes nulles et celles de u^p le sont également. La trace étant la somme des valeurs propres, nous avons alors tout de suite $\mathrm{Tr}(u^p) = 0$.

Supposons maintenant que $\mathrm{Tr}(u^p) = 0$ pour tout p. Le polynôme caractéristique (9.214) est

$$\chi_u = (-1)^n X^\alpha (X - \lambda_1)^{\alpha_1} \dots (X - \alpha_r)^{\alpha_r}. \tag{9.240}$$

où les λ_i $(i = 1, \dots, r)$ sont les valeurs propres non nulles distinctes de u.

Il est vite vu que le coefficient de X^{n-1} dans χ_u est $-\mathrm{Tr}(u)$ parce que le coefficient de X^{n-1} se calcule en prenant tous les X sauf une fois $-\lambda_i$. D'autre part le polynôme caractéristique de u^p est le même que celui de u, en remplaçant λ_i par λ_i^p ; cela est dû au fait que si v est vecteur propre de valeur propre λ, alors $u^p v = \lambda^p v$.

Par l'équation (9.240), nous voyons que le coefficient du terme X^{n-1} dans les polynôme caractéristique est

$$0 = \mathrm{Tr}(u^p) = \alpha_1 \lambda_1^p + \dots + \alpha_r \lambda_r^p. \tag{9.241}$$

Donc les nombres $(\alpha_1, \dots, \alpha_r)$ est une solution non triviale[32] du système

$$\begin{cases} \alpha_1 X_1 + \dots + \lambda_r X_r = 0 & \text{(9.242a)} \\ \quad\vdots & \text{(9.242b)} \\ \lambda_1^r X_1 + \dots + \lambda_r^r X_r = 0. & \text{(9.242c)} \end{cases}$$

Cela sont les équations (9.241) écrites avec $p = 1, \dots, r$. Le déterminant de ce système est

$$\lambda_1 \dots \lambda_r \det \begin{pmatrix} 1 & \dots & 1 \\ \lambda_1 & \dots & \lambda_1 \\ \vdots & & \vdots \\ \lambda_1^{r-1} & \dots & \lambda_r^{r-1} \end{pmatrix} \neq 0, \tag{9.243}$$

qui est un déterminant de Vandermonde (proposition 9.12) valant

$$0 = \lambda_1 \dots \lambda_r \prod_{1 \leqslant i \leqslant j \leqslant r} (\lambda_i - \lambda_j). \tag{9.244}$$

Étant donné que les λ_i sont distincts et non nuls, nous avons une contradiction et nous devons conclure que $(\alpha_1, \dots, \alpha_r)$ était une solution triviale du système (9.242). □

Proposition 9.122 ([179]).
Soit un $\mathbb{K}$-espace vectoriel E. Un endomorphisme $u \in \mathrm{End}(E)$ est nilpotent si et seulement s'il existe une base de E dans laquelle la matrice de u est strictement triangulaire supérieure.

Démonstration. $\Rightarrow$ Nous faisons la démonstration par récurrence sur la dimension de E. Lorsque $n = 1$ nous avons $u = (a)$ avec $a \in \mathbb{K}$. Vu que $a^k = 0$ pour un certain k nous avons $a = 0$ parce qu'un corps est toujours un anneau intègre[33].

32. Si $\alpha_1 = \dots = \alpha_r = 0$, alors les valeurs propres sont toutes nulles et la matrice est en réalité nulle dès le départ.
33. Lemme 1.100.

Lorsque $\dim(E) = n$ nous savons que u a un noyau non réduit au vecteur nul (parce qu'il est nilpotent). Soit donc un vecteur non nul $x \in \ker(u)$ et une base

$$\{x, e_2, \dots, e_n\} \tag{9.245}$$

donnée par le théorème de la base incomplète 4.11. La matrice de u dans cette base s'écrit

$$\left(\begin{array}{c|ccc} 0 & * & * & * \\ \hline 0 & & & \\ 0 & & A & \\ 0 & & & \end{array}\right). \tag{9.246}$$

Un tout petit peu de calcul de produit de matrice montre que la matrice de u^k est de la forme

$$\left(\begin{array}{c|ccc} 0 & * & * & * \\ \hline 0 & & & \\ 0 & & A^k & \\ 0 & & & \end{array}\right). \tag{9.247}$$

Étant donné que u est nilpotente, la matrice A l'est aussi. L'hypothèse de récurrence dit alors que A est strictement triangulaire supérieure (ou en tout cas peut le devenir par un changement de base adéquat).

$\Leftarrow$ Lorsqu'une matrice est triangulaire supérieure stricte, elle applique

$$\mathrm{Span}\{e_1, \dots, e_k\} \to \mathrm{Span}\{e_1, \dots, e_{k-1}\}. \tag{9.248}$$

Donc tout vecteur finit sur zéro si on lui applique u assez souvent.

□

Proposition 9.123 (Thème 44)**.**
Soit E un espace de Banach (espace vectoriel normé complet). Si $A \in \mathcal{L}(E, E)$ est nilpotente, alors $(\mathbb{1} - A)$ est inversible et son inverse est donné par

$$(\mathbb{1} - A)^{-1} = \sum_{k=0}^{\infty} A^k, \tag{9.249}$$

où l'infini peut évidemment être remplacé par l'ordre de nilpotence de A.

Démonstration. En ce qui concerne la convergence de la somme, elle ne fait pas de doutes parce que A étant nilpotente, la somme contient seulement une quantité finie de termes non nuls.

Montrons à présent que la somme est l'inverse de $\mathbb{1} - A$ en multipliant terme à terme :

$$\sum_{k=0}^{n} A^k(\mathbb{1} - A) = \sum_{k=0}^{n} (A^k - A^{k+1}) = \mathbb{1} - A^{n+1}. \tag{9.250}$$

Par conséquent

$$\|\mathbb{1} - \sum_{k=0}^{n} A^k(\mathbb{1} - A)\| = \|A^{n+1}\| \to 0. \tag{9.251}$$

La dernière limite est en réalité une égalité pour n assez grand. □

Proposition 9.124.
Soit $A \in \mathrm{GL}(n, \mathbb{C})$. La suite $(A^k)_{k\in\mathbb{Z}}$ est bornée si et seulement si A est diagonalisable et $\mathrm{Spec}(A) \subset \mathrm{S}^1$.

Démonstration. Si A est diagonalisable avec les valeurs propres λ_i de norme 1 dans $\mathbb{C}$, alors A^k est la matrice diagonale avec les λ_i^k sur la diagonale. Cela reste borné pour toute valeur entière de k.

En ce qui concerne l'autre sens, nous supposons encore que

$$A = \begin{pmatrix} \lambda_1 \mathbb{1} + N_1 & & \\ & \ddots & \\ & & \lambda_s \mathbb{1} + N_s \end{pmatrix}, \tag{9.252}$$

et nous regardons un des blocs. Nous voulons prouver que $N = 0$ et que $|\lambda| = 1$.

Nous commençons par regarder ce qu'implique le fait que $(\lambda\mathbb{1} + N)^n$ reste borné pour $n > 0$. En notant r l'ordre de nilpotence de N, nous avons le développement

$$(\lambda\mathbb{1} + N)^n = \sum_{k=0}^{r-1} \binom{n}{k} N^k \lambda^{n-k}. \tag{9.253}$$

Par la proposition 9.122, une matrice nilpotente s'écrit dans une base sous la forme

$$N = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix} \tag{9.254}$$

et effectuer A^k revient à décaler la diagonale de 1. Donc la famille

$$\{\mathbb{1}, N, \dots, N^{r-1}\} \tag{9.255}$$

est libre. Par conséquent la suite $(\lambda\mathbb{1} + N)^n$ restera bornée si et seulement si chacun des termes

$$\binom{n}{k} N^k \lambda^{n-k} \tag{9.256}$$

reste borné. Le premier terme étant $\lambda^n \mathbb{1}$, nous avons obligatoirement $|\lambda| \leqslant 1$. Si $|\lambda| < 1$, alors le coefficient $\binom{n}{k}\lambda^{n-k}$ tend vers zéro. Si $|\lambda| = 1$ par contre ce coefficient tend vers l'infini et la seule façon pour que (9.256) reste borné est que $N = 0$. Nous avons donc deux possibilités :

— $|\lambda| < 1$

— $|\lambda| = 1$ et $N = 0$.

Nous nous tournons maintenant sur la contrainte que $(\lambda\mathbb{1} + N)^n$ doive rester borné pour $n < 0$. Nous avons

$$\lambda\mathbb{1} + N = \lambda(\mathbb{1} + \lambda^{-1}N), \tag{9.257}$$

et nous pouvons appliquer la proposition 9.123 à l'opérateur nilpotent $-\lambda^{-1}N$ pour avoir

$$(\mathbb{1} + \lambda^{-1}N)^{-1} = \mathbb{1} + \sum_{k=1}^{\infty} (-\lambda)^{-1} N^k. \tag{9.258}$$

Ceci pour dire que $(\lambda\mathbb{1} + N)^{-1} = \lambda^{-1}(\mathbb{1} + \lambda^{-1}N')$ pour une autre matrice nilpotente N'. Le travail déjà fait, appliqué à λ^{-1} et N', nous donne deux possibilités :

— $|\lambda^{-1}| < 1$

— $|\lambda^{-1}| = 1$ et $N' = 0$.

La possibilité $|\lambda^{-1}| < 1$ est exclue parce qu'elle impliquerait $|\lambda| > 1$ qui avait déjà été exclu. Il ne reste donc que la possibilité $|\lambda| = 1$ et $N = N' = 0$. □

9.7.3 Endomorphismes diagonalisables

Définition 9.125.
*Une matrice est **diagonalisable** si elle est semblable*[34] *à une matrice diagonale.*

Lemme 9.126.
Une matrice triangulaire supérieure avec des 1 *sur la diagonale n'est diagonalisable que si elle est diagonale (c'est-à-dire si elle est la matrice unité).*

Démonstration. Si A est une matrice triangulaire supérieure de taille n telle que $A_{ii} = 1$, alors $\det(A - \lambda\mathbb{1}) = (1-\lambda)^n$, ce qui signifie que $\mathrm{Spec}(A) = \{1\}$. Pour la diagonaliser, il faudrait une matrice $P \in \mathrm{GL}(n, \mathbb{K})$ telle que $\mathbb{1} = P^{-1}AP$, ce qui est uniquement possible si $A = \mathbb{1}$. □

Lemme 9.127.
Soit F un sous-espace stable par u. Soit une décomposition du polynôme minimal

$$\mu_u = P_1^{n_1} \dots P_r^{n_r} \tag{9.259}$$

où les P_i sont des polynômes irréductibles unitaires distincts. Si nous posons $E_i = \ker P_i^{n_i}$, alors

$$F = (F \cap E_1) \oplus \dots \oplus (F \cap E_r). \tag{9.260}$$

Théorème 9.128.
Soit E, un espace vectoriel de dimension n sur le corps commutatif $\mathbb{K}$ et $u \in \mathrm{End}(E)$. Les propriétés suivantes sont équivalentes.

(1) L'endomorphisme u est diagonalisable.

(2) Il existe un polynôme $P \in \mathbb{K}[X]$ non constant, scindé sur $\mathbb{K}$ dont toutes les racines sont simples tel que $P(u) = 0$.

(3) Le polynôme minimal μ_u est scindé sur $\mathbb{K}$ et toutes ses racines sont simples[35]*.*

(4) Tout sous-espace de E possède un supplémentaire stable par u.

(5) Dans une base adaptée, la matrice de u est diagonale et les éléments diagonaux sont ses valeurs propres.

Démonstration. Plein d'implications à prouver.

(2) implique (3) Étant donné que $P(u) = 0$, il est dans l'idéal des polynômes annulateurs de u, et le polynôme minimal μ_u le divise parce que l'idéal des polynômes annulateurs est généré par μ_u par le théorème 6.37.

(3) implique (1) Étant donné que le polynôme minimal est scindé à racines simples, il s'écrit sous forme de produits de monômes tous distincts, c'est-à-dire

$$\mu_u(X) = (X - \lambda_1) \dots (X - \lambda_r) \tag{9.262}$$

où les λ_i sont des éléments distincts de $\mathbb{K}$. Étant donné que $\mu_u(u) = 0$, le théorème de décomposition des noyaux (théorème 9.88) nous enseigne que

$$E = \ker(u - \lambda_1) \oplus \dots \oplus \ker(u - \lambda_r). \tag{9.263}$$

Mais $\ker(u - \lambda_i)$ est l'espace propre $E_{\lambda_i}(u)$. Donc u est diagonalisable.

34. Définition 9.119.
35. Le polynôme *caractéristique*, lui, n'a pas spécialement ses racines simples ; il peut encore être de la forme

$$\chi_u(X) = \prod_{i=1}^{r} (X - \lambda_i)^{\alpha_i}, \tag{9.261}$$

mais alors $\dim(E_{\lambda_i}) = \alpha_i$.

(1) implique (4) Soit $\{e_1, \dots, e_n\}$ une base qui diagonalise u, soit F un sous-espace de E un $\{f_1, \dots, f_r\}$ une base de F. Par le théorème 4.15(2), nous pouvons compléter la base de F par des éléments de la base $\{e_i\}$. Le complément ainsi construit est invariant par u.

(4) implique (1) En dimension un, tout endomorphisme est diagonalisable, nous supposons donc que $\dim E = n \geqslant 2$. Nous procédons par récurrence sur le nombre de vecteurs propres connus de u. Supposons avoir déjà trouvé p vecteurs propres $e_1, \dots, e_p$ de u. Considérons H, un hyperplan qui contient les vecteurs $e_1, \dots, e_p$. Soit F un supplémentaire de H stable par u ; par construction $\dim F = 1$ et si $e_{p+1} \in F$, il doit être vecteur propre de u.

(1) implique (2) Nous supposons maintenant que u est diagonalisable. Soient $\lambda_1, \dots, \lambda_r$ les valeurs propres deux à deux distinctes, et considérons le polynôme

$$P(x) = (X - \lambda_1) \dots (X - \lambda_r). \tag{9.264}$$

Alors $P(u) = 0$. En effet si e_i est un vecteur propre pour la valeur propre λ_i,

$$P(u)e_i = \prod_{j \neq i} (u - \lambda_j) \circ (u - \lambda_i)e_i = 0 \tag{9.265}$$

par le lemme 9.87. Par conséquent $P(u)$ s'annule sur une base.

(5) implique (2) Si la matrice A est diagonale alors le polynôme $P = \prod_{i=1}^{n}(A - A_{ii}\mathbb{1})$ est annulateur de A.

(3) implique (5) le polynôme minimal de u s'écrit

$$\mu = (X - \lambda_1) \dots (X - \lambda_r), \tag{9.266}$$

et les espaces E_i du lemme 9.127 sont les espaces propres $E_i = \ker(u - \lambda_i)$. Nous avons donc une somme directe

$$E = E_1 \oplus \dots \oplus E_r. \tag{9.267}$$

Dans chacun des espaces propres, u a une matrice diagonale avec la valeur propre correspondante sur la diagonale. Une base de E constituée d'une base de chacun des espaces propres est donc une base comme nous en cherchons.

□

Corolaire 9.129.
Si u est diagonalisable et si F est une sous-espace stable par u, alors

$$F = \bigoplus_{\lambda} E_\lambda(u) \cap F \tag{9.268}$$

où $E_\lambda(u)$ est l'espace propre de u pour la valeur propre λ. En particulier la restriction de u à F, $u|_F$ est diagonalisable.

Démonstration. Par le théorème 9.128, le polynôme μ_u est scindé et ne possède que des racines simples. Notons le

$$\mu_u(X) = (X - \lambda_1) \dots (X - \lambda_r). \tag{9.269}$$

Les espaces E_i du lemme 9.127 sont maintenant les espaces propres.

En ce qui concerne la diagonalisabilité de $u|_F$, notons que nous avons une base de F composée de vecteurs dans les espaces $E_\lambda(u)$. Cette base de F est une base de vecteurs propres de u. □

Lemme 9.130.
Soit E un $\mathbb{K}$-espace vectoriel et $u \in \mathrm{End}(E)$. Si $\mathrm{Card}\big(\mathrm{Spec}(u)\big) = \dim(E)$ alors u est diagonalisable.

Démonstration. Soient $\lambda_1, \dots, \lambda_n$ les valeurs propres distinctes de u. Nous savons que les espaces propres correspondants sont en somme directe (lemme 9.85). Par conséquent $\mathrm{Span}\{E_{\lambda_i}(u)\}$ est de dimension n est u est diagonalisable. □

Voici un résultat de diagonalisation simultanée. Nous donnerons un résultat de trigonalisation simultanée dans le lemme 9.244.

Proposition 9.131 (Diagonalisation simultanée)**.**
Soit $(u_i)_{i\in I}$ une famille d'endomorphismes qui commutent deux à deux.

(1) Si $i, j \in I$ alors tout sous-espace propre de u_i est stable par u_j. Autrement dit $u_j\big(E_\lambda(u)\big) \subset E_\lambda(u)$.

(2) Si les u_i sont diagonalisables, alors ils le sont simultanément.

Démonstration. Supposons que u_i et u_j commutent et soit x un vecteur propre de u_i : $u_i x = \lambda x$. Nous montrons que $u_j x \in E_\lambda(u)$. Nous avons

$$u_i\big(u_j(x)\big) = u_j\big(u_i(x)\big) = \lambda u_j(x). \tag{9.270}$$

Par conséquent $u_j(x)$ est vecteur propre de u_i de valeur propre λ.

Montrons maintenant l'affirmation à propos des endomorphismes simultanément diagonalisables. Si $\dim E = 1$, le résultat est évident. Nous supposons également qu'aucun des u_i n'est multiple de l'identité. Nous effectuons une récurrence sur la dimension.

Soit u_0 un des u_i et considérons ses valeurs propres deux à deux distinctes $\lambda_1, \dots, \lambda_r$. Pour chaque k nous avons

$$E_{\lambda_k}(u_0) \neq E, \tag{9.271}$$

sinon u_0 serait un multiple de l'identité. Par contre le fait que u_0 soit diagonalisable permet de décomposer E en espaces propres de u_0 :

$$E = \bigoplus_k E_{\lambda_k}(u_0). \tag{9.272}$$

Ce que nous allons faire est de simultanément diagonaliser les $(u_i)_{i\in I}$ sur chacun des E_{λ_k} séparément. Par le point (1), nous avons $u_i \colon E_{\lambda_k}(u_0) \to E_{\lambda_k}(u_0)$, et nous pouvons considérer la famille d'opérateurs

$$\Big(u_i|_{E_{\lambda_k}(u_0)}\Big)_{i\in I}. \tag{9.273}$$

Ce sont tous des opérateurs qui commutent et qui agissent sur un espace de dimension plus petite. Par hypothèse de récurrence nous avons une base de $E_{\lambda_k}(u_0)$ qui diagonalise tous les u_i. □

Exemple 9.132.
Soit un espace vectoriel sur un corps $\mathbb{K}$. Un opérateur **involutif** est un opérateur différent de l'identité dont le carré est l'identité. Typiquement une symétrie orthogonale dans $\mathbb{R}^3$. Le polynôme caractéristique d'une involution est $X^2 - 1 = (X+1)(X-1)$.

Tant que $1 \neq -1$, $X^1 - 1$ est donc scindé à racines simples et les involutions sont diagonalisables (9.128). Cependant si le corps est de caractéristique 2, alors $X^2 - 1 = (X+1)^2$ et l'involution n'est plus diagonalisable.

Par exemple si le corps est de caractéristique 2, nous avons

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{9.274a}$$

$$A^1 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{9.274b}$$

Ce A est donc une involution mais n'est pas diagonalisable. △

9.7.4 Diagonalisation : cas complexe, pas toujours

Il n'est pas vrai qu'une matrice de $\mathbb{M}(n,\mathbb{C})$ soit toujours diagonalisable. En effet le théorème 9.128(3) dit qu'une matrice est diagonalisable si et seulement si son polynôme minimal est scindé à racines simples. Certes sur $\mathbb{C}$ le polynôme minimal sera scindé, mais il ne sera pas spécialement à racines simples.

Exemple 9.133.
La matrice

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{9.275}$$

a pour polynôme caractéristique $\chi_A(X) = X^2$. Cela est également son polynôme minimal, et ce n'est pas à racine simple.

Il est par ailleurs facile de voir que le seul espace propre de A est $\mathrm{Span}\{(1,0)\}$ (ici le span est sur $\mathbb{C}$). Donc l'espace $\mathbb{C}^2$ ne possède pas de base de vecteurs propres de A. △

Ce qui est vrai, c'est que le polynôme caractéristique a des racines, et que ces racines correspondent à des vecteurs propres. Mais il n'y a pas toujours autant de vecteurs propres que la multiplicité des racines.

9.7.5 Trigonalisation : généralités

Définition 9.134 ([180]).
*Une matrice dans $\mathbb{M}(n,\mathbb{K})$ est **trigonalisable** lorsqu'elle est semblable*[36] *à une matrice triangulaire supérieure.*

Proposition 9.135 (Trigonalisation et polynôme caractéristique scindé).
Soit u un endomorphisme d'un espace vectoriel E sur le corps $\mathbb{K}$. Les faits suivants sont équivalents.

(1) L'endomorphisme u est trigonalisable (auquel cas les valeurs propres sont sur la diagonale).

(2) Le polynôme caractéristique de u est scindé[37].

Démonstration. **(2)⇒(1)** Nous avons par hypothèse que

$$\chi_u(X) = \prod_{i=1}^{r} (X-\lambda_i)^{\alpha_i} \tag{9.276}$$

où les λ_i sont les valeurs propres de u. Le théorème de Cayley-Hamilton 9.115 dit que $\chi_u(u) = 0$, ce qui permet d'utiliser le théorème de décomposition des noyaux 9.88 :

$$E = \ker(X-\lambda_1)^{\alpha_1} \oplus \ldots \oplus \ker(X-\lambda_r)^{\alpha_r}. \tag{9.277}$$

Les espaces $F_{\lambda_i}(u) = \ker(X-\lambda_i)^{\alpha_i}$ sont les espaces caractéristiques de u, ce qui fait que $u-\lambda_i\mathbb{1}$ est nilpotent sur $F_{\lambda_i}(u)$. L'endomorphisme $u-\lambda_i\mathbb{1}$ est donc strictement trigonalisable supérieur sur son bloc[38]. Cela signifie que u est triangulaire supérieure avec les valeurs propres sur la diagonale.

(1)⇒(2) C'est immédiat parce que le déterminant d'une matrice triangulaire est le produit des éléments de sa diagonale.

□

Remarque 9.136.
La méthode des pivots de Gauss[39] certes permet de trigonaliser n'importe quoi, mais elle ne

36. Définition 9.119.
37. Définition 6.32.
38. Proposition 9.122.
39. Le lemme 4.111.

correspond pas à un changement de base. Autrement dit, les pivots de Gauss ne sont pas de similitudes.

C'est là qu'il faut bien avoir en tête la différence entre *équivalence* et *similarité*[40]. Lorsqu'on parle de changement de base, de matrice trigonalisable ou diagonalisable, nous parlons de similarité et non d'équivalence.

9.7.6 Trigonalisation : cas complexe

La proposition 9.135 dit déjà que tous les endomorphismes sont trigonalisables sur $\mathbb{C}$. Nous allons aller plus loin et montrer que la trigonalisation peut être effectuée à l'aide d'une matrice unitaire.

Une démonstration alternative passant par le polynôme caractéristique sera présentée dans la remarque 9.140 utilisant la proposition 9.135.

Lemme 9.137 (Lemme de Schur complexe, trigonisation[181])**.**
Si $A \in \mathbb{M}(n, \mathbb{C})$, il existe une matrice unitaire U telle que UAU^{-1} soit triangulaire supérieure[41].

Démonstration. Étant donné que $\mathbb{C}$ est algébriquement clos, nous pouvons toujours considérer un vecteur propre v_1 de A, de valeur propre λ_1. Nous pouvons utiliser un procédé de Gram-Schmidt pour construire une base orthonormée $\{v, u_2, \dots, u_n\}$ de $\mathbb{R}^n$, et la matrice (unitaire)

$$Q = \begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ v & u_2 & \cdots & u_n \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}. \tag{9.278}$$

Nous avons $Q^{-1}AQe_1 = Q^{-1}Av = \lambda Q^{-1}v = \lambda e_1$, par conséquent la matrice $Q^{-1}AQ$ est de la forme

$$Q^{-1}AQ = \begin{pmatrix} \lambda_1 & * \\ 0 & A_1 \end{pmatrix} \tag{9.279}$$

où $*$ représente une ligne quelconque et A_1 est une matrice de $\mathbb{M}(n-1, \mathbb{C})$. Nous pouvons donc répéter le processus sur A_1 et obtenir une matrice triangulaire supérieure (nous utilisons le fait qu'un produit de matrices orthogonales est une matrice orthogonale). □

En particulier les matrices hermitiennes, anti-hermitiennes et unitaires sont trigonalisables par une matrice unitaire, qui peut être choisie de déterminant 1.

Lemme 9.138.
Soit $A \in \mathbb{M}(n, \mathbb{C})$ et une matrice unitaire U telle que $A = UTU^{-1}$ où T est triangulaire.

(1) En ce qui concerne les polynômes caractéristiques, $\chi_A = \chi_T$.

(2) Pour les spectres, $\mathrm{Spec}(A) = \mathrm{Spec}(T)$.

(3) Les valeurs propres de A sont les éléments diagonaux de T.

Démonstration. Vu que U commute évidemment avec $\mathbb{1}$ nous avons

$$\chi_A(\lambda) = \det(A - \lambda\mathbb{1}) = \det(UTU^{-1} - \lambda\mathbb{1}) = \det\big(U(T - \lambda\mathbb{1})U^{-1}\big). \tag{9.280}$$

À ce niveau nous utilisons le fait que le déterminant soit multiplicatif 9.10 pour conclure :

$$\chi_A(\lambda) = \det\big(U(T - \lambda\mathbb{1})U^{-1}\big) = \det(U)\det(T - \lambda\mathbb{1})\det(U^{-1}) = \det(T - \lambda\mathbb{1}) = \chi_T(\lambda). \tag{9.281}$$

Pour les spectres, l'égalité des polynômes caractéristique implique l'égalité des spectres parce que les valeurs propres sont les racines du polynôme caractéristique par le théorème 9.111.

Les valeurs propres d'une matrice triangulaire sont les valeurs sur la diagonale. □

40. Définition 4.110.

41. « triangulaire supérieure » ne signifie pas « strictement triangulaire supérieure ». Ici, il est possible que la diagonale soit non nulle ; non seulement possible, mais même très probable en pratique.

Remarque 9.139.
Le lemme mentionne le fait que les valeurs propres de A sont les éléments diagonaux de T. Mais attention : ceci ne dit rien au niveau des multiplicités géométriques. Un nombre peut être cinq fois sur la diagonale de T alors que l'espace propre correspondant pour A n'est que de dimension 1. Exemple : la matrice

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{9.282}$$

a deux 1 sur la diagonale. Le nombre 1 est bien une valeur propre de A, mais le système

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \tag{9.283}$$

donne $y = 0$ et donc un espace propre de dimension seulement 1.

Remarque 9.140.
Si $\mathbb{K}$ est algébriquement clos (comme $\mathbb{C}$ par exemple), alors tous les polynômes sont scindés et toutes les matrices sont trigonalisables [42]. Un exemple un peu simple de cela est la matrice

$$u = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{9.284}$$

Le polynôme caractéristique est $\chi_u(X) = X^2 + 1$ et les valeurs propres sont $\pm i$. Il est vite vu que dans la base

$$\{\begin{pmatrix} i \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ i \end{pmatrix}\} \tag{9.285}$$

de $\mathbb{C}^2$, la matrice u se note $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$.

Remarque 9.141.
Cela nous donne une autre façon de prouver qu'une matrice nilpotente de $\mathbb{M}(n, \mathbb{C})$ ou $\mathbb{M}(n, \mathbb{R})$ est trigonalisable[182]. D'abord dans $\mathbb{M}(n, \mathbb{C})$, toutes les matrices sont trigonalisables [43], et les valeurs propres arrivent sur la diagonale. Mais comme les valeurs propres d'une matrice nilpotente sont zéro, elle est triangulaire stricte. Par ailleurs son polynôme caractéristique est alors X^n.

Ensuite si $u \in \mathbb{M}(n, \mathbb{R})$ nous pouvons voir u comme une matrice dans $\mathbb{M}(n, \mathbb{C})$ et y calculer son polynôme caractéristique qui sera tout de même X^n. Ce polynôme étant scindé, la proposition 9.135 nous assure que u est trigonalisable. Une fois de plus, les valeurs propres étant sur la diagonale, elle est triangulaire supérieure stricte.

Corolaire 9.142.
Le polynôme caractéristique [44] *sur $\mathbb{C}$ d'une matrice s'écrit sous la forme*

$$\chi_A(X) = \prod_{i=1}^{r} (X - \lambda_i)^{m_i} \tag{9.286}$$

où les λ_i sont les valeurs propres distinctes de A et m_i sont les multiplicités correspondantes.

Démonstration. Le lemme 9.137 nous donne l'existence d'une base de trigonalisation ; dans cette base les valeurs propres de A sont sur la diagonale et nous avons

$$\chi_A(X) = \det(A - X\mathbb{1}) = \det \begin{pmatrix} X - \lambda_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & X - \lambda_r \end{pmatrix}, \tag{9.287}$$

qui vaut bien le produit annoncé. □

42. La proposition 9.135 montre cela, et le lemme de Schur complexe 9.137 va un peu plus loin et précise que la trigonalisation peut être faite par une matrice unitaire.
43. Parce que le polynôme caractéristique est scindé, voir la proposition 9.135..
44. Définition 9.107.

Corolaire 9.143.
Si $A \in \mathbb{M}(n, \mathbb{C})$ et $k \in \mathbb{N}$ alors

$$\operatorname{Spec}(A^k) = \{\lambda^k \text{ tel que } \lambda \in \operatorname{Spec}(A)\}. \tag{9.288}$$

Démonstration. Par le lemme 9.137 nous avons une matrice unitaire U et une triangulaire T telles que $A = UTU^{-1}$. En passant à a puissance k nous avons aussi

$$A^k = UT^kU^{-1}. \tag{9.289}$$

Donc le spectre de A^k est celui de T^k (lemme 9.138 et le fait qu'une puissance d'une matrice triangulaire est encore triangulaire). Or les éléments diagonaux de T^k sont les puissances k^{e}des éléments diagonaux de T, qui sont les valeurs propres de A. □

9.7.7 Diagonalisation : cas complexe, ce qu'on a

Lemme 9.144 (Théorème spectral hermitien)**.**
Pour un opérateur hermitien[45]*,*

(1) le spectre est réel,

(2) deux vecteurs propres pour des valeurs propres distinctes sont orthogonales[46]*.*

Démonstration. Soit v un vecteur de valeur propre λ. Nous avons d'une part

$$\langle Av, v\rangle = \lambda\langle v, v\rangle = \lambda\|v\|^2, \tag{9.290}$$

et d'autre part, en utilisant le fait que A est hermitien,

$$\langle Av, v\rangle = \langle v, A^*v\rangle = \langle v, Av\rangle = \bar{\lambda}\|v\|^2, \tag{9.291}$$

par conséquent $\lambda = \bar{\lambda}$ parce que $v \neq 0$.

Soient λ_i et v_i $(i = 1, 2)$ deux valeurs propres de A avec leurs vecteurs propres correspondants. Alors d'une part

$$\langle Av_1, v_2\rangle = \lambda_1\langle v_1, v_2\rangle, \tag{9.292}$$

et d'autre part

$$\langle Av_1, v_2\rangle = \langle v_1, Av_2\rangle = \lambda_2\langle v_1, v_2\rangle. \tag{9.293}$$

Nous avons utilisé le fait que λ_2 était réel. Par conséquent, soit $\lambda_1 = \lambda_2$, soit $\langle v_1, v_2\rangle = 0$. □

Remarque 9.145.
Un opérateur de la forme A^*A est évidemment hermitien. De plus ses valeurs propres sont toutes positives parce que si $A^*Ax = \lambda v$ alors

$$0 \leqslant \langle Av, Av\rangle = \langle A^*Av, v\rangle = \lambda\langle v, v\rangle. \tag{9.294}$$

Donc $\lambda \geqslant 0$.

Définition 9.146.
*Un endomorphisme est **normal** s'il commute avec son adjoint.*

Les opérateurs normaux comprennent évidemment les opérateurs hermitiens, mais également les anti-hermitiens, et ça c'est bien parce que c'est le cas de l'algèbre associée à SU(2).

Théorème 9.147 (Théorème spectral pour les matrices normales[47][183, 184, 185])**.**
Soit $A \in \mathbb{M}(n, \mathbb{C})$ une matrice de valeurs propres $\lambda_1, \dots, \lambda_n$ (non spécialement distinctes). Alors les conditions suivantes sont équivalentes :

45. Définition 9.35.
46. Pour la forme (7.163).
47. Définition 9.146

(1) A est normale,

(2) A se diagonalise par une matrice unitaire,

(3) $\sum_{i,j=1}^{n} |A_{ij}|^2 = \sum_{j=1}^{n} |\lambda_j|^2$,

(4) il existe une base orthonormale de vecteurs propres de A.

Démonstration. Nous allons nous contenter de prouver (1)⇔(2).

Soit Q la matrice unitaire donnée par la décomposition de Schur (lemme 9.137) : $A = QTQ^{-1}$. Étant donné que A est normale nous avons

$$QTT^*Q^{-1} = QT^*TQ^{-1}, \tag{9.295}$$

ce qui montre que T est également normale. Or une matrice triangulaire supérieure normale est diagonale. En effet nous avons $T_{ij} = 0$ lorsque $i > j$ et

$$(TT^*)_{ii} = (T^*T)_{ii} = \sum_{k=1}^{n} |T_{ki}|^2 = \sum_{k=1}^{n} |T_{ik}|^2. \tag{9.296}$$

Écrivons cela pour $i = 1$ en tenant compte de $|T_{k1}|^2 = 0$ pour $k = 2, \dots, n$,

$$|T_{11}|^2 = |T_{11}|^2 + |T_{12}|^2 + \cdots + |T_{1n}|^2, \tag{9.297}$$

ce qui implique que T_{11} est le seul non nul parmi les T_{1k}. En continuant de la sorte avec $i = 2, \dots, n$ nous trouvons que T est diagonale.

Dans l'autre sens, si A se diagonalise par une matrice unitaire, $UAU^* = D$, nous avons

$$DD^* = UAA^*U^* \tag{9.298}$$

et

$$D^*D = UA^*AU^*, \tag{9.299}$$

qui ce prouve que A est normale. □

Tant que nous en sommes à parler de spectre de matrices hermitiennes... Soit une matrice inversible $A \in \mathrm{GL}(n, \mathbb{C})$. La matrice A^*A est hermitienne [48] et le théorème 9.144 nous assure que ses valeurs propres sont réelles. Par la remarque 9.145, ses valeurs propres sont même positives.

Lemme 9.148 ([186])**.**
Si A est une matrice carrée et inversible,

$$\mathrm{Spec}(A^*A) = \mathrm{Spec}(AA^*) \tag{9.300}$$

Démonstration. Nous allons montrer l'égalité des polynômes caractéristiques. D'abord une simple multiplication montre que

$$(A^*A - \lambda\mathbb{1})A^{-1} = A^{-1}(AA^* - \lambda\mathbb{1}). \tag{9.301}$$

Nous prenons le déterminant de cette égalité en utilisant les propriétés 9.10(1) et (3) :

$$\det(A^*A - \lambda\mathbb{1})\det(A^{-1}) = \det(A^{-1})\det(AA^* - \lambda\mathbb{1}). \tag{9.302}$$

En simplifiant par $\det(A^{-1})$ (qui est non nul parce que A est inversible) nous obtenons l'égalité des polynômes caractéristiques et donc l'égalité des spectres. □

48. Définition 9.35.

9.7.8 Diagonalisation : cas réel

Lemme 9.149 (Lemme de Schur réel)**.**
Soit $A \in \mathbb{M}(n, \mathbb{R})$. Il existe une matrice orthogonale Q telle que $Q^{-1}AQ$ soit de la forme

$$QAQ^{-1} = \begin{pmatrix} \lambda_1 & * & * & * & * \\ 0 & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \lambda_r & * & * \\ 0 & 0 & 0 & \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} & * \\ 0 & 0 & 0 & 0 & \begin{pmatrix} a_s & b_s \\ c_s & d_s \end{pmatrix} \end{pmatrix}. \tag{9.303}$$

Le déterminant de A est le produit des déterminants des blocs diagonaux et les valeurs propres de A sont les $\lambda_1, \dots, \lambda_r$ et celles de ces blocs.

Démonstration. Si la matrice A a des valeurs propres réelles, nous procédons comme dans le cas complexe. Cela nous fournit le partie véritablement triangulaire avec les valeurs propres $\lambda_1, \dots, \lambda_r$ sur la diagonale. Supposons donc que A n'a pas de valeurs propres réelles. Soit donc $\alpha + i\beta$ une valeur propre ($\beta \neq 0$) et $u + iv$ un vecteur propre correspondant où u et v sont des vecteurs réels. Nous avons

$$Au + iAv = A(u + iv) = (\alpha + i\beta)(u + iv) = \alpha u - \beta v + i(\alpha v + \beta v), \tag{9.304}$$

et en égalisant les parties réelles et imaginaires,

$$Au = \alpha u - \beta v \tag{9.305a}$$
$$Av = \alpha v + \beta u. \tag{9.305b}$$

Sur ces relations nous voyons que ni u ni v ne sont nuls. De plus u et v sont linéairement indépendants (sur $\mathbb{R}$), en effet si $v = \lambda u$ nous aurions $Au = \alpha u - \beta\lambda u = (\alpha - \beta\lambda)u$, ce qui serait une valeur propre réelle alors que nous avions supposé avoir déjà épuisé toutes les valeurs propres réelles.

Étant donné que u et v sont deux vecteurs réels non nuls et linéairement indépendants, nous pouvons trouver une base orthonormée $\{q_1, q_2\}$ de $\mathrm{Span}\{u, v\}$. Nous pouvons étendre ces deux vecteurs en une base orthonormée $\{q_1, q_2, q_3, \dots, q_n\}$ de $\mathbb{R}^n$. Nous considérons à présent la matrice orthogonale dont les colonnes sont formées de ces vecteurs : $Q = [q_1\, q_2\, \dots q_n]$.

L'espace $\mathrm{Span}\{e_1, e_2\}$ est stable par $Q^{-1}AQ$, en effet nous avons

$$Q^{-1}AQe_1 = Q^{-1}Aq_1 = Q^{-1}(aq_1 + bq_2) = ae_1 + be_2. \tag{9.306}$$

La matrice $Q^{-1}AQ$ est donc de la forme

$$Q^{-1}AQ = \begin{pmatrix} \begin{pmatrix} \cdot & \cdot \\ \cdot & \cdot \end{pmatrix} & C_1 \\ 0 & A_1 \end{pmatrix} \tag{9.307}$$

où C_1 est une matrice réelle $2 \times (n-1)$ quelconque et A_1 est une matrice réelle $(n-2) \times (n-2)$. Nous pouvons appliquer une récurrence sur la dimension pour poursuivre.

Notons que si A n'a pas de valeurs propres réelles, elle est automatiquement d'ordre pair parce que les valeurs propres complexes viennent par couple complexes conjuguées.

En ce qui concerne les valeurs propres, il est facile de voir en regardant (9.303) que les valeurs propres sont celles des blocs diagonaux. Étant donné que QAQ^{-1} et A ont même polynôme caractéristique, ce sont les valeurs propres de A. □

Théorème 9.150 (Théorème spectral, matrice symétrique[54])**.**
Une matrice symétrique réelle,

(1) a un spectre contenu dans $\mathbb{R}$

(2) est diagonalisable par une matrice orthogonale.

Si M *est une matrice symétrique réelle alors* $\mathbb{R}^n$ *possède une base orthonormée de vecteurs propres de* M*.*

Démonstration. Soit A une matrice réelle symétrique. Si λ est une valeur propre complexe pour le vecteur propre complexe v, alors d'une part $\langle Av, v\rangle = \lambda\langle v, v\rangle$ et d'autre part $\langle Av, v\rangle = \langle v, Av\rangle = \bar{\lambda}\langle v, v\rangle$. Par conséquent $\lambda = \bar{\lambda}$.

Le lemme de Schur réel 9.149 donne une matrice orthogonale qui trigonalise A. Les valeurs propres étant toutes réelles, la matrice QAQ^{-1} est même triangulaire (il n'y a pas de blocs dans la forme (9.303)). Prouvons que QAQ^{-1} est symétrique :

$$(QAQ^{-1})^t = (Q^{-1})^t A^t Q^t = QA^tQ^{-1} = QAQ^{-1} \tag{9.308}$$

où nous avons utilisé le fait que Q était orthogonale ($Q^{-1} = Q^t$) et que A était symétrique ($A^t = A$). Une matrice triangulaire supérieure symétrique est obligatoirement une matrice diagonale.

En ce qui concerne la base de vecteurs propres, soit $\{e_i\}_{i=1,\dots,n}$ la base canonique de $\mathbb{R}^n$ et Q une matrice orthogonale e telle que $A = Q^tDQ$ avec D diagonale. Nous posons $f_i = Q^te_i$ et en tenant compte du fait que $Q^t = Q^{-1}$ nous avons $Af_i = Q^tDQQ^te_i = Q^t\lambda_ie_i = \lambda_if_i$. Donc les f_i sont des vecteurs propres de A. De plus ils sont orthonormés parce que

$$\langle f_i, f_j\rangle = \langle Q^te_i, Q^te_j\rangle = \langle e_i, Q^tQe_j\rangle = \langle e_i, e_j\rangle = \delta_{ij}. \tag{9.309}$$

□

Le théorème spectral pour les opérateurs autoadjoints sera traité plus bas parce qu'il a besoin de choses sur les formes bilinéaires, théorème 9.267.

Remarque 9.151.
Une matrice symétrique est diagonalisable par une matrice orthogonale. Nous pouvons en réalité nous arranger pour diagonaliser par une matrice de $\mathrm{SO}(n)$. Plus généralement si A est une matrice diagonalisable par une matrice $P \in \mathrm{GL}^+(n, \mathbb{R})$ alors elle est diagonalisable par une matrice de $\mathrm{GL}^-(n, \mathbb{R})$ en changeant le signe de la première ligne de P. Et inversement.

En effet, si nous avons $P^tDP = A$, alors en notant $*$ les quantités qui ne dépendent pas de a, b ou c,

$$\begin{aligned}\begin{pmatrix} a & * & * \\ b & * & * \\ c & * & * \end{pmatrix}\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix}\begin{pmatrix} a & b & c \\ * & * & * \\ * & * & * \end{pmatrix} &= \begin{pmatrix} a & * & * \\ b & * & * \\ c & * & * \end{pmatrix}\begin{pmatrix} \lambda_1 a & \lambda_1 b & \lambda_1 c \\ * & * & * \\ * & * & * \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 a^2 + * & \lambda_1 ab + * & \lambda_1 ac + * \\ \dots & \dots & \dots \\ \dots & \dots & \dots \end{pmatrix}.\end{aligned} \tag{9.310}$$

Nous voyons donc que si nous changeons les signes de a, b et c en même temps, le résultat ne change pas.

Définition 9.152 (Matrice définie positive, opérateur définit positif)**.**
Un opérateur sur un espace vectoriel sur $\mathbb{C}$ *ou* $\mathbb{R}$ *est* ***définit positif*** *si toutes ses valeurs propres sont réelles et strictement positives. Il est* ***semi-définie positive*** *si ses valeurs propres sont réelles positives ou nulles.*

Afin d'éviter l'une ou l'autre confusion, nous disons souvent *strictement* définie positive pour positive.

9.153.
Nous nommons $S^+(n, \mathbb{R})$ l'ensemble des matrices réelles symétriques $n \times n$ et $S^{++}(n, \mathbb{R})$ le sous-ensemble de $S^+(n, \mathbb{R})$ des matrices strictement définies positives.

Remarque 9.154.
Nous ne définissons pas la notion de matrice définie positive pour une matrice non symétrique.

Proposition 9.155.
Soit M, une matrice symétrique. Nous avons

(1) $\det M > 0$ *et* $\mathrm{Tr}(M) > 0$ *implique M définie positive*[49],

(2) $\det M > 0$ *et* $\mathrm{Tr}(M) < 0$ *implique M définie négative,*

(3) $\det M < 0$ *implique ni semi-définie positive, ni définie négative*

(4) $\det M = 0$ *implique M semi-définie positive ou semi-définie négative.*

Proposition 9.156.
Une application linéaire est définie positive[50] *si et seulement si sa matrice associée l'est.*

Lorsqu'un énoncé parle d'une matrice symétrique, le premier réflexe est de la diagonaliser : considérer une matrice orthogonale T telle que $T^tMT = D$ avec D diagonale. Et les valeurs propres sur la diagonale : $D_{kl} = \delta_{kl}\lambda_k$. Les matrices symétriques définies positives ont cependant des propriétés même en dehors de leur base de diagonalisation.

Lemme 9.157.
Soit une matrice symétrique M.

(1) Elle est strictement définie positive si et seulement si $\langle x, Mx\rangle > 0$ pour tout x non nul dans $\mathbb{R}^n$.

(2) Elle est semi-définie positive si et seulement si $\langle x, Mx\rangle \geqslant 0$ pour tout x non nul dans $\mathbb{R}^n$.

(3) Si elle est seulement définie positive, alors $\langle x, Mx\rangle \geqslant \lambda\|x\|^2$ dès que $\lambda \geqslant 0$ minore toutes les valeurs propres.

Démonstration. Démonstration en trois parties.

(1) Soit $\{e_i\}_{i=1,\dots,n}$ une base orthonormée de vecteurs propres de M dont l'existence est assurée par le théorème spectral 9.150. Nous nommons x_i les coordonnées de x dans cette base. Alors,

$$\langle x, Mx\rangle = \sum_{i,j} x_i\langle e_i, x_jMe_j\rangle = \sum_{i,j} x_ix_j\langle e_i, \lambda_je_j\rangle = \sum_{ij} x_ix_j\lambda_j\delta_{ij} = \sum_i \lambda_ix_i^2 \tag{9.311}$$

où les λ_i sont les valeurs propres de M. Cela est strictement positif pour tout x si et seulement si tous les λ_i sont strictement positifs.

(2) Nous avons encore

$$\langle x, Mx\rangle = \sum_i \lambda_ix_i^2. \tag{9.312}$$

Cela est plus grand ou égal à zéro si et seulement si tous les λ_i sont plus grands ou égaux à zéro.

(3) Soit une matrice orthogonale T diagonalisant M, c'est-à-dire telle que $T^tMT = D$ avec D diagonale. Nous allons vérifier que

$$\langle Tx, Mtx\rangle \geqslant \lambda\|Tx\|^2 \tag{9.313}$$

pour tout x. Vu que T est une bijection[51], cela impliquera le résultat pour tout x. Si nous

49. Défintion 9.152.
50. Définition 9.152.
51. Une matrice orthogonale a un déterminant ± 1.

considérons la base de diagonalisation $\{e_k\}$ pour les valeurs propres λ_k, nous avons le calcul

$$\begin{aligned}
\langle Tx, MTx\rangle &= \langle x, T^tMTx\rangle && (9.314a)\\
&= \langle x, Dx\rangle && (9.314b)\\
&= \sum_k \langle x, x_k De_k\rangle && (9.314c)\\
&= \sum_k \lambda_k x_k \underbrace{\langle x, e_k\rangle}_{=x_k} && (9.314d)\\
&\geqslant \sum_k \lambda |x_k|^2 && (9.314e)\\
&= \lambda \|x\|^2 && (9.314f)\\
&= \lambda \|Tx\|^2. && (9.314g)
\end{aligned}$$

Au dernier passage nous avons utilisé le fait que T est une isométrie (proposition 9.42).

□

Les personnes qui aiment les vecteurs lignes et colonnes écriront des inégalités comme

$$x^tMx \geqslant x^tx. \tag{9.315}$$

Tout à l'autre bout du spectre des personnes névrosées des notations, on trouvera des inégalités comme

$$M(x \otimes x) \geqslant x \cdot x. \tag{9.316}$$

Le penchant personnel de l'auteur de ces lignes est la notation avec le produit tensoriel. Si vous aimez ça, vous pouvez lire la section 10.6.6 et en particulier ce qui suit (10.284).

La notation adoptée ici avec le produit scalaire $\langle x, Mx\rangle$ est entre les deux. Elle a l'avantage de n'être pas technologique comme le produit tensoriel (si vous y mettez les pieds, vous devez savoir ce que vous faites), tout en évitant de se casser la tête à savoir qui est un vecteur ligne ou un vecteur colonne.

Corolaire 9.158.
Une matrice symétrique strictement définie positive est inversible.

Démonstration. Si $Ax = 0$ alors $\langle Ax, x\rangle = 0$. Mais dans le cas d'une matrice strictement définie positive, cela implique $x = 0$ par le lemme 9.157. □

Lemme 9.159.
Pour une base quelconque, les éléments diagonaux d'une matrice symétrique semi-définie positive sont positifs. Si la matrice est strictement définie positive, alors les éléments diagonaux sont strictement positifs.

Démonstration. Il s'agit d'une application du lemme 9.157. Si A est définie positive et que $\{e_i\}$ est une base, alors

$$A_{ii} = \langle Ae_i, e_i\rangle \geqslant \lambda \|e_i\|^2 = \lambda \geqslant 0. \tag{9.317}$$

Si A est strictement définie positive, alors λ peut être choisi strictement positif. □

9.8 Formes bilinéaires et quadratiques

Plus à propos de formes bilinéaires dans le thème 49.

9.8.1 Généralités

Définition 9.160 ([187]).
*Soit un espace vectoriel E et $\mathbb{F}$ un corps de caractéristique différente de 2. Une **forme quadratique** sur E est une application $q\colon E \to \mathbb{F}$ pour laquelle il existe une forme bilinéaire symétrique $b\colon E \times E \to \mathbb{F}$ satisfaisant $q(x) = b(x, x)$ pour tout $x \in E$.*

L'ensemble des formes quadratiques réelles sur E est noté $Q(E)$.

Proposition 9.161.
*Soit une forme bilinéaire b et la forme quadratique associée q. Alors nous avons l'**identité de polarisation** :*

$$b(x, y) = \frac{1}{2}\big(q(x) + q(y) - q(x - y)\big). \tag{9.318}$$

Démonstration. Il suffit de substituer dans le membre de droite $q(x) = b(x, y)$ et d'utiliser la bilinéarité :

$$\begin{aligned} q(x) + q(y) - q(x - y) &= b(x, x) + b(y, y) - b(x - y, x - y) && \text{(9.319a)}\\ &= b(x, x) + b(y, y) - b(x) + b(x, y) + b(y, x) - b(y, y) && \text{(9.319b)}\\ &= 2b(x, y) && \text{(9.319c)} \end{aligned}$$

où nous avons utilisé le fait que b est symétrique : $b(x, y) = b(y, x)$. □

Lemme 9.162.
Si q est une forme quadratique, il existe une unique forme bilinéaire b telle que $q(x) = b(x, x)$.

Démonstration. L'existence n'est pas en cause : c'est la définition d'une forme quadratique. Pour l'unicité, étant donné une forme quadratique, la forme bilinéaire b doit forcément vérifier l'identité de polarisation de la proposition 9.161. Elle est donc déterminée par q. □

Notons la division par 2 qui est le pourquoi de la demande de la caractéristique différente de 2 pour $\mathbb{F}$ dans la définition de forme quadratique.

Définition 9.163.
*Soit une forme quadratique q sur E. Nous disons que $v, w \in E$ sont **q-orthogonaux** si $b(v, w) = 0$ la forme bilinéaire b associée à q par le lemme 9.162.*

9.8.2 Matrice associée à une forme bilinéaire

Définition 9.164.
Soit une forme bilinéaire $b\colon E \times E \to \mathbb{K}$ et une base quelconque $\{f_\alpha\}$ de E. Nous définissons les nombres

$$B_{\alpha\beta} = b(f_\alpha, f_\beta), \tag{9.320}$$

*qui forment une matrice symétrique dans $\mathbb{M}(n, \mathbb{K})$. Cette matrice est la **matrice associée** à la forme bilinéaire b.*

La matrice d'une forme quadratique est celle associée à sa forme bilinéaire associée.

Lemme 9.165.
Soit une forme bilinéaire $b\colon E \times E \to \mathbb{K}$ et une base quelconque $\{f_\alpha\}$ de E. Nous notons B la matrice de b (definition 9.164) et q la forme quadratique associée.

Alors nous avons

$$b(x, y) = \sum_{\alpha\beta} B_{\alpha\beta} x_\alpha y_\beta. \tag{9.321}$$

et

$$b(x, y) = x \cdot By. \tag{9.322}$$

où le point est le produit scalaire usuel (composante par composante).

Démonstration. Si $x = \sum_\alpha x_\alpha f_\alpha$ et $y = \sum_\beta y_\beta f_\beta$:

En utilisant la convention (4.89) et les choses autour (voir aussi -2.1),

$$b(x,y) = \sum_\alpha x_\alpha \sum_\beta B_{\alpha\beta} y_\beta = \sum_\alpha x_\alpha (By)_\alpha = x \cdot By. \tag{9.323}$$

□

Proposition 9.166.
Soit une forme quadratique $q\colon E \to \mathbb{K}$ et sa matrice[52] $(q_{ij}) \in \mathbb{M}(n,\mathbb{K})$. *Nous avons*

$$q(x) = \sum_{i=1}^n \sum_{j=1}^n q_{ij} x_i x_j \tag{9.324a}$$

$$= \sum_{i=1}^n q_{ii} x_i^2 + 2 \sum_{1 \leqslant i < j \leqslant n} q_{ij} x_i x_j. \tag{9.324b}$$

9.167.
De nombreux auteurs préfèrent écrire des choses comme $x^t By$ ou $xB^t y$ ou xBy^t et se poser de longues questions sur qui est un « vecteur colonne » et qui est un « vecteur ligne », et si la matrice B soit être transposée ou non. Toutes ces notations servent(?) à cacher un bête produit scalaire.

9.168.
Notons que la matrice associée à une forme bilinéaire (ou quadratique associée) est uniquement valable pour une base donnée. Si nous changeons de base, la matrice change. Cependant lorsque nous travaillons sur $\mathbb{R}^n$, la base canonique est tellement canonique que nous allons nous permettre de parler de « la » matrice associée à une forme bilinéaire.

9.8.3 Changement de base : matrice d'une forme bilinéaire

Proposition 9.169 (Voir la section -2.1)**.**
Soit une forme bilinéaire[53] $b\colon V \times V \to \mathbb{K}$ *dont la matrice*[54] *dans la base* $\{e_i\}$ *est* A *et celle dans la base* $\{f_\alpha\}$ *est* B. *Nous supposons que les bases sont liées par* $f_\alpha = \sum_i Q_{i\alpha} e_i$. *Alors*

$$B = Q^t A Q. \tag{9.325}$$

Démonstration. Soit $x, x' \in V$ de coordonnées (x_i) et (x'_i) dans la base $\{e_i\}$ et (y_α), (y'_α) dans la base $\{f_\alpha\}$. Par définition de la matrice associée à une forme bilinéaire,

$$b(x,x') = \sum_{ij} A_{ij} x_i x'_j = \sum_{\alpha\beta} B_{\alpha\beta} y_\alpha y'_\beta. \tag{9.326}$$

En remplaçant les x_i et x'_i par leurs valeurs en fonction de y_α et y'_β données par la proposition 4.115, nous trouvons

$$b(x,x') = \sum_{ij\alpha\beta} A_{ij} Q_{i\alpha} y_\alpha Q_{j\beta} y'_\beta \tag{9.327a}$$

$$= \sum_{\alpha\beta} (Q^t A Q)_{\alpha\beta} y_\alpha y'_\beta \tag{9.327b}$$

où Q^t désigne la transposée de la matrice Q : $Q^t_{ij} = Q_{ji}$. Vu que les nombres y_α et y'_β sont arbitraires nous déduisons[55] que $B = Q^t A Q$. □

52. Matrice associée à une forme quadratique, définition 9.164.
53. Définition 7.182
54. Définition 9.320.
55. Lemme 4.64.

Remarque 9.170.
Notons que cette « loi de transformation » n'est pas la même que celle pour une application linéaire[56]. Ici nous avons Q^t alors que pour les applications linéaires nous avions Q^{-1}.

Pour cette raison, tant que nous travaillons avec des bases orthonormées, c'est-à-dire tant que Q est orthogonale[57], nous pouvons confondre une application linéaire avec une application bilinéaire en passant par la matrice. Mais cette identification n'est pas du tout canonique : elle repose sur le fait que les bases soient orthonormées.

Il en découle que la réduction des endomorphismes et la réduction des formes bilinéaires ne sont pas tout à fait les mêmes théories. Par exemple la pseudo-diagonalisation simultanée (corollaire 9.254) est un résultat de réduction de forme bilinéaire et non d'endomorphismes.

9.8.4 Réduction de Gauss

Théorème 9.171 (Réduction de Gauss[188, 189])**.**
Soit une forme quadratique non nulle q sur l'espace vectoriel E sur le corps $\mathbb{K}$. Il existe une base $\{l_i\}_{i=1,\dots,n}$ de E^* et des coefficients $\alpha_i \in \mathbb{K}$ tels que

$$q(x) = \sum_{i=1}^{n} \alpha_i l_i(x)^2. \tag{9.328}$$

Démonstration. Notre point de départ sont les formules (9.324) pour la forme quadratique. Nous allons faire la preuve par récurrence sur la dimension de l'espace. Si $n = 1$, alors nous avons seulement

$$q(x) = \alpha x^2 \tag{9.329}$$

et donc le théorème est fait avec $l(x) = x$.

Nous supposons que le théorème est prouvé pour tout espace de dimension n. Une forme quadratique pour un espace de dimension $n+1$ s'écrit

$$q(x) = \sum_{i=1}^{n+1} m_{ii}x_i^2 + 2\sum_{1\leqslant i<j\leqslant n+1} m_{ij}x_ix_j. \tag{9.330}$$

Vu que q est non nulle, un des m_{ij} est non nul. Nous allons diviser en plusieurs cas.

— $m_{11} \neq 0$
— $m_{kk} \neq 0$ avec $k \neq 1$
— $m_{12} \neq 0$ et $m_{ii} = 0$ pour tout i.
— $m_{kl} \neq 0$ avec $(k,l) \neq (1,2)$ et $m_{ii} = 0$ pour tout i.

Ces cas ne sont pas exclusifs, mais ils couvrent toutes les possibilités.

Si $m_{11} \neq 0$ Nous écrivons q sous la forme

$$\begin{aligned}
q(x) &= m_{11}x_1^2 + \sum_{i=2}^{n+1} m_{ii}x_i^2 + 2\sum_{i=1}^{n}\Big(\sum_{j=i+1}^{n+1} m_{ij}x_ix_j\Big) && (9.331a)\\
&= m_{11}x_1^2 + \sum_{i=2}^{n+1} m_{ii}x_i^2 + 2\sum_{j=2}^{n+1} m_{1j}x_1x_j + 2\sum_{i=2}^{n}\sum_{j=i+1}^{n+1}(m_{ij}x_ix_j) && (9.331b)\\
&= m_{11}x_1^2 + 2x_1\sum_{j=2}^{n+1} m_{1j}x_k + R(x_2,\dots,x_{n+1}) && (9.331c)\\
&= m_{11}\left(x_1^2 + 2x_1\sum_{j=2}^{n+1}\frac{m_{1j}}{m_{11}}x_j\right) + R(x_2,\dots,x_{n+1}) && (9.331d)\\
&= m_{11}\big(x_1^2 + 2x_1 f(x_2,\dots,x_{n+1})\big) + R(x_2,\dots,x_{n+1}) && (9.331e)\\
&= m_{11}\big(x_1 + f(x_2,\dots,x_{n+1})\big)^2 - f(x_2,\dots,x_{n+1}) + R(x_2,\dots,x_{n+1}) && (9.331f)
\end{aligned}$$

56. Proposition 4.118.
57. Définition 4.96.

où

— R est une forme quadratique de $n-1$ variables ;

— nous avons noté $f(x_2,\dots,x_{n+1}) = \sum_{j=2}^{n+1} \frac{m_{1j}}{m_{11}} x_j$.

Maintenant, toute la partie $-f(x_2,\dots,x_{n+1})^2 + R(x_2,\dots,x_{n+1})$ est une forme quadratique de n variables. Par hypothèse de récurrence, il existe des coefficients α_i et des formes linéairement indépendantes sur $\mathbb{K}^n$ $l'_i(x_2,\dots,x_{n+1})$ telles que

$$-f(x_2,\dots,x_{n+1})^2 + R(x_2,\dots,x_{n+1}) = \sum_{i=2}^{n+1} \alpha_i l'_i(x_2,\dots,x_{n+1})^2. \tag{9.332}$$

En posant ensuite $l_j(x_1,\dots,x_{n+1}) = l'_j(x_2,\dots,x_{n+1})$, ainsi que $l_1(x_1,\dots,x_{n+1}) = x_1 + f(x_2,\dots,x_{n+1})$, nous avons

$$q(x) = m_{11} l_1(x)^2 + \sum_{j=2}^{n+1} \alpha_j l_j(x)^2. \tag{9.333}$$

Si $m_{kk} \neq 0$ avec $k \neq 1$ Nous nommons k le plus petit entier pour lequel $m_{kk} \neq 0$, et nous supposons que $k \neq 1$, parce que nous avons déjà couvert ce cas. Dans ce cas, nous avons

$$q(x) = m_{kk} x_k^2 + \sum_{j=k+1}^{n+1} m_{jj} x_j^2 + 2 \sum_{i=1}^{n} \Big(\sum_{j=i+1}^{n+1} m_{ij} x_i x_j \Big), \tag{9.334}$$

et tout tourne comme dans le premier cas.

$m_{ii} = 0$ pour tout i et $m_{12} \neq 0$ Nous écrivons q en séparant les termes m_{1k} :

$$\begin{aligned}
q(x) &= 2 \sum_{1 \leqslant i < j \leqslant n+1} m_{ij} x_i x_j && (9.335a)\\
&= 2m_{12} x_1 x_2 + 2 \sum_{2 \leqslant j \leqslant n+1} m_{1j} x_1 x_j + 2 \sum_{2 \leqslant i < j \leqslant n+1} m_{ij} x_i x_j && (9.335b)\\
&= 2m_{12} x_1 x_2 + 2x_1 \sum_{2 \leqslant j \leqslant n+1} m_{1j} x_j + 2 \sum_{3 \leqslant j \leqslant n+1} m_{2j} x_2 x_j + 2 \sum_{3 \leqslant i < j \leqslant n+1} m_{ij} x_i x_j && (9.335c)\\
&= 2m_{12} x_1 x_2 + x_1 f(x_2,\dots,x_{n+1}) + x_2 g(x_3,\dots,x_{n+1}) + T(x_3,\dots,x_{n+1}) && (9.335d)
\end{aligned}$$

où f et g sont linéaires et T est multilinéaire.

À ce moment, nous tentons de factoriser toute la partie concernant x_1 et x_2. L'idée est d'utiliser ceci :

$$(x_1 + g)(x_2 + f) = x_1 x_2 + x_1 f + x_2 g + fg, \tag{9.336}$$

mais en mettant les bons coefficients pour reproduire ce que nous avons dans (9.335d) :

$$(2m_{12} + 2g)(x_1 + \frac{f}{m_{12}}) - \frac{2fg}{m_{12}} = 2m_{12} x_1 x_2 + 2x_1 f + 2x_2 g. \tag{9.337}$$

Cela pour dire que

$$q(x) = 2(m_{12} x_1 + g)(x_2 + \frac{f}{m_{12}}) - \frac{2fg}{m_{12}} + T \tag{9.338}$$

où $-2fg/m_{12} + T$ est une forme quadratique de $x_3,\dots,x_{n+1}$, c'est à dire de $n-1$ variables. L'hypothèse de récurrence nous donne des formes linéaires $(l_i)_{i=3,\dots,n+1}$ telles que

$$\frac{2fg}{m_{12}} + T = \sum_{i=3}^{n+1} \alpha_i l_i(x)^2. \tag{9.339}$$

Nous pouvons donc déjà écrire

$$q(x) = 2l'_1(x) l'_2(x) + \sum_{i=3}^{n+1} \alpha_i l_i(x)^2 \tag{9.340}$$

où

— Les forme l_i avec $i \geqslant 3$ ne dépendent pas de x_1 et x_2, et sont donc indépendantes de l_1 et l_2.
— La forme l_1' ne dépend pas de x_2,
— La forme l_2' ne dépend pas de x_1.

Ce sont donc $n+1$ formes linéaires indépendantes. Le seul problème résiduel est que les formes l_1' et l_2' arrivent en produit l'une de l'autre. Nous en définissons donc deux de plus :

$$\begin{aligned} l_1(x) &= \frac{1}{2}(l_1' + l_2') \\ l_2(x) &= \frac{1}{2}(l_1' - l_2'), \end{aligned} \tag{9.341}$$

qui sont linéairement indépendantes l'une de l'autre et indépendantes des l_i ($i \geqslant 3$). Au final,

$$q(x) = l_1(x)^2 + l_2(x)^2 + \sum_{i=3}^{n+1} \alpha_i l_i(x)^2. \tag{9.342}$$

Si $m_{ii} = 0$ et $m_{12} = 0$ et $m_{kl} \neq 0$ avec $k < l$ Nous considérons la permutation

$$\begin{aligned} \sigma\colon \{1, \dots, n+1\} &\to \{1, \dots, n+1\} \\ i &\mapsto \begin{cases} 1 & \text{si } i = k \\ 2 & \text{si } i = l \\ k & \text{si } i = 1 \\ l & \text{si } i = 2 \\ i & \text{sinon,} \end{cases} \end{aligned} \tag{9.343}$$

c'est à dire que σ permute 1 et k ainsi que 2 et l. Ensuite nous posons

$$\begin{aligned} s\colon \mathbb{R}^{n+1} &\to \mathbb{R}^{n+1} \\ e_i &\mapsto e_{\sigma(i)}. \end{aligned} \tag{9.344}$$

Nous allons un peu considérer $q \circ s$, pour changer :

$$(q \circ s)(x) = \sum_{i,j} m_{ij} s(x)_i s(x)_j = \sum_{ij} x_{\sigma(i)} x_{\sigma(j)}. \tag{9.345}$$

parce que $s(x)_i = x_{\sigma(i)}$.
Utilisons un petit abus de notation pour considérer

$$\begin{aligned} \sigma\colon \{1, \dots, n+1\} \times \{1, \dots, n+1\} &\to \{1, \dots, n+1\} \times \{1, \dots, n+1\} \\ (i, j) &\mapsto \big(\sigma(i), \sigma(j)\big). \end{aligned} \tag{9.346}$$

Cela est une bijection ; nous pouvons utiliser le lemme 1.90 pour permuter les termes dans (9.345) :

$$(q \circ s)(x) = \sum_{ij} m_{\sigma(i)\sigma(j)} x_{\sigma\sigma(i)} x_{\sigma\sigma(j)} \tag{9.347a}$$

$$= \sum_{ij} a_{ij} x_i x_j \tag{9.347b}$$

où nous avons posé $a_{ij} = m_{\sigma(i)\sigma(j)}$ et utilisé le fait que $\sigma = \sigma^{-1}$. Le point intéressant de l'histoire est que dans (9.347b), $a_{12} = m_{kl} \neq 0$. La forme $q \circ s$ est donc dans le cas déjà traité et il existe des formes linéaires l_i' telles que

$$(q \circ s)(x) = \sum_{i=1}^{n+1} \alpha_i l_i'(x)^2. \tag{9.348}$$

En évaluant cela en $s(x)$, et en tenant compte de $s = s^{-1}$, nous trouvons

$$q(x) = \sum_i \alpha_i (l_i \circ s)(x)^2, \tag{9.349}$$

de telle sorte que $l_i = l'_i \circ s$ soit la réponse à notre théorème.

□

9.8.5 Orthogonalité

Proposition 9.172 ([189]).
Soient un espace vectoriel $(E, \mathbb{K})$ et une forme quadratique[58] *q. Une base de E est q-orthogonale*[59] *si et seulement si la matrice de q dans cette base est diagonale.*

Démonstration. La matrice de q est donnée par $Q_{ij} = b(e_i, e_j)$. Donc oui, cette matrice est diagonale si et seulement si les e_i sont orthogonaux. □

Proposition 9.173.
Soit une forme quadratique q. Si une base (e_i) de E est q-orthogonale, alors $\mathcal{B} = \{e_i$ tel que $q(e_i) = 0\}$ est une base de $\ker(q)$.

Démonstration. Nous considérons un vecteur de base e_j, et nous montrons que $q(e_j) = 0$ si et seulement si $e_j \in \ker(q)$. Nous savons par la proposition 9.172 que la matrice de q dans la base (e_i) est diagonale et que les éléments diagonaux sont les $q(e_i)$. Soit $K = \{i$ tel que $q_(e_i) = 0\}$.

<u>$\text{Span}\{e_i\}_{i\in K} \subset \ker(q)$</u> Si $x = \sum_{i\in K} x_i e_i$, alors

$$q(x) = b(x, x) = \sum_{i,j\in K} |x_i|^2 b(e_i, e_j) = \sum_{i,j\in K} |x_i|^2 \delta_{ij} q(e_i) = 0 \tag{9.350}$$

parce que $q(e_i) = 0$ dès que $i \in K$.

$\ker(q) \subset \text{Span}\{e_i\}_{i\in K}$ Soit $x \in \ker(q)$ et écrivons-le sous la forme $x = \sum_{i=1}^n x_i e_i$. Nous avons

$$0 = q(x) = \sum_i |x_i|^2 q(e_i). \tag{9.351}$$

Mais $|x_i|^2 \geqslant 0$ et $q(e_i) \geqslant 0$, donc si $q(e_i) \neq 0$, alors $x_i = 0$. Donc les seules composantes non nulles de x sont celles sur lesquelles q s'annule. En d'autres termes $x = \sum_i x_i e_i \in \text{Span}\{e_i\}_{i\in K}$.

□

Théorème 9.174 ([189, 1]).
Toute forme quadratique sur un espace vectoriel de dimension finie admet une base formée de vecteurs 2 à 2 orthogonaux (pour la forme considérée).

Démonstration. Nous considérons la base $\{l_i\}$ de E^* donnée par la réduction de Gauss (théorème 9.171). La forme quadratique q s'écrit

$$q(x) = \sum_{i=1}^n \alpha_i l_i(x)^2. \tag{9.352}$$

La base préduale[60] $\{e_i\}$ de $\{l_i\}$ répond aux conditions. Pour le vérifier, nous considérons la forme bilinéaire associée à q par l'identité de polarisation 9.161 :

$$b(e_i, e_j) = \frac{1}{2}\big(q(e_i) + q(e_j) - q(e_i - e_j)\big). \tag{9.353}$$

58. Définition 9.160.
59. Définition 9.163.
60. Définition, existence, unicité dans la proposition 4.133.

Vu que $l_k(e_i) = \delta_{ki}$, nous avons

$$q(e_i) = \sum_{k=1}^{n} \alpha_k l_k(e_i)^2 = \alpha_i. \tag{9.354}$$

En utilisant la linéarité,

$$\begin{aligned} q(e_i - e_j) &= \sum_k \alpha_k l_k(e_i - e_j)^2 && (9.355a)\\ &= \sum_k \alpha_k (\delta_{ki} - \delta_{kj})^2 && (9.355b)\\ &= \sum_k \alpha_k (\delta_{ki} + \delta_{kj} - 2\delta_{ki}\delta_{ki}) && (9.355c)\\ &= \alpha_i + \alpha_j - 2\delta_{ij}\alpha_i. && (9.355d) \end{aligned}$$

Donc

$$b(e_i, e_j) = \delta_{ij}\alpha_i. \tag{9.356}$$

Les vecteurs $\{e_i\}$ sont donc bien deux à deux q-orthogonaux. □

Notons qu'en l'absence de notion de racine carrée sur $\mathbb{K}$, il n'est pas possible de considérer $\sqrt{\alpha_i}$ et donc de base q-orthonormée.

9.8.6 Diagonalisation

Le théorème 9.174 a déjà donné une base orthogonale pour toute forme quadratique sur un espace vectoriel $(E, \mathbb{K})$ de dimension finie. Dans le cas de $\mathbb{R}^n$, nous pouvons en donner une preuve basée sur le théorème spectral, c'est la proposition 9.175.

Proposition 9.175.
Soit une forme bilinéaire symétrique b sur un $\mathbb{R}^n$. Il existe une matrice orthogonale Q telle que

(1) $D = Q^t b Q$ est diagonale

(2) $D(x, y) = b(Qx, Qy)$ pour tout $x, y \in E$.

Il existe une base $(f_i)_{i=1,\dots,n}$ qui est b-orthogonale.

Dans cet énoncé, nous mélangeons sans vergogne les formes et les matrices, en supposant qu'une base soit fixée[61]*. Par exemple*

$$D(x, y) = \sum_{ij} D_{ij} x_i y_j. \tag{9.357}$$

Démonstration. Pour la matrice diagonale, c'est le théorème spectral 9.150(2) qui joue parce que la matrice d'une forme bilinéaire symétrique est symétrique (c'est vu de la définition (9.320)).

Pour le reste c'est un calcul :

$$\begin{aligned} D(x, y) &= \sum_{ijkl} Q^t_{ik} b_{kl} Q_{lj} x_i y_j && (9.358a)\\ &= \sum_{ijkl} b_{kl} (Q_{ki} x_i)(Q_{lj} y_j) && (9.358b)\\ &= \sum_{kl} b_{kl} (Qx)_k (Qy)_l && (9.358c)\\ &= b(Qx, Qy). && (9.358d) \end{aligned}$$

Nous avons utilisé le produit matrice fois vecteur donné par (4.84).

61. Autrement dit, si vous avez en tête d'utiliser cette proposition pour $\mathbb{R}^n$ c'est bon ; mais sinon vous devez choisir une base et considérer toutes les matrices dans cette base.

En ce qui concerne l'existence d'une base b-orthogonale, vu que D est diagonale, nous avons, pour $i \neq j$ que $D(e_i, e_j) = 0$. Donc en posant $f_i = Qe_i$, nous trouvons

$$0 = D(e_i, e_j) = b(Qe_i, Qe_j) = b(f_i, f_j). \tag{9.359}$$

La base $(Qe_i)_{i=1,\dots,n}$ est donc b-orthogonale. □

9.8.7 Isométrie, forme quadratique et bilinéaire

Exemple 9.176.
La forme quadratique $q(x) = x_1^2 + x_2^2$ donne la norme euclidienne. La forme bilinéaire associée est $b(x, y) = x_1y_1 + x_2y_2$, qui est le produit scalaire usuel. △

Il ne faudrait pas déduire trop vite que la formule $\|x\|^2 = q(x)$ donne une norme dès que q est non dégénérée. En effet q peut ne pas être définie positive. La forme $q(x) = x_1^2 - x_2^2$ prend des valeurs positives et négatives. A fortiori $d(x, y) = q(x - y)$ ne donne pas toujours une distance.

Définition 9.177.
*Une **isométrie** pour la forme quadratique q est une application bijective $f\colon V \to V$ telle que*

$$q(x - y) = q\big(f(x) - f(y)\big). \tag{9.360}$$

Dans les cas où q donne une distance, alors c'est une isométrie au sens usuel.

Définition 9.178 (Thème 66)**.**
*Soit un espace vectoriel E muni d'une forme bilinéaire b. Une **isométrie** pour b est une bijection $f\colon E \to E$ telle que*

$$b\big(f(x), f(y)\big) = b(x, y) \tag{9.361}$$

pour tout $x, y \in E$.

Lemme 9.179.
Soient q une forme quadratique et b la forme bilinéaire associée par le lemme 9.162. Une application $f\colon E \to E$ telle que $f(0) = 0$ est une isométrie pour b si et seulement si elle est une isométrie pour q.

Démonstration. Pour une application bijective $f\colon E \to E$ telle que $f(0) = 0$, nous devons prouver l'équivalence des propriétés suivantes :

(1) $b\big(f(x), f(y)\big) = b(x, y)$ pour tout $x, y \in E$;
(2) $q\big(f(x) - f(y)\big) = q(x - y)$ pour tout $x, y \in E$.

Dans le sens direct, en posant $x = y$ nous trouvons tout de suite $q(f(x)) = q(x)$; ensuite en utilisant la distributivité de b,

$$\begin{aligned}
q\big(f(x) - f(y)\big) &= b\big(f(x) - f(y), f(x) - f(y)\big) && (9.362a)\\
&= q\big(f(x)\big) - 2b\big(f(x), f(y)\big) + q\big(f(y)\big) && (9.362b)\\
&= q(x) + q(y) - 2b(x, y) && (9.362c)\\
&= q(x - y). && (9.362d)
\end{aligned}$$

Dans l'autre sens, nous commençons par remarquer que l'hypothèse $f(0) = 0$ donne $q(x) = q\big(f(x)\big)$. Ensuite nous utilisons l'identité de polarisation (9.318) :

$$\begin{aligned}
b\big(f(x), f(y)\big) &= \frac{1}{2}\big[q\big(f(x)\big) + q\big(f(y)\big) - q\big(f(x - y)\big)\big] && (9.363a)\\
&= \frac{1}{2}\big[q(x) + q(y) - q(x - y)\big] && (9.363b)\\
&= b(x, y). && (9.363c)
\end{aligned}$$

□

9.9 Signature, théorème de Sylvester

Définition 9.180 (Signature[190]).
Soit une forme quadratique[62] Q *sur un espace vectoriel* E *de dimension finie* n. *L'**indice d'inertie** de* q *est le nombre*

$$q = \max\{\dim(F) \text{ tel que } Q(v) < 0 \,\forall v \in F\backslash\{0\}\}. \tag{9.364}$$

Nous définissons aussi

$$p = \max\{\dim(G) \text{ tel que } Q(v) > 0 \,\forall v \in G\backslash\{0\}\}. \tag{9.365}$$

Le couple (p, q) *est la **signature** de* Q.

Définition 9.181 (Rang d'une forme quadratique).
Si $Q\colon E \times E \to \mathbb{K}$ *est une forme quadratique, nous considérons l'application*

$$\begin{aligned} f_Q\colon E &\to E^* \\ x &\mapsto \big[y \mapsto B(x, y)\big]. \end{aligned} \tag{9.366}$$

*Le **rang** de* Q *est le rang de l'application linéaire* f_Q.

Proposition 9.182.
Le rang d'une forme quadratique est le rang de sa matrice dans n'importe quelle base.

Démonstration. Nous considérons une forme quadratique Q sur l'espace vectoriel E. Sa trace est, par définition, la trace de l'application linéaire f_Q de la définition 9.181. Or cette dernière trace ne dépend pas des bases choisies sur E et E^*. Nous la calculons donc maintenant.

Soit une base $\{e_i\}$ de E ainsi que sa base duale $\{e_i^*\}$ de E^*. Si $v = \sum_k v_k e_k \in E$, alors

$$f_Q(e_i)v = \sum_k v_k B(e_i, e_k) = \sum_k Q_{ik} v_k \tag{9.367}$$

où nous avons noté B la forme bilinéaire associée à Q et utilisé la définition 9.164 de la matrice associée à la forme quadratique Q. Nous avons donc $f_Q(ei) = \sum_k Q_{ik} e_k^*$ ou encore

$$f_Q(e_i)_k = Q_{ik}, \tag{9.368}$$

ce qui signifie, par (1) que la matrice associée à f_Q est la matrice Q^t.

Le rang de f_Q est donc celui de Q^t, qui est le même que celui de la matrice Q (ici, nous avons noté Q la matrice de la forme quadratique Q). Le rang de f_Q est celui de sa matrice par la proposition 4.106. □

Lemme 9.183 ([190]).
Soient une forme quadratique Q *ainsi que deux bases* Q*-orthogonales* $\{e_1, \dots, e_n\}$ *et* $\{e'_1, \dots, e'_n\}$. *Nous posons*

$$r = \mathrm{Card}\{e_i \text{ tel que } q(e_i) > 0\} \tag{9.369a}$$
$$r' = \mathrm{Card}\{e_i \text{ tel que } q(e'_i) > 0\} \tag{9.369b}$$
$$s = \mathrm{Card}\{e_i \text{ tel que } q(e_i) < 0\} \tag{9.369c}$$
$$s' = \mathrm{Card}\{e_i \text{ tel que } q(e'_i) < 0\} \tag{9.369d}$$

Alors $r = r'$ *et* $s = s'$.

Démonstration. Nous posons

$$I = \{i \text{ tel que } Q(e_i) > 0\} \tag{9.370a}$$
$$J = \{j \text{ tel que } Q(e_j) < 0\} \tag{9.370b}$$

62. Définition 9.160.

Nous commençons par prouver que $\{e_i\}_{i\in I} \cup \{e'_j\}_{j\in J}$ est libre. Supposons pour cela que

$$\sum_{i\in I} x_i e_i + \sum_{j\in J} y_j e'_j = 0, \tag{9.371}$$

et posons $z = \sum_{i\in I} x_i e_i$. Nous avons

$$Q(z) = \sum_{i\in I} x_i^2 Q(e_i) \geqslant 0. \tag{9.372}$$

Mais nous avons aussi $z = -\sum_{j\in J} y_j e'_j$, donc

$$Q(z) = \sum_{j\in J} y_j^2 Q(e'_j) \leqslant 0. \tag{9.373}$$

Donc $Q(z) = 0$. Vu (9.372), et le fait que $Q(e_i) > 0$, avoir $Q(z) = 0$ impose $x_i = 0$ pour tout i. La relation (9.373) nous donne aussi immédiatement que les y_j sont nuls. Donc la partie $\{e_i\}_{i\in I} \cup \{e'_j\}_{j\in J}$ est libre.

Le lemme 4.9 nous indique qu'une partie libre est toujours de cardinal plus petit ou égal à la dimension de l'espace[63]. Tout ça pour dire que

$$\underbrace{\mathrm{Card}(I)}_{=r} + \underbrace{\mathrm{Card}(J)}_{=n-r'} \leqslant n, \tag{9.374}$$

et donc $r \leqslant r'$.

Le même raisonnement, en partant de $I = \{i \text{ tel que } Q(e_i) \leqslant 0\}$ et de $J = \{j \text{ tel que } Q(e'_j) > 0\}$, prouve que $r' \leqslant r$.

La preuve de $s = s'$ est du même tonneau. □

Lemme 9.184 ([191]).
Soit une forme quadratique Q sur E. Si F est un sous-espace de E, alors

$$\dim(F) + \dim(F^\perp) \geqslant n \tag{9.375}$$

où $F^\perp$ est l'orthogonal par rapport à Q.

Démonstration. Nous posons $p = \dim(F)$. Nous considérons une base $\{f_i\}_{i=1,\dots,n}$ de E telle que $\{f_i\}_{i=1,\dots,p}$ est une base de F[64]. Nous posons

$$\begin{aligned} \phi\colon E &\to F \\ x &\mapsto \sum_{i=1}^{p} B(x, f_i) f_i \end{aligned} \tag{9.376}$$

où B est la forme bilinéaire associée à Q. Ce ϕ est une application linéaire à qui nous appliquons le théorème du rang (4.44) :

$$\dim(E) = \mathrm{rang}(\phi) + \dim\big(\ker(\phi)\big). \tag{9.377}$$

Mais vu que l'image de ϕ est dans F, nous avons $\mathrm{rang}(\phi) \leqslant \dim(F)$. De plus, $\ker(\phi) = F^\perp$. Donc (9.377) devient

$$\dim(E) \leqslant \dim(F) + \dim(F^\perp). \tag{9.378}$$

□

63. Ici nous utilisons l'hypothèse que V est de dimension finie.
64. Théorème de la base incomplète, 4.21.

Lemme 9.185 ([191]).
Soit un espace vectoriel E de dimension finie et un sous-espace F sur lequel la forme quadratique Q est strictement définie positive ou négative. Alors

$$E = F \oplus F^{\perp}. \tag{9.379}$$

Démonstration. D'abord nous montrons que $F \cap F^{\perp} = \{0\}$. Si $v \neq 0$ est dans F, alors $Q(v) > 0$, et donc v n'est pas dans $F^{\perp}$. Donc $F \cap F^{\perp} \subset \{0\}$. L'inclusion inverse est immédiate.

Nous avons vu dans le lemme 9.184 que

$$\dim(E) \leqslant \dim(F) + \dim(F^{\perp}). \tag{9.380}$$

Vu que F et $F^{\perp}$ n'ont pas d'intersection autre que $\{0\}$, nous avons

$$\dim(E) \geqslant \dim(F \oplus F^{\perp}) = \dim(F) + \dim(F^{\perp}) \geqslant \dim(E). \tag{9.381}$$

Toutes ces inégalités sont donc des égalités et $\dim(E) = \dim(F) + \dim(F^{\perp})$. □

Théorème 9.186 (de Sylvester[190]).
Soit Q une forme quadratique réelle de signature[65] *(p, q). Alors pour toute base Q-orthogonale $\{e_i\}$ de $\mathbb{R}^{p+q}$ nous avons les propriétés suivantes.*

(1) Les nombres p et q sont donnée par

$$p = \mathrm{Card}\{i \text{ tel que } Q(e_i) > 0\} \tag{9.382a}$$
$$q = \mathrm{Card}\{i \text{ tel que } Q(e_i) < 0\}. \tag{9.382b}$$

(2) Si A est la matrice de Q dans une base, alors il existe une matrice inversible P telle que

$$P^t A P = \begin{pmatrix} -\mathbb{1}_q & & \\ & \mathbb{1}_p & \\ & & 0 \end{pmatrix}. \tag{9.383}$$

(3) Le rang de Q est $p + q$.

Démonstration. Soit F un sous-espace de dimension maximale q sur lequel Q est définie négative. Le fait que la dimension de F soit q est la définition 9.180 de la signature. Nous notons $F^{\perp}$ sont Q-orthogonal, c'est à dire que

$$F^{\perp} = \{v \in E \text{ tel que } B(v, x) = 0 \,\forall x \in F\}. \tag{9.384}$$

Le lemme 9.185 nous assure que $E = F \oplus F^{\perp}$.

Le théorème 9.174 sur l'existence de bases Q-orthogonales nous permet de considérer une base Q-orthogonale de F et une de $F^{\perp}$. En réunissant les deux, nous avons une base de E. Nous la notons $\{f_1, \dots, f_n\}$ avec

— La partie $\{f_1, \dots, f_q\}$ est une base de F,
— La partie $\{f_{q+1}, \dots, f_n\}$ est une base de $F^{\perp}$,
— Remarquez cependant qu'il n'est pas dit que $n = q + p$.

Notons que pour $i > q$, nous avons $Q(f_i) \geqslant 0$, sinon la maximalité de F serait contredite par $\mathrm{Span}\{f_1, \dots, f_q, f_i\}$.

Cela prouve que

$$\mathrm{Card}\{i \text{ tel que } Q(f_i) > 0\} = p. \tag{9.385}$$

Le lemme 9.183 nous dit alors que

$$\mathrm{Card}\{i \text{ tel que } Q(e_i) > 0\} = \mathrm{Card}\{i \text{ tel que } Q(f_i) > 0\} = p. \tag{9.386}$$

65. Définition 9.180.

C'est l'égalité (9.382b). L'égalité (9.382a) se prouve de la même façon, en prenant F maximal pour la propriété que Q y est strictement définie positive.

Le point (1) est prouvé.

Dans une base Q-orthogonale, la matrice de Q est diagonale, et contient sur la diagonale les valeurs de $Q(e_i)$. Parmi celles-ci, on en a p strictement positives et q strictement négatives. Les $n-p-q$ autres sont nulles. Vu que Q est à valeur réelle, nous avons une notion de racine carré, et nous pouvons considérer $e_i/\sqrt{|Q(e_i)|}$ au lieu de e_i. De cette façon, $Q(e_i)$ est normalisé. Avec ça, la matrice de Q est

$$D = \begin{pmatrix} \mathbb{1}_p & & \\ & -\mathbb{1}_q & \\ & & 0 \end{pmatrix}. \tag{9.387}$$

Nous venons de prouver qu'il existe une base $\{e_i\}$ dans laquelle la matrice de Q est (9.387). Si A est la matrice de Q dans une base quelconque $\{f_i\}$ et si P est la matrice de changement de base $f_j = \sum_i P_{ij} e_i$, la proposition 9.169 donne $D = P^t A P$.

Le point (2) est prouvé.

Pour (3), la proposition 9.182 nous permet de calculer le rang de Q par le rang de sa matrice dans n'importe quelle base. Nous choisissons la base qui donne la matrice (9.387). Le rang est alors bien $p+q$. □

Définition 9.187 (Équivalence de forme quadratique[187])**.**
*Deux formes quadratiques Q et Q' sont **équivalentes** si il existe une application linéaire inversible ϕ telle que $Q' = Q \circ \phi$.*

Proposition 9.188 ([190])**.**
Deux formes quadratiques sont équivalentes[66] *si et seulement si elles ont même signature.*

9.9.1 Invariance de la trace

Proposition 9.189 ([1])**.**
Soit une application linéaire f. Si la matrice de f dans une base est A et est B dans une autre base, alors

$$\mathrm{Tr}(A) = \mathrm{Tr}(B). \tag{9.388}$$

Démonstration. Les matrices A et B sont liées par la proposition 4.118 : $B = Q^{-1}AQ$ où Q est la matrice qui lie les vecteurs des deux bases. L'invariance cyclique de la trace donnée en le lemme 4.62 implique que

$$\mathrm{Tr}(B) = \mathrm{Tr}(Q^{-1}AQ) = \mathrm{Tr}(QQ^{-1}A) = \mathrm{Tr}(A). \tag{9.389}$$

□

9.10 Fonctions

Soient $(V, \|.\|_V)$ et $(W, \|.\|_W)$ deux espaces vectoriels normés, et une fonction f de V dans W. Il est maintenant facile de définir les notions de limites et de continuité pour de telles fonctions en copiant les définitions données pour les fonctions de $\mathbb{R}$ dans $\mathbb{R}$ en changeant simplement les valeurs absolues par les normes sur V et W.

La caractérisation suivante est un recopiage de la définition 7.71 lorsque la topologie est donnée par des boules.

Proposition 9.190.
Soit $f\colon V \to W$ une fonction de domaine $\mathrm{Dom}(f) \subset V$ et soit a un point d'accumulation de

66. Définition 9.187.

$\mathrm{Dom}(f)$. *La fonction f admet une limite en $a \in V$ si et seulement s'il existe un élément $\ell \in W$ tel que pour tout $\varepsilon > 0$, il existe un $\delta > 0$ tel que pour tout $x \in \mathrm{Dom}(f)$,*

$$0 < \|x - a\|_V < \delta \Rightarrow \|f(x) - \ell\|_W < \varepsilon. \tag{9.390}$$

Dans ce cas, nous écrivons $\lim_{x\to a} f(x) = \ell$ et nous disons que ℓ est la ***limite*** *de f lorsque x tend vers a.*

Remarque 9.191.
Le fait que nous limitions la formule (9.390) aux x dans le domaine de f n'est pas anodin. Considérons la fonction $f(x) = \sqrt{x^2 - 4}$, de domaine $|x| \geqslant 2$. Nous avons

$$\lim_{x\to 2} \sqrt{x^2 - 4} = 0. \tag{9.391}$$

Nous ne pouvons pas dire que cette limite n'existe pas en justifiant que la limite à gauche n'existe pas. Les points $x < 2$ sont hors du domaine de f et ne comptent dons pas dans l'appréciation de l'existence de la limite.

Vous verrez plus tard que ceci provient de la topologie induite de $\mathbb{R}$ sur l'ensemble $[2, \infty[$.

9.11 Sous espaces caractéristiques

Lorsqu'un opérateur n'est pas diagonalisable, les valeurs propres jouent quand même un rôle important.

Définition 9.192.
Soit E un $\mathbb{K}$-espace vectoriel $f \in \mathrm{End}(E)$. Pour $\lambda \in \mathbb{K}$ nous définissons

$$F_\lambda(f) = \{v \in E \text{ tel que } (f - \lambda\mathbb{1})^n v = 0, n \in \mathbb{N}\} \tag{9.392}$$

et nous appelons ça un ***sous-espace caractéristique*** *de f.*

L'espace $F_\lambda(f)$ est l'ensemble de nilpotence de l'opérateur $f - \lambda\mathbb{1}$ et

Lemme 9.193.
L'ensemble $F_\lambda(f)$ est non vide si et seulement si λ est une valeur propre de f. L'espace $F_\lambda(f)$ est invariant sous f.

Démonstration. Si $F_\lambda(f)$ est non vide, nous considérons $v \in F_\lambda(f)$ et n le plus petit entier non nul tel que $(f - \lambda)^n v = 0$. Alors $(f - \lambda)^{n-1} v$ est un vecteur propre de f pour la valeur propre λ. Inversement si v est une valeur propre de f pour la valeur propre λ, alors $v \in F_\lambda(f)$.

En ce qui concerne l'invariance, remarquons que f commute avec $f - \lambda\mathbb{1}$. Si $x \in F_\lambda(f)$ il existe n tel que $(f - \lambda\mathbb{1})^n x = 0$. Nous avons aussi

$$(f - \lambda\mathbb{1})^n f(x) = f\big((f - \lambda\mathbb{1})^n x\big) = 0, \tag{9.393}$$

par conséquent $f(x) \in F_\lambda(f)$. □

Remarque 9.194.
Toute matrice sur $\mathbb{C}$ n'est pas diagonalisable : nous en avons déjà donné une exemple simple en 9.243. Nous en voyons maintenant un moins simple. Considérons en effet l'endomorphisme f donné par la matrice

$$\begin{pmatrix} a & \alpha & \beta \\ 0 & a & \gamma \\ 0 & 0 & b \end{pmatrix} \tag{9.394}$$

où $a \neq b$, $\alpha \neq 0$, β et γ sont des nombres complexes quelconques. Son polynôme caractéristique est

$$\chi_f(\lambda) = (a - \lambda)^2 (b - \lambda), \tag{9.395}$$

et les valeurs propres sont donc a et b. Nous trouvons les vecteurs propres pour la valeur a en résolvant

$$\begin{pmatrix} a & \alpha & \beta \\ 0 & a & \gamma \\ 0 & 0 & b \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ax \\ ay \\ az \end{pmatrix}. \tag{9.396}$$

L'espace propre $E_a(f)$ est réduit à une seule dimension générée par $(1, 0, 0)$. De la même façon l'espace propre correspondant à la valeur propre b est donné par

$$\begin{pmatrix} \frac{1}{b-a}\left(\beta + \frac{\alpha\gamma}{b-a}\right) \\ \frac{\gamma}{b-a} \\ 1 \end{pmatrix}. \tag{9.397}$$

Il n'y a donc pas trois vecteurs propres linéairement indépendants, et l'opérateur f n'est pas diagonalisable.

Par contre nous pouvons voir que

$$(f - \alpha\mathbb{1})^2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a & \alpha & \beta \\ 0 & a & \gamma \\ 0 & 0 & b \end{pmatrix} \begin{pmatrix} \alpha \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} a\alpha \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \tag{9.398}$$

de telle sorte que le vecteur $(0, 1, 0)$ soit également dans l'espace caractéristique $F_a(f)$.

Dans cet exemple, la multiplicité algébrique de la racine a du polynôme caractéristique vaut 2 tandis que sa multiplicité géométrique vaut seulement 1.

9.11.1 Théorèmes de décomposition

Théorème 9.195 (Théorème spectral, décomposition primaire)**.**
Soit E espace vectoriel de dimension finie sur le corps algébriquement clos $\mathbb{K}$ et $f \in \mathrm{End}(E)$. Alors

$$E = F_{\lambda_1}(f) \oplus \ldots \oplus F_{\lambda_k}(f) \tag{9.399}$$

où la somme est sur les valeurs propres distinctes de f.

Les projecteurs sur les espaces caractéristique forment un système complet et orthogonal.

Démonstration. Soit P le polynôme caractéristique de f et une décomposition

$$P = (f - \lambda_1)^{\alpha_1} \ldots (f - \lambda_r)^{\alpha_r} \tag{9.400}$$

en facteurs irréductibles. La le théorème de noyaux (9.88) nous avons

$$E = \ker(f - \lambda_1)^{\alpha_1} \oplus \ldots \oplus \ker(f - \lambda_r)^{\alpha_r}. \tag{9.401}$$

Les projecteurs sont des polynômes en f et forment un système orthogonal. Il nous reste à prouver que $\ker(f - \lambda_i)^{\alpha_i} = F_{\lambda_i}(f)$. L'inclusion

$$\ker(f - \lambda_i)^{\alpha_i} \subset F_{\lambda_i}(f) \tag{9.402}$$

est évidente. Nous devons montrer l'inclusion inverse. Prouvons que la somme des $F_{\lambda_i}(f)$ est directe. Si $v \in F_{\lambda_i}(f) \cap F_{\lambda_j}(f)$, alors il existe $v_1 = (f - \lambda_i)^n v \neq 0$ avec $(f - \lambda_i)v_1 = 0$. Étant donné que $(f - \lambda_i)$ commute avec $(f - \lambda_j)$, ce v_1 est encore dans $F_{\lambda_j}(f)$ et par conséquent il existe $w = (f - \lambda_j)^m v_1$ non nul tel que

$$\begin{cases} (f - \lambda_i)w = 0 & \text{(9.403a)} \\ (f - \lambda_j)w = 0. & \text{(9.403b)} \end{cases}$$

Ce w serait donc un vecteur propre simultané pour les valeurs propres λ_i et λ_j, ce qui est impossible parce que les espaces propres sont linéairement indépendants. Les espaces F_{λ_i} sont donc en somme directe et

$$\sum_i \dim F_{\lambda_i}(f) \leqslant \dim E. \tag{9.404}$$

En tenant compte de l'inclusion (9.402) nous avons même

$$\dim E = \sum_i \dim \ker(f-\lambda_i)^{\alpha_i} \leqslant \sum_i F_{\lambda_i}(f) \leqslant \dim E. \tag{9.405}$$

Par conséquent nous avons $\dim \ker(f-\lambda_i)^{\alpha_i} = \dim F_{\lambda_i}(f)$ et l'égalité des deux espaces. □

Problèmes et choses à faire

Dans le cas où le corps n'est pas algébriquement clos, il paraît qu'il faut remplacer « diagonalisable » par « semi-simple ».

Si l'espace vectoriel est sur un corps algébriquement clos, alors les endomorphismes semi-simples[67] sont les endomorphismes diagonaux.

Théorème 9.196 (Décomposition de Dunford)**.**
Soit E un espace vectoriel sur le corps algébriquement clos $\mathbb{K}$ et $u \in \mathrm{End}(E)$ un endomorphisme de E.

(1) L'endomorphisme u se décompose de façon unique sous la forme

$$u = s + n \tag{9.406}$$

où s est diagonalisable, n est nilpotent et $[s,n]=0$.

(2) Les endomorphismes s et n sont des polynômes en u et commutent avec u.

(3) Les parties s et n sont données par

$$s = \sum_i \lambda_i p_i \tag{9.407a}$$

$$n = \sum_i (s-\lambda_i \mathbb{1}) p_i \tag{9.407b}$$

où les sommes sont sur les valeurs propres distinctes[68] de f et où $p_i\colon E \to F_{\lambda_i}(u)$ est la projection de E sur $F_{\lambda_i}(u)$.

Démonstration. Le théorème spectral 9.195 nous indique que

$$E = \bigoplus_i F_{\lambda_i}(f). \tag{9.408}$$

Nous considérons l'endomorphisme s de E qui consiste à dilater d'un facteur λ l'espace caractéristique $F_\lambda(f)$:

$$s = \sum_i \lambda_i p_i \tag{9.409}$$

où $p_i\colon E \to F_{\lambda_i}(u)$ est la projection de E sur $F_{\lambda_i}(u)$.

Nous allons prouver que $[s,f]=0$ et $n = f - s$ est nilpotent. Cela impliquera que $[s,n]=0$.

Si $x \in F_\lambda(f)$, alors nous avons $sf(x) = \lambda f(x)$ parce que $f(x) \in F_\lambda(f)$ tandis que $fs(x) = f(\lambda x) = \lambda f(x)$. Par conséquent f commute avec s.

Pour montrer que $f - s$ est nilpotent, nous en considérons la restriction

$$f - s\colon F_\lambda(f) \to F_\lambda(f). \tag{9.410}$$

Cet opérateur est égal à $f - \lambda\mathbb{1}$ et est par conséquent nilpotent.

67. Définition 9.102.
68. C'est-à-dire sur les sous-espaces caractéristiques.

Prouvons à présent l'unicité. Soit $u = s' + n'$ une autre décomposition qui satisfait aux conditions : s' est diagonalisable, n' est nilpotent et $[n', s'] = 0$. Commençons par prouver que s' et n' commutent avec u. En multipliant $u = s' + n'$ par s' nous avons

$$s'u = s'^2 + s'n' = s'^2 + n's' = (s' + n')s' = us', \tag{9.411}$$

par conséquent $[u, s'] = 0$. Nous faisons la même chose avec n' pour trouver $[u, n'] = 0$. Notons que pour obtenir ce résultat nous avons utilisé le fait que n' et s' commutent, mais pas leur propriétés de nilpotence et de diagonalisibilité.

Si $s' + n' = s + n$ est une autre décomposition, s' et n' commutent avec u, et par conséquent avec tous les polynômes en u. Ils commutent en particulier avec n et s. Les endomorphismes s et s' sont alors deux endomorphismes diagonalisables qui commutent. Par la proposition 9.131, ils sont simultanément diagonalisables. Dans la base de simultanée diagonalisation, la matrice de l'opérateur $s' - s = n - n'$ est donc diagonale. Mais $n - n'$ est également nilpotent, en effet si A et B sont deux opérateurs nilpotents,

$$(A + B)^n = \sum_{k=0}^{n} \binom{k}{n} A^k B^{n-k}. \tag{9.412}$$

Si n est assez grand, au moins un parmi A^k ou B^{n-k} est nul.

Maintenant que $n - n'$ est diagonal et nilpotent, il est nul et $n = n'$. Nous avons alors immédiatement aussi $s = s'$.

□

9.11.2 Diverses conséquences

Théorème 9.197.
Soit une matrice $A \in \mathbb{M}(n, \mathbb{C})$. On a que la suite $(A^k x)$ tend vers zéro pour tout x si et seulement si $\rho(A) < 1$ où $\rho(A)$ est le rayon spectral de A

Démonstration. Dans le sens direct, il suffit de prendre comme x, un vecteur propre de A. Dans ce cas nous avons $A^k x = \lambda^k x$. Mais $\lambda^k x$ ne tend vers zéro que si $\lambda < 1$. Donc toutes les valeurs propres de A doivent être plus petite que 1 et $\rho(A) < 1$.

Pour l'autre sens nous utilisons la décomposition de Dunford (théorème 9.196) : il existe une matrice inversible P telle que

$$A = P^{-1}(D + N)P \tag{9.413}$$

où D est diagonale, N est nilpotente et $[D, N] = 0$. Étant donné que $D + N$ est triangulaire, son polynôme caractéristique que

$$\chi_{D+N}(\lambda) = \prod_i D_{ii} - \lambda. \tag{9.414}$$

Par similitude, c'est le même polynôme caractéristique que celui de A et nous savons alors que la diagonale de D contient les valeurs propres de A.

Par ailleurs nous avons

$$A^k = P^{-1}(D + N)^k P \tag{9.415a}$$

$$= P^{-1} \sum_{j=0}^{k} \binom{j}{k} D^{j-k} N^j P \tag{9.415b}$$

$$= P^{-1} \sum_{j=0}^{n-1} \binom{j}{k} D^{j-k} N^j P \tag{9.415c}$$

où nous avons utilité le fait que D et N commutent ainsi que $N^{n-1} = 0$ parce que N est nilpotente. Nous utilisons la norme matricielle usuelle, pour laquelle $\|D\| = \rho(D) = \rho(A)$. Nous avons alors

$$\|(D + N)^k\| \leqslant \sum_{j=0}^{k} \binom{j}{k} \rho(D)^{k-j} \|N\|^j. \tag{9.416}$$

Du coup si $\rho(D) < 1$ alors $\|(D+N)^k\| \to 0$ (et c'est même un si et seulement si). □

Une application de la décomposition de Jordan est l'existence d'un logarithme pour les matrices. La proposition suivant va d'une certaine manière donner un logarithme pour les matrices inversibles complexes. Dans le cas des matrices réelles m telles que $\|m - \mathbb{1}\| < 1$, nous donnerons au lemme 14.117 une formule pour le logarithme sous forme d'une série ; ce logarithme sera réel.

9.11.3 Valeurs singulières

Définition 9.198.
*Soit M une matrice $m \times n$ sur $\mathbb{K}$ ($\mathbb{K}$ est $\mathbb{R}$ ou $\mathbb{C}$). Un nombre réel σ est une **valeur singulière** de M s'il existent des vecteurs unitaires $u \in \mathbb{K}^m$, $v \in \mathbb{K}^n$ tels que*

$$Mv = \sigma u \tag{9.417a}$$

$$M^* u = \sigma v. \tag{9.417b}$$

Théorème 9.199 (Décomposition en valeurs singulières)**.**
Soit $M \in \mathbb{M}(m \times n, \mathbb{K})$ où $\mathbb{K} = \mathbb{R}, \mathbb{C}$. Alors M se décompose en

$$M = ADB \tag{9.418}$$

où il existe deux matrices unitaires $A \in \mathrm{U}(m \times m)$, $B \in \mathrm{U}(n \times n)$ et une matrice (pseudo)diagonale $D \in \mathbb{M}(m \times n)$ tels que

(1) $A \in \mathrm{U}(m \times m)$, $B \in \mathrm{U}(n \times n)$ sont deux matrices unitaires ;,
(2) D est (pseudo)diagonale,
(3) les éléments diagonaux de D sont les valeurs singulières de M,
(4) le nombre d'éléments non nuls sur la diagonale de D est le rang[69] de M.

Corolaire 9.200.
Soit $M \in \mathbb{M}(n, \mathbb{C})$. Il existe un isomorphisme $f \colon \mathbb{C}^n \to \mathbb{C}^n$ tel que fM soit autoadjoint.

Démonstration. Si $M = ADB$ est la décomposition de M en valeurs singulières, alors nous pouvons prendre $f = \overline{B}^t A^{-1}$ qui est une matrice inversible. Pour la vérification que ce f répond bien à la question, ne pas oublier que D est réelle, même si M ne l'est pas. □

9.12 Extension du corps de base

Nous avons discuté dans la section 6.4 de ce qui arrive au corps lorsqu'on l'étend. Dans cette sections nous allons étudier ce qui arrive aux applications linéaires entre deux $\mathbb{K}$-espaces vectoriels lorsque nous étendons le corps $\mathbb{K}$ en un corps $\mathbb{L}$.

Soit donc un corps $\mathbb{K}$ et deux $\mathbb{K}$-espaces vectoriels E et F, et entrons dans le vif du sujet[70]. Soit $\mathbb{K}$ un corps (commutatif) et une extension $\mathbb{L}$ de $\mathbb{K}$. Soient E et F, des $\mathbb{K}$-espaces vectoriels de dimension finie.

9.12.1 Extension des applications linéaires

Définition 9.201 ([192])**.**
*L'espace vectoriel obtenu par **extension du corps de base** de E est l'espace vectoriel*

$$E_{\mathbb{L}} = \mathbb{L} \otimes_{\mathbb{K}} E. \tag{9.419}$$

Ce dernier est le quotient $\mathbb{L} \otimes_{\mathbb{K}} E = (\mathbb{L} \times E)/\sim$ par la relation d'équivalence

$$(\lambda, v) \sim \big(a\lambda, \frac{1}{a} v\big) \tag{9.420}$$

pour tout $a \in \mathbb{K}$. Nous noterons $[\lambda, v]$ ou $\lambda \otimes v$ ou encore $\lambda \otimes_{\mathbb{K}} v$ la classe de (λ, v).

69. Définition 4.38.
70. Le sujet étant le corps étendu.

Un élément de $E_{\mathbb{L}}$ est de la forme $\sum_k [\lambda_k, v_k]$ avec $\lambda_k \in \mathbb{L}$ et $v_k \in E$. Si $f\colon E \to F$ est une applications linéaire nous définissons

$$\begin{aligned} f_{\mathbb{L}}\colon E_{\mathbb{L}} &\to F_{\mathbb{L}} \\ [\lambda, v] &\mapsto [\lambda, f(v)]. \end{aligned} \tag{9.421}$$

Remarque 9.202.
Si deux vecteurs de $E_{\mathbb{L}}$ sont linéairement indépendants pour $\mathbb{K}$, ils ne le sont pas spécialement pour $\mathbb{L}$. Par exemple si $\mathbb{C}$ est vu comme $\mathbb{R}$-espace vectoriel, alors $\{1, i\}$ est une partie libre. Mais dans $\mathbb{C}$ vu comme $\mathbb{C}$-espace vectoriel, la partie $\{1, i\}$ n'est pas libre.

Nous définissons aussi l'injection canonique

$$\begin{aligned} \iota\colon E &\to E_{\mathbb{L}} \\ v &\mapsto [1, v]. \end{aligned} \tag{9.422}$$

Proposition 9.203 ([80]).
Injectivité et surjectivité respectées.

(1) L'application $f_{\mathbb{L}}$ est injective si et seulement si f est injective.

(2) L'application $f_{\mathbb{L}}$ est surjective si et seulement si f est surjective.

Démonstration. Supposons pour commencer que $f_{\mathbb{L}}$ est injective. Le diagramme

$$\begin{tikzcd} E \arrow[r, "f"] \arrow[d, "\tau"'] & F \arrow[d, "\tau"] \\ E_{\mathbb{L}} \arrow[r, "f_{\mathbb{L}}"'] & F_{\mathbb{L}} \end{tikzcd} \tag{9.423}$$

est un diagramme commutatif. En effet

$$(\tau \circ f)(v) = [1, f(v)] \tag{9.424}$$

tandis que

$$(f_{\mathbb{L}\circ\tau})(v) = f_{\mathbb{L}}[1, v] = [1, f(v)]. \tag{9.425}$$

Donc si $f(v) = 0$ avec $v \neq 0$ nous aurions $(\tau \circ f)(v) = 0$ et donc aussi $(f_{\mathbb{L}} \circ \tau)(v) = 0$, alors que $\tau(v) \neq 0$ dans $E_{\mathbb{L}}$.

Réciproquement, nous supposons que f est injective et nous prouvons que $f_{\mathbb{L}}$ est injective. Par le lemme 4.53(1), nous savons qu'il existe $g\colon F \to E$ telle que $f \circ g = \mathrm{Id}\,|_F$. Nous en déduisons que $f_{\mathbb{L}} \circ g_{\mathbb{L}} = \mathrm{Id}\,|_{F_{\mathbb{L}}}$ parce que si $[\lambda, v] \in F_{\mathbb{L}}$ alors

$$(f_{\mathbb{L}} \circ g_{\mathbb{L}})[\lambda, v] = f_{\mathbb{L}}[\lambda, g(v)] = [\lambda, (f \circ g)(v)] = [\lambda, v]. \tag{9.426}$$

Notons que g est injective, donc $g_{\mathbb{L}}$ est injective et l'égalité $f_{\mathbb{L}} \circ g_{\mathbb{L}} = \mathrm{Id}\,|_{F_{\mathbb{L}}}$ implique que $f_{\mathbb{L}}$ est également injective. □

Proposition 9.204 ([1, 193]).
Soit $\{e_i\}_{i=1,\dots,p}$ une base de E. Alors $\{1 \otimes e_i\}_i$ est une base de $E_{\mathbb{L}} = \mathbb{L} \otimes_{\mathbb{K}} E$.

Démonstration. L'espace vectoriel E peut être écrit comme somme directe $E = \bigoplus_i \mathbb{K}e_i$. Si $\lambda \in \mathbb{L}$ et $k \in \mathbb{K}$ nous avons

$$\lambda \otimes k e_i = \frac{\lambda}{k} \otimes e_i = \frac{\lambda}{k}(1 \otimes e_i). \tag{9.427}$$

Cela pour introduire que l'application

$$\begin{aligned} \psi\colon \mathbb{L} \otimes_{\mathbb{K}} E &\to \bigoplus_i \mathbb{L}(1 \otimes e_i) \\ \sum_k \lambda_k \otimes v_k &\mapsto \oplus_i \sum_k (\lambda_k v_{ik})(1 \otimes e_i) \end{aligned} \tag{9.428}$$

où $v_k = \sum_i v_{ik} e_i$ avec $v_{ik} \in \mathbb{K}$ est un isomorphisme de $\mathbb{L}$-espaces vectoriels. La surjectivité est facile. En ce qui concerne l'injectivité, si

$$\sum_i \sum_k (\lambda_k v_{ik})(1 \otimes e_i) = 0 \tag{9.429}$$

alors les choses suivantes sont nulles également :

$$\sum_i \sum_k (\lambda_k v_{ik})(1 \otimes e_i) = \sum_{ik} (\lambda_k \otimes v_{ik} e_i) = \sum_k (\lambda_k \otimes \sum_i v_{ik} e_i) = \sum_k (\lambda_k \otimes v_k). \tag{9.430}$$

Le dernier est l'argument de ψ. Le fait que ce soit nul implique que ψ est injective. □

Remarque 9.205.
Nous n'avons pas dû prouver que chacun des $\lambda_k \otimes v_k$ était nul. Et encore heureux, parce que cela pouvait très bien être faux, vu qu'il y a plusieurs façons de noter un élément de $E_{\mathbb{L}}$ sous la forme de tels termes.

Corolaire 9.206.
La $\mathbb{L}$-dimension de $E_{\mathbb{L}}$ est égale à la $\mathbb{K}$-dimension de E.

9.12.2 Projections

Problèmes et choses à faire

Nous allons définir $\texttt{proj}\colon \mathcal{L}(E_{\mathbb{L}}, F_{\mathbb{L}}) \to \mathcal{L}(E, F)$ en faisant appel à des bases et en prouvant que les choses définies ne dépendent pas des bases choisies. Il y a surement une façon plus « intrinsèque » de faire.

Nous savons que $\mathbb{L}$ est un $\mathbb{K}$-espace vectoriel dans lequel nous pouvons voir $\mathbb{K}$ comme un sous-espace (lemme 6.53). Dans cette optique nous choisissons dans $\mathbb{L}$ un supplémentaire de $\mathbb{K}$, c'est-à-dire un sous-espace vectoriel de $\mathbb{L}$ tel que

$$\mathbb{L} = \mathbb{K} \oplus V. \tag{9.431}$$

Nous avons alors naturellement une projection $\texttt{proj}\colon \mathbb{L} \to \mathbb{K}$.

Soit $\{e_i\}$ une base de E et $\{e_a\}$ une de$F$. Nous noterons également e_i et e_a les éléments τe_i et τe_a correspondants. Grâce à la proposition 9.204, ce sont des bases de $E_{\mathbb{L}}$ et $F_{\mathbb{L}}$. Si la fonction $f\colon E_{\mathbb{L}} \to F_{\mathbb{L}}$ s'écrit dans ce ces bases comme

$$f(e_i) = \sum_a f_{ai} e_a \tag{9.432}$$

alors nous définissons $\texttt{proj}(f)$ par

$$(\texttt{proj}\, f) e_i = \sum_a \texttt{proj}(f_{ai}) e_a. \tag{9.433}$$

Proposition 9.207 ([1]).
L'application `proj` *définie en* (9.433) *est indépendante du choix des bases.*

Démonstration. Notons que dans ce qui suit, les sommes sur a ou b et celles sur i ou j ne vont pas jusqu'au même indice (dimensions de E et F). De plus nous manipulons deux choses qui se notent `proj`. La première est la projection $\texttt{proj}\colon \mathbb{L} \to \mathbb{K}$ qui ne dépend que d'un choix de supplémentaire et que nous supposons fixée ici. D'autre part il y a $\texttt{proj}\colon E_{\mathbb{L}} \to E$ qui dépend à priori des bases choisies.

Nous choisissons de nouvelles bases qui sont liées aux anciennes bases par

$$\begin{cases} e'_b = \sum_a B_{ab} e_a & \text{(9.434a)} \\ e'_i = \sum_j A_{ji} e_j. & \text{(9.434b)} \end{cases}$$

Les matrices A et B sont dans $\mathrm{GL}(\mathbb{K})$. Nous allons écrire l'opérateur $\mathtt{proj}'$ qui correspond à ces bases et montrer que pour toute application linéaire $f\colon E_{\mathbb{L}} \to F_{\mathbb{L}}$ nous avons $\mathtt{proj}(f) = \mathtt{proj}'(f)$. Nous avons :

$$f(e'_j) = \sum_i A_{ji} f(e_i) \tag{9.435a}$$
$$= \sum_a \sum_b \sum_i A_{ji} f_{ai} (B^{-1})_{ba} e'b \tag{9.435b}$$
$$= \sum_b \Big(\sum_{ai} A_{ji} f_{ai} (B^{-1})_{ba} \Big) e'b, \tag{9.435c}$$

ce qui fait que

$$(\mathtt{proj}'\, f) e'_j = \sum_b \Big(\mathtt{proj}\, \big(A_{ji} f_{ai} (B^{-1})_{ba} \big) \Big) e'_b. \tag{9.436}$$

Nous calculons maintenant $(\mathtt{proj}'\, f)e_j$ en substituant $e_j = \sum_l (A^{-1})_{lj} e'_l$ et en utilisant (9.436) et la linéarité de $\mathtt{proj}'$ et la $\mathbb{K}$-linéarité de $\mathtt{proj}\colon \mathbb{L} \to \mathbb{K}$:

$$(\mathtt{proj}'\, f)\Big(\sum_l (A^{-1})_{lj} e'_l \Big) = \sum_l (A^{-1})_{lj} \sum_b \sum_{ai} \mathtt{proj}\, \big(A_{li} f_{ai} (B^{-1})_{ba} \big) e_b \tag{9.437a}$$
$$= \sum_a \mathtt{proj}(f_{aj}) e_a \tag{9.437b}$$
$$= (\mathtt{proj}\, f) e_j. \tag{9.437c}$$

Donc $\mathtt{proj} = \mathtt{proj}'$. □

Note au passage comme toujours : il y a un abus systématique de notation entre $e_i \in E$ et $\tau(e_i) = 1 \otimes e_i \in E_{\mathbb{L}}$.

Remarque 9.208 ([1]).
L'opération $\mathtt{proj}\colon \mathcal{L}(E_{\mathbb{L}}, F_{\mathbb{L}}) \to \mathcal{L}(E, F)$ ne dépend pas des bases choisies un peu partout. Mais elle dépend de l'application $pr\colon \mathbb{L} \to \mathbb{K}$ déjà construite. Et celle-là dépend du choix d'un supplémentaire V qui fournit $\mathbb{L} = \mathbb{K} \oplus V$.

Si $\mathtt{proj}(\lambda) = 0$ pour un de ces choix, cela n'implique nullement que $\lambda = 0$. Penser à $i \in \mathbb{C}$ si la projection $\mathtt{proj}\colon \mathbb{C} \to \mathbb{R}$ est l'application $(x + iy) \mapsto x$ parallèle à l'axe des imaginaires.

Par contre si $\mathtt{proj}(\lambda) = 0$ pour tout choix de V, alors nous avons bien $\lambda = 0$. Dans la suit nous « fixons » un choix de V générique, et lorsque nous rencontrerons l'égalité $\mathtt{proj}(\lambda) = 0$ nous en déduirons $\lambda = 0$.

Proposition 9.209.
Si $f\colon E \to F$ et si $f_{\mathbb{L}} e_j = \sum_a (f_{\mathbb{L}})_{aj} e_a$ et si $f(e_j) = \sum_a f_{aj} e_a$ alors

(1) $\mathtt{proj}\, f_{\mathbb{L}} = f$,

(2) $(f_{\mathbb{L}})_{ja} = f_{ja} \in \mathbb{K}$.

Démonstration. Nous avons

$$f_{\mathbb{L}}(e_i) = \sum_a f_{ai} (1 \otimes e_a) = \sum_a f_{ai} \tau(e_a), \tag{9.438}$$

donc

$$(\mathtt{proj}\, f_{\mathbb{L}}) e_i = \sum_a \mathtt{proj}(f_{ai}) e_a = \sum_a f_{ai} e_a = f(e_i). \tag{9.439}$$

Cela prouve que $\mathtt{proj}\, f_{\mathbb{L}} = f$.

Par ailleurs,

$$f_{\mathbb{L}}(\tau e_i) = f_{\mathbb{L}}(1 \otimes e_i) = 1 \otimes f(e_i) = \tau\big(f(e_i)\big) = \sum_a f_{ai} \tau(e_a) \tag{9.440}$$

alors que par définition,

$$f_{\mathbb{L}}(\tau e_i) = \sum_a (f_{\mathbb{L}})_{ai} \tau(e_a). \tag{9.441}$$

Les éléments $\tau(e_a)$ formant une base[71], la comparaison de (9.440) avec (9.441) donne $(f_{\mathbb{L}})_{ai} = f_{ai} \in \mathbb{K}$. □

Lemme 9.210.
Soient

(1) Une base $\{e_i\}$ de E et une application linéaire $f\colon E \to F$;

(2) une base $\{e_a\}$ de F et une application linéaire $g\colon G \to F$;

(3) une base $\{e_\alpha\}$ de G et une application linéaire $\tilde{h}\colon G_{\mathbb{L}} \to E_{\mathbb{L}}$.

Alors nous avons

$$\mathtt{proj}(f_{\mathbb{L}} \circ \tilde{h}) = \mathtt{proj}(f_{\mathbb{L}}) \circ \mathtt{proj}(\tilde{h}). \tag{9.442}$$

Démonstration. Pour écrire $\mathtt{proj}(f_{\mathbb{L}} \circ \tilde{h})$ à partir de la définition (9.433) nous commençons par écrire

$$(f_{\mathbb{L}} \circ \tilde{h})e_\alpha = \sum_a (f_{\mathbb{L}} \circ \tilde{h})_{a\alpha} e_a = \sum_{ai} (f_{\mathbb{L}})_{ai} (\tilde{h})_{i\alpha} e_a = \sum_a \Big(\sum_i f_{ai} (\tilde{h})_{i\alpha} \Big) e_a \tag{9.443}$$

où nous avons utilisé le fait que $(f_{\mathbb{L}})_{ai} = f_{ai}$. Donc, en utilisant la $\mathbb{K}$-linéarité de $\mathtt{proj}$,

$$\mathtt{proj}(f_{\mathbb{L}} \circ \tilde{h})e_\alpha = \sum_a \sum_i \mathtt{proj}\Big(f_{ai} (\tilde{h})_{i\alpha} \Big) e_a = \sum_a \sum_i f_{ai}\, \mathtt{proj}\Big((\tilde{h})_{i\alpha} \Big) e_a. \tag{9.444}$$

D'autre part,

$$\begin{aligned} \mathtt{proj}(f_{\mathbb{L}}) \circ \mathtt{proj}(\tilde{h}) e_\alpha &= \mathtt{proj}(f_{\mathbb{L}}) \sum_i \mathtt{proj}\Big((\tilde{h})_{i\alpha} \Big) e_i \\ &= \sum_i \mathtt{proj}\Big((\tilde{h})_{i\alpha} \Big) \sum_a f_{ai} e_a \\ &= \sum_{ai} \mathtt{proj}\Big((\tilde{h})_{i\alpha} \Big) f_{ai} e_a, \end{aligned} \tag{9.445}$$

et c'est égal à (9.444). □

Remarque 9.211.
Nous n'avons en général pas $\mathtt{proj}(xy) = \mathtt{proj}(x)\,\mathtt{proj}(y)$ pour tout $x, y \in \mathbb{L}$. Par exemple si $\mathbb{K} = \mathbb{R}$ et $\mathbb{L} = \mathbb{C}$ avec la projection canonique,

$$\mathtt{proj}(i \cdot i) = \mathtt{proj}(-1) = -1 \tag{9.446}$$

alors que $\mathtt{proj}(i) = 0$.

Proposition 9.212.
Soient $f \in \mathcal{L}(E, F)$ et $g \in \mathcal{L}(F, E)$. Alors il existe $h\colon G \to E$ tel que $f \circ h = g$ si et seulement s'il existe $\tilde{g}\colon G_{\mathbb{L}} \to E_{\mathbb{L}}$ tel que $f_{\mathbb{L}} \circ \tilde{g} = g_{\mathbb{L}}$.

Démonstration. Dans le sens direct, il suffit de poser $\tilde{h} = h_{\mathbb{L}}$.

Dans le sens inverse, si nous avons $\tilde{h}\colon G_{\mathbb{L}} \to E_{\mathbb{L}}$ tel que $f_{\mathbb{L}} \circ \tilde{h} = g_{\mathbb{L}}$ alors en appliquant $\mathtt{proj}$ des deux côtés et en utilisant le lemme 9.210,

$$\mathtt{proj}(f_{\mathbb{L}}) \circ \mathtt{proj}(\tilde{h}) = \mathtt{proj}(g_{\mathbb{L}}) \tag{9.447}$$

c'est-à-dire

$$f \circ \mathtt{proj}(\tilde{h}) = g, \tag{9.448}$$

c'est-à-dire que l'application $\mathtt{proj}\,\tilde{h}\colon G \to E$ est la réponse à la proposition. □

71. Encore la proposition 9.204.

9.12.3 Rang, polynôme minimal, polynôme caractéristique

Proposition 9.213 (Stabilité du rang par extension des scalaires[80]).
Si $f\colon E \to F$ est linéaire alors nous avons

$$\operatorname{rang}(f) = \operatorname{rang}(f_{\mathbb{L}}). \tag{9.449}$$

où à droite nous considérons le rang de l'application $\mathbb{L}$-linéaire $f_{\mathbb{L}}\colon E_{\mathbb{L}} \to F_{\mathbb{L}}$.

Démonstration. Il existe un supplémentaire V tel que $E = \ker(f) \oplus V$ avec $\dim(V) = \operatorname{rang}(f)$. Nous pouvons factoriser f en

$$f = f_2 \circ f_1 \tag{9.450}$$

avec $f_1\colon E \to V$ est la projection parallèle à $\ker(f)$ et est surjective (vers V) parce que $\dim(V) = \operatorname{rang}(f) = \dim\big(\operatorname{Image}(f)\big)$. De plus $f_2\colon V \to F$ est injective parce que si $v \in V$ est tel que $f_2(v) = 0$ alors on aurait

$$f(v) = (f_2 \circ f_1)(v) = f_2(v) = 0. \tag{9.451}$$

Cela donne $v \in \ker(f) \cap V = \{0\}$. Par la proposition 9.203, les applications $(f_1)_{\mathbb{L}}$ et $(f_2)_{\mathbb{L}}$ sont respectivement surjective et injective.

L'application $(f_2)_{\mathbb{L}}\colon V_{\mathbb{L}} \to F_{\mathbb{L}}$ est forcément surjective sur son image, donc

$$(f_2)_{\mathbb{L}}\colon V_{\mathbb{L}} \to \operatorname{Image}(f_{\mathbb{L}}) \tag{9.452}$$

est un isomorphisme de $\mathbb{L}$-espaces vectoriels. Nous avons alors les égalités

$$\dim_{\mathbb{L}}(V_{\mathbb{L}}) = \dim_{\mathbb{L}}\big(\operatorname{Image}(f_{\mathbb{L}})\big) = \operatorname{rang}(f_{\mathbb{L}}). \tag{9.453}$$

Mais aussi, par les définitions posées plus haut,

$$\dim(V) = \operatorname{rang}(f) = \dim\big(\operatorname{Image}(f)\big). \tag{9.454}$$

Mais le corolaire 9.206 nous dit que $\dim_{\mathbb{L}}(V_{\mathbb{L}}) = \dim_{\mathbb{K}}(V)$. Donc il y a égalité des deux lignes (9.453) et (9.454) donne $\operatorname{rang}(f) = \operatorname{rang}(f_{\mathbb{L}})$. □

Proposition 9.214.
Nous avons

(1) $\det(f) = \det(f_{\mathbb{L}})$

(2) $\chi_f = \chi_{f_{\mathbb{L}}}$.

Démonstration. Dès que l'on a des bases nous avons $(f_{\mathbb{L}})_{ai} = f_{ai}$ par la proposition 9.209(2). Le nombre $\det(f) \in \mathbb{K}$ est un polynôme en les f_{ai}. Entendons nous : il existe un polynôme indépendant de f et de $\mathbb{K}$ et de $\mathbb{L}$ donnant le déterminant de n'importe quelle matrice. Donc $\det(f) = \det(f_{\mathbb{L}})$.

Même chose pour le polynôme caractéristique (définition 9.107) : les coefficients de ce polynôme sont des polynômes en les f_{ai} qui sont indépendants de $\mathbb{L}$, de $\mathbb{K}$ et de f.

Notons que $\chi_{f_{\mathbb{L}}}$ est un polynôme à coefficients dans $\mathbb{K}$. □

La situation est très différente avec le polynôme minimal [72]. Autant il existe une « recette » pour créer le polynôme caractéristique, il n'en n'existe pas pour le polynôme minimal (ou en tout cas, il ne suffit pas d'appliquer des polynômes en les coefficients de la matrice). La proposition suivante montre que le polynôme minimal est conservé par extension de corps, mais que pour le voir, il faut travailler plus.

Proposition 9.215 ([80, 1]).
Soit $\mathbb{L}$ une extension du corps $\mathbb{K}$ et une application linéaire $f\colon E \to F$ entre deux $\mathbb{K}$-espaces vectoriels. Alors $\mu_f = \mu_{f_{\mathbb{L}}}$.

72. Définition 6.57.

Démonstration. Nous allons montrer que l'application

$$\begin{aligned}\tilde{g}\colon \frac{\mathbb{L}[X]}{(\mu)} &\to \operatorname{End}(E_{\mathbb{L}})\\ \bar{P} &\mapsto P(f_{\mathbb{L}})\end{aligned} \tag{9.455}$$

est bien définie et injective. La proposition 9.97 nous dira alors que μ est le polynôme minimal de $f_{\mathbb{L}}$.

Pour prouver que l'application $\tilde{g}$ est bien définie, nous commençons par prouver que $P(f_{\mathbb{L}}) = P(f)_{\mathbb{L}}$:

$$P(f_{\mathbb{L}})\lambda\otimes v = \sum_k a_k f_{\mathbb{L}}^k \lambda\otimes v \tag{9.456a}$$
$$= \lambda\otimes\sum_k a_k f^k(v) \tag{9.456b}$$
$$= \lambda\otimes P(f)v \tag{9.456c}$$
$$= P(f)_{\mathbb{L}}\lambda\otimes v. \tag{9.456d}$$

Par conséquent $\mu(f_{\mathbb{L}}) = 0$ et l'application est bien définie.

Sur $\mathbb{L}[X]/(\mu)$ nous considérons la base $\{1, \bar{X}, \dots, \bar{X}^{\deg(\mu)-1}\}$, et $\operatorname{End}(E_{\mathbb{L}})$ nous considérons une base qui commence [73] par $\{f_{\mathbb{L}}^k\}_{k=0,\dots,\deg(\mu)-1}$. Montrons tout de même que cette partie est libre (sinon le théorème de la base incomplète ne s'applique pas) : si $\sum_k \lambda_k f_{\mathbb{L}}^k = 0$ alors

$$\sum_k \texttt{proj}\left(\lambda_k f_{\mathbb{L}}^k\right) = 0. \tag{9.457}$$

Pour détailler ce que cela implique, nous calculons ceci :

$$(\lambda f_{\mathbb{L}})(\tau e_i) = \lambda f_{\mathbb{L}}(\tau e_i) = \sum_a \lambda f_{ia} e_a, \tag{9.458}$$

par conséquent $\texttt{proj}(\lambda f_{\mathbb{L}})e_i = \sum_a \texttt{proj}(\lambda f_{ia})e_a$, et comme $\texttt{proj}$ est $\mathbb{K}$-linéaire et que $f_{ai}\in\mathbb{K}$,

$$\texttt{proj}(\lambda f_{\mathbb{L}})e_i = \texttt{proj}(\lambda)\sum_a f_{ai}e_a = \texttt{proj}(\lambda)\,\texttt{proj}(f_{\mathbb{L}})e_i = \texttt{proj}(\lambda)f(e_i). \tag{9.459}$$

Appliquer la projection $\texttt{proj}$ à l'équation (9.457) donne alors $\sum_k \texttt{proj}(\lambda)_k f^k = 0$. Mais comme les f^k sont linéairement indépendantes sur $\mathbb{K}$ nous avons pour tout k : $\texttt{proj}(\lambda_k) = 0$ (égalité dans $\mathbb{K}$). En nous souvenant de la remarque 9.208 nous en déduisons $\lambda_k = 0$ dans $\mathbb{L}$.

Dans les choix de bases faits, l'application $\tilde{g}$ a la forme

$$\tilde{g} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \\ * & * & * \\ * & * & * \\ * & * & * \end{pmatrix}, \tag{9.460}$$

qui est injective.

Vu que $\tilde{g}$ est injective, μ est le polynôme minimal de $f_{\mathbb{L}}$ et donc $\mu = \mu_{\mathbb{L}}$. □

9.13 Frobenius et Jordan

9.13.1 Matrice compagnon

Définition 9.216.
*Soit le polynôme $P = X^n - a_{n-1}X^{n-1} - \dots - a_1X - a_0$ dans $\mathbb{K}[X]$. La **matrice compagnon** de*

73. Théorème de la base incomplète 4.11(2).

P est la matrice donnée par

$$C(P) = \begin{pmatrix} 0 & \cdots & \cdots & 0 & a_0 \\ 1 & 0 & & \vdots & a_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & a_{n-2} \\ 0 & \cdots & 0 & 1 & a_{n-1} \end{pmatrix} \tag{9.461}$$

si $n \geqslant 2$ et par (a_0) si $n = 1$.

Une matrice est dite compagnon si elle a cette forme.

Proposition 9.217.
Si f est l'endomorphisme associé à la matrice $C(P)$ nous avons

$$f(e_i) = \begin{cases} e_{i+1} & \text{si } i < n \\ (a_0, \dots, a_{n-1}) & \text{si } i = n. \end{cases} \tag{9.462}$$

De plus l'endomorphisme f vérifie $P(f)e_1 = 0$.

Lemme 9.218 ([194])**.**
Un polynôme sur un corps commutatif est le polynôme caractéristique de sa matrice compagnon. En d'autres termes nous avons $\chi_{C(P)} = P$.

Démonstration. Nous notons f l'endomorphisme associé à $C(P)$. La propriété $P(f)e_1 = 0$ nous indique que le polynôme minimal ponctuel de f en e_1 divise P. L'ensemble des puissances de f appliquées à e_1, $\big(f^i(e_1)\big)_{i=1,\dots,n-1}$ est libre, donc le polynôme minimal ponctuel en e_1 est de degré n au minimum. En reprenant les notations du théorème 6.37, nous avons $I_{e_1} = (P)$ parce que P est de degré minimum dans I_{e_1} et $\chi_f \in I_{e_1}$.

Donc P divise χ_f et est de degré égal à celui de χ_f. Étant donné qu'ils sont tous deux unitaires, ils sont égaux. □

Remarque 9.219.
Les matrices compagnons ne sont pas les seules dont le polynôme caractéristique est égal au polynôme minimal. En fait les matrices dont le polynôme caractéristique est égale au polynôme minimal sont denses dans les matrices. En effet une matrice dont le polynôme minimal n'est pas égal au polynôme caractéristique a un polynôme caractéristique avec une racine double. Il est possible, en modifiant arbitrairement peu la matrice de séparer la racine double en deux racines distinctes.

9.13.2 Réduction de Frobenius

Lemme 9.220.
Soit un endomorphisme $f\colon E \to E$ sur l'espace vectoriel de dimension finie n. Nous notons μ et χ les polynômes minimal et caractéristique. Si f est cyclique, alors $\mu = \chi$.

Le théorème 9.233 donnera une version plus complète de ce lemme.

Démonstration. Soit v un vecteur cyclique de f, c'est-à-dire que $\{f^k(v)\}_{k=0,\dots,n-1}$ est libre. Donc si P est un polynôme de degré jusqu'à $n-1$ nous ne pouvons pas avoir $P(f) = 0$ parce que, appliqué à v, ce serait une combinaisons nulle non triviale des $f^k(v)$. Donc le polynôme minimal est au minimum de degré n. Mais le polynôme caractéristique est annulateur de degré n (Cayley-Hamilton 9.115), donc il est le polynôme minimal. □

Théorème 9.221 (Réduction de Frobenius [195, 196, 197])**.**
Soit E, un $\mathbb{K}$-espace vectoriel, et $f \in \mathrm{End}(E)$. Alors il existe une suite de sous-espaces $E_1, \dots, E_r$ stables par f tels que

(1) $E = \bigoplus_{i=1}^{r} E_i$ *;*

(2) pour chaque E_i*, l'endomorphisme restreint* $f_i = f|_{E_i}$ *est cyclique ;*

(3) si μ_i *est le polynôme minimal de* f_i *alors* μ_{i+1} *divise* μ_i *;*

Une telle décomposition vérifie automatiquement $\mu_1 = \mu_f$ *et* $\mu_1 \cdots \mu_r = \chi_f$*, et la suite* $(\mu_i)_{i=1,\dots,r}$ *ne dépend que de* f *et non du choix de la décomposition du point (1).*

Les polynômes μ_i sont les **invariants de similitude** de l'endomorphisme f.

Démonstration. Nous commençons par montrer que si une telle décomposition existe, alors

$$\chi_f = \prod_{i=1}^{r} \mu_i \tag{9.463a}$$

$$\mu_f = \mu_1 \tag{9.463b}$$

où χ_f est le polynôme caractéristique de f et μ_f est le polynôme minimal. D'abord le polynôme caractéristique de f devra être égal au produit des polynômes caractéristique des $f|_{E_i}$, mais ces derniers endomorphismes étant cycliques[74], leurs polynôme caractéristiques sont égaux à leurs polynômes minimaux (lemme 9.220). Cela prouve l'égalité (9.463a). Ensuite tous les μ_i doivent diviser le polynôme minimal, donc $\operatorname{ppcm}(\mu_1, \dots, \mu_r)$ divise μ_f. Cependant le polynôme minimal doit contenir une et une seule fois chacun des facteurs irréductibles du polynôme caractéristique, et chacun de ces facteurs sont dans les polynômes μ_i. Par conséquent $\operatorname{ppcm}(\mu_1, \dots, \mu_r) = \mu_f$. Mais par ailleurs $\mu_1 = \operatorname{ppcm}(\mu_1, \dots, \mu_r)$ parce qu'on a supposé $\mu_{i+1} \mid \mu_i$, donc $\mu_1 = \mu_f$.

Soit d, le degré du polynôme minimal de f et $y \in E$ tel que $\mu_f = \mu_{f,y}$ (voir lemme 9.98). Le plus petit espace stable sous f contenant y est

$$E_y = \operatorname{Span}\{y, f(y), \dots, f^{d-1}(y)\}. \tag{9.464}$$

Nous notons $e_i = f^{i-1}(y)$. Notons que les vecteurs donnés forment bien une base de E_y parce que si les e_i n'était pas linéairement indépendants, alors nous aurions des a_k tels que $\sum_k a_k e_k = 0$ et avec lesquels

$$\big(\sum_k a_k X^k\big)(f)y = 0, \tag{9.465}$$

ce qui contredirait la minimalité de $\mu_{f,y}$.

La difficulté du théorème est de trouver un complément de E_y qui soit également stable sous f. Nous commençons par étendre[75] $\{e_1, \dots, e_d\}$ en une base $\{e_1, \dots, e_n\}$ de E. Ensuite nous allons montrer que

$$E = E_y \oplus F \tag{9.466}$$

avec

$$F = \{x \in E \text{ tel que } e_d^*\big(f^k(x)\big) = 0 \forall k \in \mathbb{N}\}. \tag{9.467}$$

Par construction, F est invariant sous f. Montrons pour commencer que $E_y \cap F = \{0\}$. Un élément de E_y s'écrit

$$z = a_1 e_1 + \dots + a_k e_k \tag{9.468}$$

avec $k \leqslant d$. Étant donné que f décale les vecteurs de base, nous avons $e_d^*\big(f^{d-k}(z)\big) = a_k$. Du coup $z \in F$ si et seulement si $a_1 = \dots = a_d = 0$, c'est-à-dire que $E_y \cap F = \{0\}$.

Nous montrons maintenant que $\dim F = n - d$. Pour cela nous considérons l'application

$$\begin{aligned} T\colon \mathbb{K}[F] &\to E^* \\ g &\mapsto e_d^* \circ g. \end{aligned} \tag{9.469}$$

Cette application est injective. En effet un élément général de $\mathbb{K}[f]$ est

$$g = a_1 \operatorname{Id} + a_2 f + \dots + a_p f^{p-1} \tag{9.470}$$

74. Définition 9.99.

75. Pour autant que j'aie compris, cette extension manque dans [195]. Corrigez-moi si je me trompe.

avec $p \leqslant d$. Si $T(g) = 0$, alors nous avons en particulier

$$0 = T(g)e_{d-p+1} = e_d^*(a_1 e_{d-p+1} + a_2 e_{d-p+2} + \cdots + a_p e_d) = a_p. \tag{9.471}$$

Donc $a_p = 0$ et en appliquant maintenant $T(g)$ à e_{d-p} nous obtenons $a_{p-1} = 0$. Au final nous trouvons que $g = 0$ et donc que T est injective.

Étant donné que $\dim \mathbb{K}[f] = d$ et que T est injective, $\dim \mathrm{Image}(T) = d$. Nous regardons l'orthogonal de l'image :

$$\begin{aligned}
(\mathrm{Image}(T))^\perp &= \{x \in E \text{ tel que } T(g)x = 0 \forall g \in \mathbb{K}[f]\} && (9.472a)\\
&= \{x \in E \text{ tel que } e_d^*\big(g(x)\big) = 0 \forall g \in \mathbb{K}[f]\} && (9.472b)\\
&= F. && (9.472c)
\end{aligned}$$

Par conséquent $F^\perp = \mathrm{Image}(T)$. Vu que $\dim \mathrm{Image}(T) = d$, nous avons donc $\dim F = n - d$ et il est établi que $E = E_y \oplus F$.

Nous avons donc trouvé F, stable par f et tel que $E = E_y \oplus F$. Nous devons maintenant nous assurer que cette décomposition tombe bien pour les polynômes minimaux. Si P_1 est le polynôme minimal de $f|_{E_y j}$, alors par le lemme 9.100 nous avons $P_1 = \mu_{f,y} = \mu_f$ parce que $f|_{E_y}$ est cyclique sur E_y. Mettons P_2, le polynôme minimal de $f|_F$. Étant attendu que F est stable par f, le polynôme P_2 divise P_1. En recommençant la construction sur F, nous construisons un nouvel espace F' stable sous F et vérifiant $\mu_{f|_{F'}} = P_2$, etc.

Nous passons maintenant à la partie unicité du théorème. Soient deux suites $F_1, \ldots, F_r$ et $G_1, \ldots, G_s$ de sous-espaces stables par f et vérifiant

(1) $E = \bigoplus_{i=1}^r F_i$,

(2) $f|_{F_i}$ est cyclique,

(3) $\mu_{f|_{F_{i+1}}}$ divise $\mu_{f|_{F_i}}$,

et, *mutatis mutandis*, les mêmes conditions pour la famille $\{G_i\}$. Nous posons $P_i = \mu_{f_{F_i}}$ et $Q_i = \mu_{f|_{G_i}}$. Nous allons montrer par récurrence que $P_i = Q_i$ et $\dim F_i = \dim G_i$. Il ne sera cependant pas garanti que $F_i = G_i$. D'abord, $P_1 = Q_1$ parce qu'ils sont tous deux égaux à μ_f par les relations (9.463). Nous supposons que $P_i = Q_i$ pour $i \leqslant 1 \leqslant j - 1$ et nous tentons de montrer que $P_j = Q_j$.

Nous avons

$$P_j(f) = P_j(f)|_{F_1} \oplus \ldots \oplus P_j(f)|_{F_{j-1}}. \tag{9.473}$$

En effet étant donné que P_{j+k} divise P_j, nous avons [76] $P_j(f) = A(f) \circ P_{j+k}(f)$, mais $P_{j+k}(f)F_{j+k} = 0$, donc $P_j(f)F_{j+k} = 0$. Les espaces G_i n'ayant à priori aucun rapport avec les polynômes P_i, nous écrivons

$$P_j(f) = P_j(f)|_{G_1} \oplus \ldots \oplus P_j(f)|_{G_{j-1}} \oplus P_j(f)|_{G_j} \oplus \ldots \oplus P_j(f)|_{G_s}. \tag{9.474}$$

Pour $1 \leqslant i \leqslant j - 1$, nous avons supposé $P_i = Q_i$. Étant donné que $f|_{F_i}$ est semblable à C_i et $f|_{G_i}$ est semblable à C_{Q_i}, la matrice de $f|_{E_i}$ est semblable à la matrice de $f|_{G_i}$. En particulier,

$$\dim P_j(f)F_i = \dim P_j(f)G_i. \tag{9.475}$$

En prenant les dimensions des images dans les égalités (9.473) et (9.474), nous trouvons que

$$P_j(f)|_{G_j} = \ldots = P_j(f)|_{G_s} = 0. \tag{9.476}$$

Par conséquent $P_j \in I_{f|G_j}$ et donc P_j divise Q_j, qui est générateur de $I_{f|_{G_j}}$. La situation étant symétrique entre P et Q, nous montrons de même que Q_j divise P_j et donc que $P_j = Q_j$.

Ceci achève la démonstration du théorème de réduction de Frobenius.

□

76. En vertu du lemme 9.87.

Remarque 9.222.
Sous forme matricielle, ce théorème dit que toute matrice est semblable à une matrice de la forme bloc-diagonale

$$f = \begin{pmatrix} C_{\mu_1} & & \\ & \ddots & \\ & & C_{\mu_r} \end{pmatrix} \tag{9.477}$$

où les C_{μ_i} sont les matrices compagnon (définition 9.216).

En particulier, et ceci est très important, deux applications sont semblables si et seulement si elles ont même suite d'invariants de similitude.

Remarque 9.223.
Si nous travaillons sur $\mathbb{R}$, la réduite de Frobenius restera une matrice réelle, même si les valeurs propres sont complexes. En effet le procédé de Frobenius ne regarde absolument pas les valeurs propres, mais seulement les facteurs irréductibles du polynôme caractéristique. La réduite de Frobenius ne tente pas de résoudre ces polynômes, mais se contente d'en utiliser les matrices compagnon.

La situation sera différente dans le cas de la forme normale de Jordan.

9.13.3 Forme normale de Jordan

Il existe une preuve directe de la réduction de Jordan ne nécessitant pas la réduction de Frobenius[183]. Cette dernière passe par les espaces caractéristiques[77] et est à mon avis plus compliquée que la démonstration de Frobenius elle-même. Nous allons donc nous contenter de donner la réduction de Jordan comme un cas particulier de Frobenius.

Théorème 9.224 (Réduction de Jordan)**.**
Soit E un espace vectoriel sur $\mathbb{K}$, et $f \in \mathrm{End}(E)$ un endomorphisme dont le polynôme caractéristique χ_f est scindé[78]*. Il existe une base de E dans laquelle la matrice de f s'écrit sous la forme*

$$M = \begin{pmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_k}(\lambda_k) \end{pmatrix} \tag{9.478}$$

où les λ_i sont les valeurs propres de f (avec éventuelle répétitions) et $J_n(\lambda)$ représente le bloc $n \times n$

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \lambda & & \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}. \tag{9.479}$$

En d'autres termes, $J_n(\lambda)_{ii} = \lambda$ et $J_n(\lambda)_{i-1,i} = 1$.

Démonstration. Nous commençons par le cas où f est nilpotente ; nous notons M sa matrice. Dans ce cas la seule valeur propre est zéro et le polynôme caractéristique est X^m pour un certain m. Nous savons par le lemme 9.218 que (la matrice de) f est semblable à sa matrice compagnon. En l'occurrence pour f nous avons

$$C_{X^m} = \begin{pmatrix} 0 & & & 0 \\ 1 & \ddots & & \vdots \\ & \ddots & \ddots & \vdots \\ & & 1 & 0 \end{pmatrix}. \tag{9.480}$$

77. Aussi appelés « espaces propres généralisés ».
78. C'est pour cette hypothèse que $\mathbb{K} = \mathbb{R}$ n'est pas le bon cadre.

Ensuite le changement de base (qui est une similitude) $(e_1, \dots, e_n) \mapsto (e_n, \dots, e_1)$ montre que C_{X^m} est semblable à un bloc de Jordan $J_m(0)$.

Supposons à présent que f ne soit pas nilpotente. Par l'hypothèse de polynôme caractéristique scindé, nous supposons que f a m valeurs propres distinctes et que son polynôme caractéristique est

$$\chi_f = (X - \lambda_1)^{l_1} \dots (X - \lambda_m)^{l_m}. \tag{9.481}$$

Le lemme des noyaux (théorème 9.88) nous enseigne que

$$E = \bigoplus_{i=1}^{m} \underbrace{\ker(f - \mu_i \mathbb{1})^{l_i}}_{F_i}. \tag{9.482}$$

La restriction de $f - \lambda_i \mathbb{1}$ à F_i est par construction un endomorphisme nilpotent, et donc peut s'écrire comme un bloc de Jordan avec des zéros sur la diagonale. En utilisant la décomposition

$$f|_{F_i} = (f - \lambda_i \mathbb{1})|_{F_i} + \lambda_i \mathbb{1}_{F_i}, \tag{9.483}$$

nous voyons que $f|_{F_i}$ s'écrit comme un bloc de Jordan avec λ_i sur la diagonale. □

Remarque 9.225.
Nous pouvons calculer la forme normale de Jordan pour une matrice complexe ou réelle, mais dans les deux cas nous devons nous attendre à obtenir une matrice complexe parce que les valeurs propres d'une matrice réelle peuvent être complexes. Cependant nous demandons que le polynôme caractéristique de f soit scindé sur $\mathbb{K}$. En pratique, la décomposition de Jordan n'est garantie que sur les corps algébriquement clos, c'est-à-dire sur $\mathbb{C}$.

La suite des invariants de similitude sur laquelle repose Frobenius, elle, est disponible sur tout corps, y compris $\mathbb{R}$.

9.14 Commutant et endomorphismes cycliques

9.14.1 Endomorphisme cyclique

Lemme 9.226.
Si A est la matrice de l'endomorphisme f alors nous avons équivalence des propriétés suivantes :

(1) La matrice A est cyclique.

(2) L'endomorphisme f est cyclique.

Si f est un endomorphisme de l'espace vectoriel E et si $x \in E$, nous notons

$$E_{f,x} = \mathrm{Span}\{f^k(x) \text{ tel que } k \in \mathbb{N}\}. \tag{9.484}$$

Définition 9.227.
Soit E un espace vectoriel de dimension finie sur un corps $\mathbb{K}$ et un endomorphisme $f\colon E \to E$. Le ***commutant*** *de f est l'ensemble des endomorphismes de E qui commutent avec f :*

$$\mathcal{C}(f) = \{g \in \mathcal{L}(E, E) \text{ tel que } g \circ f = f \circ g\}. \tag{9.485}$$

Il n'est pas très compliqué de vérifier que $\mathcal{C}(f)$ est un sous-espace vectoriel de $\mathcal{L}(E, E)$.

Notons l'inclusion évidente $\mathbb{K}[f] \subset \mathcal{C}(f)$. L'inclusion inverse va un peu nous occuper durant les prochaines pages.

9.14.2 Commutant : cas diagonalisable

Proposition 9.228 ([198]).
Si f est diagonalisable, alors

$$\dim\big(\mathcal{C}(f)\big) = \sum_{\lambda\in\mathrm{Spec}(f)} \dim(E_\lambda)^2. \tag{9.486}$$

où les E_λ sont les espaces propres de f.

Démonstration. D'abord su $g \in \mathcal{C}(f)$ alors E_λ est stable par g. En effet si $v \in E_\lambda$ alors $f\big(g(v)\big) = g\big(f(v)\big) = g(\lambda v) = \lambda g(v)$, ce qui montre que $g(v)$ est un vecteur propre de f pour la valeur propre λ, et donc que $g(v) \in E_\lambda$.

Nous considérons ensuite l'application

$$\begin{aligned}\psi\colon \mathcal{C}(f) &\to \mathrm{End}(E_1)\times\ldots\times\mathrm{End}(E_r)\\ g &\mapsto g|_{E_1}\times\ldots\times g|_{E_r}\end{aligned} \tag{9.487}$$

qui est bien définie parce que g se restreint aux espaces propres de f. Nous allons noter $\psi(g)_\lambda$ la restriction de g à E_λ.

ψ est injective Supposons que $g, h \in \mathcal{C}(f)$ tels que $\psi(g) = \psi(h)$. Vu que f est diagonalisable nous pouvons décomposer $x \in E$ en ses composantes sur les espaces propres [79] :

$$x = \sum_{\lambda\in\mathrm{Spec}(f)} x_\lambda \tag{9.488}$$

avec $x_\lambda \in E_\lambda$. Nous avons alors

$$g(x) = \sum_\lambda g(x_\lambda) = \sum_\lambda \psi(g)_\lambda(x_\lambda). \tag{9.489}$$

Vu que nous avons $\psi(g)_\lambda = \psi(h)_\lambda$, nous avons aussi

$$g(x) = \sum_\lambda \psi(g)_\lambda(x_\lambda) = \sum_\lambda \psi(h)_\lambda(x_\lambda) = \sum_\lambda h(x_\lambda) = h(x). \tag{9.490}$$

Cela prouve $g = h$ et donc que ψ est injective.

ψ est surjective Si nous avons pour chaque $\lambda \in \mathrm{Spec}(f)$ un endomorphisme g_λ de E_λ alors en posant

$$g(x) = \sum_{\lambda\in\mathrm{Spec}(f)} g_\lambda(x_\lambda) \tag{9.491}$$

alors nous avons bien

$$\psi(g) = \big(g_{\lambda_1},\ldots,g_{\lambda_r}\big). \tag{9.492}$$

Nous pouvons donc conclure en écrivant

$$\dim\big(\mathcal{C}(f)\big) = \sum_{\lambda\in\mathrm{Spec}(f)} \dim\big(\mathrm{End}(E_\lambda)\big) = \sum_{\lambda\in\mathrm{Spec}(f)} \dim(E_\lambda)^2. \tag{9.493}$$

□

Remarque 9.229.
Nous avons alors immédiatement

$$\dim\big(\mathcal{C}(f)\big) \geqslant \dim(E) \tag{9.494}$$

lorsque f est diagonalisable.

79. Théorème 9.128(5).

En suivant la notation (9.181), un endomorphisme est cyclique lorsqu'il existe $x \in E$ tel que $E_x = E$.

Proposition 9.230 ([198]).
Si f est un endomorphisme diagonalisable d'un espace vectoriel E de dimension n. Nous avons équivalence entre les faits suivants.

(1) Le polynôme minimal est égal au polynôme caractéristique : $\mu_f = \chi_f$
(2) L'endomorphisme f est cyclique.
(3) $\mathcal{C}(f) = \mathbb{K}[f]$.
(4) $\dim\big(\mathcal{C}(f)\big) = n$
(5) L'endomorphisme f possède n valeurs propres distinctes.
(6) $\dim\big(\mathbb{K}[f]\big) = n$

Démonstration. Le point important de cette proposition sont les équivalences (1)-(3). Les autres sont des intermédiaires. En particulier, dans le cas diagonalisable, nous allons voir que le point (5) est essentiellement une reformulation de (1).

(4) implique (5) Par la formule (9.486), les espaces propres de f ont dimension 1. Par conséquent f possède n valeurs propres distinctes.

(5) implique (6) Le théorème 9.128 nous dit que le polynôme minimal est scindé à racines simples. Vu que f possède n valeurs propres distinctes, μ est de degré n. Par l'isomorphisme $\mathbb{K}[f] = \mathbb{K}[X]/(\mu)$ de la proposition 9.106 nous avons $\dim\big(\mathbb{K}[f]\big) = \deg(\mu) = n$ par la proposition 6.38.

(6) implique (1) Par l'isomorphisme $\mathbb{K}[f] = \mathbb{K}[X]/(\mu)$ de la proposition 9.106 et la proposition 6.38 nous avons $n = \dim\big(\mathbb{K}[f]\big) = \deg(\mu)$. Vu que χ est un polynôme annulateur (Caley-Hamilton 9.115), il est divisé par μ. Maintenant μ et χ sont des polynômes unitaires de degré n et μ divise χ. Ils sont donc égaux.

(1) implique (2) Le fait que f soit diagonalisable permet d'utiliser le théorème 9.128 pour dire que μ est scindé à racines simples. L'égalisation avec χ nous permet de dire que f possède n valeurs propres distinctes. Soient $\{e_1, \dots, e_n\}$ une base de diagonalisation, et prouvons que le vecteur $v = e_1 + \cdots + e_n$ est cyclique. Nous avons

$$f^k(v) = \sum_{i=1}^{n} \lambda_i^k e_i. \tag{9.495}$$

Pour prouver que cette famille (avec $k = 0, \dots, n-1$) est libre[80] nous en prenons une combinaison linéaire nulle et nous prouvons que les coefficients sont tous nuls. Soit donc

$$0 = \sum_{l=0}^{n-1} a_l f^l(v) = \sum_{l=0}^{n-1} a_l \sum_{i=1}^{n} \lambda_i^l e_i = \sum_{i=1}^{n} \Big(\sum_{l=0}^{n-1} a_l \lambda_i^l \Big) e_i. \tag{9.496}$$

Vu que cela est nul, nous avons pour tout i :

$$\sum_{l=0}^{n-1} a_l \lambda_i^l = 0. \tag{9.497}$$

En posant la matrice $A_{ij} = \lambda_i^j$, cela revient à étudier le système $\sum_j A_{ij} a_j = 0$. Ce système n'a des solutions non nulles que si $\det(A) = 0$; sinon il possède une unique solution et elle est $a_j = 0$ pour tout j. Nous devons donc calculer le déterminant

$$\det \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \lambda_n & \lambda_n^2 & \cdots & \lambda_n^{n-1} \end{pmatrix}. \tag{9.498}$$

80. Ce sera alors une base parce que n vecteurs libres dans un espace de dimension n est toujours une base, proposition 4.16.

Il s'agit du déterminant de Vandermonde déjà étudié par la proposition 9.12. Nous avons $\det(A) = \prod_{1\leqslant i<j\leqslant n}(\lambda_j - \lambda_i)$. Cela est bien non nul du fait que toutes les valeurs propres soient distinctes.

(2) implique (3) Soit v un vecteur cyclique de f. Un endomorphisme g donne lieu à un polynôme par le fait suivant : il existe des uniques a_k $(k = 0, \dots, n-1)$ tels que

$$g(v) = \sum_{k=0}^{n-1} a_k f^k(v). \tag{9.499}$$

Cela donne une application linéaire

$$\begin{aligned} \psi\colon \mathcal{C}(f) &\to \mathbb{K}[f] \\ g &\mapsto P \text{ tel que } P(f)v = g(v). \end{aligned} \tag{9.500}$$

C'est une application injective parce que si $\psi(g) = 0$ alors $g(v) = 0$ et pour tout k nous avons $g\big(f^k(v)\big) = f^k\big(g(v)\big) = 0$. L'endomorphisme g s'annulant sur une base, est nul.

(3) implique (4) Si $n_1, \dots, n_r$ sont les dimensions des différents espaces propres de f, nous avons les inégalités

$$\dim\big(\mathbb{K}[f]\big) = \deg(\mu) \leqslant n = n_1 + \cdots + n_r \leqslant n_1^2 + \cdots + n_r^2 = \dim\big(\mathcal{C}(f)\big). \tag{9.501}$$

Par hypothèse d'égalité entre le premier et le dernier terme de cette suite d'inégalités, toutes les inégalités sont des égalités et en particulier $\dim\big(\mathcal{C}(f)\big) = n$.

Nous avons fini de prouver toutes les équivalences demandées. □

Exemple 9.231.
Pour mieux comprendre pourquoi le fait d'avoir n valeurs propres distinctes est équivalent à être cyclique, notons que si deux valeurs propres sont identiques, alors un morceau de la matrice de f serait par exemple $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, et dans ce cas n'importe quelle combinaison $ae_i + be_j$ reste proportionnelle à elle-même après application de f. Si nous avons des valeurs propres différentes par contre, nous avons par exemples dans $\mathbb{R}^2$ la matrice $\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ qui donne $f(e_1 + e_2) = e_1 + 2e_2$. La partie $\{e_1 + e_2, e_1 + 2e_2\}$ est une base. △

9.14.3 Commutant : cas général

Nous considérons encore un espace vectoriel E de dimension finie n et un endomorphisme $f\colon E \to E$. Nous notons μ sont polynôme minimal et μ_x le polynôme minimal ponctuel en x.

Lemme 9.232 ([199, 195, 80]).
Nous avons

$$\dim\big(\mathcal{C}(f)\big) \geqslant \dim(E) \tag{9.502}$$

Démonstration. Si f est donnée, l'espace $\mathcal{C}(f)$ est l'espace des solutions de $fg = gf$. Supposons avoir choisi une base de E et notons A la matrice de f et X celle de g. L'équation est $AX - XA = 0$.

Si A est trigonalisable Nous supposons avoir choisi la base de telle sorte que A soit triangulaire supérieure, et nous allons nous contenter de chercher les solutions X qui sont également triangulaires supérieure. S'il y en a déjà plus que n, a fortiori le résultat sera vrai.

Le produit de deux matrices triangulaires supérieures étant une matrice triangulaire supérieure, l'équation $AX - XA$ contient, pour les coefficients de X, $n(n+1)/2$ équations. Mais il se fait que les termes diagonaux ne sont pas de vraies équations parce que

$$(AX - XA)_{kk} = \sum_i \big(A_{ki}X_{ik} - X_{ki}A_{ik}\big) = \sum_{k\leqslant i\leqslant k} (A_{ki}X_{ik} - X_{ki}A_{ik}) = 0. \tag{9.503}$$

Nous avons donc au maximum

$$\frac{n(x+1)}{2} - n \tag{9.504}$$

équations linéairement indépendantes pour un minimum de $n(n+1)/2$ inconnues. L'espace des solutions est donc de dimension au minimum n.

Cela a l'air d'être une majoration assez large, mais il existe des cas d'égalité.

Si A n'est pas trigonalisable La preuve que nous donnons ici est valable même pour les endomorphismes trigonalisables.

Nous considérons le résultat de Frobenius 9.221. Nous avons donc la structure suivante :

— une décomposition en somme directe $E = E_1 \oplus \ldots \oplus E_r$,

— les espaces E_i sont fixés par f,

— les endomorphismes $f_i = f|_{E_i}$ sont cycliques

— le polynôme minimal de f_i est μ_i et $\prod_{i=1}^r \mu_i = \chi_f$.

Les endomorphismes f_i^k commutent évidemment avec f_j, et la partie $\{f_i^k\}_{k=0,\ldots,\deg(\mu_i)-1}$ est libre. Libre en tout cas en tant que partie de $\mathrm{End}(E_i)$. Mais en prolongeant par 0 sur E, ça reste libre en tant que partie de $\mathrm{End}(E)$.

Bien entendu les f_j^k et les f_i^k $(i \neq j)$ sont linéairement indépendants dans $\mathrm{End}(E)$ parce qu'ils n'agissent pas sur les mêmes vecteurs. Donc les endomorphismes $f_i^{k_i}$ avec $k_i = 0, \ldots, \deg(\mu_i)-1$ forment une partie libre de $\mathrm{End}(E)$ composée d'endomorphismes qui commutent avec f. Il y en a en tout

$$\sum_{i=1}^r \deg(\mu_i) = \deg(\chi_f) = n. \tag{9.505}$$

Par conséquent $\dim\big(\mathcal{C}(f)\big) \geqslant \dim(E)$.

□

Théorème 9.233 ([177]).
Soit un endomorphisme $f\colon E \to E$ sur l'espace vectoriel de dimension finie n. Nous notons μ et χ les polynômes minimal et caractéristique. Nous avons équivalence entre les faits suivants :

(1) $\mu = \chi$,

(2) f est cyclique,

(3) $\mathcal{C}(f) = \mathbb{K}[f]$.

Démonstration. Plusieurs implications. Notons que (1) implique (1) a déjà été démontré par le lemme 9.220.

(1) implique (2) Conformément à ce que nous permet le lemme 9.98 nous choisissons[81] $a \in E$ de telle sorte à avoir $\mu_a = \mu$. De plus pour $x \in E$ nous considérons l'application

$$\begin{aligned} \varphi_x\colon \mathbb{K}[X] &\to E \\ P &\mapsto P(f)x. \end{aligned} \tag{9.506}$$

Nous avons $\varphi_a(P) = P(f)a$. Étant donné queE_a est engendré par les $f^k(a)$ nous avons $\varphi_a\big(\mathbb{K}[X]\big) = E_a$. De plus l'application φ_a passe aux classes pour (μ_a). Pour rappel, un élément de $\mathbb{K}[X]/(\mu_a)$ est de la forme

$$\bar{P} = \{P + Q\mu_a\}_{Q \in \mathbb{K}[X]}. \tag{9.507}$$

Nous considérons donc l'application

$$\varphi_a\colon \frac{\mathbb{K}[X]}{(\mu_a)} \to E_a \tag{9.508}$$

et nous prouvons que c'est un isomorphisme d'espace vectoriel.

81. Dans toute la suite, nous devrions écrire μ_f et $\mu_{f,a}$ mais nous omettons d'indiquer explicitement la dépendance en f.

Linéaire Parce que $(\lambda P + Q)(f) = (\lambda P)(f) + Q(f)$.

Injectif Si $\varphi_a(\bar{P}) = 0$ alors $\varphi_a(P) = 0$ (dans la deuxième, φ_a est l'application définie sur les polynômes et non sur les classes), ce qui montrer que P est annulateur de a. Mais par définition 9.94 du polynôme minimal ponctuel, μ_a est générateur de $\ker(\varphi_a)$; donc il existe $Q \in \mathbb{K}[X]$ tel que $P = Q\mu_a$. En d'autres termes, du point de vue du quotient, $\bar{P} = 0$.

Surjectif Si $x \in E_a$ alors il existe des coefficients $x_k \in \mathbb{K}$ tels que $x = \sum_{k=0}^{\deg(\mu_a)-1} x_k f^k(a)$, c'est-à-dire $x = P(f)a = \varphi_a(P)$.

Mais par hypothèse et par choix de a nous avons $\mu_a = \mu = \chi$, donc en fait $E_a = \mathbb{K}[X]/(\chi)$. Mais nous savons que $\deg(\chi) = \dim(E)$ et que $\dim\big(\mathbb{K}[X]/P\big) = \deg(P)$ par la proposition 9.106. Au final nous avons $\dim(E_a) = \deg(\chi) = \dim(E)$. Et par conséquent $E_a = E$. Cela prouver que a est un vecteur cyclique pour f.

(2) implique (3) Soit $g \in \mathcal{C}(f)$; nous devons prouver que g est un polynôme de f. Par hypothèse nous avons un vecteur cyclique que nous notons v. Nous avons un polynôme P (dépendant de g) tel que $g(v) = P(f)v$. Nous allons voir que $g = P(f)$. Soient $y \in E$ et Q un polynôme tels que $y = Q(f)v$; en notant que g commute avec $P(f)$ nous avons

$$g(y) = g\big(Q(f)v\big) = Q(f)\big(g(v)\big) = Q(f)\big(P(f)v\big) = P(f)Q(f)v = P(f)y. \tag{9.509}$$

Donc $g = P(f)$.

(3) implique (1) Nous avons les inégalités :

$$n \leqslant \dim\big(\mathcal{C}(f)\big) = \dim\big(\mathbb{K}[f]\big) = \deg(\mu) \leqslant \deg(\chi) = n. \tag{9.510}$$

La première est le lemme 9.232. Toutes les inégalités sont des égalités. En particulier $\deg(\mu) = n$, ce qui signifie que $\mu = \chi$ parce que μ est un polynôme diviseur de χ, de même degré que χ et unitaire tout comme χ.

□

Corolaire 9.234 ([80]).
En suivant les notations sur les extensions de corps de base de la section 9.12, l'endomorphisme $f\colon E \to F$ est cyclique si et seulement si l'endomorphisme $f_{\mathbb{L}}\colon E_{\mathbb{L}} \to F_{\mathbb{L}}$ est cyclique.

Démonstration. Nous savons par le théorème 9.233 qu'un endomorphisme est cyclique si et seulement si son polynôme minimal est égal à son polynôme caractéristique. Or par les propositions 9.214 et 9.215, nous savons que ces polynômes sont identiques pour f et pour $f_{\mathbb{L}}$. □

Théorème 9.235 (Similitude et extension de corps[80]).
Les applications linéaires $f, g\colon E \to E$ sont semblables si et seulement si $f_{\mathbb{L}}$ et $g_{\mathbb{L}}$ le sont.

Démonstration. En ce qui concerne le sens direct, s'il existe $m \in \mathrm{GL}(E)$ tel que $f = mgm^{-1}$ alors il suffit d'appliquer le lemme 9.210 pour avoir $f_{\mathbb{L}} = m_{\mathbb{L}} g_{\mathbb{L}} m_{\mathbb{L}}^{-1}$.

Nous considérons les invariants de similitude de f du théorème 9.221. Il existe une unique suite de polynômes unitaires μ_i $(i = 1, \dots, s)$ tels que $\mu_i \mid \mu_{i+1}$ et pour laquelle nous avons une décomposition $E = E_1 \oplus \dots \oplus E_s$ pour laquelle $f|_{E_i}\colon E_1 \to E_i$ est cyclique et de polynôme minimal μ.

Nous avons aussi $E_{\mathbb{L}} = (E_1)_{\mathbb{L}} \oplus \dots \oplus (E_s)_{\mathbb{L}}$ et les $(E_i)_{\mathbb{L}}$ sont stables sous $f_{\mathbb{L}}$ qui y sera également cyclique (corolaire 9.234). De plus le polynôme minimal de $f_{\mathbb{L}}|_{(E_i)_{\mathbb{L}}}$ est également μ_i.

Autrement dit, la suite μ_i est également la suite des invariants de similitude de $f_{\mathbb{L}}$. La remarque 9.222 nous permet de conclure que f et g sont semblables si et seulement si $f_{\mathbb{L}}$ et $g_{\mathbb{L}}$ le sont. □

9.15 Hyperplans et formes linéaires

Définition 9.236.
Si E est un espace vectoriel de dimension n, un ***hyperplan*** *de E est un sous-espace vectoriel de dimension $n-1$.*

Proposition 9.237 ([200]).
À propos d'hyperplans et de formes linéaires sur un espace vectoriel E sur le corps $\mathbb{K}$.

(1) Si φ est une forme linéaire non nulle, alors $\ker(\varphi)$ est un hyperplan.

(2) Si H est un hyperplan de E, il existe une forme linéaire dont H est le noyau :

$$H = \ker(\varphi). \tag{9.511}$$

Démonstration. En deux parties.

(1) Soit un supplémentaire A de H. Nous considérons la restriction $\varphi_A \colon A \to \mathbb{K}$. Vu que les éléments non nuls de A sont hors de H, nous avons $\varphi(x) \neq 0$ dès que x est non nul dans A. Cela implique que φ_A est surjective.

D'autre part, φ_A est également injective : si $\varphi_A(x) = \varphi_A(y)$, alors $\varphi_A(x-y) = 0$, ce qui signifie que $x - y = 0$ ou encore que $x = y$.

Donc φ_A est un isomorphisme de $\mathbb{K}$-espaces vectoriels ; nous en déduisons par le corolaire 4.37 que A est de dimension 1 sur $\mathbb{K}$, parce que $\mathbb{K}$ est de dimension 1.

(2) Nous utilisons le théorème de la base incomplète 4.11(4) pour considérer une base $\{e_i\}_{i=1,\dots,n}$ de E telle que $\operatorname{Span}\{e_1,\dots,e_{n-1}\} = H$. Nous pouvons alors considérer la forme linéaire définie par

$$\varphi(e_i) = \begin{cases} 0 & \text{si } i = 1,\dots,n-1 \\ 1 & \text{si } i = n. \end{cases} \tag{9.512}$$

Cette forme vérifie $\ker(\varphi) = H$.

□

Proposition 9.238 ([79]).
Soit un espace vectoriel E de dimension finie $n \geqslant 2$. Soit un sous-espace vectoriel V de E de dimension s. Alors V est une intersection de $n-s$ hyperplans de E.

Démonstration. Nous considérons une base de V que nous complétons[82] en une base de E : si $x = \sum_{i=1}^n x_i e_i$, nous avons $x \in V$ si et seulement si $x_{s+1} = \dots = x_n = 0$. Nous considérons les formes linéaires

$$\begin{aligned} \varphi_i \colon E &\to \mathbb{R} \\ x &\mapsto x_i, \end{aligned} \tag{9.513}$$

et nous considérons les parties $H_i = \ker(\varphi_i)$ qui sont de hyperplans par la proposition 9.237. Les H_i avec $s+1 \leqslant i \leqslant n$ sont une famille de $n-s$ hyperplans qui vérifient

$$V = \bigcap_{i=s+1}^{n} \ker(\varphi_i) \tag{9.514}$$

parce que $x \in \ker(\varphi_i)$ si et seulement si $x_i = 0$.

Donc V peut être écrit comme intersection de $n-s$ hyperplans de E. □

Proposition 9.239 ([79]).
Soit un $\mathbb{K}$-espace vectoriel E de dimension finie $n \geqslant 2$. Si H_i sont des hyperplans de E, alors

$$\dim\Big(\bigcup_{i=1}^{m} H_i\Big) \geqslant n - m. \tag{9.515}$$

82. Théorème de la base incomplète, 4.11(4).

Démonstration. N'oubliez pas de prouver que $\bigcap_{i=1}^{m} H_i$ est un espace vectoriel. À part ça, nous faisons une petite récurrence.

Pour $m = 2$ Nous savons déjà par la proposition 4.40 que

$$\dim(H_1 \cap H_2) = \dim(H_1) + \dim(H_2) - \dim(H_1 \cap H_2). \tag{9.516}$$

De plus $\dim(H_1 + H_2) \leqslant n$. En remplaçant, par les valeurs,

$$\begin{aligned}
\dim(H_1 \cap H_2) &= \dim(H_1) + \dim(H_2) - \dim(H_1 \cap H_2) && (9.517a)\\
&= n - 1 + n - 1 - \dim(H_1 + H_2) && (9.517b)\\
&\geqslant 2n - 2 - n && (9.517c)\\
&= n - 2. && (9.517d)
\end{aligned}$$

Donc $\dim(H_1 \cap H_2) \geqslant n - 2$.

La récurrence Nous calculons $\dim(H_1 \cap \ldots \cap H_m \cap H_{m+1})$ en commençant encore par la proposition 4.40 :

$$\begin{aligned}
\dim(H_1 \cap \ldots \cap H_m \cap H_{m+1}) &= \underbrace{\dim(H_1 \cap \ldots \cap H_m)}_{\leqslant n-m} + \dim(H_{m+1}) && (9.518a)\\
&\quad - \underbrace{\dim\big((H_1 \cap \ldots H_m) + H_{m+1}\big)}_{\leqslant n} && (9.518b)\\
&\geqslant n - m + (n-1) - n && (9.518c)\\
&= n - m - 1. && (9.518d)
\end{aligned}$$

C'est bon pour la récurrence.

□

9.15.1 Trouver la matrice d'une symétrie donnée

Les notions de déterminants, produit scalaire et vectoriels [83] donnent une bonne intuition géométrique des matrices. Nous pouvons alors chercher les matrices de quelques symétriques dans $\mathbb{R}^2$ ou $\mathbb{R}^3$.

9.15.1.1 Symétrie par rapport à un plan

Comment trouver par exemple la matrice A qui donne la symétrie autour du plan $z = 0$? La définition d'une telle symétrie est que les vecteurs du plan $z = 0$ ne bougent pas, tandis que les vecteurs perpendiculaires changent de signe. Ces informations vont permettre de trouver comment A agit sur une base de $\mathbb{R}^3$. En effet :

(1) Le vecteur $\begin{pmatrix}1\\0\\0\end{pmatrix}$ est dans le plan $z = 0$, donc il ne bouge pas,

(2) le vecteur $\begin{pmatrix}0\\1\\0\end{pmatrix}$ est également dans le plan, donc il ne bouge pas non plus,

(3) et le vecteur $\begin{pmatrix}0\\0\\1\end{pmatrix}$ est perpendiculaire au plan $z = 0$, donc il va changer de signe.

83. Définitions 9.8, 7.186 et 7.204.

Cela nous donne directement les valeurs de A sur la base canonique et nous permet d'écrire

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{9.519}$$

Pour écrire cela, nous avons juste mit en colonne les images des vecteurs de base. Les deux premiers n'ont pas changé et le troisième a changé.

Et si maintenant on donne un plan moins facile que $z = 0$? Le principe reste le même : il faudra trouver deux vecteurs qui sont dans le plan (et dire qu'ils ne bougent pas), et puis un vecteur qui est perpendiculaire au plan[84], et dire qu'il change de signe.

Voyons ce qu'il en est pour le plan $x = -z$. Il faut trouver deux vecteurs linéairement indépendants dans ce plan. Prenons par exemple

$$f_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}. \tag{9.520}$$

Nous avons

$$\begin{aligned} Af_1 &= f_1 \\ Af_2 &= f_2. \end{aligned} \tag{9.521}$$

Afin de trouver un vecteur perpendiculaire au plan, calculons le produit vectoriel :

$$f_3 = f_1 \times f_2 = \begin{vmatrix} e_1 & e_2 & e_3 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{vmatrix} = -e_1 - e_3 = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix}. \tag{9.522}$$

Nous avons

$$Af_3 = -f_3. \tag{9.523}$$

Afin de trouver la matrice A, il faut trouver Ae_1, Ae_2 et Ae_3. Pour ce faire, il faut d'abord écrire $\{e_1, e_2, e_3\}$ en fonction de $\{f_1, f_2, f_3\}$. La première des équations (9.520) dit que

$$f_1 = e_2. \tag{9.524}$$

Ensuite, nous avons

$$\begin{aligned} f_2 &= e_1 - e_3 \\ f_3 &= -e_1 - e_3. \end{aligned} \tag{9.525}$$

La somme de ces deux équations donne $-2e_3 = f_2 + f_3$, c'est-à-dire

$$e_3 = -\frac{f_2 + f_3}{2} \tag{9.526}$$

Et enfin, nous avons

$$e_1 = \frac{f_2 - f_3}{2}. \tag{9.527}$$

Maintenant nous pouvons calculer les images de e_1, e_2 et e_3 en faisant

$$\begin{aligned} Ae_1 &= \frac{Af_2 - Af_3}{2} = \frac{1}{2}\begin{pmatrix} 0 \\ 0 \\ -2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \\ Ae_2 &= Af_1 = f_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \\ Ae_3 &= -\frac{f_2 - f_3}{2} = -\frac{1}{2}\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}. \end{aligned} \tag{9.528}$$

84. Pour le trouver, penser au produit vectoriel.

La matrice A s'écrit maintenant en mettant les trois images trouvées en colonnes :

$$A = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}. \tag{9.529}$$

9.15.1.2 Symétrie par rapport à une droite

Le principe est exactement le même : il faut trouver trois vecteurs f_1, f_2 et f_3 sur lesquels on connaît l'action de la symétrie. Ensuite il faudra exprimer e_1, e_2 et e_3 en termes de f_1, f_2 et f_3.

Le seul problème est de trouver les trois vecteurs f_i. Le premier est tout trouvé : c'est n'importe quel vecteur sur la droite. Pour les deux autres, il faut un peu ruser parce qu'il faut impérativement qu'ils soient perpendiculaire à la droite. Pour trouver f_2, on peut écrire

$$f_2 = \begin{pmatrix} 1 \\ 0 \\ x \end{pmatrix}, \tag{9.530}$$

et puis fixer le x pour que le produit scalaire de f_2 avec f_1 soit nul. S'il n'y a pas moyen (genre si f_1 a sa troisième composante nulle), essayer avec $\begin{pmatrix} x \\ 1 \\ 0 \end{pmatrix}$. Une fois que f_2 est trouvé (il y a des milliards de choix possibles), trouver f_3 est super facile : prendre le produit vectoriel entre f_1 et f_2.

9.15.1.3 En résumé

La marche à suivre est

(1) Trouver trois vecteurs f_1, f_2 et f_3 sur lesquels on connaît l'action de la symétrie. Typiquement : des vecteurs qui sont sur l'axe ou le plan de symétrie, et puis des perpendiculaires. Pour la perpendiculaire, penser au produit scalaire et au produit vectoriel.

(2) Exprimer la base canonique e_1, e_2 et e_3 en termes de f_1, f_2, f_3.

(3) Trouver Ae_1, Ae_2 et Ae_3 en utilisant leur expression en termes des f_i, et le fait que l'on connaisse l'action de A sur les f_i.

(4) La matrice s'obtient en mettant les images des e_i en colonnes.

9.16 Théorème de Burnside

Lemme 9.240.
Soit P, un polynôme sur $\mathbb{K}$. Une racine de P est une racine simple si et seulement si elle n'est pas racine de P'.

Théorème 9.241.
Toute représentation[85] *d'un groupe abélien d'exposant fini sur $\mathbb{C}^n$ a une image finie.*

Démonstration. Étant donné que G est d'exposant fini, il existe $\alpha \in \mathbb{N}^*$ tel que $g^\alpha = e$ pour tout $g \in G$. Le polynôme $P(X) = X^\alpha - 1$ est scindé à racines simples. En effet tout polynôme sur $\mathbb{C}$ est scindé. Le fait qu'il soit à racines simples provient du lemme 9.240 parce que si $a^\alpha = 1$, alors il n'est pas possible d'avoir $\alpha a^{\alpha-1} = 0$.

Par ailleurs $P(g) = 0$. Le fait que nous ayons un polynôme annulateur de g scindé à racines simples implique que g est diagonalisable (théorème 9.128). Le fait que G soit abélien montre

85. Définition 4.149.

qu'il existe une base de $\mathbb{C}^n$ dans laquelle tous les éléments de G sont diagonaux. Nous devons par conséquent montrer qu'il existe un nombre fini de matrices de la forme

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}. \tag{9.531}$$

Nous savons que $\lambda_i^\alpha = 1$ parce que $g^\alpha = \mathbb{1}$, par conséquent chacun des λ_i est une racine de l'unité dont il n'existe qu'un nombre fini. □

Théorème 9.242 (Burnside[71, 201]).
Un sous-groupe de $\mathrm{GL}(n, \mathbb{C})$ *est fini si et seulement s'il est d'exposant*[86] *fini.*

Démonstration. Soit G un sous-groupe de $\mathrm{GL}(n, \mathbb{C})$. Si G est fini, l'ordre de ses éléments divise $|G|$ (corolaire 2.35 au théorème de Lagrange) et l'exposant est le PPCM qui est donc fini également. Le théorème est déjà démontré dans un sens.

Dans l'autre sens, nous notons $e < \infty$ l'exposant de G, et nous allons prouver que l'ensemble G est fini. Nous commençons par remarquer que tous les éléments de G sont des racines du polynôme $X^e - 1$, et ensuite nous nous lançons dans le travail.

Générateurs Le groupe G est une partie de $\mathbb{M}(n, \mathbb{C})$ dont nous considérons l'algèbre engendrée[87] $\mathcal{G}$. Soit $C_1, \dots, C_r$ une famille génératrice de $\mathcal{G}$ constituée d'éléments de G et la fonction

$$\begin{aligned} \tau \colon G &\to \mathbb{C}^r \\ A &\mapsto \big(\mathrm{Tr}(AC_1), \dots, \mathrm{Tr}(AC_r)\big). \end{aligned} \tag{9.532}$$

τ est injective Soient $A, B \in G$ tels que $\tau(A) = \tau(B)$. Si C_i est un générateur de G, nous avons $\mathrm{Tr}(AC_i) = \mathrm{Tr}(BC_i)$ et par la linéarité de la trace, nous avons

$$\mathrm{Tr}(AM) = \mathrm{Tr}(BM) \tag{9.533}$$

pour tout $M \in G$. Notons par ailleurs

$$N = AB^{-1} - \mathbb{1}, \tag{9.534}$$

qui est diagonalisable parce que $AB^{-1} \in G$ et donc est annulé par le polynôme $X^e - 1$ qui est scindé à racines simples. Du coup AB^{-1} est diagonalisable ; posons $PAB^{-1}P^{-1} = D$, alors $P\big(AB^{-1} - \mathbb{1}\big)P^{-1} = D - \mathbb{1}$ qui est encore diagonale. Donc N est diagonalisable.

Par ailleurs nous avons

$$\begin{aligned} \mathrm{Tr}\left((AB^{-1})^p\right) &= \mathrm{Tr}\left(AB^{-1}(AB^{-1})^{p-1}\right) && (9.535a) \\ &= \mathrm{Tr}\left(BB^{-1}(AB^{-1})^{p-1}\right) \quad (9.533) && (9.535b) \\ &= \mathrm{Tr}\left((AB^{-1})^{p-1}\right). && (9.535c) \end{aligned}$$

En continuant nous obtenons

$$\mathrm{Tr}\left((AB^{-1})^p\right) = \mathrm{Tr}(\mathbb{1}) = n. \tag{9.536}$$

D'autre part,

$$N^k = (AB^{-1} - \mathbb{1})^k = \sum_{p=0}^{k} \binom{p}{k} (-1)^{k-p} (AB^{-1})^p \tag{9.537}$$

86. Définition 2.18.
87. Définition 3.85.

En prenant la trace, et en tenant compte du fait que $\mathrm{Tr}\left((AB^{-1})^p\right) = n$,

$$\mathrm{Tr}(N^k) = \sum_{p=0}^{k} \binom{p}{k} (-1)^{k-p} n = n(1-1)^k = 0. \tag{9.538}$$

Donc la trace de N^k est nulle et le lemme 9.121 nous enseigne que N est alors nilpotente. Étant donné qu'elle est aussi diagonalisable, elle est nulle. Nous en concluons que $AB^{-1} = \mathbb{1}$ et donc que $A = B$. La fonction τ est donc injective.

Nombre fini de valeurs Les éléments de G sont annulés par $X^e - 1$ qui est un polynôme scindé à racines simples. Dons le polynôme minimal d'un élément de G est (a fortiori) scindé à racines simples et le théorème 9.128 nous assure alors que ces éléments sont diagonalisables. Du coup les valeurs propres des matrices de G sont des racines $e$ièmes de l'unité. Par conséquent les traces des éléments de G ne peuvent prendre qu'un nombre fini de valeurs : toutes les sommes de n racines $e$ièmes de l'unité. Mais vu que les C_i sont dans G, nous avons

$$\mathrm{Image}(\tau) = \{\mathrm{Tr}(A) \text{ tel que } A \in G\}^r, \tag{9.539}$$

qui est un ensemble fini. Par conséquent G est fini parce que τ est injective.

□

9.16.1 Théorème de Lie-Kolchin

Contrairement à ce que l'on peut parfois croire, il n'est pas vrai que toute matrice à coefficient réel est diagonalisable, même pas sur $\mathbb{C}$. La raison est qu'une telle matrice peut très bien avoir des valeurs propres multiples.

Exemple 9.243.
Le théorème 9.128 nous donne une façon simple de trouver des matrices non diagonalisables sur $\mathbb{C}$: il suffit que le polynôme minimal ne soit pas scindé à racines simples. Par exemple

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{9.540}$$

dont le polynôme caractéristique est $\chi_A = (1-X)^2$. Ce polynôme n'a manifestement pas des racines simples. Nous pouvons faire le calcul explicite pour montrer que A n'est pas diagonalisable. D'abord l'unique valeur propre de A est 1 et nous pouvons sans peine résoudre

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \tag{9.541}$$

qui revient au système

$$\begin{cases} x + y = x & \text{(9.542a)} \\ y = y. & \text{(9.542b)} \end{cases}$$

La première équation donne directement $y = 0$. Le seul espace propre est de dimension 1 et est engendré par $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. △

La remarque 9.194 donne un exemple un peu plus avancé, qui montre la multiplicité algébrique et géométrique d'une racine d'un polynôme caractéristique.

Lemme 9.244 (Trigonalisation simultanée)**.**
Une famille de matrices de $\mathrm{GL}(n, \mathbb{C})$ *commutant deux à deux est simultanément trigonalisable.*

Démonstration. Commençons par enfoncer une porte ouverte par la proposition 9.135 : toutes les matrices de $\mathrm{GL}(n,\mathbb{C})$ sont trigonalisables parce que tous les polynômes sont scindés.

Nous effectuons la démonstration par récurrence sur la dimension. Si $n = 1$ alors toutes les matrices sont triangulaires et nous ne nous posons pas de questions. Nous supposons donc $n > 1$.

Soit la famille $(A_i)_{i\in I}$ dans $\mathrm{GL}(n,\mathbb{C})$ et A_0 un de ses éléments. Nous nommons $\lambda_1,\dots,\lambda_r$ les valeurs propres distinctes de A_0. Le théorème de décomposition primaire 9.195 nous donne la somme directe d'espaces caractéristiques [88]

$$E = F_{\lambda_1}(A_0)\oplus\dots\oplus F_{\lambda_r}(A_0). \tag{9.543}$$

Nous pouvons supposer que cette somme n'est pas réduite à un seul terme. En effet si tel était le cas, A_0 serait un multiple de l'identité parce que A_0 n'aurait qu'une seule valeur propre et les sommes dans la décomposition de Dunford 9.196(3) se réduisent à un seul terme (et $p_i = \mathrm{Id}$). En particulier les dimensions des espaces $F_\lambda(A_0)$ sont strictement plus petites que n.

Vu que tous les A_i commutent avec A_0, les espaces $F_\lambda(A_0)$ sont stables par les A_i et nous pouvons trigonaliser les A_i simultanément sur chacun des $F_\lambda(A_0)$ en utilisant l'hypothèse de récurrence. □

Théorème 9.245 (Lie-Kolchin[202]).
Tout sous-groupe connexe et résoluble de $\mathrm{GL}(n,\mathbb{C})$ *est conjugué à un groupe de matrices triangulaires.*

Démonstration. Soit G un sous-groupe connexe et résoluble de $\mathrm{GL}(n,\mathbb{C})$.

Si sous-espace non trivial stable par G Nous commençons par voir ce qu'il se passe s'il existe un sous-espace vectoriel non trivial V de $\mathbb{C}^n$ stabilisé par G. Pour cela nous considérons une base de $\mathbb{C}^n$ dont les premiers éléments forment une base de V (base incomplète, théorème 4.11). Les éléments de G s'écrivent, dans cette base,

$$\begin{pmatrix} g_1 & * \\ 0 & g_2 \end{pmatrix}. \tag{9.544}$$

Les matrices g_1 et g_2 sont carrés. Nous considérons alors l'application ψ définie par

$$\begin{aligned} \psi\colon G &\to \mathrm{GL}(V) \\ g &\mapsto g_1. \end{aligned} \tag{9.545}$$

Cela est un morphisme de groupes parce que

$$\begin{pmatrix} g_1 & * \\ 0 & g_2 \end{pmatrix}\begin{pmatrix} h_1 & * \\ 0 & h_2 \end{pmatrix} = \begin{pmatrix} g_1h_1 & * \\ 0 & g_2h_2 \end{pmatrix}, \tag{9.546}$$

de telle sorte que $\psi(gh) = \psi(g)\psi(h)$.

Le groupe $\psi(G)$ est connexe et résoluble. En effet $\psi(G)$ est connexe en tant qu'image d'un connexe par une application continue (proposition 7.90). Et il est résoluble en tant qu'image d'un groupe résoluble par un homomorphisme par la proposition 2.45. Vu que $\psi(G)$ est un sous-groupe résoluble et connexe de $\mathrm{GL}(V)$ et que la dimension de V est strictement plis petite que celle de $\mathbb{C}^n$, une récurrence sur la dimension indique que $\psi(G)$ est conjugué à un groupe de matrices triangulaires. C'est-à-dire qu'il existe une base de V dans laquelle toutes les matrices g_1 (avec $g\in G$) sont triangulaires supérieures.

On fait de même avec l'application $g\mapsto g_2$, ce qui donne une base du supplémentaire de V dans laquelle les matrices g_2 sont triangulaires.

En couplant ces deux bases, nous obtenons une base de $\mathbb{C}^n$ dans laquelle toutes les matrices (9.544) (c'est-à-dire toutes les matrices de G) sont triangulaires supérieures.

88. Définition 9.192.

Sinon Nous supposons à présent que $\mathbb{C}^n$ n'a pas de sous-espaces non triviaux stables sous G. Nous posons $m = \min\{k \text{ tel que } D^k(G) = \{e\}\}$, qui existe parce que G et résoluble et que sa suite dérivée termine sur e (proposition 2.44).

Si $m = 1$ Si $m = 1$ alors G est abélien et il existe une base de G dans laquelle toutes les matrices de G sont triangulaires (lemme 9.244). Le premier vecteur d'une telle base serait stable par G, mais comme nous avons supposé qu'il n'y avait pas de sous-espaces non triviaux stabilisés par G, il faut déduire que ce vecteur stable est à lui tout seul non trivial, c'est-à-dire que $n = 1$. Dans ce cas, le théorème est démontré.

Si $m > 1$ Nous devons maintenant traiter le cas où $m > 1$. Nous posons $H = D^{m-1}(G)$; cela est un sous-groupe normal et abélien de G. Encore une fois le résultat de trigonalisation simultanée 9.244 donne une base dans laquelle tous les éléments de H sont triangulaires. En particulier le premier élément de cette base est un vecteur propre commun à toutes les matrices de H.

Soit V le sous-espace engendré par tous les vecteurs propres communs de H. Nous venons de voir que V n'est pas vide. Nous allons montrer que V est stable par G. Soient $h \in H$, $v \in V$ et $g \in G$:

$$h\big(g(v)\big) = g\,\underbrace{g^{-1}hg}_{\in H}(v) = g(\lambda v) = \lambda g(v) \tag{9.547}$$

parce que v est vecteur propre de $g^{-1}hg$. Ce que le calcul (9.547) montre est que $g(v)$ est vecteur propre de h pour la valeur propre λ. Donc $g(v) \in V$ et V est stabilisé par G. Mais comme il n'existe pas d'espaces non triviaux stabilisés par G, nous en déduisons que $V = \mathbb{C}^n$. Donc tous les vecteurs de $\mathbb{C}^n$ sont vecteurs propres communs de H. Autrement dit on a une base de diagonalisation simultanée de H.

H est dans le centre de G Montrons à présent que H est dans le centre de G, c'est-à-dire que pour tout $g \in G$ et $h \in H$ il faut $ghg^{-1} = h$. D'abord ghg^{-1} est une matrice diagonale (parce que elle est dans H) ayant les mêmes valeurs propres que h. En effet si λ est valeur propre de ghg^{-1} pour le vecteur propre v, alors

$$(ghg^{-1})(v) = \lambda v \tag{9.548a}$$
$$h\big(g^{-1}v\big) = \lambda\big(g^{-1}v\big), \tag{9.548b}$$

c'est-à-dire que λ est également valeur propre de h, pour le vecteur propre $g^{-1}v$. Mais comme h a un nombre fini de valeurs propres, il n'y a qu'un nombre fini de matrices diagonales ayant les mêmes valeurs propres que h. L'ensemble $\mathbf{Ad}(G)h$ est donc un ensemble fini. D'autre part, l'application $g \mapsto g^{-1}hg$ est continue, et G est connexe, donc l'ensemble $\mathbf{Ad}(G)h$ est connexe. Un ensemble fini et connexe dans $\mathrm{GL}(n, \mathbb{C})$ est nécessairement réduit à un seul point. Cela prouve que $ghg^{-1} = h$ pour tout $g \in G$ et $h \in H$.

Espaces propres stables pour tout G Soit $h \in H$ et W un espace propre de h (ça existe non vide parce que H est triangularisé, voir plus haut). Alors nous allons prouver que W est stable pour tous les éléments de G. En effet si $w \in W$ avec $h(w) = \lambda w$ alors en permutant g et h,

$$hg(w) = g(hw) = \lambda g(w), \tag{9.549}$$

donc $g(w)$ est aussi vecteur propre de h pour la valeurs propre λ, c'est-à-dire que $g(w) \in W$. Vu que nous supposons que $\mathbb{C}^n$ n'a pas d'espaces invariants non triviaux, nous devons conclure que $W = \mathbb{C}^n$, c'est-à-dire que H est composé d'homothéties. C'est-à-dire que pour tout $h \in H$ nous avons $h = \lambda_h \mathbb{1}$.

Contradiction sur la minimalité de m Les éléments d'un groupe dérivé sont de déterminant 1 parce que $\det(g_1 g_2 g_1^{-1} g_2^{-1}) = 1$. Par conséquent pour tout h, le nombre λ_h est une racine n^{e}de l'unité. Vu qu'il n'y a qu'une quantité finie de racines n^{e}de l'unité, le groupe H est fini et connexe et donc une fois de plus réduit à un élément, c'est-à-dire $H = \{e\}$. Cela contredit la minimalité de m et donc produit une contradiction. Nous devons donc avoir $m = 1$.

Conclusion Nous avons vu que si $\mathbb{C}^n$ avait un sous-espace non trivial fixé par G alors le théorème était démontré. Par ailleurs si $\mathbb{C}^n$ n'a pas un tel sous-espace, soit $m = 1$ (et alors le théorème est également prouvé), soit $m > 1$ et alors on a une contradiction.

Bref, le théorème est prouvé sous peine de contradiction.

□

9.17 Retour sur les formes bilinéaires et quadratiques

9.17.1 Dégénérescence d'une forme bilinéaire

Soit b, une forme bilinéaire symétrique non dégénérée sur l'espace vectoriel E de dimension n sur $\mathbb{K}$ où $\mathbb{K}$ est un corps de caractéristique différente de 2. Nous notons q la forme quadratique associée.

Définition 9.246.
Une forme bilinéaire est ***non dégénérée*** $b(x, z) = 0$ *pour tout* z *implique* $x = 0$.

Lemme 9.247.
Soit b *une forme bilinéaire non dégénérée. Si* x *et* y *sont tels que* $b(x, z) = b(y, z)$ *pour tout* z*, alors* $x = y$.

Démonstration. C'est immédiat du fait de la linéarité en le premier argument et de la non-dégénérescence : si $b(x, z) - b(y, z) = 0$ alors

$$b(x - y, z) = 0 \tag{9.550}$$

pour tout z, ce qui implique $x - y = 0$. □

Proposition 9.248.
Une forme bilinéaire est non-dénénérée[89] *si et seulement si sa matrice associée est inversible.*

Démonstration. Nous savons que la matrice associée est symétrique et qu'elle peut donc être diagonalisée (théorème 9.150). En nous plaçant dans une base de diagonalisation, nous devons prouver que la forme est non-dégénérée si et seulement si les éléments diagonaux de la matrice sont tous non nuls.

Écrivons $b(x, z)$ en choisissant pour z le vecteur de base e_k de composantes $(e_k)_j = \delta_{kj}$:

$$b(x, e_k) = \sum_{ij} x_i (e_k)_j = \sum_i b_{ik} x_i = b_{kk} x_k. \tag{9.551}$$

Si b est dégénérée et si x est un vecteur non nul (disons que la composante x_i est non nulle) de E tel que $b(x, z) = 0$ pour tout $z \in E$, alors $b_{ii} = 0$, ce qui montre que la matrice de b n'est pas inversible.

Réciproquement si la matrice de b est inversible, alors tous les b_{kk} sont différents de zéro, et le seul vecteur x tel que $b_{kk} x_k = 0$ pour tout k est le vecteur nul. □

9.17.2 Isométries

Voici un théorème pas toujours bien énoncé dans les cours de physique qui font de la relativité. Au moment de « prouver » les transformations de Lorentz[90], beaucoup oublient de justifier pourquoi elles devraient être linéaires.

Théorème 9.249 ([203])**.**
Une isométrie d'une forme bilinéaire[91] *non dégénérée est linéaire.*

89. Définition 9.246.
90. Théorème 17.168.
91. Définition 9.178.

Démonstration. Soient une forme bilinéaire non-dégénérée b sur l'espace vectoriel E ainsi qu'une isométrie f pour icelle. Soit $z \in E$; étant donné que f est bijective nous pouvons considérer l'élément $f^{-1}(z) \in E$ et calculer

$$\begin{aligned} b\big(f(x+y), z\big) &= b\big(f(x+y), f(f^{-1}(z))\big) && (9.552a)\\ &= b(x+y, f^{-1}(z)) && (9.552b)\\ &= b(x, f^{-1}(z)) + b(y, f^{-1}(z)) && (9.552c)\\ &= b(f(x), z) + b(f(y), z) && (9.552d)\\ &= b\big(f(x) + f(y), z\big), && (9.552e) \end{aligned}$$

donc $f(x+y) = f(x) + f(y)$ par le lemme 9.247.

De la même façon on trouve $b\big(f(\lambda x), z\big) = b\big(\lambda f(x), z\big)$ qui prouve que $f(\lambda x) = \lambda f(x)$ et donc que f est linéaire. □

Exemple 9.250.
Une isométrie peut ne pas être linéaire quand la forme bilinéaire est dégénérée. Par exemple pour la forme bilinéaire sur $\mathbb{R}^2$ donnée par

$$b\big((a,b),(x,y)\big) = ax, \tag{9.553}$$

nous pouvons faire

$$f(x,y) = \begin{pmatrix} x \\ \lambda(x,y) \end{pmatrix} \tag{9.554}$$

où λ est n'importe quoi. △

Théorème 9.251.
Soit un espace vectoriel E muni d'une forme quadratique q. Soit une isométrie $f\colon E \to E$ pour q. Alors

(1) si $f(0) = 0$, alors f est linéaire ;

(2) si $f(0) \neq 0$ alors f est affine[92].

Démonstration. Nous considérons la forme bilinéaire associée b. Si $f(0) = 0$, nous savons par le lemme 9.179 que $b\big(f(x), f(y)\big) = b(x,y)$. La proposition 9.249 nous dit alors que f est linéaire.

Si $f(0) \neq 0$, alors nous posons $g(x) = f(x) - f(0)$ qui vérifie $g(0) = 0$ et

$$q\big(g(x) - g(y)\big) = q\big(f(x) - f(0) - f(y) + f(0)\big) = q(x-y). \tag{9.555}$$

Nous pouvons donc appliquer le premier point à g, déduire que g est linéaire et donc que f est affine. C'est la caractérisation du lemme 8.54 des fonctions affines. □

Nous pouvons maintenant particulariser tout cela au cas de $\mathbb{R}^n$ muni du produit scalaire usuel et de la norme associée pour voir quel résultat nous avons à peine prouvé.

Lemme 9.252 ([204])**.**
Une isométrie d'un espace vectoriel normé de dimension finie est bijective.

Démonstration. Si $f\colon E \to E$ est une isométrie, elle est linéaire par le théorème 9.249. Elle vérifie également $\|f(x)\| = \|x\|$, et donc $f(x) = 0$ si et seulement si $x = 0$, c'est-à-dire que f est injective. Elle est alors bijective par le corolaire 4.41 du théorème du rang. □

Nous notons ici $T(n)$ le groupe des translations sur $\mathbb{R}^n$. Un élément de $T(n)$ est une translation τ_v donnée par un vecteur v et agissant sur $\mathbb{R}^n$ par

$$\begin{aligned} \tau_v\colon \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto x + v. \end{aligned} \tag{9.556}$$

92. Définition 8.9.

Ce groupe est isomorphe au groupe abélien $(\mathbb{R}^n, +)$, et nous allons souvent identifier τ_v à v.

Vous savez par culture générale que les isométries de $\mathbb{R}^n$ pour le produit scalaire usuel sont les matrices orthogonales. En voici une petite généralisation (pensez à $\eta = \mathbb{1}$ dans le cas du produit scalaire usuel).

Proposition 9.253.
Soit une forme bilinéaire b sur $\mathbb{R}^n$ de matrice symétrique η. Si A est la matrice d'une application linéaire $\mathbb{R}^n \to \mathbb{R}^n$ telle que

$$b(Ax, Ay) = b(x, y) \tag{9.557}$$

pour tout $x, y \in \mathbb{R}^n$, alors

$$A^t \eta A = \eta. \tag{9.558}$$

Démonstration. En suivant la formule générale (9.321),

$$b(Ax, Ay) = \sum_{ij} \eta_{ij} (Ax)_i (Ay)_j = \sum_{ijkl} \eta_{ij} A_{ik} A_{jl} x_k y_l. \tag{9.559}$$

En imposant que ce soit égal à $\sum_{kl} \eta_{kl} \eta_{kl} x_k y_l$ pour tout x, y nous avons la contrainte

$$\sum_{ij} \eta_{ij} A_{ik} A_{jl} = \eta_{kl} \tag{9.560}$$

qui signifie exactement $A^t \eta A = \eta$. □

9.17.3 Pseudo-réduction simultanée

Corolaire 9.254 (Pseudo-réduction simultanée[205])**.**
Soient $A, B \in \mathrm{S}(n, \mathbb{R})$ avec A définie positive[93]*. Alors il existe $Q \in \mathrm{GL}(n, \mathbb{R})$ telle que $Q^t B Q$ soit diagonale et $Q^t A Q = \mathbb{1}$.*

Démonstration. Nous allons noter $x \cdot y$ le produit scalaire usuel de $\mathbb{R}^n$ et $\{e_i\}_{i=1,\dots,n}$ sa base canonique.

Vu que A est définie positive, l'expression $\langle x, y \rangle = x \cdot Ay$ donne un produit scalaire sur $\mathbb{R}^n$. Nous avons donc deux produits scalaires sur $\mathbb{R}^n$, et nous allons travailler avec les deux.

La proposition 7.216 appliquée à l'espace euclidien $(\mathbb{R}^n, \langle ., . \rangle)$ dit qu'il existe une base de $\mathbb{R}^n$ orthonormée $(f_i)_{i=1,\dots,n}$ pour ce produit scalaire. Nous considérons l'application linéaire P définie par

$$Pe_i = f_i. \tag{9.561}$$

Nous démontrons à présent que $P^t A P = \mathbb{1}$. Pour cela, nous calculons

$$\begin{aligned}
\delta_{ij} &= \langle f_i, f_j \rangle && \text{(9.562a)}\\
&= f_i \cdot A f_j && \text{(9.562b)}\\
&= P e_i \cdot A P e_j && \text{(9.562c)}\\
&= e_i \cdot P^t A P e_j && \text{(9.562d)}\\
&= (P^t A P)_{ij}. && \text{(9.562e)}
\end{aligned}$$

Justifications :

— Pour (9.562a), la base (f_j) est orthonormée pour le produit scalaire $\langle ., . \rangle$.

— Pour (9.562d), la proposition 4.136 sur la transposée.

— Pour (9.562e), la formule du produit scalaire usuel pour avoir les éléments de matrice, proposition 4.123.

93. Définition 9.152.

La matrice P^tBP est une matrice symétrique, donc le théorème spectral 9.150 nous donne une matrice $R \in \mathrm{O}(n,\mathbb{R})$ telle que R^tP^tBPR soit diagonale. En posant maintenant $Q = PR$ nous avons la matrice cherchée. □

Remarque 9.255.
Plusieurs remarques

(1) Nous n'avons pas prouvé l'existence d'une matrice P telle que $P^{-1}BP$ et $P^{-1}AP$ soient diagonales. Au contraire, nous avons Q^tBQ et Q^tAQ qui sont diagonales. Tant que Q n'est pas orthogonales, ce n'est pas la même chose.

Autrement dit, nous n'avons pas ici une réelle diagonalisation, parce que les matrices A et B ne sont pas semblables à des matrices diagonales. Voir les définitions 9.125 (diagonalisable) et 9.119 (semblable).

C'est pour cela que nous parlons de *pseudo*-diagonalisation.

(2) Dans le même ordre d'idée, la démonstration de la pseudo-diagonalisation simultanée parle clairement de formes bilinéaires, et non d'endomorphismes. Or en comparant les lois de transformations (4.238) et (9.325), nous voyons bien que la réduction en passant par Q^tAQ est bien une réduction de forme bilinéaire et non une réduction d'endomorphismes.

(3) Nous avons prouvé la pseudo-réduction simultanée comme corolaire du théorème de diagonalisation des matrices symétriques 9.150. Il aurait aussi pu être vu comme un corolaire du théorème spectral 9.267 sur les opérateurs autoadjoints via son corolaire 9.268.

9.17.4 Topologie

La topologie considérée sur $Q(E)$ est celle de la norme

$$N(q) = \sup_{\|x\|_E=1} |q(x)|, \tag{9.563}$$

qui du point de vue de $S(n,\mathbb{R})$ est

$$N(A) = \sup_{\|x\|_E=1} |x^tAx|. \tag{9.564}$$

Notons que à droite, c'est la valeur absolue usuelle sur $\mathbb{R}$.

Proposition 9.256.
Soit $\{e_i\}$ une base de E. L'application

$$\begin{aligned} \phi\colon Q(E) &\to S(n,\mathbb{R}) \\ q &\mapsto \big(b(e_i,e_j)\big)_{i,j} \end{aligned} \tag{9.565}$$

où b est forme bilinéaire associée à q est une bijection linéaire et continue[94].

Démonstration. Si $\phi(q) = \phi(q')$; alors

$$q(x) = \sum_{i,j} \phi(q)_{ij}x_ix_j = \sum_{i,j} \phi(q')_{ij}x_ix_j = q'(x). \tag{9.566}$$

Donc $q = q'$. L'application ϕ est donc injective

De plus elle est surjective parce que si $B \in S(n,\mathbb{R})$ alors la forme quadratique

$$q(x) = \sum_{i,j} B_{ij}x_ix_j \tag{9.567}$$

a évidemment B comme matrice associée. L'application ϕ est donc surjective.

94. Pour les topologies des normes (9.563) et (9.564).

Notre application ϕ est de plus linéaire parce que l'association d'une forme quadratique à la forme bilinéaire associée est linéaire.

En ce qui concerne la continuité, nous la prouvons en zéro en considérant une suite convergente $q_n \xrightarrow{Q(E)} 0$. C'est-à-dire que

$$\sup_{\|x\|=1} |q_n(x)| \to 0. \tag{9.568}$$

Nous rappelons l'identité de polarisation :

$$b_n(x,y) = \frac{1}{2}\big(q_n(x-y) - q(x) - q(y)\big). \tag{9.569}$$

En ce qui concerne deux des trois termes, il n'y a pas de problèmes :

$$\big|\phi(q_n)_{ij}\big| = \big|b_n(e_i,e_j)\big| \leqslant \frac{1}{2}\big|b_n(e_i-e_j)\big| + \frac{1}{2}\big|q_n(e_i)\big| + \frac{1}{2}\big|q_n(e_j)\big|. \tag{9.570}$$

Si n est assez grand, nous avons tout de suite

$$\big|\phi(q_n)_{ij}\big| \leqslant \frac{1}{2}\big|q_n(e_i-e_j)\big| + \epsilon. \tag{9.571}$$

Nous définissons e_{ij} et α_{ij} de telle sorte que $e_i - e_j = \alpha_{ij}e_{ij}$ avec $\|e_{ij}\| = 1$. Si $\alpha = \max\{\alpha_{ij}, 1\}$ alors nous avons

$$q_n(e_i - e_j) = \alpha_{ij}^2 q_n(e_{ij}) \leqslant \alpha^2 q_n(e_{ij}). \tag{9.572}$$

Il suffit maintenant de prendre n assez grand pour avoir $\sup_{\|x\|=1}|q_n(x)| \leqslant \frac{\epsilon}{\alpha^2}$ pour avoir

$$\big|\phi(q_n)_{ij}\big| \leqslant \frac{\epsilon}{2} + \frac{\epsilon}{\alpha^2}. \tag{9.573}$$

□

9.17.5 Diagonalisation

Proposition 9.257.
Dans la base de diagonalisation de sa matrice associée, une forme quadratique a la forme

$$q(x) = \sum_i \lambda_i x_i^2 \tag{9.574}$$

où les λ_i sont les valeurs propres de la matrice associée à q.

Démonstration. Soit q une forme quadratique et b la forme bilinéaire associée. Si $\{f_i\}$ est une base de diagonalisation[95] de la matrice de b alors dans cette base nous avons

$$q(x) = b(x,x) = \sum_{ij} x_i x_j b(f_i, f_j) = \sum_i \lambda_i x_i^2 \tag{9.575}$$

où les λ_i sont les valeurs propres de la matrice de b. □

Notons que si nous choisissons une autre base de diagonalisation, les λ_i ne changement pas (à part l'ordre éventuellement). Cela pour dire que nous nous permettrons de parler des **valeurs propres** d'une forme quadratique comme étant les valeurs propres de la matrice associée.

95. Qui existe parce que la matrice est symétrique, théorème 9.150.

9.17.6 Isotropie

Définition 9.258 (Isotropie)**.**
*Un vecteur est **isotrope** pour b s'il est perpendiculaire à lui-même ; en d'autres termes, x est isotrope si et seulement si $b(x,x)=0$. Un sous-espace $W\subset E$ est **totalement isotrope** si pour tout $x,y\in W$, nous avons $b(x,y)=0$.*

*Le **cône isotrope** de b est l'ensemble de ses vecteurs isotropes :*

$$C(b)=\{x\in E \text{ tel que } b(x,x)=0\}. \tag{9.576}$$

Nous introduisons quelques notations. D'abord pour $y\in E$ nous notons

$$\begin{aligned}\Phi_y\colon E&\to\mathbb{R}\\ x&\mapsto b(x,y)\end{aligned} \tag{9.577}$$

et ensuite

$$\begin{aligned}\Phi\colon E&\to E^*\\ y&\mapsto \Phi_y.\end{aligned} \tag{9.578}$$

Définition 9.259.
*Le fait pour une forme bilinéaire b d'être dégénérée signifie que l'application Φ n'est pas injective. Le **noyau** de la forme bilinéaire est celui de Φ, c'est-à-dire*

$$\ker(b)=\{z\in E \text{ tel que } b(z,y)=0\,\forall y\in E\}. \tag{9.579}$$

Autrement dit, $\ker(b)=E^\perp$ où le perpendiculaire est pris par rapport à b.

Notons tout de même que nous utilisons la notation $\perp$ même si b est dégénérée et éventuellement pas positive ; c'est-à-dire même si la formule $(x,y)\mapsto b(x,y)$ ne fournit pas un produit scalaire.

Proposition 9.260 ([206])**.**
Soit b une forme bilinéaire et symétrique. Alors

(1) $\ker(b)\subset C(b)$ (cône d'isotropie, définition 9.258)

(2) si b est positive alors $\ker(b)=C(b)$.

Démonstration. (1) Si $z\in\ker(b)$ alors pour tout $y\in E$ nous avons $b(z,y)=0$. En particulier pour $y=z$ nous avons $b(z,z,)=0$ et donc $z\in C(b)$.

(2) Soit b positive et $x\in C(b)$. Par l'inégalité de Cauchy-Schwarz (proposition 7.191) nous avons

$$|b(x,y)|\leqslant\sqrt{b(x,x)b(y,y)}=0. \tag{9.580}$$

Donc pour tout y nous avons $b(x,y)=0$.

□

9.17.7 Inégalité de Minkowski

Ce qui est couramment nommé « inégalité de Minkowski » est la proposition 26.37 dans les espaces L^p. Nous allons en donner ici un cas très particulier.

Proposition 9.261.
Si q est une forme quadratique sur $\mathbb{R}^n$ et si $x,y\in\mathbb{R}^n$ alors

$$\sqrt{q(x+y)}\leqslant\sqrt{q(x)}+\sqrt{q(y)}. \tag{9.581}$$

Démonstration. La proposition 9.257 nous permet de « diagonaliser » la forme quadratique q. Quitte à ne plus avoir une base orthonormale, nous pouvons renormaliser les vecteurs de base pour avoir

$$q(x)=\sum_i x_i^2. \tag{9.582}$$

Le résultat n'est donc rien d'autre que l'inégalité triangulaire pour la norme euclidienne usuelle, laquelle est démontrée dans la proposition 7.192. □

9.17.8 Ellipsoïde

Lemme 9.262.
Toute matrice peut être décomposée de façon unique en une partie symétrique et une partie antisymétrique. Cette décomposition est donnée par

$$S = \frac{M + M^t}{2}, \quad A = \frac{M - M^t}{2} \tag{9.583}$$

Démonstration. L'existence est une vérification immédiate de $S + A = M$ en utilisant (9.583). Pour l'unicité, si $S + A = S' + A'$ alors $S - S' = A - A'$. Mais $S - S'$ est symétrique et $A - A'$ est antisymétrique ; l'égalité implique l'annulation des deux membres, c'est-à-dire $S = S'$ et $A = A'$. □

Définition 9.263.
*Un **ellipsoïde** dans $\mathbb{R}^n$ centré en v est le lieu des points x vérifiant l'équation*

$$\langle x - v, M(x - v)\rangle = 1 \tag{9.584}$$

où M est une matrice symétrique strictement définie positive[96].

Lorsque nous parlons d'ellipsoïde plein, *il suffit de changer l'égalité en une inégalité.*

Remarque 9.264.
Le fait que M soit symétrique n'est pas tout à fait obligatoire ; la chose important est que toutes les valeurs propres soient strictement positives. En effet si M a toutes ses valeurs propres strictement positives, nous nommons S la partie symétrique de M et A la partie antisymétrique (lemme 9.262). Alors pour tout $x \in \mathbb{R}^n$ nous avons

$$x^t Ax = \langle x, Ax\rangle = \langle A^t x, x\rangle = -\langle Ax, x\rangle = -\langle x, Ax\rangle, \tag{9.585}$$

donc $x^t Ax = 0$. L'équation $x^t Mx = 1$ est donc équivalente à $x^t Sx = 1$ (elles ont les mêmes solutions).

De plus S reste strictement définie positive parce que pour tout $x \in \mathbb{R}^n$ nous avons

$$0 < x^t Mx = x^t Sx. \tag{9.586}$$

Proposition 9.265.
Si $\mathcal{E}$ est un ellipsoïde centrée à l'origine, il existe une base de $\mathbb{R}^n$ dans laquelle son équation est :

$$\sum_{i=1}^{n} \frac{x_i^2}{a_i^2} = 1. \tag{9.587}$$

Démonstration. Nous avons une matrice symétrique strictement définie positive S telle que l'équation soit $\langle x, Sx\rangle = 1$. Le théorème spectral 9.150 nous fournit une base orthonormale $\{e_i\}$ dans laquelle $Se_i = \lambda_i e_i$ avec $\lambda_i > 0$. En substituant dans l'équation $\langle x, Sx\rangle = 1$ nous trouvons l'équation

$$\sum_i \lambda_i x_i^2 = 1. \tag{9.588}$$

En posant $a_i = \frac{1}{\sqrt{\lambda_i}}$, nous trouvons le résultat. Cette définition des a_i est toujours possible parce que $\lambda_i > 0$. □

Corolaire 9.266.
Un ellipsoïde plein centré en l'origine admet une équation de la forme $q(x) \leqslant 1$ où q est une forme quadratique strictement définie positive.

Pour rappel de notation, l'ensemble des formes quadratiques strictement définies positives sur l'espace vectoriel E est noté $Q^{++}(E)$.

96. Définition 9.152.

Démonstration. Soit $\{e_i\}$ une base de $\mathbb{R}^n$ telle que l'ellipsoïde $\mathcal{E}$ ait pour équation

$$\sum_{i=1}^{n} \frac{x_i^2}{a_i^2} \leqslant 1. \tag{9.589}$$

Nous considérons la forme quadratique

$$\begin{aligned} q\colon \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto \sum_{i=1}^{n} \frac{\langle x, e_i\rangle^2}{a_i^2}. \end{aligned} \tag{9.590}$$

Nous avons évidemment $\mathcal{E} = \{x \in \mathbb{R}^n \text{ tel que } q(x) \leqslant 1\}$. De plus la forme q est strictement définie positive parce que dès que $x \neq 0$, au moins un des produits scalaires $\langle x, e_i\rangle$ est non nul et $q(x) > 0$. □

9.18 Théorème spectral autoadjoint

Théorème 9.267 (Théorème spectral autoadjoint)**.**
Un endomorphisme autoadjoint d'un espace euclidien

(1) est diagonalisable dans une base orthonormée,

(2) a son spectre réel.

Démonstration. Nous procédons par récurrence sur la dimension de E, et nous commençons par $n = 1$ [97]. Soit donc $f\colon E \to E$ avec $\langle f(x), y\rangle = \langle x, f(y)\rangle$. Étant donné que f est également linéaire, il existe $\lambda \in \mathbb{R}$ tel que $f(x) = \lambda x$ pour tout $x \in E$. Tous les vecteurs de E sont donc vecteurs propres de f.

Passons à la récurrence. Nous considérons $\dim(E) = n + 1$ et $f \in \mathrm{S}(E)$. Nous considérons la forme bilinéaire symétrique Φ_f et la forme quadratique associée ϕ_f. Pour rappel,

$$\Phi_f(x, y) = \langle x, f(y)\rangle \tag{9.591a}$$

$$\phi_f(x) = \Phi_f(x, x). \tag{9.591b}$$

Et nous allons laisser tomber les indices f pour noter simplement Φ et ϕ. Étant donné que $\overline{B(0,1)}$ est compacte et que ϕ est continue, il existe $x_0 \in \overline{B(0,1)}$ tel que

$$\lambda = \phi(x_0) = \sup_{x \in \overline{B(0,1)}} \phi(x). \tag{9.592}$$

Notons aussi que $\|x_0\| = 1$: le maximum est pris sur le bord. Nous posons

$$g = \lambda \,\mathrm{Id} - f \tag{9.593}$$

ainsi que

$$\Phi_1(x, y) = \langle x, g(y)\rangle. \tag{9.594}$$

Cela est une forme bilinéaire et symétrique parce que

$$\Phi_1(y, x) = \langle y, g(x)\rangle = \langle g(y), x\rangle = \langle x, g(y)\rangle = \Phi_1(x, y) \tag{9.595}$$

où nous avons utilisé le fait que g était autoadjoint et la symétrie du produit scalaire. De plus Φ_1 est semi-définie positive parce que

$$\Phi_1(x, x) = \langle x, \lambda x - f(x)\rangle = \lambda\|x\|^2 - \phi(x). \tag{9.596}$$

97. Dans [54], l'auteur commence avec $n = 0$ mais moi je n'en ai pas le courage..

Vu que λ est le maximum, nous avons tout de suite $\Phi_1(x) \geqslant 0$ tant que $\|x\| = 1$. Et si x n'est pas de norme 1, c'est le même prix parce qu'on se ramène à $\|x\| = 1$ en multipliant par un nombre positif. Attention cependant :

$$\Phi_1(x_0, x_0) = \lambda\|x_0\|^2 - \phi(x_0) = 0. \tag{9.597}$$

Donc Φ_1 a un noyau contenant x_0 par la proposition 9.260. Nous en déduisons que $\text{Image}(g) \neq E$ en effet, $x_0 \in \text{Image}(g)^\perp$, mais nous avons la proposition 4.135 sur les dimensions :

$$\dim E = \dim(\text{Image}(g)) + \dim(\text{Image}(g)^\perp). \tag{9.598}$$

Vu que $\text{Image}(g)^\perp$ est un espace vectoriel non réduit à $\{0\}$, la dimension de $\text{Image}(g)$ ne peut pas être celle de E. L'endomorphisme g n'étant pas surjectif, il ne peut pas être injectif non plus parce que nous sommes en dimension finie ; il existe donc $e_1 \in E$ tel que $g(e_1) = 0$ et tant qu'à faire nous choisissons $\|e_1\| = 1$ (ici la norme est bien celle de l'espace euclidien considéré). Par définition,

$$f(e_1) = \lambda e_1, \tag{9.599}$$

c'est-à-dire que $\lambda \in \text{Spec}(f)$. Et ϕ étant une forme quadratique réelle nous avons $\lambda \in \mathbb{R}$.

Nous posons à présent $H = \text{Span}\{e_1\}^\perp$. C'est un sous-espace stable par f parce que si $x \in H$ alors

$$\langle e_1, f(x)\rangle = \langle f(e_1 j), x\rangle = \lambda\langle e_1, x\rangle = 0. \tag{9.600}$$

Nous pouvons donc considérer la restriction de f à H : $f_H \colon H \to H$. Cet endomorphisme est bilinéaire et symétrique sur l'espace H de dimension inférieure à celle de E, donc la récurrence nous donne une base orthonormée

$$\{e_2, \dots, e_n\} \tag{9.601}$$

de vecteurs propres de f_H. De plus les valeurs propres sont réelles, toujours par récurrence. Donc

$$\text{Spec}(f) = \{\lambda\} \cup \text{Spec}(f_H) \subset \mathbb{R}. \tag{9.602}$$

Notons pour être complet que si $i \geqslant 2$ alors

$$\langle e_1, e_i\rangle = 0 \tag{9.603}$$

parce que le vecteur e_i est par construction choisi dans l'espace $H = e_1^\perp$. Nous avons donc bien une base orthonormée de E construite sur des vecteurs propres de f. □

Corolaire 9.268.
Soit E un espace vectoriel ainsi que ϕ et ψ des formes quadratiques sur E avec ψ définie positive. Alors il existe une base ψ-orthonormale dans laquelle ϕ est diagonale.

Démonstration. Il suffit de considérer l'espace euclidien E muni du produit scalaire $\langle x, y\rangle = \psi(x, y)$. Ensuite nous diagonalisons la matrice (symétrique) de ϕ pour ce produit scalaire à l'aide du théorème 9.267. □

Définition 9.269.
*Dans le cas de $V = \mathbb{R}^m$ nous avons un produit scalaire canonique. Soient u et v, deux vecteurs de $\mathbb{R}^m$. Le **produit scalaire** de u et v, noté $\langle u, v\rangle$ ou $u \cdot v$ est le réel*

$$\langle u, v\rangle = \sum_{k=1}^{m} u_k v_k = u_1 v_1 + u_2 v_2 + \cdots + u_m v_n. \tag{9.604}$$

Calculons par exemple le produit scalaire de deux vecteurs de la base canonique : $\langle e_i, e_j\rangle$. En utilisant la formule de définition et le fait que $(e_i)_k = \delta_{ik}$, nous avons

$$\langle e_i, e_j\rangle = \sum_{k=1}^{m} \delta_{ik}\delta_{jk}. \tag{9.605}$$

Nous pouvons effectuer la somme sur k en remarquant qu'à cause du δ_{ik}, seul le terme avec $k = i$ n'est pas nul. Effectuer la somme revient donc à remplacer tous les k par des i :

$$\langle e_i, e_j \rangle = \delta_{ii}\delta_{ji} = \delta_{ji}. \tag{9.606}$$

Une des propriétés intéressantes du produit scalaire est qu'il permet de décomposer un vecteur dans une base, comme nous le montre la proposition suivante.

Proposition 9.270.
Si nous notons v_i les composantes du vecteur v, c'est-à-dire si $v = \sum_{i=1}^m v_i e_i$, alors nous avons $v_j = \langle v, e_j \rangle$.

Démonstration.

$$v \cdot e_j = \sum_{i=1}^m \langle v_i e_i, e_j \rangle = \sum_{i=1}^m v_i \langle e_i, e_j \rangle = \sum_{i=1}^m v_i \delta_{ij} \tag{9.607}$$

En effectuant la somme sur i dans le membre de droite de l'équation (9.607), tous les termes sont nuls sauf celui où $i = j$; il reste donc

$$v \cdot e_j = v_j. \tag{9.608}$$

□

Le produit scalaire ne dépend en réalité pas de la base orthogonale choisie.

Lemme 9.271.
Si $\{e_i\}$ est la base canonique, et si $\{f_i\}$ est une autre base orthonormale, alors si u et v sont deux vecteurs de $\mathbb{R}^m$, nous avons

$$\sum_i u_i v_j = \sum_i u'_i v'_j \tag{9.609}$$

où u_i sont les composantes de u dans la base $\{e_i\}$ et u'_i sont celles dans la base $\{f_i\}$.

Démonstration. La preuve demande un peu d'algèbre linéaire. Étant donné que $\{f_i\}$ est une base orthonormale, il existe une matrice A orthogonale ($AA^t = \mathbb{1}$) telle que $u'_i = \sum_j A_{ij} u_j$ et idem pour v. Nous avons alors

$$\begin{aligned}
\sum_i u'_i v'_j &= \sum_i \left(\sum_j A_{ij} u_j \right) \left(\sum_k A_{ik} v_k \right) \\
&= \sum_{ijk} A_{ij} A_{ik} u_j v_k \\
&= \sum_{jk} \underbrace{\sum_i (A^t)_{ji} A_{ik}}_{=\delta_{jk}} u_j v_k \\
&= \sum_{jk} \delta_{jk} u_j v_k \\
&= \sum_k u_j v_k.
\end{aligned} \tag{9.610}$$

□

Cette proposition nous permet de réellement parler du produit scalaire entre deux vecteurs de façon intrinsèque sans nous soucier de la base dans laquelle nous regardons les vecteurs.

Nous dirons que deux vecteurs sont **orthogonaux** lorsque leur produit scalaire est nul. Nous écrivons que $u \perp v$ lorsque $\langle u, v \rangle = 0$.

Définition 9.272.
*La **norme euclidienne** d'un élément de $\mathbb{R}^m$ est définie par $\|u\| = \sqrt{u \cdot u}$.*

Cette définition est motivée par le fait que le produit scalaire $u \cdot u$ donne exactement la norme usuelle donnée par le théorème de Pythagore :

$$u \cdot u = \sum_{i=1}^{m} u_i u_i = \sum_{i=1}^{m} u_i^2 = u_1^2 + u_2^2 + \cdots + u_m^2. \tag{9.611}$$

Le fait que $e_i \cdot e_j = \delta_{ij}$ signifie que la base canonique est **orthonormée**, c'est-à-dire que les vecteurs de la base canonique sont orthogonaux deux à deux et qu'ils ont tout 1 comme norme.

Lemme 9.273.
Pour tout $u \in \mathbb{R}^m$, il existe un $\xi \in \mathbb{R}^m$ tel que $\|u\| = \xi \cdot u$ et $\|\xi\| = 1$.

Démonstration. Vérifions que le vecteur $\xi = u/\|u\|$ ait les propriétés requises. D'abord $\|\xi\| = 1$ parce que $u \cdot u = \|u\|^2$. Ensuite

$$\xi \cdot u = \frac{u \cdot u}{\|u\|} = \frac{\|u\|^2}{\|u\|} = \|u\|. \tag{9.612}$$

□

9.19 Système d'équations linéaires : méthode de Gauss

Pour résoudre un système d'équations linéaires, on procède comme suit :

(1) Écrire le système sous forme matricielle.

$$\text{p.ex.} \begin{cases} 2x + 3y &= 5 \\ x + 2y &= 4 \end{cases} \Leftrightarrow \left(\begin{array}{cc|c} 2 & 3 & 5 \\ 1 & 2 & 4 \end{array}\right)$$

(2) Se ramener à une matrice avec un maximum de 0 dans la partie de gauche en utilisant les transformations admissibles :

(2a) Remplacer une ligne par elle-même + un multiple d'une autre ;

$$\text{p.ex.} \left(\begin{array}{cc|c} 2 & 3 & 5 \\ 1 & 2 & 4 \end{array}\right) \overset{L_1 - 2.L_2 \mapsto L_1'}{\Longrightarrow} \left(\begin{array}{cc|c} 0 & -1 & -3 \\ 1 & 2 & 4 \end{array}\right)$$

(2b) Remplacer une ligne par un multiple d'elle-même ;

$$\text{p.ex.} \left(\begin{array}{cc|c} 0 & -1 & -3 \\ 1 & 2 & 4 \end{array}\right) \overset{-L_1 \mapsto L_1'}{\Longrightarrow} \left(\begin{array}{cc|c} 0 & 1 & 3 \\ 1 & 2 & 4 \end{array}\right)$$

(2c) Permuter des lignes.

$$\text{p.ex.} \left(\begin{array}{cc|c} 0 & 1 & 3 \\ 1 & 0 & -2 \end{array}\right) \overset{L_1 \mapsto L_2' \text{ et } L_2 \mapsto L_1'}{\Longrightarrow} \left(\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 3 \end{array}\right)$$

(3) Retransformer la matrice obtenue en système d'équations.

$$\text{p.ex.} \left(\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 3 \end{array}\right) \Leftrightarrow \begin{cases} x &= -2 \\ y &= 3 \end{cases}$$

Remarque 9.274. — Si on obtient une ligne de zéros, on peut l'enlever :

$$\text{p.ex.} \left(\begin{array}{ccc|c} 3 & 4 & -2 & 2 \\ 4 & -1 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \Leftrightarrow \left(\begin{array}{ccc|c} 3 & 4 & -2 & 2 \\ 4 & -1 & 3 & 0 \end{array}\right)$$

— Si on obtient une ligne de zéros suivie d'un nombre non-nul, le système d'équations n'a pas de solution :

$$\text{p.ex.} \left(\begin{array}{ccc|c} 3 & 4 & -2 & 2 \\ 4 & -1 & 3 & 0 \\ 0 & 0 & 0 & 7 \end{array}\right) \Leftrightarrow \begin{cases} \cdots \\ \cdots \\ 0x + 0y + 0z = 7 \end{cases} \Rightarrow \textbf{Impossible}$$

— Si on a moins d'équations que d'inconnues, alors il y a une infinité de solutions qui dépendent d'un ou plusieurs paramètres :

$$\text{p.ex.} \left(\begin{array}{ccc|c} 1 & 0 & -2 & 2 \\ 0 & 1 & 3 & 0 \end{array}\right) \Leftrightarrow \begin{cases} x - 2z = 2 \\ y + 3z = 0 \end{cases} \Leftrightarrow \begin{cases} x = 2 + 2\lambda \\ y = -3\lambda \\ z = \lambda \end{cases}$$

Chapitre 10

Espaces vectoriels normés

Plusieurs choses sur les espaces vectoriels normés (dont la définition 7.113) ont déjà été vues dans la section 7.16. Voir aussi le thème 7.

On fixe maintenant une définition largement utilisée dans la suite.

Définition 10.1.
*Soient U et V, deux ouverts d'un espace vectoriel normé. Une application f de U dans V est un **difféomorphisme** si elle est bijective, différentiable et dont l'inverse $f^{-1} : V \to U$ est aussi différentiable.*

Remarque 10.2.
Il n'est pas possible d'avoir une application inversible d'un ouvert de $\mathbb{R}^m$ vers un ouvert de $\mathbb{R}^n$ si $m \neq n$. Il n'y a donc pas de notion de difféomorphismes entre ouverts de dimensions différentes.

10.1 Norme opérateur

La proposition suivante donne une norme (au sens de la définition 7.113) sur $\mathcal{L}(V, W)$ dès que V et W sont des espaces vectoriels normés.

Proposition-définition 10.3 (Norme opérateur[207], thème 43)**.**
Soit une application linéaire $T\colon V \to W$, et le nombre

$$\|T\|_{\mathcal{L}} = \sup_{\substack{x\in V\\ x\neq 0}} \frac{\|T(x)\|_W}{\|x\|_V}. \tag{10.1}$$

(1) Si V est de dimension finie, alors $\|T\|_{\mathcal{L}} < \infty$.

(2) L'application $T \mapsto \|T\|_{\mathcal{L}}$ est une norme sur l'espace vectoriel des applications linéaires $V \to W$.

(3) Nous avons la formule

$$\|T\|_{\mathcal{L}} = \sup_{x\in V} \frac{\|T(x)\|_W}{\|x\|_V} = \sup_{\|x\|_V=1} \|T(x)\|_W \tag{10.2}$$

*Le nombre $\|T\|_{\mathcal{L}}$ est la **norme opérateur** de T. Nous disons que cette norme est **subordonnée** aux normes choisies sur V et W.*

Démonstration. Si V est de dimension finie alors l'ensemble $\{\|x\| = 1\}$ est compact par le théorème de Borel-Lebesgue 7.127. Alors la fonction

$$x \mapsto \frac{\|T(x)\|}{\|x\|} \tag{10.3}$$

est une fonction continue sur un compact. Le corolaire 7.300 nous dit alors qu'elle est bornée. Le supremum est donc un nombre réel fini.

Nous vérifions que l'application $\|.\|$ de $\mathcal{L}(V, W)$ dans $\mathbb{R}$ ainsi définie est effectivement une norme.

(1) $\|T\|_{\mathcal{L}} = 0$ signifie que $\|T(x)\| = 0$ pour tout x dans V. Comme $\| \cdot \|_W$ est une norme nous concluons que $T(x) = 0_n$ pour tout x dans V, donc T est l'application nulle.

(2) Pour tout a dans $\mathbb{R}$ et tout T dans $\mathcal{L}(V, W)$ nous avons

$$\|aT\|_{\mathcal{L}} = \sup_{\|x\|_V \leqslant 1} \|aT(x)\|_W = |a| \sup_{\|x\|_V \leqslant 1} \|T(x)\|_W = |a| \|T\|_{\mathcal{L}}. \tag{10.4}$$

(3) Pour tous T_1 et T_2 dans $\mathcal{L}(V, W)$ nous avons

$$\begin{aligned} \|T_1 + T_2\|_{\mathcal{L}} &= \sup_{\|x\| \leqslant 1} \|T_1(x) + T_2(x)\| \leqslant \\ &\leqslant \sup_{\|x\| \leqslant 1} \|T_1(x)\| + \sup_{\|x\| \leqslant 1} \|T_2(x)\| \\ &= \|T_1\| \|T_2\|. \end{aligned}$$

Enfin nous prouvons la formule alternative (10.2). Nous allons montrer que les ensembles sur lesquels ont prend le supremum sont en réalité les mêmes :

$$\underbrace{\left\{ \frac{\|Ax\|}{\|x\|} \right\}_{x \neq 0}}_{A} = \underbrace{\{\|Ax\| \text{ tel que } \|x\| = 1\}}_{B}. \tag{10.5}$$

Attention : ce sont des sous-ensembles de réels ; pas de sous-ensembles de $\mathbb{M}(\mathbb{R})$ ou des sous-ensembles de $\mathbb{R}^n$.

Pour la première inclusion, prenons un élément de A, et prouvons qu'il est dans B. C'est-à-dire que nous prenons $x \in V$ et nous considérons le nombre $\|Ax\|/\|x\|$. Le vecteur $y = x/\|x\|$ est un vecteur de norme 1, donc la norme de Ay est un élément de B, mais

$$\|Ay\| = \frac{\|Ax\|}{\|x\|}. \tag{10.6}$$

Nous avons donc $A \subset B$.

L'inclusion $B \subset A$ est immédiate. □

En d'autres termes, il y a autant de normes opérateur sur $\mathcal{L}(E, F)$ qu'il y a de paires de choix de normes sur E et F. En particulier, cela donne lieu à toutes les normes $\|A\|_p$ qui correspondent aux normes $\|.\|_p$ sur $\mathbb{R}^n$.

Exemple 10.4.
Voyons la norme opérateur subordonnée à la norme $\|x\|_\infty = \max_i |x_i|$ sur $\mathbb{C}^n$. Par définition (et surtout par la propriété 10.3(3)),

$$\|A\|_\infty = \sup_{\|x\|_\infty = 1} = \|Ax\|_\infty. \tag{10.7}$$

Vu que $(Ax)_i = \sum_k A_{ik} x_k$, lorsque $\|x\|_\infty \leqslant 1$ nous avons $|(Ax)_i| \leqslant \sum_k |A_{ik}|$. Donc nous avons toujours

$$\|A\|_\infty \leqslant \max_i \sum_k |A_{ik}|. \tag{10.8}$$

△

Définition 10.5.
*La **topologie forte** sur l'espace des opérateurs est la topologie de la norme opérateur.*

Lorsque nous considérons un espace vectoriel d'applications linéaires, nous considérons toujours [1] dessus la topologie liée à cette norme.

1. Sauf lorsque les événements nous forceront à trahir.

Il existe aussi la **topologie faible** donnée par la notion de convergence[2] $A_i \to A$ si et seulement si $A_i x \to Ax$ pour tout $x \in E$.

Problèmes et choses à faire

Je crois, mais demande confirmation, que la topologie faible est celle des semi-normes $\{p_v\}_{v\in E}$ données par $p_v(A) = \|A\|$. En effet la notion de convergence associée par la proposition 7.329 est $A_i \to A$ si et seulement si $p_v(A_i - A) \to 0$. Cette condition signifie $\|A_i(v) - A(v)\| \to 0$, c'est-à-dire $A_i(v) \to A(v)$.

Si le lecteur veut parler de cela au jury d'un concours, il est évident qu'il devra être capable d'ajouter des petits symboles au-dessus de toutes les flèches « $\to$ » du paragraphe précédent pour indiquer pour quelles topologies sont les convergences dont on parle.

Remarque 10.6.
Il faut noter que la topologie faible n'est pas une topologie métrique. Cela même si la condition $A_i x \to Ax$, elle, est métrique vu qu'elle est écrite dans E.

Dans le cas où E est de dimension infinie, la topologie faible est réellement différente de la topologie forte. Nous verrons à la sous-section 24.3.6 que dans le cas des projections sur un espaces de Hilbert, l'égalité

$$\sum_{i=1}^{\infty} \mathtt{proj}_{u_i} = \mathrm{Id} \tag{10.9}$$

est vraie pour la topologie faible, mais pas pour la topologie forte.

10.1.1 Norme d'algèbre

Définition 10.7 (Norme d'algèbre[207])**.**
Si A est une algèbre[3]*, une **norme d'algèbre** sur A est une norme telle que pour toute $x, y \in A$,*

$$\|xy\| \leqslant \|x\|\|y\|. \tag{10.10}$$

La norme opérateur est une norme d'algèbre, comme nous le verrons dans le lemme 10.13.

Un des intérêts d'utiliser une norme d'algèbre est que l'on a l'inégalité $\|x^k\| \leqslant \|x\|^k$. Cela sera particulièrement utile lors de l'étude des séries entières, voir par exemple 14.12.

Définition 10.8 ([208])**.**
*Le **rayon spectral** d'une matrice carrée A, noté $\rho(A)$, est défini de la manière suivante :*

$$\rho(A) = \max_i |\lambda_i| \tag{10.11}$$

où les λ_i sont les valeurs propres de A.

10.9.
Quelques remarques sur la définition du rayon spectral.

- Même si A est une matrice réelle, les valeurs propres sont dans $\mathbb{C}$. Donc dans (10.11), $|\lambda_i|$ est le module dans $\mathbb{C}$ de λ_i.
- Vu que les valeurs propres de A sont les racines de son polynôme caractéristique (théorème 9.111), il y en a un nombre fini et le maximum est bien défini.
- La définition s'applique uniquement pour les espaces de dimension finie.

Lemme 10.10.
Soient des espaces vectoriels normés E et F, sur les corps $\mathbb{R}$ ou $\mathbb{C}$. Pour tout $A \in \mathcal{L}(E, F)$, et pour tout $u \in E$ nous avons la majoration

$$\|Au\| \leqslant \|A\|\|u\| \tag{10.12}$$

où la norme sur A est la norme opérateur subordonnée à la norme sur u.

2. Est-ce qu'on peut décrire cette topologie à partir de ses ouverts ? Facilement ?
3. Définition 3.83.

Démonstration. Si $u \in E$ alors, étant donné que le supremum d'un ensemble est plus grand ou égal à chacun de éléments qui le compose,

$$\|A\| = \sup_{x \in E} \frac{\|Ax\|}{\|x\|} \geqslant \frac{\|Au\|}{\|u\|}, \tag{10.13}$$

donc le résultat annoncé : $\|Au\| \leqslant \|A\|\|u\|$. □

Le lemme suivant est valable en dimension infinie. Nous en toucherons un mot dans l'exemple 10.31.

Lemme 10.11.
Soient des espaces vectoriels normés E et F. Soit $x \in E$. Alors l'application d'évaluation

$$\begin{aligned} ev_x \colon \mathcal{L}(E,F) &\to F \\ f &\mapsto f(x) \end{aligned} \tag{10.14}$$

est continue.

Démonstration. Si $x = 0$, alors par linéarité de f nous avons $ev_0(f) = 0$ pour tout f. Donc d'accord pour la continuité.

Soit une suite convergente $f_k \xrightarrow{\mathcal{L}(E,F)} f$. Nous voulons prouver que $ev_x(f_k) \xrightarrow{F} ev_x(f)$, c'est-à-dire que

$$\lim_{k \to \infty} \|f_k(x) - f(x)\| = 0. \tag{10.15}$$

Par hypothèse si k est grand, alors $\|f_k - f\|_{\mathcal{L}(E,F)} \leqslant \epsilon$, c'est-à-dire que[4]

$$\sup_{y \in E} \frac{\|f_k(y) - f(y)\|}{\|y\|} \leqslant \epsilon. \tag{10.16}$$

En particulier pour notre x nous avons

$$\frac{\|f_k(x) - f(x)\|}{\|x\|} \leqslant \epsilon, \tag{10.17}$$

c'est-à-dire $\|f_k(x) - f(x)\| \leqslant \|x\|\epsilon$. Vu que $\|x\|$ est une simple constante et que ϵ est arbitraire, cela implique $f_k(x) \to f(x)$. □

10.1.2 Matrices, spectre et norme

La lien entre la norme opérateur d'une matrice et son spectre sera entre autres utilisé pour étudier le conditionnement de problèmes numériques. Voir la définition 33.105 et par exemple son lien avec la résolution numérique de systèmes linéaires dans la proposition 33.109.

Proposition 10.12 ([208]).
Soit une matrice $A \in \mathbb{M}(n, \mathbb{C})$ de rayon spectral $\rho(A)$. Soit une norme $\|.\|$ sur $\mathbb{C}^n$ et la norme opérateur correspondante. Alors

$$\rho(A) \leqslant \|A^k\|^{1/k} \tag{10.18}$$

pour tout $k \in \mathbb{N}$.

Démonstration. Soit $v \in \mathbb{C}^n$ et $\lambda \in \mathbb{C}$ un couple vecteur-valeur propre. Nous avons $\|Av\| = |\lambda|\|v\|$ et aussi

$$|\lambda|^k \|v\| = \|\lambda^k v\| = \|A^k v\| \leqslant \|A^k\|\|v\|. \tag{10.19}$$

La dernière inégalité est due au fait que nous avons choisi sur $\mathbb{M}(n, \mathbb{C})$ la norme subordonnée à celle choisie sur $\mathbb{C}^n$, via le lemme 10.10. Nous simplifions par $\|v\|$ et obtenons $|\lambda| \leqslant \|A^k\|^{1/k}$. Étant donné que $\rho(A)$ est la maximum de tous les λ possibles, la majoration passe au maximum :

$$\rho(A) \leqslant \|A^k\|^{1/k}. \tag{10.20}$$

□

4. Définition 10.3 de la norme sur $\mathcal{L}(E, F)$.

Lemme 10.13 (La norme opérateur est une norme d'algèbre[1])**.**
Soient des espaces vectoriels normés E, F et G. Soient des opérateurs linéaires bornés $B\colon E \to F$, $A\colon F \to G$. Alors

$$\|AB\| \leqslant \|A\|\|B\|. \tag{10.21}$$

C'est à dire que la norme opérateur est une norme d'algèbre[5]*.*

Démonstration. Nous avons les (in)égalités suivantes :

$$\begin{aligned}
\|AB\| &= \sup_{x\in E} \frac{\|ABx\|_G}{\|x\|_E} && (10.22a)\\
&= \sup_{\substack{x\in E\\ Bx\neq 0}} \frac{\|ABx\|}{\|x\|}\frac{\|Bx\|_F}{\|Bx\|_F} && (10.22b)\\
&= \sup_{\substack{x\in E\\ Bx\neq 0}} \frac{\|ABx\|}{\|Bx\|}\frac{\|Bx\|}{\|x\|} && (10.22c)\\
&\leqslant \underbrace{\sup_{\substack{x\in E\\ Bx\neq 0}} \frac{\|ABx\|}{\|Bx\|}}_{\leqslant\|A\|} \underbrace{\sup_{\substack{y\in E\\ By\neq 0}} \frac{\|Bx\|}{\|y\|}}_{=\|B\|} && (10.22d)\\
&\leqslant \|A\|\|B\|. && (10.22e)
\end{aligned}$$

La dernière inégalité provient que dans $\sup_{\substack{x\in E\\ Bx\neq 0}} \|ABx\|/\|x\|$, le supremum est pris sur un ensemble plus petit que celui sur lequel porte la définition de la norme de A : seulement l'image de B au lieu de tout l'espace de départ de A. □

Proposition 10.14.
Soient deux espaces vectoriels normés E et V. Soient des applications continues $f, g\colon E \to \mathrm{End}(V)$. Alors l'application

$$\begin{aligned}
\psi\colon E &\to \mathrm{End}(V)\\
x &\mapsto f(x)\circ g(x)
\end{aligned} \tag{10.23}$$

est continue.

Démonstration. Soit une suite $x_k \xrightarrow{E} x$. Nous devons montrer que $\psi(x_k) \xrightarrow{\mathrm{End}(V)} \psi(x)$. Pour cela nous utilisons le lemme 10.13 qui indique que la norme opérateur est une norme d'algèbre. Nous avons :

$$\begin{aligned}
\|\psi(x_k) - \psi(x)\| &= \|f(x_k)\circ g(x_k) - f(x)\circ g(x)\| && (10.24a)\\
&\leqslant \|f(x_k)\circ g(x_k) - f(x_k)\circ g(x)\| + \|f(x_k)\circ g(x) - f(x)\circ g(x)\| && (10.24b)\\
&= \|f(x_k)\circ \big(g(x_k)\circ g(x)\big)\| + \|\big(f(x_k) - f(x)\big)\circ g(x)\| && (10.24c)\\
&\leqslant \|f(x_k)\|\|g(x_k) - g(x)\| + \|f(x_k) - f(x)\|\|g(x)\|. && (10.24d)
\end{aligned}$$

Pour $k \to \infty$ nous avons $\|f(x_k) \to \|f(x)\|\|$, $\|f(x_k) - f(x)\| \to 0$ (parce que f est continue) et similaire avec g. Donc le tout tend vers zéro. □

10.1.3 Rayon spectral

La chose impressionnante dans la proposition suivante est que $\rho(A)$ est définit indépendamment du choix de la norme sur $\mathbb{M}(n, \mathbb{K})$ ou sur $\mathbb{K}$. Lorsque nous écrivons $\|A\|$, nous disons implicitement qu'une norme a été choisie sur $\mathbb{K}$ et que nous avons pris la norme subordonnée sur $\mathbb{M}(n, \mathbb{K})$.

Proposition 10.15 ([209])**.**
Soit A une matrice de $\mathbb{M}(n, \mathbb{R})$ ou $\mathbb{M}(n, \mathbb{C})$. Alors

$$\rho(A) \leqslant \|A\|. \tag{10.25}$$

5. Définition 10.7.

Démonstration. Nous devons séparer les cas suivant que le corps de base soit $\mathbb{R}$ ou $\mathbb{C}$.

Pour $A \in \mathbb{M}(n, \mathbb{C})$ Soit λ une valeur propre de A telle que $|\lambda|$ soit la plus grande. Nous avons donc $\rho(A) = |\lambda|$. Soit un vecteur propre $u \in \mathbb{C}^n$ pour la valeur propre λ. En prenant la norme sur l'égalité $Au = \lambda u$, et en utilisant le lemme 10.10,

$$|\lambda|\|u\| = \|Au\| \leqslant \|A\|\|u\|. \tag{10.26}$$

Donc $|\lambda| \leqslant \|A\|$ et $\rho(A) \leqslant \|A\|$.

Pour $A \in \mathbb{M}(n, \mathbb{R})$ L'endroit qui coince dans le raisonnement fait pour $\mathbb{M}(n, \mathbb{C})$ est que certes $A \in \mathbb{M}(n, \mathbb{R})$ possède une plus grande valeur propre en module et qu'un vecteur propre lui est associé. Mais ce vecteur propre est à priori dans $\mathbb{C}^n$, et non dans $\mathbb{R}^n$. Nous pouvons donc écrire $Au = \lambda u$, mais pas $\|Au\| = |\lambda|\|u\|$ parce que nous ne savons pas quelle norme prendre sur $\mathbb{C}^n$.

Il n'est pas certain que nous ayons une norme sur $\mathbb{C}^n$ qui se réduit sur $\mathbb{R}^n$ à celle choisie implicitement dans l'énoncé. Nous allons donc ruser un peu.

Soit une norme N sur $\mathbb{C}^n$ [6]. Nous nommons également N la norme subordonnée sur $\mathbb{M}(n, \mathbb{C})$ et la norme restreinte sur $\mathbb{M}(n, \mathbb{R})$. Vu que N est une norme sur $\mathbb{M}(n, \mathbb{R})$ et que ce dernier est de dimension finie, le théorème 7.224 nous indique que N est équivalente à $\|.\|$. Il existe donc $C > 0$ tel que

$$N(B) \leqslant C\|B\| \tag{10.27}$$

pour tout $B \in \mathbb{M}(n, \mathbb{R})$. Nous avons maintenant

$$\rho(A)^m \leqslant N(A^m) \leqslant C\|A^m\| \leqslant C\|A\|^m. \tag{10.28}$$

Justifications

— Par la proposition 10.12.

— Parce que $A^m \in \mathbb{M}(n, \mathbb{R})$ et la relation (10.27).

— Par itération du lemme 10.13.

Nous avons donc $\rho(A) \leqslant C^{1/m}\|A\|$ pour tout $m \in \mathbb{N}$. En prenant $m \to \infty$ et en tenant compte de $C^{1/m} \to 1$ nous trouvons $\rho(A) \leqslant \|A\|$.

□

Lemme 10.16 ([209]).
Soit $A \in \mathbb{M}(n, \mathbb{K})$ avec $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. Soit $\epsilon > 0$. Il existe une norme algébrique sur $\mathbb{M}(n, \mathbb{K})$ telle que

$$N(A) \leqslant \rho(A) + \epsilon. \tag{10.29}$$

Démonstration. Soit par le lemme 9.137 une matrice inversible U telle que $T = UAU^{-1}$ soit triangulaire supérieure, avec les valeurs propres sur la diagonale. Notons que même si $A \in \mathbb{M}(n, \mathbb{R})$, les matrices U et T sont à priori complexes.

Soit $s \in \mathbb{R}$ ainsi que les matrices

$$D_s = \operatorname{diag}(1, s^{-1}, s^{-2}, \ldots, s^{1-n}) \tag{10.30}$$

et $T_s = D_s T D_s^{-1}$. Nous fixerons un choix de s plus tard.

La norme que nous considérons est :

$$N(B) = \|(D_s U)B(D_s U)^{-1}\|_\infty \tag{10.31}$$

où $\|.\|_\infty$ est la norme sur $\mathbb{M}(, n\mathbb{K})$ subordonnée à la norme $\|.\|_\infty$ sur $\mathbb{K}^n$ dont nous avons déjà parlé dans l'exemple 10.4. Cela est bien une norme parce que

6. Il y en a plein, par exemple celle du produit scalaire $\langle x, y\rangle = \sum_k x_k \bar{y}_k$.

— Nous avons $\|B\|_\infty = 0$ si et seulement si $B = 0$, et vu que (D_sU) est inversible nous avons $(D_sU)B(D_sU)^{-1} = 0$ si et seulement si $B = 0$.
— $N(\lambda B) = |\lambda| N(B)$.
— Pour l'inégalité triangulaire :

$$N(B+C) = \|(D_sU)B(D_sU)^{-1} + (D_sU)C(D_sU)^{-1}\|_\infty \tag{10.32a}$$
$$\leqslant \|(D_sU)B(D_sU)^{-1}\|_\infty + \|(D_sU)C(D_sU)^{-1}\|_\infty \tag{10.32b}$$
$$= N(B) + N(C). \tag{10.32c}$$

En ce qui concerne la matrice A elle-même, nous avons

$$N(A) = \|(D_sU)A(D_sU)^{-1}\|_\infty = \|T_s\|_\infty. \tag{10.33}$$

C'est le moment de se demander comment se présente la matrice T_s. En tenant compte du fait que $(D_s)_{ik} = \delta_{ik}s^{1-i}$ nous avons

$$(T_s)_{ij} = \sum_{kl}(D_s)_{ik}T_{kl}(D_s^{-1})_{lj} = T_{ij}s^{j-i}. \tag{10.34}$$

La matrice T est encore triangulaire supérieure avec les valeurs propres de A sur la diagonale. Les éléments au-dessus de la diagonale sont tous multipliés par au moins s. Il est donc possible de choisir s suffisamment petit pour avoir [7]

$$\sum_{j=i+1}^{n} |(T_s)_{ij}| < \epsilon \tag{10.35}$$

Avec ce choix, la formule 10.8 donne

$$N(T_s) \leqslant \max_i \sum_k |(T_s)_{ik}| \leqslant \epsilon + \rho(A). \tag{10.36}$$

En effet le ϵ vient de la somme sur toute la ligne sauf la diagonale (c'est-à-dire la partie $k \neq i$) et du choix (10.35) pour s. Le $\rho(A)$ provient du dernier terme de la somme (le terme sur la diagonale) qui est une valeur propre de A, donc majorable par $\rho(A)$.

Nous devons encore prouver que N est une norme algébrique. Pour cela nous allons montrer qu'elle est subordonnée à la norme

$$\begin{aligned} n\colon \mathbb{K}^n &\to \mathbb{R}^+ \\ v &\mapsto \|(UD_s)v\|_\infty. \end{aligned} \tag{10.37}$$

Cela sera suffisant pour avoir une norme algébrique par le lemme 10.13. La norme n sur $\mathbb{K}^n$ produit la norme suivante sur $\mathbb{M}(n,\mathbb{K})$:

$$n(B) = \sup_{v\neq 0}\frac{n(B)}{n(v)} = \sup_{v\neq 0}\frac{\|(UD_s)Bv\|_\infty}{\|UD_sv\|_\infty}. \tag{10.38}$$

Vu que UD_s est inversible nous pouvons effectuer le changement de variables $v \mapsto (UD_s)^{-1}v$ pour écrire

$$n(B) = \sup_{v\neq 0}\frac{\|(UD_s)B(UD_s)^{-1}v\|_\infty}{\|(UD_s)(UD_s)^{-1}v\|_\infty} = \sup_{v\neq 0}\frac{\|(UD_s)B(UD_s)^{-1}v\|_\infty}{\|v\|_\infty} = \|(UD_s)B(UD_s)^{-1}\|_\infty = N(B). \tag{10.39}$$

□

Proposition 10.17.
Si $A \in \mathbb{M}(n,\mathbb{R})$ alors $\rho(A)^m = \rho(A^m)$ pour tout $m \in \mathbb{N}$.

7. Il me semble qu'il manque un module dans [209].

Démonstration. La matrice A peut être vue dans $\mathbb{M}(n,\mathbb{C})$ et nous pouvons lui appliquer le corollaire 9.143 :

$$\operatorname{Spec}(A^k) = \{\lambda^k \text{ tel que } \lambda \in \operatorname{Spec}(A)\}. \tag{10.40}$$

À noter qu'il n'y a pas de magie : le spectre de la matrice réelle A est déjà défini en voyant A comme matrice complexe. Le spectre dont il est question dans (10.40) est bien celui dont on parle dans la définition du rayon spectral.

Nous avons ensuite :

$$\begin{aligned}
\rho(A^k) &= \max\{|\lambda| \text{ tel que } \lambda \in \operatorname{Spec}(A^k)\} && (10.41a)\\
&= \max\{|\lambda^k| \text{ tel que } \lambda \in \operatorname{Spec}(A)\} && (10.41b)\\
&= \max\{|\lambda|^k \text{ tel que } \lambda \in \operatorname{Spec}(A)\} && (10.41c)\\
&= \rho(A)^k. && (10.41d)
\end{aligned}$$

□

Proposition 10.18 (Bornée si et seulement si continue[210])**.**
Soient E et F des espaces vectoriels normés. Une application linéaire $E \to F$ est bornée si et seulement si elle est continue.

Démonstration. Nous commençons par supposer que A est bornée. Par le lemme 10.13, pour tout $x, y \in E$, nous avons

$$\|A(x) - A(y)\| = \|A(x-y)\| \leqslant \|A\|\|x-y\|. \tag{10.42}$$

En particulier si $x_n \xrightarrow{E} x$ alors

$$0 \leqslant \|A(x_n) - A(x)\| \leqslant \|A\|\|x_n - x\| \to 0 \tag{10.43}$$

et A est continue en vertu de la caractérisation séquentielle de la continuité, proposition 7.80.

Nous supposons maintenant que $\|A\|$ n'est pas borné : l'ensemble $\{\|A(x)\| \text{ tel que } \|x\| = 1\}$ contient des valeurs arbitrairement grandes. Alors pour tout $k \geqslant 1$ il existe $x_k \in B(0,1)$ tel que $\|A(x_k)\| > k$. La suite x_k/k tend vers zéro parce que $\|x_k\| = 1$, mais $\|A(x_k)\| \geqslant 1$ pour tout k. Cela montre que A n'est pas continue. □

Définition 10.19 ([211])**.**
Soient E et F deux espaces vectoriels normés.

- *L'ensemble des applications linéaires $E \to F$ est noté $\mathcal{L}(E,F)$.*
- *Un **morphisme** est une application linéaire $E \to F$ continue pour la topologie de la norme opérateur. Nous avons vu dans la proposition 10.18 que la continuité était équivalente à être bornée. L'ensemble des morphismes est noté $L(E,F)$.*
- *Un **isomorphisme** est un morphisme continu inversible dont l'inverse est continu. Nous notons $\mathrm{GL}(E,F)$ l'ensemble des isomorphismes entre E et F.*

Le point important de la définition 10.19 est la continuité. En dimension infine, la continuité n'est par exemple pas équivalente à l'inversibilité (penser à $e_k \mapsto ke_k$).

Si V est un espace vectoriel normé, nous avons déjà défini son dual topologique V' comme étant l'ensemble des applications linéaires continues $V \to \mathbb{C}$ ou $V \to \mathbb{R}$ selon le corps de base de V. C'est la définition 4.129.

Proposition 10.20.
Soient un espace vectoriel normé V et un élément $v \in V$ vérifiant $\|v\| = 1$. Il existe une forme $\varphi \in V'$ telle que $\|\varphi\| = 1$ et $\varphi(v) = 1$.

Démonstration. Nous allons utiliser le théorème de la base incomplète 4.21. Pour cela nous considérons $I = V$ et la partie clairement génératrice $G = \{e_i = i\}_{i\in I}$ (si vous avez bien suivi, $G = V$ en

fait ; rien de bien profond). Nous considérons ensuite $I_0 = \{v\}$. Le théorème de la base incomplète nous donne l'existence de I_1 tel que $I_0 \subset I_I \subset I$ et tel que $B = \{e_i\}_{i\in I_1}$ est une base.

Tout cela pour dire que $B = \{e_i\}_{i\in I_1}$ est une base contenant v. Nous allons aussi éventuellement redéfinir la norme de e_i pour avoir $\|e_i\| = 1$. Cette renormalisation n'affecte pas le fait que $v \in B$.

Nous passons maintenant à la définition de $\varphi\colon V \to \mathbb{K}$. Pour $x \in V$ nous commençons par écrire

$$x = \sum_{j\in J} x_j e_j \tag{10.44}$$

et nous posons

$$\varphi(x) = \begin{cases} x_v & \text{si } v \in J \\ 0 & \text{sinon.} \end{cases} \tag{10.45}$$

Cette définition a un sens par la partie unicité de la proposition 4.6 de décomposition d'un élément dans une base.

Nous devons calculer la norme de φ. Par la proposition 10.3(3) nous avons

$$\|\varphi\| = \sup_{\|x\|=1} |\varphi(x)|. \tag{10.46}$$

Avec $x = v$ nous avons $\varphi(x) = 1$ et donc $\|\varphi\| \geqslant 1$.

Nous devons encore montrer que $\|\varphi\| \leqslant 1$. Un élément $x \in V$ s'écrit toujours sous la forme

$$x = \sum_{i\in J} x_j e_j \tag{10.47}$$

pour un certain J fini dans I_1 et pour certains $x_j \in \mathbb{K}$. Pour un tel x nous avons $\varphi(x) = x_v$. Si $|\varphi(x)| \geqslant 1$, alors $|x_v| \geqslant 1$, mais alors

$$\|x\| \leqslant \sum_{j\in J} |x_j|\|e_j\| = \sum_{j\in J} |x_j| \geqslant |x_v| > 1, \tag{10.48}$$

ce qui fait que ce x ne participe pas au supremum (10.46).

Notons que φ est continue (et donc bien dans V') parce qu'elle est bornée (proposition 10.18). □

10.1.4 Normes de matrices et d'applications linéaires

Théorème 10.21 (Norme matricielle et rayon spectral[212])**.**
La norme 2 d'une matrice est liée au rayon spectral de la façon suivante :

$$\|A\|_2 = \sqrt{\rho(A^t A)} \tag{10.49}$$

ou plus généralement par $\|A\|_2 = \sqrt{\rho(A^* A)}$.

Lemme 10.22.
Soit une matrice $A \in \mathbb{M}(n, \mathbb{R})$ qui est symétrique, strictement définie positive. Soient λ_{min} et λ_{max} les plus petites et plus grandes valeurs propres. Alors

$$\|A\|_2 = \lambda_{max} \qquad \text{et} \qquad \|A^{-1}\|_2 = \frac{1}{\lambda_{min}}. \tag{10.50a}$$

Démonstration. Soient les vecteurs $v_1, \dots, v_n$ formant une base orthonormée de vecteurs propres[8] de A. Nous notons v_{max} celui de λ_{max}. Nous avons :

$$\|A\|_2 \geqslant \|Av_{max}\| = |\lambda_{max}|\|v_{max}\| = |\lambda_{max}| = \lambda_{max}. \tag{10.51}$$

8. Possible par le théorème spectral 9.150.

Voilà l'inégalité dans un sens. Montrons l'inégalité dans l'autre sens. Soit $x = \sum_i x_i v_i$ avec $\|x\|_2 = 1$. Alors

$$\|Ax\| = \|\sum_i x_i \lambda_i v_i\| \leqslant \sqrt{\sum_i x_i^2 \lambda_i^2} \leqslant \lambda_{max} \sqrt{\sum_i x_i^2} = \lambda_{max}. \tag{10.52}$$

En ce qui concerne l'affirmation pour la norme de A^{-1}, il suffit de remarquer que ses valeurs propres sont les inverses des valeurs propres de A. □

Proposition 10.23.
La fonction

$$\begin{aligned} f\colon \mathbb{M}(n,\mathbb{R}) \times \mathbb{M}(n,\mathbb{R}) &\to \mathbb{R} \\ (X,Y) &\mapsto \mathrm{Tr}(X^tY) \end{aligned} \tag{10.53}$$

est un produit scalaire sur $\mathbb{M}(n,\mathbb{R})$.

Démonstration. Il faut vérifier la définition 7.186.

— La bilinéarité est la linéarité de la trace.

— La symétrie de f est le fait que $\mathrm{Tr}(A^t) = \mathrm{Tr}(A)$.

— L'application f est définie positive parce que si $X \in \mathbb{M}$, alors X^tX est symétrique définie positive, donc diagonalisable avec des nombres positifs sur la diagonale. La trace étant un invariant de similitude, nous avons $f(X,X) = \mathrm{Tr}(X^tX) \geqslant 0$. De plus si $\mathrm{Tr}(X^tX) = 0$, alors $X^tX = 0$ (pour la même raison de diagonalisation). Mais alors $\|Xu\| = 0$ pour tout $u \in E$, ce qui signifie que $X = 0$.

□

Exemple 10.24.
Soit $m = n$, un point λ dans $\mathbb{R}$ et T_λ l'application linéaire définie par $T_\lambda(x) = \lambda x$. La norme de T_λ est alors

$$\|T_\lambda\|_{\mathcal{L}} = \sup_{\|x\|_{\mathbb{R}^m} \leqslant 1} \|\lambda x\|_{\mathbb{R}^n} = |\lambda|.$$

Notez que T_λ n'est rien d'autre que l'homothétie de rapport λ dans $\mathbb{R}^m$. △

Exemple 10.25.
Toutes les isométries ont norme 1. En effet si T est une isométrie, $\|Tx\| = \|x\|$. En ce qui concerne la norme de T nous avons alors

$$\|T\| = \sup_{x \in \mathbb{R}^2} \frac{\|T(x)\|}{\|x\|} = \sup_{x \in \mathbb{R}^2} \frac{\|x\|}{\|x\|} = 1. \tag{10.54}$$

△

Exemple 10.26.
Soit $m = n$, un point b dans $\mathbb{R}^m$ et T_b l'application linéaire définie par $T_b(x) = b \cdot x$ (petit exercice : vérifiez qu'il s'agit vraiment d'une application linéaire). La norme de T_b satisfait les inégalités suivantes

$$\|T_b\|_{\mathcal{L}} = \sup_{\|x\|_{\mathbb{R}^m} \leqslant 1} \|b \cdot x\|_{\mathbb{R}^n} \leqslant \sup_{\|x\|_{\mathbb{R}^m} \leqslant 1} \|b\|_{\mathbb{R}^n} \|x \cdot x\|_{\mathbb{R}^n} \leqslant \|b\|_{\mathbb{R}^n},$$

$$\|T_b\|_{\mathcal{L}} = \sup_{\|x\|_{\mathbb{R}^m} \leqslant 1} \|b \cdot x\|_{\mathbb{R}^n} \geqslant \left\| b \cdot \frac{b}{\|b\|_{\mathbb{R}^n}} \right\|_{\mathbb{R}^n} = \|b\|_{\mathbb{R}^n},$$

donc $\|T_b\|_{\mathcal{L}} = \|b\|_{\mathbb{R}^n}$. △

Proposition 10.27.
Une application linéaire de $\mathbb{R}^m$ *dans* $\mathbb{R}^n$ *est continue.*

Démonstration. Soit x un point dans $\mathbb{R}^m$. Nous devons vérifier l'égalité

$$\lim_{h\to 0_m} T(x+h) = T(x). \tag{10.55}$$

Cela revient à prouver que $\lim_{h\to 0_m} T(h) = 0$, parce que $T(x+h) = T(x) + T(h)$. Nous pouvons toujours majorer $\|T(h)\|_n$ par $\|T\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)}\|h\|_{\mathbb{R}^m}$ (lemme 10.10). Quand h s'approche de 0_m sa norme $\|h\|_m$ tend vers 0, ce que nous permet de conclure parce que nous savons que de toutes façons, $\|T\|_{\mathcal{L}}$ est fini. □

Note : dans un espace de dimension infinie, la linéarité ne suffit pas pour avoir la continuité : il faut de plus être borné (ce que sont toutes les applications linéaires $\mathbb{R}^m \to \mathbb{R}^n$). Voir la proposition 10.18.

10.1.5 Application linéaire continue et bornée

Nous avons vu dans la proposition 10.18 que pour une application linéaire, être bornée est équivalent à être continue. Nous allons maintenant voir un certain nombre d'exemples illustrant ce fait.

Exemple 10.28 (Une application linéaire non continue)**.**
Soit V l'espace vectoriel normé des suites *finies* de réels muni de la norme usuelle $\|c\| = \sqrt{\sum_{i=0}^{\infty} |c_i|^2}$ où la somme est finie. Nous nommons $\{e_k\}_{k\in\mathbb{N}}$ la base usuelle de cet espace, et nous considérons l'opérateur $f\colon V \to V$ donnée par $f(e_k) = ke_k$. C'est évidemment linéaire, mais ce n'est pas continu en zéro. En effet la suite $u_k = e_k/k$ converge vers 0 alors que $f(u_k) = e_k$ ne converge pas. △

Cet exemple aurait pu également être donnée dans un espace de Hilbert, mais il aurait fallu parler de domaine.

Exemple 10.29 (Une autre application linéaire non continue[213])**.**
En dimension infinie, une application linéaire n'est pas toujours continue. Soit E l'espace des polynômes à coefficients réels sur $[0,1]$ muni de la norme uniforme. L'application de dérivation $\varphi\colon E \to E$, $\varphi(P) = P'$ n'est pas continue.

Pour la voir nous considérons la suite $P_n = \frac{1}{n}X^n$. D'une part nous avons $P_n \to 0$ dans E parce que $P_n(x) = \frac{x^n}{n}$ avec $x \in [0,1]$. Mais en même temps nous avons $\varphi(P_n) = X^{n-1}$ et donc $\|\varphi(P_n)\| = 1$.

Nous n'avons donc pas $\lim_{n\to\infty} \varphi(P_n) = \varphi(\lim_{n\to\infty} P_n)$ et l'application φ n'est pas continue en 0. Elle n'est donc continue nulle part par linéarité.

Nous avons utilisé le critère séquentiel de la continuité, voir la définition 7.77 et la proposition 7.80. △

Remarque 10.30.
Cette proposition permet de retrouver l'exemple 10.28 plus simplement. Si $\{e_k\}_{k\in\mathbb{N}}$ est une base d'un espace vectoriel normé formée de vecteurs de norme 1, alors l'opérateur linéaire donné par $u(e_k) = ke_k$ n'est pas borné et donc pas continu.

C'est également ce résultat qui montre que le produit scalaire est continu sur un espace de Hilbert par exemple.

Exemple 10.31.
Nous avons vu dans le lemme 10.11 que pour un $x \in E$ donné, l'application

$$\begin{aligned} ev_x\colon \mathcal{L}(E,F) &\to F \\ f &\mapsto f(x) \end{aligned} \tag{10.56}$$

est continue. Vu que ev_x est linéaire, la proposition 10.18 nous indique que ev_x est bornée. Vérifions-le directement. Le calcul n'est pas très compliqué :

$$\|ev_x\| = \sup_{\|f\|=1} \|ev_x(f)\| = \sup_{\|f\|=1} \|f(x)\| \leqslant \sup_{\|f\|=1} \|x\|\|f\| = \|x\| \tag{10.57}$$

où nous avons utilisé le lemme 10.10 en passant. Donc la norme de ev_x est majorée par $\|x\|$.

Elle est même égale à $\|x\|$. En effet, pour chaque $f \in \mathcal{L}(E,F)$ tel que $\|f\| = 1$, nous avons

$$\|ev_x\| \geqslant \|ev_x(f)\| = \|f(x)\|. \tag{10.58}$$

En prenant $f = \mathrm{Id}$ nous trouvons $\|ev_x\| \geqslant \|x\|$. △

Définition 10.32.
*Soit un espace vectoriel E sur le corps $\mathbb{K}$. Son **dual topologique**, noté E', est l'ensemble des formes linéaires continues de E vers $\mathbb{K}$.*

Lemme 10.33.
Soit F un espace de Banach et deux suites $A_k \to A$ et $B_k \to B$ dans $\mathcal{L}(F,F)$. Alors $A_k \circ B_k \to A \circ B$ dans $\mathcal{L}(F,F)$, c'est-à-dire

$$\lim_{n\to\infty}(A_kB_k) = \left(\lim_{n\to\infty} A_k\right)\left(\lim_{n\to\infty} B_k\right). \tag{10.59}$$

Démonstration. Il suffit d'écrire

$$\|A_kB_k - AB\| \leqslant \|A_kB_k - A_kB\| + \|A_kB - AB\|. \tag{10.60}$$

Le premier terme tend vers zéro pour $k \to \infty$ parce que

$$\begin{aligned}
\|A_kB_k - A_kB\| &= \|A_k(B_k - B)\| && (10.61a)\\
&\leqslant \|A_k\|\|B_k - B\| \to \|A\| \cdot 0 && (10.61b)\\
&= 0 && (10.61c)
\end{aligned}$$

où nous avons utilisé la propriété fondamentale de la norme opérateur : la proposition 10.18. Le second terme tend également vers zéro pour la même raison. □

Proposition 10.34 (Distributivité de la somme infinie)**.**
Soient E un espace normé, une suite (u_k) dans $\mathrm{GL}(E)$ ainsi que $a \in \mathrm{GL}(E)$. Pourvu que la série $\sum_{n=0}^{\infty} u_k$ converge nous avons

$$\left(\sum_{k=0}^{\infty} u_k\right) a = \sum_{k=0}^{\infty}(u_k a). \tag{10.62}$$

Démonstration. Par définition de la somme infinie,

$$\spadesuit = \left(\sum_{k=0}^{\infty} u_k\right) a = \left(\lim_{n\to\infty}\sum_{k=0}^{n} u_k\right) a. \tag{10.63}$$

Le lemme 10.33 appliqué à $n \mapsto \sum_{k=0}^{n} u_k$ et à la suite constante a nous donne

$$\spadesuit = \lim_{n\to\infty}\left(\sum_{k=0} u_k a\right), \tag{10.64}$$

ce que nous voulions par distributivité de la somme finie : dans (10.64), le a est dans ou hors de la somme, au choix. L'important est qu'il soit dans la limite. □

Proposition 10.35 ([214])**.**
Soient des espaces vectoriels normés V et W ainsi qu'une forme sesquilinéaire $\phi\colon V \times W \to \mathbb{C}$. Il y a équivalence des faits suivants.

(1) ϕ est continue.

(2) ϕ est continue en $(0,0)$

(3) ϕ est bornée

(4) Il existe $C \geqslant 0$ telle que $|\phi(x,y)| \leqslant C\|x\|\|y\|$ pour tout $(x,y) \in V \times W$.

De plus la norme de ϕ est alors donnée par

$$\|\phi\| = \min\{C \geqslant 0 \text{ tel que } |\phi(x,y)| \leqslant C\|x\|\|y\| \forall (x,y) \in V \times W\}. \tag{10.65}$$

On remarque tout de suite que la norme $\|.\|_\infty$ sur $\mathbb{R}^2$ est la norme de l'espace produit $\mathbb{R} \times \mathbb{R}$. En outre cette définition nous permet de trouver plusieurs nouvelles normes dans les espaces $\mathbb{R}^p$. Par exemple, si nous écrivons $\mathbb{R}^4$ comme $\mathbb{R}^2 \times \mathbb{R}^2$ on peut munir $\mathbb{R}^4$ de la norme produit

$$\|(x_1, x_2, x_3, x_4)\|_{\infty,2} = \max\{\|(x_1, x_2)\|_\infty, \|(x_3, x_4)\|_2\}.$$

Les applications de projection de l'espace produit $V \times W$ vers les espaces «facteurs», V W sont notées $\mathtt{proj}_V$ et $\mathtt{proj}_W$ et sont définies par

$$\begin{aligned} \mathtt{proj}_V \colon V \times W &\to V \\ (v,w) &\mapsto v \end{aligned} \tag{10.66}$$

et

$$\begin{aligned} \mathtt{proj}_W \colon V \times W &\to W \\ (v,w) &\mapsto w. \end{aligned} \tag{10.67}$$

Les inégalités suivantes sont évidentes

$$\begin{aligned} \|\mathtt{proj}_V(v,w)\|_V &\leqslant \|(v,w)\|_{V\times W} \\ \|\mathtt{proj}_W(v,w)\|_W &\leqslant \|(v,w)\|_{V\times W}. \end{aligned} \tag{10.68}$$

La topologie de l'espace produit est induite par les topologies des espaces «facteurs». La construction est faite en deux passages : d'abord nous disons que une partie $A \times B$ de $V \times W$ est ouverte si A et B sont des parties ouvertes de V et de W respectivement. Ensuite nous définissons que une partie quelconque de $V \times W$ est ouverte si elle est une intersection finie ou une réunion de parties ouvertes de $V \times W$ de la forme $A \times B$.

Ce choix de topologie donne deux propriétés utiles de l'espace produit

(1) Les projections sont des **applications ouvertes**. Cela veut dire que l'image par $\mathtt{proj}_V$ (respectivement $\mathtt{proj}_W$) de toute partie ouverte de $V \times W$ est une partie ouverte de V (respectivement W).

(2) Pour toute partir A de V et B de W, nous avons $\operatorname{Int}(A \times B) = \operatorname{Int} A \times \operatorname{Int} B$.

Une propriété moins facile a prouver est que pour toute partie A de V et B de W nous avons $\overline{A \times B} = \bar{A} \times \bar{B}$. Voir le lemme 10.38.

Ce que nous avons dit jusqu'ici est valable pour tout produit d'un nombre fini d'espaces vectoriels normés. En particulier, pour tout $m > 0$ l'espace $\mathbb{R}^m$ peut être considéré comme le produit de m copies de $\mathbb{R}$.

Exemple 10.36.
Si V et W sont deux espaces vectoriels, nous pouvons considérer le produit $E = V \times W$. Les projections $\mathtt{proj}_V$ et $\mathtt{proj}_W$, définies dans la section 7.11, sont des applications linéaires.

En effet, la projection $\mathtt{proj}_V \colon V \times W \to V$ est donnée par $\mathtt{proj}_V(v,w) = v$. Alors,

$$\begin{aligned} \mathtt{proj}_V\big((v,w) + (v',w')\big) &= \mathtt{proj}_V\big((v+v'),(w+w')\big) \\ &= v + v' \\ &= \mathtt{proj}_V(v,w) + \mathtt{proj}_V(v',w'), \end{aligned} \tag{10.69}$$

et

$$\mathtt{proj}_V\big(\lambda(v,w)\big) = \mathtt{proj}_V\big((\lambda v, \lambda w)\big) = \lambda v = \lambda\, \mathtt{proj}_V(v,w). \tag{10.70}$$

Nous laissons en exercice le soin d'adapter ces calculs pour montrer que $\mathtt{proj}_W$ est également une projection. △

Proposition 10.37.
Si $\mathcal{O}$ est un voisinage de (a,b) dans $V\times W$ alors $\mathcal{O}$ contient un ouvert de la forme $B(a,r)\times B(b,r)$.

Démonstration. Vu que $\mathcal{O}$ est un voisinage, il contient un ouvert et donc une boule

$$B\big((a,b),r\big)=\{(v,w)\in V\times W \text{ tel que } \max\{\|v-a\|,\|w-b\|\}<r\}. \tag{10.71}$$

Évidemment l'ensemble $B(a,r)\times B(b,r)$ est dedans. □

10.1.6 Suites

Nous allons maintenant parler de suites dans $V\times W$. Nous noterons (v_n,w_n) la suite dans $V\times W$ dont l'élément numéro n est le couple (v_n,w_n) avec $v_n\in V$ et $w_n\in W$. La notions de convergence de suite découle de la définition de la norme via la proposition 7.130. Il se fait que dans le cas des produits d'espaces, la convergence d'une suite est équivalente à la convergence des composantes. Plus précisément, nous avons le lemme suivant.

Lemme 10.38.
La suite (v_n,w_n) converge vers (v,w) dans $V\times W$ si et seulement les suites (v_n) et (w_n) convergent séparément vers v et w respectivement dans V et W.

Démonstration. Pour le sens direct, nous devons étudier le comportement de la norme de $(v_n,w_n)-(v,w)$ lorsque n devient grand. En vertu de la définition de la norme dans $V\times W$ nous avons

$$\Big\|(v_n,w_n)-(v,w)\Big\|_{V\times W}=\max\big\{\|v_n-v\|_V,\|w_n-w\|_W\big\}. \tag{10.72}$$

Soit $\varepsilon>0$. Par définition de la convergence de la suite (v_n,w_n), il existe un $N\in\mathbb{N}$ tel que $n>N$ implique

$$\max\big\{\|v_n-v\|_V,\|w_n-w\|_W\big\}<\varepsilon, \tag{10.73}$$

et donc en particulier les deux inéquations

$$\|v_n-v\|<\varepsilon \tag{10.74a}$$
$$\|w_n-w\|<\varepsilon. \tag{10.74b}$$

De la première, il ressort que $(v_n)\to v$, et de la seconde que $(w_n)\to w$.

Pour le sens inverse, nous avons pour tout ε un N_1 tel que $\|v_n-v\|_V\leqslant\varepsilon$ pour tout $n>N_1$ et un N_2 tel que $\|w_n-w\|_W\leqslant\varepsilon$ pour tout $n>N_2$. Si nous posons $N=\max\{N_1,N_2\}$ nous avons les deux inégalités simultanément, et donc

$$\max\big\{\|v_n-v\|_V,\|w_n-w\|_W\big\}<\varepsilon, \tag{10.75}$$

ce qui signifie que la suite (v_n,w_n) converge vers (v,w) dans $V\times W$. □

Proposition 10.39 ([1])**.**
Soit un espace E muni d'un produit scalaire à valeurs dans $\mathbb{K}$ (si $\mathbb{K}=\mathbb{C}$ nous supposons le produit hermitien, mais ce n'est pas très important ici). Alors l'application

$$\begin{aligned} a\colon E\times E&\to\mathbb{K}\\ (x,y)&\mapsto\langle x,y\rangle \end{aligned} \tag{10.76}$$

est continue.

Démonstration. Nous ne disons pas que l'espace $V\times V$ est muni d'un produit scalaire. Mais en tout cas c'est un espace métrique, et $\mathbb{K}$ l'est aussi. Donc a est une application entre deux espaces métriques et elle sera continue si et seulement si elle est séquentiellement continue (proposition 7.807.264).

Soit donc une suite convergente dans $E \times E$, c'est-à-dire $(x_k, y_k) \xrightarrow{E\times E} (x, y)$. Nous devons démontrer que $\langle x_k, y_k\rangle \xrightarrow{\mathbb{R}} \langle x, y\rangle$. Les majorations usuelles donnent

$$\begin{aligned}\big|\langle x_k, y_k\rangle - \langle x, y\rangle\big| &\leqslant \big|\langle x_k, y_k\rangle - \langle x, y_k\rangle\big| + \big|\langle x, y_k\rangle - \langle x, y\rangle\big| && (10.77a)\\ &= \big|\langle x_k - x, y_k\rangle\big| + \big|\langle x, y_k - y\rangle\big|. && (10.77b)\end{aligned}$$

Nous savons du lemme 10.38 que les suites (x_k) et (y_k) sont séparément convergentes : $x_k \xrightarrow{E} x$ et $y_k \xrightarrow{E} y$. En utilisant l'inégalité de Cauchy-Schwarz 7.165 nous trouvons

$$\big|\langle x_k - x, y_k\rangle\big| \leqslant \|x_k - x\|\|y_k\|. \tag{10.78}$$

Nous avons $\|x_k - x\| \to 0$ et $\|y_k\| \to \|y\|$, et par la règle du produit de limites dans $\mathbb{R}$ nous avons que $\big|\langle x_k - x, y_k\rangle\big| \to 0$. □

Remarque 10.40.
Il faut remarquer que la norme (7.278) est une norme *par défaut*. C'est la norme qu'on met quand on ne sait pas quoi mettre. Or il y a au moins un cas d'espace produit dans lequel on sait très bien quelle norme prendre : les espaces $\mathbb{R}^m$. La norme qu'on met sur $\mathbb{R}^2$ est

$$\|(x, y)\| = \sqrt{x^2 + y^2}, \tag{10.79}$$

et non la norme « par défaut » de $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ qui serait

$$\|(x, y)\| = \max\{|x|, |y|\}. \tag{10.80}$$

Les théorèmes que nous avons donc démontré à propos de $V \times W$ ne sont donc pas immédiatement applicables au cas de $\mathbb{R}^2$.

Cette remarque est valables pour tous les espaces $\mathbb{R}^m$. À moins de mention contraire explicite, nous ne considérons jamais la norme par défaut (7.278) sur un espace $\mathbb{R}^m$.

Étant donné la remarque 10.40, nous ne savons pas comment calculer par exemple la fermeture du produit d'intervalle $]0, 1, [\times [4, 5[$. Il se fait que, dans $\mathbb{R}^m$, les fermetures de produits sont quand même les produits de fermetures.

Proposition 10.41.
Soit $A \subset \mathbb{R}^m$ et $B \subset \mathbb{R}^m$. Alors dans $\mathbb{R}^{m+n}$ nous avons $\overline{A \times B} = \bar{A} \times \bar{B}$.

10.1.7 Continuité du produit de matrices

Nous avons introduit des normes sur $\mathbb{M}(n, \mathbb{K})$, entre autres la norme opérateur de la définition 10.3. Qui dit norme dit topologie. Il advient alors la question évidente : est-ce que des opérations aussi élémentaires que le produit de matrices sont continues pour ces topologies ?

Une façon simple de répondre à cela est d'introduire sur $\mathbb{M}(n, \mathbb{K})$ une nouvelle norme très simple : celle de $\mathbb{K}^n$. C'est la topologie par composante. Pour cette topologie, il est simple de voir que le produit matriciel est continu parce que les éléments de AB sont des polynômes en les éléments de A B. Ensuite il suffit d'invoquer l'équivalence de toutes les normes (théorème 7.224).

Voyons comment montrer cela de façon plus directe (bien que le raisonnement précédent soit une démonstration qui devrait déjà avoir convaincu les plus sceptiques). La preuve suivante va donc s'amuser à bien préciser les topologies et caractérisations utilisées.

Lemme 10.42.
Si $\|.\|$ est une norme algébrique sur $\mathbb{M}(n, \mathbb{K})$ ($\mathbb{K}$ est $\mathbb{R}$ ou $\mathbb{C}$) alors l'application

$$\begin{aligned} p\colon \mathbb{M}(n, \mathbb{K}) \times \mathbb{M}(n, \mathbb{K}) &\to \mathbb{M}(n, \mathbb{K})\\ (A, B) &\mapsto AB \end{aligned} \tag{10.81}$$

est continue.

Démonstration. L'espace $\mathbb{M}(n,\mathbb{K}) \times \mathbb{M}(n,\mathbb{K})$ est métrique (définition 7.230), donc la caractérisation séquentielle de la continuité (proposition 7.268) s'applique. Nous considérons donc une suite (A_k, B_k) dans $\mathbb{M}(n,\mathbb{K}) \times \mathbb{M}(n,\mathbb{K})$ convergente vers AB.

Nous savons que la topologie sur $\mathbb{M}(n,\mathbb{K}) \times \mathbb{M}(n,\mathbb{K})$ est la topologie produit (lemme 7.230) et que celle-ci donne la convergence composante par composante dès que nous avons convergence d'une suite ; c'est la proposition 7.38. Nous avons donc $A_k \xrightarrow{\mathbb{M}(n,\mathbb{K})} A$ et $B_k \xrightarrow{\mathbb{M}(n,\mathbb{K})} B$.

Voilà pour le contexte. Maintenant, la preuve de la continuité. Nous effectuons les majorations suivantes :

$$\begin{aligned}
\|p(A_k, B_K) - AB\| &\leqslant \|p(A_k, B_k) - p(A_k, B)\| + \|p(A_k, B) - AB\| && \text{(10.82a)}\\
&= \|A_K b_k - A_k B\| + \|A_k B - AB\| && \text{(10.82b)}\\
&= \|A_k(B_k - B)\| + \|(A_k - A)B\| && \text{(10.82c)}\\
&\leqslant \underbrace{\|A_k\|}_{\to\|A\|}\underbrace{\|B_k - B\|}_{\to 0} + \underbrace{\|A_k - A\|}_{\to 0}\|B\|. && \text{(10.82d)}
\end{aligned}$$

□

10.2 Applications multilinéaires

Définition 10.43 (Application multilinéaire)**.**
*Une application $T : \mathbb{R}^{m_1} \times \ldots \times \mathbb{R}^{m_k} \to \mathbb{R}^p$ est dite k-**linéaire** si pour tout $X = (x_1, \ldots, x_k)$ dans $\mathbb{R}^{m_1} \times \ldots \times \mathbb{R}^{m_k}$ les applications $x_i \mapsto T(x_1, \ldots, x_i, \ldots, x_k)$ sont linéaires pour tout i dans $\{1, \ldots, k\}$, c'est-à-dire*

$$\begin{aligned}
T(\ \cdot\ , x_2, \ldots, x_i, \ldots, x_k) &\in \mathcal{L}(\mathbb{R}^{m_1}, \mathbb{R}^p),\\
T(x_1,\ \cdot\ , \ldots, x_i, \ldots, x_k) &\in \mathcal{L}(\mathbb{R}^{m_2}, \mathbb{R}^p),\\
&\vdots\\
T(x_1, \ldots, x_i, \ldots, x_{k-1},\ \cdot\ ) &\in \mathcal{L}(\mathbb{R}^{m_k}, \mathbb{R}^p).
\end{aligned} \tag{10.83}$$

*En particulier lorsque $k = 2$, nous parlons d'applications **bilinéaires**. Vous pouvez deviner ce que sont les applications* tri*linéaire ou* quadri*linéaire.*

L'ensemble des applications k-linéaires de $\mathbb{R}^{m_1} \times \ldots \times \mathbb{R}^{m_k}$ dans $\mathbb{R}^p$ est noté $\mathcal{L}(\mathbb{R}^{m_1} \times \ldots \times \mathbb{R}^{m_k}, \mathbb{R}^p)$ ou $\mathcal{L}(\mathbb{R}^{m_1}, \ldots, \mathbb{R}^{m_k}; \mathbb{R}^p)$.

Exemple 10.44.
Soit A une matrice avec m lignes et n colonnes. L'application bilinéaire de $\mathbb{R}^m \times \mathbb{R}^n$ dans $\mathbb{R}$ associée à A est définie par

$$T_A(x, y) = x^T A y = \sum_{i,j} a_{i,j} x_i y_j, \qquad \forall x \in \mathbb{R}^m,\ y \in \mathbb{R}^n.$$

△

Nous énonçons la proposition suivante dans le cas d'espaces vectoriels normés [9] parce que nous allons l'utiliser dans ce cas, mais le cas particulier $E_i = \mathbb{R}^{m_i}$ et $F = \mathbb{R}^p$ est important.

Proposition 10.45.
Soient des espaces vectoriels normés E_i et F. Une application n-linéaire

$$T\colon E_1 \times \ldots \times E_n \to F \tag{10.84}$$

est est continue si et seulement s'il existe un réel $L \geqslant 0$ tel que

$$\|T(x_1, \ldots, x_n)\|_F \leqslant L\|x_1\|_{F_1} \cdots \|x_n\|_{F_n}, \qquad \forall x_i \in E_i. \tag{10.85}$$

9. Sans hypothèses sur la dimension.

Démonstration. Pour simplifier l'exposition nous nous limitons au cas $n = 2$ et nous notons $T(x, y) = x * y$

Supposons que l'inégalité (10.85) soit satisfaite.

$$\begin{aligned} \|x * y - x_0 * y_0\| &= \|(x - x_0) * y - x_0 * (y - y_0)\| \\ &\leqslant \|(x - x_0) * y\| + \|x_0 * (y - y_0)\| \\ &\leqslant L\|x - x_0\|\|y\| + L\|x_0\|\|y - y_0\|. \end{aligned} \tag{10.86}$$

Si $x \to x_0$ et $y \to y_0$ on voit que T est continue en passant à la limite aux deux côtés de l'inégalité (10.86).

Soit T continue en $(0, 0)$. Évidemment [10] $0 * 0 = 0$, donc il existe $\delta > 0$ tel que si $x \in B_{E_1}(0, \delta)$ et $y \in B_{E_2}(0, \delta)$ alors $\|x * y\| \leqslant 1$. En particulier si $(x, y) \in B_{E_1 \times E_2}(0, \delta)$ nous sommes dans ce cas. Soient maintenant $x \in E_1 \backslash \{0\}$ et $y \in E_2 \backslash \{0\}$

$$x * y = \left(\frac{\|x\|}{\delta} \frac{\delta x}{\|x\|} \right) * \left(\frac{\|y\|}{\delta} \frac{\delta y}{\|y\|} \right) = \frac{\|x\|\|y\|}{\delta^2} \left(\frac{\delta x}{\|x\|} \right) * \left(\frac{\delta y}{\|y\|} \right). \tag{10.87}$$

On remarque que $\delta x / \|x\|_m$ est dans la boule de rayon δ centrée en 0_m et que $\delta y / \|y\|_n$ est dans la boule de rayon δ centrée en 0_n. On conclut

$$x * y \leqslant \frac{\|x\|_m \|y\|_n}{\delta^2}.$$

Il faut prendre $L = 1/\delta^2$. □

La norme de T est alors définie comme la plus petite constante L qui fait fonctionner la proposition 10.45.

Définition 10.46.
La norme sur l'espace $\mathcal{L}(E_1 \times \cdots \times E_n, F)$ des applications k-linéaires et continues est

$$\|T\|_{E_1 \times \ldots \times E_n} = \sup\{\|T(u_1, \ldots, u_k)\|_F \mid \|u_i\|_{E_i} \leqslant 1, i = 1, \ldots, k\}. \tag{10.88}$$

Nous avons donc automatiquement

$$\|T(u, v)\| \leqslant \|T\|\|u\|\|v\|. \tag{10.89}$$

Et nous notons que cette norme est uniquement définie pour les applications linéaires continues. Ce n'est pas très grave parce qu'alors nous définissons $\|T\| = \infty$ si T n'est pas continue. Cela pour retrouver le principe selon lequel on est continue si et seulement si on est borné.

Proposition 10.47.
On définit les fonctions

$$\begin{array}{rlcl} \omega_g : & \mathcal{L}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^p) & \to & \mathcal{L}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)), \\ \omega_d : & \mathcal{L}(\mathbb{R}^m \times \mathbb{R}^n, \mathbb{R}^p) & \to & \mathcal{L}(\mathbb{R}^n, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^p)), \end{array} \tag{10.90}$$

par

$$\omega_g(T)(x) = T(x, \ \cdot \ ), \qquad \forall x \in \mathbb{R}^m,$$

et

$$\omega_d(T)(y) = T(\ \cdot \ , y), \qquad \forall y \in \mathbb{R}^n.$$

Les fonctions ω_g et ω_d sont des isomorphismes qui préservent les normes.

10. Dans la formule suivante, les trois zéros sont les zéros de trois espaces différents.

10.3 Séries

Pour une définition plus générale de somme indexée par un ensemble infine, voir la définition 10.92.

Définition 10.48.
Soit (a_k) une suite dans un espace vectoriel normé $(V, \|.\|)$. La suite des ***sommes partielles*** *associée est la suite (s_k) définie par*

$$s_k = \sum_{i=0}^{k} a_i \tag{10.91}$$

La ***série*** *associée est la limite des sommes partielles*

$$\sum_{n=0}^{\infty} a_k = \lim_{k\to\infty} \sum_{k=0}^{n} a_k \tag{10.92}$$

si elle existe.

Si une telle limite existe nous disons que $\sum_{k=0}^{\infty} a_k$ ***converge*** *dans V. Si la limite de la suite des sommes partielles n'existe pas nous disons que la série* ***diverge****.*

Remarque 10.49.
Si la limite de la suite des sommes partielles n'existe pas dans V, alors elle peut parfois exister dans des extensions de V. Par exemple une série de rationnels convergeant vers $\sqrt{2}$ dans $\mathbb{R}$ ne converge pas dans $\mathbb{Q}$. Autre exemple : avec une bonne topologie sur $\bar{\mathbb{R}}$, une série peut ne pas converger dans $\mathbb{R}$ mais converger vers $\pm\infty$ dans $\bar{\mathbb{R}}$.

Dans le cas des espaces de fonctions, nous avons une norme importante : la norme uniforme définie par $\|f\|_\infty = \sup\{f(x)\}$ où le supremum est pris sur l'ensemble de définition de f.

Lemme 10.50.
Soit une suite (a_k) dans un espace métrique complet[11] *dont la série converge.*

(1) Pour tout N nous avons

$$\sum_{k=0}^{\infty} a_k = \sum_{k=0}^{N} a_k + \sum_{k=N+1}^{\infty} a_k. \tag{10.93}$$

(2) La suite des queues de série converge vers 0, c'est-à-dire que

$$\lim_{N\to\infty} \sum_{k=N}^{\infty} a_k = 0. \tag{10.94}$$

Démonstration. Voici un petit calcul :

$$\lim_{n\to\infty} \sum_{k=0}^{n} a_k = \lim_{n\to\infty} \Big(\sum_{k=0}^{N} a_k + \sum_{k=N+1}^{n} a_k \Big) \tag{10.95a}$$

$$= \lim_{n\to\infty} \sum_{k=0}^{N} a_k + \lim_{n\to\infty} \sum_{k=N+1}^{n} a_k \tag{10.95b}$$

$$= \sum_{k=0}^{N} a_k + \sum_{k=N+1}^{\infty} a_k. \tag{10.95c}$$

Justifications :

— Pour (10.95a). Pour chaque n, la somme est finie et nous pouvons la décomposer. Si vous voulez vraiment couper les cheveux en quatre, vous devez fixer un ϵ, et un n de telle sorte à avoir $n > N$, parce que N est fixé dans l'énoncé du lemme.

11. Définition 7.277.

— Pour (10.95b). Nous sommes dans un cas $\lim_{n\to\infty}(u_n + v_n)$ où (u_n) est constante et où $(u_n + v_n)$ converge. Nous pouvons donc permuter limite et somme [12].

Voila que (1) est prouvé.

Nous écrivons $s_n = \sum_{k=0}^{n} a_k$; l'hypothèse est que la suite (s_n) est une suite convergente dans un espace métrique. Elle est donc de Cauchy par la proposition 7.282.

Soit $\epsilon > 0$. Il existe $N \in \mathbb{N}$ tel que pour tout $p, q > N$, nous ayons $\|s_p - s_q\| \leqslant \epsilon$. Soit $p > N$. Pour tout $n \geqslant 0$ nous avons

$$\epsilon > \|s_{p+n} - s_{p+1}\| = \|\sum_{k=p}^{p+n} a_k\|. \tag{10.96}$$

En prenant la limite $n \to \infty$ nous avons

$$\|\sum_{k=p}^{\infty} a_k\| \leqslant \epsilon. \tag{10.97}$$

Nous avons donc démontré qu'il existe N tel que si $p > N$, alors $\|\sum_{k=p}^{\infty} a_k\| \leqslant \epsilon$. Cela signifie exactement que $\lim_{n\to\infty} \sum_{k=n}^{\infty} a_k = 0$. □

10.3.1 Les trois types de convergence

Trois notions de convergence à ne pas confondre :

(1) La convergence absolue,

(2) la convergence normale. C'est la même que la convergence absolue, mais dans le cas particulier d'un espace de fonctions muni de la norme uniforme.

(3) la convergence uniforme.

Voici les définitions.

Définition 10.51 (Convergence absolue)**.**
Nous disons que la série $\sum_{n=0}^{\infty} a_n$ dans l'espace vectoriel normé V ***converge absolument*** *si la série $\sum_{n=0}^{\infty} \|a_n\|$ converge dans $\mathbb{R}$.*

Définition 10.52 (Convergence normale)**.**
Une série de fonctions $\sum_{n\in\mathbb{N}} u_n$ converge ***normalement*** *si la série de nombres $\sum_n \|u_n\|_\infty$ converge. C'est-à-dire si la série converge absolument pour la norme $\|f\|_\infty$.*

Définition 10.53 (Convergence uniforme)**.**
La somme $\sum_n f_n$ ***converge uniformément*** *vers la fonction F si la suite des sommes partielles converge uniformément, c'est-à-dire si*

$$\lim_{N\to\infty} \|\sum_{n=1}^{N} f_n - F\|_\infty = 0. \tag{10.98}$$

Proposition 10.54.
Une série convergeant absolument dans un espace de Banach [13] y converge au sens usuel.

Démonstration. Soit (a_k) une suite dans un espace vectoriel normé complet dont la série converge absolument. Nous allons montrer que la suite des sommes partielles est de Cauchy. Cela suffira à montrer sa convergence par hypothèse de complétude.

Nous avons

$$\|s_p - s_l\| = \|\sum_{k=l+1}^{p} a_k\| \leqslant \sum_{k=l+1}^{p} \|a_k\| = \bar{s}_p - \bar{s}_l \tag{10.99}$$

12. Pour rappel, la proposition 7.133 demande la convergence des deux suites pour fonctionner.
13. Un espace vectoriel normé complet. Typiquement $\mathbb{R}$.

où $\bar{s}_n = \sum_{k=0}^{n} \|a_k\|$ est la suite des sommes partielles de la série des normes (qui converge). Vu que la suite $(\bar{s}_n)$ converge dans $\mathbb{R}$, elle y est de Cauchy par la proposition 1.115. Donc il existe un N tel que $p, l > N$ implique

$$\|s_p - s_l\| = \bar{s}_p - \bar{s}_l \leqslant \epsilon. \tag{10.100}$$

Cela signifie que (s_n) est une suite de Cauchy et donc convergente. □

Exemple 10.55 (Si l'espace n'est pas complet[1])**.**
Dans un espace pas complet, il est possible de construire un série qui converge absolument sans converger au sens usuel.

Nous allons trouver dans $\mathbb{Q}$ une série qui converge simplement vers $\sqrt{2}$ (et donc ne converge pas dans $\mathbb{Q}$) mais absolument vers 4.

La base est que si $A, B \in \mathbb{Q}$ avec $A < B$ il est possible de résoudre

$$\begin{cases} r_1 + r_2 = A & (10.101a) \\ |r_1| + |r_2| = B & (10.101b) \end{cases}$$

pour $r_1, r_2 \in \mathbb{Q}$. Ce n'est pas très compliqué : la solution est $r_1 = (A + B)/2$ et $r_2 = (A - B)/2$.

Nous considérons l'espace $\mathbb{Q}$ qui n'est pas complet dans $\mathbb{R}$. Soit une série (a_k) dans $\mathbb{Q}$ qui converge vers $\sqrt{2}$ (convergence dans $\mathbb{R}$) nous nommons (s_k) la suite des ses sommes partielles. Soit aussi la suite (b_k) qui converge vers 4 (zéro serait encore plus facile mais bon, juste pour faire un peu de généralité).

Nous supposons que $a_k < b_k$ pour tout k et que les deux suites sont constituées de rationnels positifs. Nous nommons (s_k) et (s'_k) les sommes partielles. En particulier $s_n < s'_n$ et ce sont des suites croissantes.

Nous savons comment trouver $r_1, r_2 \in \mathbb{Q}$ tels que $r_1 + r_2 = s_1$ et $|r_1| + |r_2| = s'_1$. Par récurrence, si nous savons $r_1, \ldots, r_k$ tels que

$$\begin{cases} r_1 + \ldots + r_k = s_n & (10.102a) \\ |r_1| + \ldots + |r_k| = s'_n & (10.102b) \end{cases}$$

(avec, soit dit en passant $k = 2n$), alors nous pouvons trouver des rationnels r_{k+1}, r_{k+2} tels que

$$\begin{cases} r_1 + \ldots + r_k + r_{k+1} + r_{k+2} = s_{n+1} & (10.103a) \\ |r_1| + \ldots + |r_k| + |r_{k+1}| + |r_{k+2}| = s'_{n+1}, & (10.103b) \end{cases}$$

en effet il s'agit de résoudre

$$\begin{cases} r_{k+1} + r_{k+2} = s_{n+1} - r_1 - \ldots - r_k = s_{n+1} - s_n > 0 & (10.104a) \\ |r_{k+1}| + |r_{k+2}| = s'_{n+1} - |r_1| - \ldots - |r_k| = s'_{n+1} - s'_n > 0. & (10.104b) \end{cases}$$

Cela se résout comme plus haut. Au final nous pouvons construire une suite (r_k) dans $\mathbb{Q}$ telle que

$$\sum_{k=0}^{2n} r_k = s_n \tag{10.105}$$

et

$$\sum_{k=0}^{2n} |r_k| = s'_n. \tag{10.106}$$

△

Remarque 10.56.
Nous savons que sur les espaces vectoriels de dimension finie toutes les normes sont équivalentes (théorème 7.221). La notion de convergence de série ne dépend alors pas du choix de la norme. Il n'en est pas de même sur les espaces de dimension infinie. Une série peut converger pour une norme mais pas pour une autre.

Lorsque nous verrons la convergence de séries, nous verrons que la convergence normale est la convergence absolue pour la norme uniforme.

Lemme 10.57.
Si E et F sont des espaces de Banach[14]*, l'espace $\mathcal{L}(E,F)$ est également de Banach.*

Démonstration. Soit (u_n) une suite de Cauchy dans $\mathcal{L}(E,F)$; si $x \in E$ il existe N tel que si $l,m > N$ alors $\|u_l - u_m\| < \epsilon$, c'est-à-dire que pour tout $\|x\| = 1$ on a $\|u_l(x) - u_n(x)\| < \epsilon$. Cela signifie que $u_n(x)$ est une suite de Cauchy dans l'espace complet F. Cette suite converge et nous pouvons définir l'application $u\colon E \to F$ par

$$u(x) = \lim_{n\to\infty} u_n(x). \tag{10.107}$$

Il suffit maintenant de prouver que u est linéaire, ce qui est une conséquence directe de la linéarité de la limite :

$$u(\alpha x + \beta y) = \lim_{n\to\infty} \big(\alpha u_n(x) + \beta u_n(y)\big). \tag{10.108}$$

□

Proposition 10.58.
Si une série converge dans un espace complet, la norme de son terme général converge vers 0.

Démonstration. Soit une suite (a_n) dont la série converge vers s. Soit $\epsilon > 0$. La suite des sommes partielles (s_n) est de Cauchy et converge vers s : $s_n \to s$. En particulier il existe un N tel que si $n > N$, nous avons $\|s_n - s_{n-1}\| < \epsilon$. Pour de telles valeurs de n nous avons :

$$\|a_n\| = \|s_n - s_{n-1}\| \leqslant \epsilon. \tag{10.109}$$

Cela prouve que $a_n \to 0$. □

Dans le même ordre d'idée nous avons la convergence des queues de suites.

Lemme 10.59.
Si $\sum_{k=0}^{\infty} a_k$ est finie, alors

$$\lim_{n\to\infty} \sum_{k=n}^{\infty} a_k = 0. \tag{10.110}$$

Proposition 10.60.
Si la série converge alors la somme est associative : $\sum_k (a_k + b_k) = \sum_k a_k + \sum_k b_k$.

Démonstration. Associativité. Supposons que $\sum_k a_k$ et $\sum_k b_k$ convergent tous deux. Alors nous avons pour tout N :

$$\sum_{k=0}^{N} (a_k + b_k) = \sum_{k=0}^{N} a_k + \sum_{k=0}^{N} b_k. \tag{10.111}$$

Mais si deux limites existent alors la somme commute avec la limite. C'est le cas pour la limite $N \to \infty$, donc

$$\lim_{N\to\infty} \sum_{k=1}^{\infty} (a_k + b_k) = \lim_{N\to\infty} \sum_{k=0}^{\infty} a_k + \lim_{N\to\infty} \sum_{k=0}^{\infty} b_k. \tag{10.112}$$

□

10.4 Série réelle

La notion de série formalise le concept de somme infinie[15]. L'absence de certaines propriétés de ces objets (problèmes de commutativité et même d'associativité) incite à la prudence et montre à quel point une définition précise est importante.

14. Je crois qu'il ne faut pas que E soit complet.
15. La définition d'une somme infinie est la définition 10.91.

10.4.1 Critères de convergence absolue

Étant donné le terme général d'une série, il est souvent –dans les cas qui nous intéressent– difficile de déterminer la somme de la série. L'exemple de la série géométrique est particulier [16], puisqu'on connaît une formule pour chaque somme partielle, mais pour l'exemple des séries de Riemann il n'y a aucune formule simple pour un α général. D'où l'intérêt d'avoir des critères de convergence ne nécessitant aucune connaissance de l'éventuelle limite de la série.

Lemme 10.61 (Critère de comparaison)**.**
Soient $\sum_i a_i$ et $\sum_j b_j$ deux séries à termes positifs vérifiant

$$0 \leqslant a_i \leqslant b_i$$

alors

(1) si $\sum_i a_i$ diverge, alors $\sum_j b_j$ diverge,

(2) si $\sum_j b_j$ converge, alors $\sum_i a_i$ converge (absolument).

Proposition 10.62 (Critère d'équivalence[147])**.**
Soient $\sum_i a_i$ et $\sum_j b_j$ deux séries à termes positifs. Supposons l'existence de la limite (éventuellement infinie) suivante

$$\lim_{i\to\infty} \frac{a_i}{b_i} = \alpha \tag{10.113}$$

avec $\alpha \in \mathbb{R} \cup \{+\infty\}$. Alors

(1) si $\alpha \neq 0$ et $\alpha \neq \infty$, alors

$$\sum_i a_i \text{ converge} \iff \sum_j b_j \text{ converge,} \tag{10.114}$$

(2) si $\alpha = 0$ et $\sum_j b_j$ converge, alors $\sum_i a_i$ converge (absolument),

(3) si $\alpha = +\infty$ et $\sum_j b_j$ diverge, alors $\sum_i a_i$ diverge.

Démonstration. (1) Le fait que la suite a_n/b_n converge vers α signifie que tant sa limite supérieure que sa limite inférieure convergent vers α. En particulier la suite $\frac{a_n}{b_n}$ est bornée vers le haut et vers le bas. À partir d'un certain rang N, il existe M tel que

$$\frac{a_n}{b_n} < M \tag{10.115}$$

et il existe m tel que

$$\frac{a_n}{b_n} > m. \tag{10.116}$$

Nous avons donc $a_n < Mb_n$ et $a_n > mb_n$. La série de (a_n) converge donc si et seulement si la série de (b_n) converge.

(2) Si $\alpha = 0$, cela signifie que pour tout ϵ, il existe un rang tel que $\frac{a_n}{b_n} < \epsilon$, et donc tel que $a_n < \epsilon b_k$. La suite de (a_i) converge donc dès que la suite de (b_i) converge.

(3) Pour tout M, il existe un rang dans la suite à partir duquel on a $\frac{a_i}{b_i} > M$, et donc $a_k > Mb_k$. Si la série de (b_k) diverge, la série de (a_k) doit également diverger.

□

Proposition 10.63 (Critère du quotient[215])**.**
Soit $\sum_i a_i$ une série. Supposons l'existence de la limite (éventuellement infinie) suivante

$$\lim_{i\to\infty} \left|\frac{a_{i+1}}{a_i}\right| = L \tag{10.117}$$

avec $L \in \mathbb{R} \cup \{+\infty\}$. Alors

16. Voir l'exemple 10.67.

(1) si $L < 1$, la série converge absolument,

(2) si $L > 1$, la série diverge,

(3) si $L = 1$ le critère échoue : il existe des exemples de convergence et des exemples de divergence.

Démonstration. (1) Soit b tel que $L < b < 1$. À partir d'un certain rang K, on a $\left|\frac{a_{i+1}}{a_i}\right| < b$. En particulier,

$$|a_{K+1}| < b|a_K|, \tag{10.118}$$

et pour a_{K+2} nous avons

$$|a_{K+2}| < b|a_{K+1}| < b^2|a_K|. \tag{10.119}$$

Au final,

$$|a_{K+n}| < b^n|a_K|. \tag{10.120}$$

Étant donné que la série $\sum_{n\geqslant K} b^n$ converge (parce que $b < 1$), la queue de suite $\sum_{i\geqslant K} a_i$ converge, et par conséquent la suite au complet converge.

(2) Si $L > 1$, on a

$$|a_K| < |a_{K+1}| < |a_{K+2}| < \dots \tag{10.121}$$

Il est donc impossible que la suite (a_i) converge vers zéro. La série ne peut donc pas converger.

(3) Par exemple la suite harmonique $a_n = \frac{1}{n}$ vérifie $L = 1$, mais la série ne converge pas. Par contre, la suite $a_n = \frac{1}{n^2}$ vérifie aussi le critère avec $L = 1$ tandis que la série $\sum_n \frac{1}{n^2}$ converge.

□

Proposition 10.64 (Critère de la racine[147])**.**
Soit $\sum_i a_i$ une série, et considérons

$$\limsup_{i\to\infty} \sqrt[i]{|a_i|} = L$$

avec $L \in \mathbb{R} \cup \{+\infty\}$. Alors

(1) si $L < 1$, la série converge absolument,

(2) si $L > 1$, la série diverge,

(3) si $L = 1$ le critère échoue.

Démonstration. (1) Si $L < 1$, il existe un $r \in]0, 1[$ tel que $|a_n|^{1/n} < r$ pour les grands n. Dans ce cas, $|a_n| < r^n$, et la série converge absolument parce que la série $\sum_n r^n$ converge du fait que $r < 1$.

(2) Si $L > 1$, il existe un $r > 1$ tel que $|a_n|^{1/n} > r > 1$. Cela fait que $|a_n|$ prend des valeurs plus grandes que n pour une infinité de termes. Le terme général a_n ne peut donc pas être une suite convergente. Par conséquent la suite diverge au sens où elle ne converge pas.

□

10.4.2 Critères de convergence simple

Les critères de comparaison, d'équivalence, du quotient et de la racine sont des critères de convergence absolue. Pour conclure à une convergence simple qui n'est pas une convergence absolue, le critère d'Abel sera notre outil principal.

10.4.2.1 Critère d'Abel

Proposition 10.65 (Critère d'Abel)**.**
Soit la série $\sum_i c_i z_i$ avec

(1) (c_i) est une suite réelle décroissante qui tend vers zéro,

(2) (z_i) *est une suite dans* $\mathbb{C}$ *dont la suite des sommes partielles est bornée dans* $\mathbb{C}$*, c'est-à-dire qu'il existe un* $M > 0$ *tel que pour tout* n*,*

$$\left|\sum_{i=1}^{n} z_i\right| \leqslant M. \tag{10.122}$$

Alors la série $\sum_i c_i z_i$ *est convergente.*

Remarquons que ce critère ne donne pas de convergence absolue.

10.4.3 Quelques séries usuelles

Exemple 10.66 (Série harmonique)**.**
La **série harmonique** est

$$\sum_{i=k}^{\infty} \frac{1}{k} = +\infty. \tag{10.123}$$

△

Exemple 10.67 (Série géométrique)**.**
La **série géométrique** de raison $q \in \mathbb{C}$ est

$$\sum_{i=0}^{\infty} q^i. \tag{10.124}$$

Étudions la somme partielle $S_N = 1+q+q^2+\cdots+q^N$. Nous avons évidemment $S_N - qS_N = 1-q^{N+1}$ et donc

$$S_N = \sum_{n=0}^{N} q^n = \frac{1-q^{N+1}}{1-q}. \tag{10.125}$$

La limite $\lim_{N\to\infty} S_N$ existe si et seulement si $|q| \leqslant 1$ et dans ce cas nous avons

$$\sum_{n=0}^{\infty} q^n = \frac{1}{1-q}. \tag{10.126}$$

La convergence est absolue.

Si la somme commence en $n = 1$ au lieu de $n = 0$ alors

$$\sum_{n=1}^{\infty} q^n = \frac{1}{1-q} - 1 = \frac{q}{1-q}. \tag{10.127}$$

△

Un cas particulier de la formule (10.125) est le calcul de $\sum_{j=1}^{N} q^{-j}$ bien utile lorsque l'on joue avec des nombres binaires (voir l'exemple 33.12). Nous avons

$$\sum_{j=1}^{N} q^{-j} = \sum_{j=0}^{N} q^{-j} - 1 = \frac{1-q^{-N}}{q-1}. \tag{10.128}$$

Exemple 10.68 (Série de Riemann)**.**
Pour $\alpha \in \mathbb{R}$, la **série de Riemann**

$$\sum_{i=1}^{\infty} \frac{1}{i^\alpha} \tag{10.129}$$

converge (absolument, puisque réelle et positive) si et seulement si $\alpha > 1$, et diverge sinon. △

Exemple 10.69 (Série exponentielle)**.**
La série exponentielle est la série (pour $t \in \mathbb{R}$)

$$\exp(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!}. \tag{10.130}$$

Nous montrons qu'elle converge pour tout $t \in \mathbb{R}$. Si $a_k = t^k/k!$ alors $\frac{a_{k+1}}{a_k} = \frac{t}{k}$ dont la limite $k \to \infty$ est zéro (quel que soit t). En vertu du critère du quotient 10.63 la série exponentielle converge (absolument) pour tout $t \in \mathbb{R}$.

Pour tout savoir de l'exponentielle et de ses variations, voir le thème 58. △

Exemple 10.70 (Série arithmético-géométrique[216])**.**
Une **suite arithmético-géométrique** est une suite vérifiant pour tout n la relation

$$u_{n+1} = au_n + b \tag{10.131}$$

avec a et b non nuls. Si elle possède une limite, cette dernière doit résoudre $l = al + b$, et donc être donnée par

$$l = \frac{b}{1-a}. \tag{10.132}$$

Comportement amusant : la limite peut exister pour certains valeurs de a_0 et pas pour d'autres. Mais elle ne dépend pas de a_0 parmi ceux pour lesquelles la limite existe.

Il n'est pas très compliqué de trouver le terme général de la suite en fonction de a et de b. Il suffit de considérer la suite $v_n = u_n - r$, et de remarquer que cette suite est géométrique :

$$v_{n+1} = av_n. \tag{10.133}$$

Par conséquent $v_n = a^n v_0$, ce qui donne pour la suite (u_n) la formule

$$u_n = a^n(u_0 - r) + r. \tag{10.134}$$

△

Lemme 10.71 ([217])**.**
Nous avons :

$$\sum_{k=1}^{N} \frac{1}{k(k+1)} = \frac{N}{N+1}. \tag{10.135}$$

et

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1. \tag{10.136}$$

Démonstration. Nous posons

$$f(n) = \sum_{k=1}^{n} \frac{1}{k(k+1)} \tag{10.137a}$$

$$g(n) = \frac{n}{n+1} \tag{10.137b}$$

et nous montrons par récurrence que $f(n) = g(n)$. Pour $n = 1$ nous avons $f(1) = g(1) = \frac{1}{2}$.

Nous supposons que $f(n) = g(n)$ et nous prouvons que $f(n+1) = g(n+1)$. Facile :

$$f(n+1) = f(n) + \frac{1}{(n+1)(n+2)} \tag{10.138a}$$
$$= \frac{n}{n+1} + \frac{1}{(n+1)(n+2)} \tag{10.138b}$$
$$= \frac{n(n+2)+1}{(n+1)(n+2)} \tag{10.138c}$$
$$= \frac{n^2+2n+1}{(n+1)(n+2)} \tag{10.138d}$$
$$= \frac{(n+1)^2}{(n+1)(n+2)} \tag{10.138e}$$
$$= \frac{n+1}{n+2} \tag{10.138f}$$
$$= g(n+1). \tag{10.138g}$$

En ce qui concerne la seconde formule, par définition [17]

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k(k+1)} = \lim_{n\to\infty} \frac{n}{n+1} = 1. \tag{10.139}$$

□

10.4.4 Séries alternées

Théorème 10.72 (Critère des séries alternées[133])**.**
Si $(a_n)_{n\in\mathbb{N}}$ est une suite positive décroissante à limite nulle, alors

(1) Si nous notons (S_n) la suite des sommes partielles, les sous-suites (S_{2n}) et (S_{2n+1}) sont adjacentes [18].

(2) La série $\sum_n (-1)^n a_n$ converge.

(3) Si nous considérons le reste

$$R_n = \sum_{k=n+1}^{\infty} (-1)^k a_k, \tag{10.140}$$

nous avons

$$\operatorname{sgn}(R_n) = (-1)^{n+1} \tag{10.141a}$$
$$|R_n| \leqslant a_{n+1}. \tag{10.141b}$$

Démonstration. En termes de notations, nous allons écrire (S_n) la suite des sommes partielles de $\sum_{k=0}^{\infty}(-1)^k a_k$. Nous notons (S_{2n}) la suite des termes pairs de cette suite. C'est donc la suite $n \mapsto S_{2n}$. Nous divisons en plusieurs morceaux.

S_{2n} est croissante Nous avons simplement

$$S_{2n+2} - S_{2n} = a_{2n+2} - a_{2n+1} \leqslant 0. \tag{10.142}$$

(S_{2n+1}) est décroissante Même calcul.

Les suites (S_{2n}) et S_{2n+1} sont adjacentes Nous avons simplement

$$S_{2n+1} - S_{2n} = a_{2n+1} \to 0. \tag{10.143}$$

Nous concluons par le théorème des suites adjacentes 7.143 que les sous-suites des termes pairs et impairs sont convergentes et convergent vers la même limite.

17. Définition d'une série, 10.48.
18. Définition 7.142.

C'est le moment d'utiliser la proposition 7.144 qui convaincra la lectrice que (S_n) converge vers la même limite, que nous notons S. Le théorème des suites adjacentes nous dit encore que

$$S_{2n+1} \leqslant S \leqslant S_{2n} \tag{10.144}$$

et donc que $R_{2n} = S - S_{2n} \leqslant 0$. Cela donne la majoration

$$|R_{2n}| = |S - S_n| = S_{2n} - S \leqslant S_{2n} - S_{2n+1} = a_{2n+1}. \tag{10.145}$$

Nous faisons le même genre de majorations pour R_{2n+1}. □

10.4.5 Moyenne de Cesaro

Définition 10.73.
*Si $(a_n)_{n\in\mathbb{N}}$ est une suite dans $\mathbb{R}$ ou $\mathbb{C}$, alors sa **moyenne de Cesaro** est la limite (si elle existe) de la suite*

$$c_n = \frac{1}{n}\sum_{k=1}^{n} a_k. \tag{10.146}$$

En un mot, c'est la limite des moyennes partielles.

Lemme 10.74.
Si la suite (a_n) converge vers la limite ℓ alors la suite admet une moyenne de Cesaro qui vaudra ℓ.

Démonstration. Soit $\epsilon > 0$ et $N \in \mathbb{N}$ tel que $|a_n - \ell| < \epsilon$ pour tout $n > N$. En remarquant que

$$\frac{1}{n}\sum_{k=1}^{n} k - \ell = \frac{1}{n}\sum_{k=1}^{n}(a_k - \ell), \tag{10.147}$$

nous avons

$$\begin{aligned}
|\frac{1}{n}\sum_{k=1}^{n} a_k - \ell| &\leqslant |\frac{1}{n}\sum_{k=1}^{N}|a_k - \ell|| + |\frac{1}{n}\sum_{k=N+1}^{n}\underbrace{|a_k - \ell|}_{\leqslant\epsilon}| && \text{(10.148a)}\\
&\leqslant \epsilon + \frac{n-N-1}{n}\epsilon && \text{(10.148b)}\\
&\leqslant 2\epsilon. && \text{(10.148c)}
\end{aligned}$$

Dans ce calcul nous avons redéfinit N de telle sorte que le premier terme soit inférieur à ϵ. □

10.4.6 Écriture décimale d'un nombre

10.75.
Soit $b \geqslant 2$ un entier qui sera la base dans laquelle nous allons écrire les nombres. Nous considérons l'ensemble $\mathbb{D}_b$ des suites dans $\{0, 1, \ldots, b-1\}$ qui n'ont pas une queue de suite uniquement formée de $b-1$. Autrement dit une suite (c_n) est dans $\mathbb{D}_b$ lorsque pour tout N, il existe $k > N$ tel que $c_k \neq b-1$. Associé à cet ensemble nous considérons la fonction

$$\begin{aligned}
\varphi_b\colon \mathbb{D}_b &\to [0,1[\\
c &\mapsto \sum_{n=1}^{\infty}\frac{c_n}{b^n}.
\end{aligned} \tag{10.149}$$

Lemme 10.76.
La fonction φ_b est bien définie au sens où elle converge et prend ses valeurs dans $[0,1[$.

Démonstration. Tout se base sur la somme de la série géométrique (10.126) sous la forme

$$\sum_{k=0}^{\infty}\frac{1}{b^k} = \frac{b}{b-1}. \tag{10.150}$$

La somme (10.149) est donc majorée par $\sum_n \frac{b-1}{b^n}$ qui converge.

Pour prouver que l'image de φ_b est bien $[0,1[$, nous savons qu'au moins un des c_n (en fait une infinité) est plus petit que $b-1$, donc nous avons la majoration stricte [19]

$$\varphi_b(c) < \sum_{n=1}^{\infty} \frac{b-1}{b^n} = (b-1)\left(\sum_{n=1}^{\infty} \frac{1}{b^n} - 1\right) = 1 \tag{10.151}$$

□

Le fait d'introduire l'ensemble $\mathbb{D}$ au lieu de l'ensemble de toutes les suites est justifié par la proposition suivante. Elle explique pourquoi un nombre possède au maximum deux écritures décimales distinctes et que ces deux sont obligatoirement de la forme, par exemple en base 10 :

$$0.34599999999\ldots = 0.34600000\ldots \tag{10.152}$$

mais qu'un nombre commençant par 0.347 ne peut pas être égal. C'est pour cela que dans la définition de $\mathbb{D}_b$ nous avons exclu les suites qui terminent par tout des $b-1$.

Proposition 10.77.
Soit la fonction

$$\begin{aligned} \varphi\colon \{0,\ldots,b-1\}^{\mathbb{N}} &\to [0,1[\\ x &\mapsto \sum_{n=1}^{\infty} \frac{x_n}{b^n}. \end{aligned} \tag{10.153}$$

Si $\varphi(x) = \varphi(y)$ et si n_0 est le plus petit entier tel que $x_{n_0} \neq y_{n_0}$ alors soit

$$x_{n_0} - y_{n_0} = 1 \tag{10.154}$$

et $x_n = 0$, $y_n = b-1$ pour tout $n > n_0$, soit le contraire : $y_{n_0} - x_{n_0} = 1$ avec $y_n = 0$ et $x_n = b-1$ pour tout $n > n_0$.

Démonstration. Nous nous basons sur la formule (facilement dérivable depuis (10.150)) suivante :

$$\sum_{k=n_0+1}^{\infty} \frac{1}{b^k} = \frac{1}{b^{n_0+1}} \frac{b}{b-1}. \tag{10.155}$$

Nous avons

$$0 = \varphi(x) - \varphi(y) = \frac{x_{n_0} - y_{n_0}}{b^{n_0}} + \sum_{n=n_0+1}^{\infty} \frac{x_n - y_n}{b^n} \geqslant \frac{x_{n_0} - y_{n_0}}{b^{n_0}} - \sum_{n=n_0+1}^{\infty} \frac{b-1}{b^n} = \frac{x_{n_0} - y_{n_0} - 1}{b^{n_0}}. \tag{10.156}$$

Le dernier terme étant manifestement positif [20], il est nul et nous avons $x_{n_0} - y_{n_0} = 1$.

Nous avons donc maintenant

$$0 = \varphi(x) - \varphi(y) = \frac{1}{b^{n_0}} + \sum_{n=n_0+1}^{\infty} \frac{x_n - y_n}{b^n}. \tag{10.157}$$

Nous majorons la dernière somme de la façon suivante, en supposant que $|x_n - y_n| \neq b-1$ pour un certain $n > n_0$:

$$\left| \sum_{n=n_0+1}^{\infty} \frac{x_n - y_n}{b^n} \right| \leqslant \sum_{n=n_0+1}^{\infty} \frac{|x_n - y_n|}{b^n} < \sum_{n=n_0+1}^{\infty} \frac{b-1}{b^n} = \frac{1}{b^{n_0}}. \tag{10.158}$$

19. Notez que la somme (10.149) commence à un tandis que la série géométrique (10.150) commence à zéro.

20. C'est ici qu'intervient la subdivision entre le cas $x_{n_0} - y_{n_0} = 1$ ou le contraire. En effet si « ce dernier terme était manifestement *négatif* », il aurait fallu majorer avec de $1-b$ au lieu de $1-b$.

Étant donné cette inégalité stricte, l'équation (10.157) ne peut pas être correcte (valoir zéro). Nous avons donc $|x_n - b_n| = b - 1$ pour tout $n > n_0$. Donc pour chaque $n > n_0$ nous avons soit $x_n = 0$ et $y_n = b - 1$, soit $a_n = b - 1$ et $b_n = 0$. Pour conclure il faut encore prouver que le choix doit être le même pour tout n.

Nous nous mettons dans le cas $x_{n_0} - y_{n_0} = 1$; dans ce cas nous avons bien l'égalité (10.157) sans petites nuances de signes. Nous écrivons

$$\sum_{n=n_0+1}^{\infty} \frac{x_n - y_n}{b^n} = (b-1) \sum_{n=n_0+1}^{\infty} \frac{(-1)^{s_n}}{b^n} \tag{10.159}$$

où s_n est pair ou impair suivant que $x_n = 0$, $y_n = b - 1$ ou le contraire. Si un des $(-1)^{s_n}$ est pas -1 alors nous avons l'inégalité stricte

$$(b-1) \sum_{n=n_0+1}^{\infty} \frac{(-1)^{s_n}}{b^n} > (b-1) \sum_{n=n_0+1}^{\infty} \frac{-1}{b^n} = -\frac{1}{b^{n_0}}. \tag{10.160}$$

Dans ce cas il est impossible d'avoir $\varphi(x) - \varphi(y) = 0$. Nous en concluons que $(-1)^{s_n}$ est toujours -1, c'est-à-dire $x_n - y_n = 1 - b$, ce qui laisse comme seule possibilité $x_n = 0$ et $y_n = b - 1$. □

Théorème 10.78.
L'application $\varphi_b \colon \mathbb{D}_b \to [0, 1[$ est bijective.

Démonstration. En ce qui concerne l'injection, nous savons de la proposition 10.77 que si $\varphi_b(x) = \varphi_b(y)$ pour $x, y \in \{0, \dots, b-1\}^{\mathbb{N}}$, alors soit x soit y a une queue de suite composée uniquement de $b - 1$, ce qui est exclu dans $\mathbb{D}_b$. Nous en déduisons que φ_b est bien injective en prenant $\mathbb{D}_b$ comme ensemble départ.

La partie lourde est la surjectivité. Nous prenons $x \in [0, 1[$ et nous allons construire par récurrence une suite $a \in \mathbb{D}_b$ telle que $\varphi_b(a) = x$. Si il existe $a_1 \in \{0, \dots, b-1\}$ tel que $x = a_1/b$ alors nous prenons la suite $(a_1, 0, \dots,)$ et nous avons évidemment $\varphi(a) = x$. Sinon il existe $a_1 \in \{0, \dots, b-1\}$ tel que

$$\frac{a_1}{b} < x < \frac{a_1 + 1}{b} \tag{10.161}$$

parce que les autres possibilités pour x sont dans l'ensemble $[0, 1[\backslash \{\frac{k}{b}\}_{k=0,\dots,b-1}$ que nous subdivisons en

$$]0, \frac{1}{b}[\cup]\frac{1}{b}, \frac{2}{b}[\cup \dots \cup]\frac{b-1}{b}, 1[. \tag{10.162}$$

Pour la récurrence nous supposons avoir trouvé $a_1, \dots, a_n$ tels que

$$\sum_{k=1}^{n} \frac{a_k}{b^k} < x < \sum_{k=1}^{n-1} \frac{a_k}{b^k} + \frac{a_n + 1}{b^n}. \tag{10.163}$$

Encore une fois s'il existe $a_{n+1} \in \{0, \dots, b-1\}$ tel que $\sum_{k=1}^{n+1} \frac{a_k}{b^k} = x$ alors nous prenons ce a_{n+1} et nous complétons la suite avec des zéros pour avoir $\varphi(a) = x$. Sinon , pour simplifier les notations nous notons $x' = x - \sum_{k=1}^{n} \frac{a_k}{b^k}$ et nous avons

$$0 < x' < \frac{a_n + 1}{b^n}. \tag{10.164}$$

Le nombre x' est forcément dans un des intervalles

$$]\frac{s}{b^{n+1}}, \frac{s+1}{b^{n+1}}[\tag{10.165}$$

avec $s \in \{0, \dots, b-1\}$. Nous prenons le s correspondant à x' comme a_{n+1}. Dans ce cas nous avons

$$\sum_{k=1}^{n+1} \frac{a_k}{b^k} < x < \sum_{k=1}^{n+1} \frac{a_k}{b^k} + \frac{1}{b^{n+1}}. \tag{10.166}$$

Note : les deux inégalités sont strictes. La première parce que s'il y avait égalité, nous nous serions déjà arrêté en complétant avec des zéros. La seconde parce que

$$\sum_{k=n+2}^{\infty} \frac{a_k}{b^k} \leqslant \sum_{k=n+2}^{\infty} \frac{b-1}{b^k} = \frac{1}{b^{n+1}} \tag{10.167}$$

où l'égalité n'est possible que si $a_k = b-1$ pour tout $k \geqslant n+2$. Dans ce cas nous aurions eu

$$x = \sum_{k=1}^{n} \frac{a_k}{b^k} + \frac{a_{n+1}+1}{b^{n+1}} \tag{10.168}$$

et nous aurions choisi le nombre a_{n+1} autrement et complété la suite par des zéros à partir de là. Notons que cela prouve au passage que la suite que nous sommes en train de construire est bien dans $\mathbb{D}_b$ parce qu'elle ne contiendra pas de queue de suite composée de $b-1$.

Ceci termine la construction par récurrence de la suite $a \in \mathbb{D}_b$. Par construction nous avons pour tout $N \geqslant 1$,

$$\sum_{k=1}^{N} \frac{a_k}{b^k} \leqslant x \leqslant \sum_{k=1}^{N} \frac{a_k}{b^k} + \frac{1}{b^{N+1}}, \tag{10.169}$$

autrement dit : $\varphi_b(a_1, \dots, a_N) \in B(x, \frac{1}{b^{N+1}})$. Nous avons donc bien convergence

$$\lim_{N\to\infty} \varphi_b(a_1, \dots, a_N) = x \tag{10.170}$$

et l'application φ_b est surjective. □

L'application $\varphi_b^{-1} \colon [0,1[\to \mathbb{D}_b$ est la **décomposition décimale** en base b des nombres de $[0,1[$.

Tout cela nous permet de montrer entre autres que $\mathbb{R}$ n'est pas dénombrable. Vu qu'il y a une bijection entre $[0,1[$ et $\mathbb{D}_b$, il suffit de prouver que $\mathbb{D}_b$ est non dénombrable. De plus il suffit de démontrer que $\mathbb{D}_b$ est non dénombrable pour un entier $b \geqslant 2$ donné.

Proposition 10.79 ([218]).
Il n'existe pas de surjection $\mathbb{N} \to \mathbb{D}_b$. Autrement dit $\mathbb{D}_b$ est non dénombrable.

Démonstration. Nous prenons $b \neq 2$ pour des raisons qui seront claires plus tard. Soit $f \colon \mathbb{N} \to \mathbb{D}_b$. Pour $i \in \mathbb{N}$ nous notons

$$f(n) = (c_i^{(n)})_{i\geqslant 1}, \tag{10.171}$$

et nous définissons la suite

$$c_k = \begin{cases} 0 & \text{si } c_k^{(k)} \neq 0 \\ 1 & \text{si } c_k^{(k)} = 0. \end{cases} \tag{10.172}$$

Cela est une suite dans $\mathbb{D}_b$ parce que $b \neq 2$ et que la suite ne contient que des 0 et des 1. Mais nous n'avons $f(n) = c$ pour aucun $n \in \mathbb{N}$ parce que nous avons $c_n \neq f(n)_n$.

Si $b = 2$ alors nous savons que $\mathbb{D}_2 \sim [0,1[\sim \mathbb{D}_3$. Donc $\mathbb{D}_2 \sim \mathbb{D}_3$ et $\mathbb{D}_2$ ne peut pas plus être mis en bijection avec $\mathbb{N}$ que $\mathbb{D}_3$. □

Remarque 10.80.
Le cas de la base $b = 2$ doit être fait à part parce que rien n'empêche d'avoir une queue de 1. Il y a alors toutefois moyen de se débrouiller en construisant la suite c de façon plus subtile. Si $b = 2$ et $n \in \mathbb{N}$ alors $f(n)$ est une suite de 0 et 1 contenant une infinité de 0 (parce qu'il n'y a pas de queue de suite ne contenant que des 1). Nous construisons alors c de la façon suivante : d'abord nous recopions $f(0)$ jusqu'à son *deuxième* zéro que nous changeons en 1 ; nommons n_0 le rang de ce deuxième zéro. Ensuite nous recopions les éléments de $f(1)$ à partir du rang $n_0 + 1$ jusqu'au second zéro que nous changeons en 1, etc.

Le fait de prendre le deuxième zéro nous garantit que la suite c n'aura pas de queue de suite ne contenant que des 1.

Notons que cette construction s'adapte à tout b ; il suffit de prendre le second terme qui n'est pas $b-1$ et le remplacer par $b-1$.

Corolaire 10.81.
L'ensemble $[0,1[$ *n'est pas dénombrable.*

Démonstration. L'ensemble $[0,1[$ est en bijection avec $\mathbb{D}_b$ que nous venons de prouver n'être pas dénombrable. □

10.4.7 Théorème de Banach-Steinhaus

Lemme 10.82 ([219])**.**
Soient des espaces vectoriels normés X et Y ainsi qu'une application linéaire bornée $T\colon X\to Y$. Pour tout $a\in X$ et pour tout $r>0$ nous avons

$$\sup_{x\in B(a,r)}\|Tx\|\geqslant r\|T\| \tag{10.173}$$

Démonstration. Nous commençons avec $a=0$. En utilisant la définition 10.3 de la norme opérateur,

$$\|T\|=\sup_{x\in X}\frac{\|Tx\|}{\|x\|}=\sup_{x\in B(0,r)}\frac{\|Tx\|}{\|x\|}\leqslant\frac{1}{r}\sup_{x\in B(0,r)}\|Tx\|. \tag{10.174}$$

Donc

$$\sup_{x\in B(0,r)}\|Tx\|\geqslant r\|T\|. \tag{10.175}$$

Il y a maintenant une astuce. Nous considérons un maximum :

$$\begin{aligned}\max\{\|T(a+x),\|T(a-x)\|\|\}&\geqslant\frac{1}{2}\big(\|T(a+x)\|+\|T(a-x)\|\big) &(10.176a)\\ &\geqslant\frac{1}{2}\big(\|T(a+x)-T(a+x)\|\big) &(10.176b)\\ &=\frac{1}{2}\|T(2x)\| &(10.176c)\\ &=\|Tx\|. &(10.176d)\end{aligned}$$

Justifications :

— Pour (10.176a), la moyenne est plus petite que le maximum.

— Pour (10.176b), inégalité triangulaire : $\|\alpha-\beta\|\leqslant\|\alpha\|+\|\beta\|$.

Si maintenant $y\in B(a,r)$, nous avons $y=a+x$ pour un certain $x\in B(0,r)$, donc

$$\begin{aligned}\sup_{y\in B(a,r)}\|Ty\|&=\sup_{x\in B(0,r)}\|T(a+x)\| &(10.177a)\\ &=\sup_{x\in B(0,r)}\max\{\|T(a+x)\|,\|T(a-x)\|\} &(10.177b)\\ &\geqslant\sup_{x\in B(0,r)}\|Tx\| &(10.177c)\\ &\geqslant r\|T\|. &(10.177d)\end{aligned}$$

Pour (10.177b), l'ensemble sur lequel nous prenons le supremum n'est pas modifié fondamentalement si nous regroupons les éléments deux à deux en prenant le maximum : les éléments exclus sont majorés. □

Théorème 10.83 (Théorème de Banach-Steinhaus[219])**.**
Soient un espace de Banach[21] *X et un espace vectoriel normé Y. Soit une famille $\mathcal{F}$ d'opérateurs linéaire bornés. Si pour tout $x\in X$,*

$$\sup_{T\in\mathcal{F}}\|Tx\|<\infty, \tag{10.178}$$

alors

$$\sup_{T\in\mathcal{F}}\|T\|<\infty. \tag{10.179}$$

21. Définition 7.279.

Démonstration. Nous supposons que $\sup_{T\in\mathcal{F}}\|T\| = \infty$, de telle sorte que nous pouvons choisir une suite (T_n) dans $\mathcal{F}$ telle que $\|T_n\| \to \infty$. Cette suite peut diverger arbitrairement vite, et nous fixerons exactement cela plus tard.

Soit par ailleurs une suite $\alpha_n > 0$ d'éléments petits et tels que $\alpha_n \to 0$. Nous supposons que $\sum_{n=0}^{\infty}\alpha_n < \infty$.

Si $a \in X$, le lemme 10.82 dit que

$$\sup_{x\in B(a,\alpha_n)} \|T_n x\| \geqslant \|T_n\|\alpha_n. \tag{10.180}$$

En posant $x_0 = 0$, nous construisons une suite (x_n) par récurrence en imposant

(1) $x_n \in B(x_{n=1}, \alpha_n)$

(2) $\|T_n x_n\| \geqslant \|T_n\|\alpha_n$.

En utilisant une série télescopique et l'inégalité triangulaire $\|x_k - x_{k+1}\| \leqslant \alpha_n$ à chaque étage,

$$\|x_p - x_q\| \leqslant \sum_{k=p}^{q}\alpha_k \leqslant \sum_{k=p}^{\infty}\alpha_k. \tag{10.181}$$

Mais vu que la somme des α_n converge, la suite des queues de somme converge vers zéro[22] : $\lim_{p\to\infty}\sum_{k=p}^{\infty}\alpha_n = 0$. Cela implique que (x_n) est une suite de Cauchy[23]. Vu que X est de Banach, la suite (x_n) a une limite dans X. Soit x cette limite.

Nous avons $\beta_n = \|x_n - x\| \to 0$. Il y aurait moyen de calculer β_n en fonction de α_n (surtout si nous avions donné une forme explicite à α_n), mais c'est sans importance ici. L'important est que c'est une suite qui tend vers zéro.

Nous avons

$$x \in B(x_n, \beta_n), \tag{10.182}$$

et donc il existe $a_n \in B(0, \beta_n)$ tel que $x = x_n + a_n$. Avec cela, pour chaque n nous avons :

$$\begin{aligned}
\|T_n x\| &= \|T_n(x_n + a_n)\| && (10.183a)\\
&\geqslant \|T_n x_n\| - \|T_n a_n\| && (10.183b)\\
&\geqslant \|T_n x_n\| - \|T_n\|\beta_n && (10.183c)\\
&\geqslant \|T_n\|\alpha_n - \|T_n\|\beta_n && (10.183d)\\
&= \|T_n\|(\alpha_n - \beta_n). && (10.183e)
\end{aligned}$$

Pour 10.183c, nous avons utilisé $\|T_n a_n\| \leqslant \|T_n\|\beta_n$. En résumé,

$$\|T_n x\| \geqslant \|T_n\|(\alpha_n - \beta_n). \tag{10.184}$$

Il suffit de choisir $\|T_n\|$ suffisamment rapidement croissant pour que[24]

$$\|T_n\|(\alpha_n - \beta_n) \to \infty, \tag{10.185}$$

et nous avons $\|T_n x\| \to \infty$, qui est contraire aux hypothèses. □

Théorème 10.84 (Théorème de Banach-Steinhaus[54, 220]).
Soit E un espace de Banach[25] *et F un espace vectoriel normé. Nous considérons une partie $H \subset \mathcal{L}_c(E, F)$ (espace des fonctions linéaires continues). Alors H est uniformément borné si et seulement s'il est simplement borné.*

22. Lemme 10.50(2).
23. Proposition 7.282.
24. Le point important ici est que α_n (et donc β_n) est choisi sans référence à $\|T_n\|$.
25. Définition 7.279.

Démonstration. Si H est uniformément borné, il est borné ; pas besoin de rester longtemps sur ce sens de l'équivalence. Supposons donc que H soit borné. Pour chaque $k \in \mathbb{N}^*$ nous considérons l'ensemble

$$\Omega_k = \{x \in E \text{ tel que } \sup_{f \in H} \|f(x)\| > k\}. \tag{10.186}$$

Les Ω_k sont ouverts Soit $x_0 \in \Omega_k$; nous avons alors une fonction $f \in H$ telle que $\|f(x_0)\| > k$, et par continuité de f il existe $\rho > 0$ tel que $\|f(x)\| > k$ pour tout $x \in B(x_0, \rho)$. Par conséquent $B(x_0, \rho) \subset \Omega_k$ et Ω_k est ouvert par le théorème 7.4.

Les Ω_k ne sont pas tous denses dans E Nous supposons que les ensembles Ω_k soient tous dense dans E. Le théorème de Baire 7.339 nous indique que E est un espace de Baire (parce que de Banach) et donc que

$$\overline{\bigcap_{k \in \mathbb{N}} \Omega_k} = E. \tag{10.187}$$

En particulier l'intersection des Ω_k n'est pas vide. Soit $x_0 \in \bigcap_{k \in \mathbb{N}} \Omega_k$. Nous avons alors

$$\sup_{f \in H} \|f(x)\| = \infty, \tag{10.188}$$

ce qui est contraire à l'hypothèse. Donc les ouverts Ω_k ne sont pas tous denses dans E.

La majoration Il existe $k \geqslant 0$ tel que Ω_k ne soit pas dense dans E, et nous voulons prouver que $\{\|f\| \text{ tel que } f \in H\}$ est un ensemble borné. Soit donc $k \geqslant 0$ tel que Ω_k ne soit pas dense dans E ; il existe un $x_0 \in E$ et $\rho > 0$ tels que

$$B(x_0, \rho) \cap \Omega_k = \varnothing. \tag{10.189}$$

Si $x \in B(x_0, \rho)$ alors x n'est pas dans Ω_k et donc

$$\sup_{f \in H} \|f(x)\| \leqslant k. \tag{10.190}$$

Afin d'évaluer $\|f\|$ nous devons savoir ce qu'il se passe avec les vecteurs sur une boule autour de 0. Pour tout $x \in B(0, \rho)$ et pour tout $f \in H$, la linéarité de f donne

$$\|f(x)\| = \|f(x + x_0) - f(x_0)\| \leqslant \|f(x + x_0) + f(x_0)\| \leqslant 2k. \tag{10.191}$$

Par continuité nous avons alors $\|f(x)\| \leqslant 2k$ pour tout $x \in \overline{B(0, \rho)}$. Si maintenant $x \in F$ vérifie $\|x\| = 1$ nous avons

$$\|f(x)\| = \frac{1}{\rho}\|f(\rho x)\| \leqslant \frac{2k}{\rho}, \tag{10.192}$$

et donc $\|f\| \leqslant \frac{2k}{\rho}$, ce qui montre que $2k/\rho$ est un majorant de l'ensemble $\{\|f\| \text{ tel que } f \in H\}$.

□

Une application du théorème de Banach-Steinhaus est l'existence de fonctions continues et périodiques dont la série de Fourier ne converge pas. Ce sera l'objet de la proposition 27.18.

10.4.8 Convergence forte

Lorsque nous avons une suite d'opérateurs linéaires, nous pouvons considérer la convergence d'une suite pour la norme opérateur : $A_k \to A$ lorsque $\|A_k - A\| \to 0$.

Définition 10.85 ([221]).
Soient un espace vectoriel E et un espace vectoriel normé V. Nous disons que la suite d'opérateur $T_k \colon E \to V$ ***converge fortement*** *vers l'opérateur T si pour tout $x \in E$ nous avons*

$$\|T_k x - Tx\| \to 0. \tag{10.193}$$

Cette notion s'appelle *forte* par opposition à la convergence *faible* dont nous ne parlerons pas. Elle est cependant moins forte que la convergence en norme dont nous avons déjà parlé.

Proposition 10.86.
Soient des espaces vectoriels normés E et F et une suite d'opérateurs $T_k\colon E \to F$ convergeant vers T[26]*. Alors cette suite converge également fortement.*

Démonstration. Soit $x \in E$ que nous supposons non nul. Soit $\lambda \in \mathbb{C}$ tel que $x = \lambda y$ avec $\|y\| = 1$. Nous avons

$$\|T_k x - Tx\| = |\lambda|\|T_k y - Ty\| \leqslant |\lambda| \sup_{\|z\|=1} \|T_k z - Tz\| = |\lambda|\|T_k - T\| \to 0. \tag{10.194}$$

La dernière étape est la convergence en norme $T_k \to T$. □

Proposition 10.87.
Soient E et F, des espaces vectoriels normés de dimension finie. Soit une suite (A_n) d'applications linéaires $E \to F$. Si elle converge fortement vers A, alors elle converge en norme vers A.

Démonstration. En plusieurs coups.

Si une sous-suite converge Commençons par montrer que si (B_n) est une sous-suite de (A_n) qui converge vers B, alors $B = A$. Autrement dit, A est le seul candidat limite pour A_n.
Soit $\|x\| = 1$. Nous avons

$$\|B_n x - Bx\| \leqslant \|B_n - B\|\|x\| = \|B_n - B\|, \tag{10.195}$$

mais pour la sous-suite (B_n) nous avons supposé $\|B_n - B\| \to 0$. Donc $\|B_n x - Bx\| \to 0$, ce qui signifie que $B_n x \to Bx$. Mais par hypothèse, $B_n x \to Ax$. Par unicité de la limite, $Bx = Ax$ pour tout x de norme 1. Pour les autres x, c'est la linéarité qui conclu.

Utilisation de deux gros résultats Par l'hypothèse de convergence, pour chaque x nous avons $\sup_n \|A_n x\| < \infty$. Le théorème de Banach-Steinhaus 10.83 nous indique alors que l'ensemble $\mathcal{F} = \{A_n\}_{n\in\mathbb{N}}$ est borné. Il existe donc $M > 0$ tel que $\|A_n\| < M$ pour tout n.
Nous utilisons à présent l'hypothèse de dimension finie en disant que l'espace des applications linéaires $E \to F$ est de dimension finie, de telle sorte que ses boules fermées soient compactes. Donc la suite (A_n) est contenue dans un compact.

Les sous-suite convergentes La suite (A_n) est contenue dans un compact. Toutes ses sous-suites sont dans ce compact et possèdent donc une sous-suite convergente (théorème 7.103). Toutes ces sous-sous-suites convergent nécessairement vers A par ce que nous avons dit dans la première étape de la preuve. Le lemme 7.39 nous dit alors que $A_n \to A$.

□

10.5 Sommes de familles infinies

10.5.1 Convergence commutative

Définition 10.88.
Soit x_k une suite dans un espace vectoriel normé E. Nous disons que la suite ***converge commutativement*** *vers $x \in E$ si $\lim_{n\to\infty} \|x_n - x\| = 0$ et si pour toute bijection $\tau\colon \mathbb{N} \to \mathbb{N}$ nous avons aussi*

$$\lim_{n\to\infty} \|x_{\tau(k)} - x\| = 0. \tag{10.196}$$

La notion de convergence commutative est surtout intéressante pour les séries. La somme

$$\sum_{k=0}^{\infty} x_k \tag{10.197}$$

26. Sans précisions, ce sera toujours la convergence en norme.

converge commutativement vers x si $\lim_{N\to\infty} \|x-\sum_{k=0}^{N} x_k\| = 0$ *et si pour toute bijection* $\tau\colon \mathbb{N}\to\mathbb{N}$ *nous avons*

$$\lim_{N\to\infty} \|x - \sum_{k=0}^{N} x_{\tau(k)}\| = 0. \tag{10.198}$$

Nous démontrons maintenant qu'une série converge réelle commutativement si et seulement si elle converge absolument.

Proposition 10.89.
Soit $(a_i)_{i\in\mathbb{N}}$ *une suite absolument convergente*[27] *dans* $\mathbb{C}$. *Alors elle converge commutativement.*

Démonstration. Soit $\epsilon > 0$. Nous posons $\sum_{i=0}^{\infty} a_i = a$ et nous considérons N tel que

$$|\sum_{i=0}^{N} a_i - a| < \epsilon. \tag{10.199}$$

Étant donné que la série des $|a_i|$ converge, il existe N_1 tel que pour tout $p, q > N_1$ nous ayons $\sum_{i=p}^{q} |a_i| < \epsilon$. Nous considérons maintenant une bijection $\tau\colon \mathbb{N}\to\mathbb{N}$. Prouvons que la série $\sum_{i=0}^{\infty} |a_{\tau(i)}|$ converge. Nous choisissons M de telle sorte que pour tout $n > M$, $\tau(n) > N_1$. Si s_k est la somme partielle de la suite $(a_{\tau(i)})_{i\in\mathbb{N}}$ et si $M < p < q$ nous avons

$$|s_q - s_p| = |\sum_{i=p}^{q} a_{\tau(i)}| \leqslant \sum_{i=p}^{q} |a_{\tau(i)}| < \epsilon. \tag{10.200}$$

Cela montre que (s_k) est une suite de Cauchy. Elle est alors convergente et nous en déduisons que la série

$$\sum_{i=0}^{\infty} a_{\tau(i)} \tag{10.201}$$

converge. Nous devons montrer à présent qu'elle converge vers la même limite que la somme « usuelle » $\lim_{N\to\infty} \sum_{i=0}^{N} a_i$.

Soit $n > \max\{M, N\}$. Alors

$$\sum_{k=0}^{n} a_{\tau(k)} - \sum_{k=0}^{n} a_k = \sum_{k=0}^{M} a_{\tau(k)} - \sum_{k=0}^{N} a_k + \underbrace{\sum_{M+1}^{n} a_{\tau(k)}}_{<\epsilon} - \underbrace{\sum_{k=N+1}^{n} a_k}_{<\epsilon}. \tag{10.202}$$

Par construction les deux derniers termes sont plus petits que ϵ parce que M et N sont les constantes de Cauchy pour les séries $\sum a_{\tau(i)}$ et $\sum a_i$. Afin de traiter les deux premiers termes, quitte à redéfinir M, nous supposons que $\{1, \dots, N\} \subset \tau\{1, \dots, M\}$; par conséquent tous les a_i avec $i < N$ sont atteints par les $a_{\tau(i)}$ avec $i < M$. Dans ce cas, les termes qui restent dans la différence

$$\sum_{k=0} a_{\tau(k)} - \sum_{k=0}^{N} a_k \tag{10.203}$$

sont des a_k avec $k > N$. Cette différence est donc en valeur absolue plus petite que ϵ, et nous avons en fin de compte que

$$\left|\sum_{k=0}^{n} a_{\tau(k)} - \sum_{k=0}^{n} a_k\right| < \epsilon. \tag{10.204}$$

□

Proposition 10.90.
Soit $\sum_{k=0}^{\infty} a_k$ *une série réelle qui converge mais qui ne converge pas absolument. Alors pour tout* $b \in \mathbb{R}$, *il existe une bijection* $\tau\colon \mathbb{N}\to\mathbb{N}$ *telle que* $\sum_{i=0}^{\infty} a_{\tau(i)} = b$.

27. Définition 10.51.

Pour une preuve, voir chez Gilles Dubois.

Les propositions 10.89 et 10.90 disent entre autres qu'une série dans $\mathbb{C}$ est commutativement sommable si et seulement si elle est absolument sommable.

Soit $(a_i)_{i\in I}$ une famille de nombres complexes indexée par un ensemble I quelconque. Nous allons nous intéresser à la somme $\sum_{i\in I} a_i$.

Soit $\{a_i\}_{i\in I}$ des nombres positifs. Nous définissons la somme

$$\sum_{i\in I} a_i = \sup_{J \text{ fini}} \sum_{j\in J} a_j. \tag{10.205}$$

Notons que cela est une définition qui ne fonctionne bien que pour les sommes de nombres positifs. Si $a_i = (-1)^i$, alors selon la définition nous aurions $\sum_i (-1)^i = \infty$. Nous ne voulons évidemment pas un tel résultat.

Dans le cas de familles de nombres réels positifs, nous avons une première définition de la somme.

Définition 10.91.
Soit $(a_i)_{i\in I}$ une famille de nombres réels positifs indexés par un ensemble quelconque I. Nous définissons

$$\sum_{i\in I} a_i = \sup_{J \textit{ fini dans } I} \sum_{j\in J} a_j. \tag{10.206}$$

Définition 10.92.
Si $\{v_i\}_{i\in I}$ est une famille de vecteurs dans un espace vectoriel normé indexée par un ensemble quelconque I. Nous disons que cette famille est ***sommable*** *de somme v si pour tout $\epsilon > 0$, il existe un J_0 fini dans I tel que pour tout ensemble fini K tel que $J_0 \subset K$ nous avons*

$$\| \sum_{j\in K} v_j - v\| < \epsilon. \tag{10.207}$$

Notons que cette définition implique la convergence commutative.

Exemple 10.93.
La suite $a_i = (-1)^i$ n'est pas sommable parce que quel que soit J_0 fini dans $\mathbb{N}$, nous pouvons trouver J fini contenant J_0 tel que $\sum_{j\in J}(-1)^j > 10$. Pour cela il suffit d'ajouter à J_0 suffisamment de termes pairs. De la même façon en ajoutant des termes impairs, on peut obtenir $\sum_{j\in J'}(-1)^i < -10$. △

Exemple 10.94.
De temps en temps, la somme peut sortir d'un espace. Si nous considérons l'espace des polynômes $[0,1] \to \mathbb{R}$ muni de la norme uniforme, la somme de l'ensemble

$$\{1, -1, \pm\frac{x^n}{n!}\}_{n\in\mathbb{N}} \tag{10.208}$$

est zéro.

Par contre la somme de l'ensemble $\{1, \frac{x^n}{n!}\}_{n\in\mathbb{N}}$ est l'exponentielle qui n'est pas un polynôme. △

10.5.2 Somme non dénombrables

Nous allons voir que les sommes non dénombrables ne sont pas intéressantes : si le nombre de valeurs non nulles parmi les $(x_i)_{i\in I}$ est non dénombrable, alors la somme est infinie. La bonne généralisation de somme infinie dans le cas non dénombrable est l'intégrale qui viendra seulement avec la définition 13.151 et la mesure de Lebesgue 13.127.

Lemme 10.95.
Si A est non dénombrable dans $\mathbb{R}$, alors il existe $\delta > 0$ tel que $A \cap \{|x| \geqslant \delta\}$ est non dénombrable.

Démonstration. Nous y allons par l'absurde, et nous supposons que A ne contient pas zéro (sinon il faut ajouter zéro aux A_n ci-dessous, et ça alourdit les notations). Nous supposons donc que les parties

$$A_n = A \cap \{|x| \geqslant \frac{1}{n}\} \tag{10.209}$$

sont dénombrables. Mais

$$A \subset \bigcup_{n=1}^{\infty} A_n. \tag{10.210}$$

Une union dénombrable d'ensembles dénombrables est dénombrable[28]. Vu qu'un ensemble non dénombrable ne peut être inclus à un ensemble dénombrable[29], nous avons une contradiction. □

Lemme 10.96.
Soit un ensemble I et une « suite » $(x_i)_{i\in I}$ avec $x_i \geqslant 0$ pour tout i. Si l'ensemble

$$F = \{i \in I \text{ tel que } x_i > 0\} \tag{10.211}$$

est non dénombrable, alors

$$\sum_{i\in I} x_i = \infty. \tag{10.212}$$

Démonstration. Nous considérons l'ensemble des valeurs non nulles atteintes par x :

$$V = \{x_i \text{ tel que } i \in F\}. \tag{10.213}$$

Il y a deux possibilités : soit V est dénombrable (ou fini), soit il est non dénombrable.

<u>V **est fini ou dénombrable**</u> Dans ce cas, l'application $x\colon F \to [0,\infty[$ est une application d'un ensemble indénombrable vers un ensemble dénombrable. Le lemme 1.49 nous indique qu'il existe $y \in \mathbb{R}$ tel que $x^{-1}(y)$ est indénombrable et en particulier infini. La somme $\sum_{i\in x^{-1}(y)} x_i$ est une somme indénombrable de termes tous égaux et strictement positifs. Elle est infinie.

<u>V **est indénombrable**</u> La partie V de $\mathbb{R}$ est non dénombrable ; elle est donc sujette au lemme 10.95 : il existe $\delta > 0$ tel que $W = V \cap \{x \geqslant \delta\}$ est indénombrable. Vu que $x_i \geqslant \delta$ pour tout i dans $x^{-1}(W)$ nous avons

$$\sum_{i\in x^{-1}(W)} x_i = \infty. \tag{10.214}$$

□

10.5.3 Sommes dénombrables

Nous avons vu que les sommes non dénombrables ne sont pas intéressantes. La notion 10.91 de sommes est par contre réellement plus utile que la notion de somme sur $\mathbb{N}$ parce que $\mathbb{N}$ a un ordre. En effet une somme sur $\mathbb{N}$ peut être définie par les sommes partielles avec un ordre sans réelle discussions, alors que l'ordre de sommation sur $\mathbb{Z}$ est déjà plus discutable. Bref, nous allons voir maintenant quelque propriétés de la somme 10.91 dans le cas dénombrable.

Exemple 10.97.
Au sens de la définition 10.92 la famille

$$\frac{(-1)^n}{n} \tag{10.215}$$

n'est pas sommable. En effet la somme des termes pairs est ∞ alors que la somme des termes impairs est $-\infty$. Quel que soit $J_0 \in \mathbb{N}$, nous pouvons concocter, en ajoutant des termes pairs, un J avec $J_0 \subset J$ tel que $\sum_{j\in J}(-1)^j/j$ soit arbitrairement grand. En ajoutant des termes négatifs, nous pouvons également rendre $\sum_{j\in J}(-1)^j/j$ arbitrairement petit. △

28. Proposition 1.44.
29. Proposition 1.48.

Proposition 10.98.
Si (a_{ij}) est une famille de nombres positifs indexés par $\mathbb{N} \times \mathbb{N}$ alors

$$\sum_{(i,j)\in\mathbb{N}^2} a_{ij} = \sum_{i=1}^{\infty}\Big(\sum_{j=1}^{\infty} a_{ij}\Big) \tag{10.216}$$

où la somme de gauche est celle de la définition 10.91.

Démonstration. Nous considérons $J_{m,n} = \{0,\dots,m\} \times \{0,\dots,n\}$ et nous avons pour tout m et n :

$$\sum_{(i,j)\in\mathbb{N}^2} a_{ij} \geqslant \sum_{(i,j)\in J_{m,n}} a_{ij} = \sum_{i=1}^{m}\Big(\sum_{j=1}^{n} a_{ij}\Big). \tag{10.217}$$

Si nous fixons m et que nous prenons la limite $n \to \infty$ (qui commute avec la somme finie sur i) nous trouvons

$$\sum_{(i,j)\in\mathbb{N}^2} a_{ij} \geqslant= \sum_{i=1}^{m}\Big(\sum_{j=1}^{\infty} a_{ij}\Big). \tag{10.218}$$

Cela étant valable pour tout m, c'est encore valable à la limite $m \to \infty$ et donc

$$\sum_{(i,j)\in\mathbb{N}^2} a_{ij} \geqslant \sum_{i=1}^{\infty}\Big(\sum_{j=1}^{\infty} a_{ij}\Big). \tag{10.219}$$

Pour l'inégalité inverse, il faut remarquer que si J est fini dans $\mathbb{N}^2$, il est forcément contenu dans $J_{m,n}$ pour m et n assez grand. Alors

$$\sum_{(i,j)\in J} a_{ij} \leqslant \sum_{(i,j)\in J_{m,n}} a_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} \leqslant \sum_{i=1}^{\infty}\Big(\sum_{j=1}^{\infty} a_{ij}\Big). \tag{10.220}$$

Cette inégalité étant valable pour tout ensemble fini $J \subset \mathbb{N}^2$, elle reste valable pour le supremum. □

La définition générale de la somme 10.92 est compatible avec la définition usuelle dans les cas où cette dernière s'applique.

Proposition 10.99 (commutative sommabilité)**.**
Soit I un ensemble dénombrable et une bijection $\tau\colon \mathbb{N} \to I$. Soit $(a_i)_{i\in I}$ une famille dans un espace vectoriel normé. Si $\sum_{i\in I} a_i$ existe, alors il est donné par

$$\sum_{i\in I} a_i = \lim_{N\to\infty}\sum_{k=0}^{N} a_{\tau(k)}. \tag{10.221}$$

Démonstration. Nous posons $a = \sum_{i\in I} a_i$. Soit $\epsilon > 0$ et J_0 comme dans la définition. Nous choisissons

$$N > \max_{j\in J_0}\{\tau^{-1}(j)\}. \tag{10.222}$$

En tant que sommes sur des ensembles finis, nous avons l'égalité

$$\sum_{k=0}^{N} a_{\tau(k)} = \sum_{j\in J_0} a_j \tag{10.223}$$

où J est un sous-ensemble de I contenant J_0. Soit J fini dans I tel que $J_0 \subset J$. Nous avons alors

$$\Big\|\sum_{k=0}^{N} a_{\tau(k)} - a\Big\| = \Big\|\sum_{j\in J} a_j - a\Big\| < \epsilon. \tag{10.224}$$

Nous avons prouvé que pour tout ϵ, il existe N tel que $n > N$ implique $\|\sum_{k=0}^{n} a_{\tau(k)} - a\| < \epsilon$. □

La réciproque n'est pas vraie. Même en supposant que $\lim_{N\to\infty}\sum_{n=0}^{N} a_n$ existe, il n'est pas forcé que $\sum_{n\in\mathbb{N}} a_n$ existe. Cela est une conséquence de l'exemple 10.97.

Corolaire 10.100.
Nous pouvons permuter une somme dénombrable et une fonction linéaire continue. C'est-à-dire que si f est une fonction linéaire continue sur l'espace vectoriel normé E et $(a_i)_{i\in I}$ une famille sommable dans E alors

$$f\left(\sum_{i\in I} a_i\right) = \sum_{i\in I} f(a_i). \tag{10.225}$$

Problèmes et choses à faire

À mon avis, ce corolaire est faux parce qu'il manque l'hypothèse que la famille $f(a_i)$ est sommable. Voir la proposition 10.102.

Démonstration. En utilisant une bijection τ entre I et $\mathbb{N}$ avec la proposition 10.99 ainsi que le résultat connu à propos des sommes sur $\mathbb{N}$, nous avons

$$f\left(\sum_{i\in I} a_i\right) = f\left(\sum_{k=0}^{\infty} a_{\tau(k)}\right) \tag{10.226a}$$

$$= \sum_{k=0}^{\infty} f(a_{\tau(k)}) \tag{10.226b}$$

$$= \sum_{i\in I} f(a_i). \tag{10.226c}$$

Notons que le passage à (10.226b) n'est pas du tout une trivialité à deux francs cinquante. Il s'agit d'écrire la somme comme la limite des sommes partielles, et de permuter f avec la limite en invoquant la continuité, puis de permuter f avec la somme partielle en invoquant sa linéarité.

Ah, tiens et tant qu'on y est-à-dire qu'il y a des choses évidentes qui ne le sont pas, oui, il existe des applications linéaires non continues, voir le thème 22. □

La proposition suivante nous enseigne que les sommes infinies peuvent être manipulée de façon usuelle.

Proposition 10.101.
Soit I un ensemble dénombrable. Soient $(a_i)_{i\in I}$ et $(b_i)_{i\in I}$, deux familles de réels positifs telles que $a_i < b_i$ et telles que (b_i) est sommable. Alors (a_i) est sommable.

Si $(a_i)_{i\in I}$ est une famille de complexes telle que $(|a_i|)$ est sommable, alors (a_i) est sommable.

Proposition 10.102 ([1]).
Soit un espace vectoriel normé E et une famille sommable[30] *$\{v_i\}_{i\in I}$ d'éléments de E. Soit $f\colon E\to\mathbb{C}$ une application sur laquelle nous supposons*

(1) f est linéaire et continue ;

(2) la partie $\{f(v_i)_{i\in I}\}$ est sommable.

Alors nous pouvons permuter la somme et f :

$$f\big(\sum_{i\in I} v_i\big) = \sum_{i\in I} f(v_i). \tag{10.227}$$

Démonstration. Soit $\epsilon > 0$; vu que les familles $\{v_i\}_{i\in I}$ et $\{f(v_i)\}_{i\in I}$ sont sommables, nous pouvons considérer les parties finies J_1 et J_2 de I telles que

$$\Big\| \sum_{j\in J_1} v_j - \sum_{i\in I} v_i \Big\| \leqslant \epsilon \tag{10.228}$$

et

$$\Big\| \sum_{j\in J_2} f(v_j) - \sum_{i\in I} f(v_i) \Big\| \leqslant \epsilon \tag{10.229}$$

30. Définition 10.92.

Ensuite nous posons $J = J_1 \cup J_2$. Avec cela nous calculons un peu avec les majorations usuelles :

$$\|f(\sum_{i\in I} v_i) - \sum_{i\in I} f(v_i)\| \leqslant \|f(\sum_{i\in I} v_i) - f(\sum_{j\in J} v_j)\| + \|f(\sum_{j\in J} v_j) - \sum_{i} \in If(v_i)\|. \tag{10.230}$$

Le second terme est majoré par ϵ, tandis que le premier, en utilisant la linéarité de f possède la majoration

$$\|f(\sum_{i\in I} v_i) - f(\sum_{j\in J} v_j)\| = \|f(\sum_{i\in I} v_i - \sum_{j\in J} v_j)\| \leqslant \|f\|\|\sum_{i\in I} v_i - \sum_{j\in J} v_j\| \leqslant \epsilon\|f\|. \tag{10.231}$$

Donc pour tout $\epsilon > 0$ nous avons

$$\|f(\sum_{i\in I} v_i) - \sum_{i\in I} f(v_i)\| \leqslant \epsilon(1 + \|f\|). \tag{10.232}$$

D'où l'égalité (10.227). □

10.6 Produit tensoriel d'espaces vectoriels

Si vous êtes pressés, vous pouvez aller lire la définition 10.112 de produit tensoriel d'espaces vectoriels. Mais si vous étiez vraiment pressés, vous ne seriez pas en train de lire des choses sur le produit tensoriel (il vous suffit de croire que $x \otimes y$ n'est finalement que la concatenation de x et y).

Définition 10.103.
Soient un espace vectoriel V et un sous-espace N. Le ***quotient*** *de V par N, noté V/N est l'ensemble des classes d'équivalence pour la relation $x \sim y$ si et seulement si $x - y \in N$.*

Proposition 10.104.
Soient un espace vectoriel V et un sous-espace vectoriel N de V. Les définitions

(1) $[v] + [w] = [v + w]$

(2) $\lambda[v] = [\lambda v]$

ont un sens et définissent une structure d'espace vectoriel sur V/N.

Démonstration. Un élément général de la classe $[v]$ est de la forme $v + n$ avec $n \in N$. Le calcul suivant montre que la somme fonctionne :

$$[v + n_1] + [w + n_2] = [v + w + n_1 + n_2] = [v + w] \tag{10.233}$$

parce que $n_1 + n_2 \in N$. De même,

$$\lambda[v + n] = [\lambda v + \lambda n] = [\lambda v] \tag{10.234}$$

toujours parce que $\lambda n \in N$.

Notons que nous avons utilisé de façon on ne peut plus cruciale le fait que N soit un sous-espace vectoriel. □

Proposition 10.105.
Si $\{e_i\}$ est une base de V et si N est un sous-espace de V, alors $\{[e_i]\}$ est une partie génératrice de V/N.

Démonstration. Si $x = \sum_k x_k e_k$, alors $[x] = \sum_k x_k [e_k]$, donc oui. □

10.6.1 Somme directe d'espaces vectoriels

Si V et W sont des espaces vectoriels, ce que nous notons $V \oplus W$ n'est rien d'autre que l'espace vectoriel de l'ensemble $V \times W$.

Proposition-définition 10.106 ([222]).
Si V et W sont des espaces vectoriels sur le même corps $\mathbb{K}$, alors les définitions

(1) $(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2)$

(2) $\lambda(v, w) = (\lambda v, \lambda w)$

donnent une structure d'espace vectoriel sur $V \times W$.

Cet espace sera noté $V \oplus W$ et est appelé ***somme directe*** *de V et W.*

Proposition 10.107 ([223]).
Soient un espace vectoriel de dimension finie V et deux sous-espaces M_1 et M_2 satisfaisant

(1) $M_1 \cap M_2 = \{0\}$,

(2) $\dim(M_1) + \dim(M_2) \geqslant \dim(V)$.

Alors $V = M_1 \oplus M_2$.

Démonstration. Soient une base $\{e_i\}_{i \in I}$ de M_1 et $\{f_\alpha\}$ de M_2. Nous commençons par prouver que la partie $B = \{e_i\} \cup \{f_\alpha\}$ est libre.

Supposons en effet avoir des coefficients a_i et b_α tels que

$$\sum_i a_i e_i + \sum_\alpha b_\alpha f_\alpha. \tag{10.235}$$

Cela implique que $\sum_i a_i e_i = -\sum_\alpha b_\alpha f_\alpha$. Or $\sum_i a_i e_i \in M_1$ et $-\sum_\alpha b_\alpha f_\alpha \in M_2$. Donc les éléments $\sum_i a_i e_i$ et $\sum_\alpha b_\alpha f_\alpha$ sont dans $M_1 \cap M_2 = \{0\}$. Nous avons alors les égalités

$$\sum_i a_i e_i = 0 \tag{10.236}$$

et

$$\sum_\alpha b_\alpha f_\alpha = 0. \tag{10.237}$$

La première implique $a_i = 0$ pour tout i et la seconde implique $b_\alpha = 0$ pour tout α.

Donc B est une partie libre de V contenant $\dim(M_1) + \dim(M_2) \geqslant \dim(V)$ éléments. La proposition 4.16(2) nous indique alors qu'en réalité $\dim(M_1) + \dim(M_2) = \dim(V)$. Vu que B est une partie libre contenant $\dim(V)$ éléments, c'est une base par la proposition 4.16(2). □

La proposition suivante est une version plus « pragmatique » de la proposition 4.135.

Proposition 10.108 ([223]).
Soient un espace euclidien[31] *de dimension finie V ainsi qu'un sous-espace M. Nous posons*

$$M^\perp = \{x \in V \text{ tel que } x \cdot y = 0 \forall y \in M\}. \tag{10.238}$$

Alors $M \oplus M^\perp = V$.

Démonstration. D'abord si $x \in M \cap M^\perp$, alors $x \cdot x = 0$ et donc $x = 0$. Donc nous avons déjà $M \cap M^\perp = \{0\}$. Nous considérons une base $\{b_1, \dots, b_k\}$ de M, et nous définissons l'application linéaire

$$\begin{aligned} f \colon V &\to \mathbb{R}^k \\ x &\mapsto (x \cdot b_1, \dots, x \cdot b_k). \end{aligned} \tag{10.239}$$

Nous avons que $M^\perp = \ker(f)$. Le théorème du rang 4.39 nous indique que

$$\dim(V) = \dim\big(\ker(f)\big) + \dim\big(\mathrm{Image}(f)\big) \leqslant \dim(M^\perp) + k = \dim(M^\perp) + \dim(M). \tag{10.240}$$

31. Qui possède un produit scalaire, définition 7.188.

Une justification : vu que f prend ses valeurs dans $\mathbb{R}^k$, la dimension de son image est majorée par k.

Nous en déduisons que

$$\dim(M) + \dim(M^\perp) \geqslant \dim(V), \tag{10.241}$$

et la proposition 10.107 nous permet de conclure que $M \oplus M^\perp = V$. □

10.6.2 Les produits tensoriels

Nous allons procéder en deux temps. D'abord nous allons définir ce qu'est *un* produit tensoriel entre deux espaces vectoriels V et W, et nous allons montrer que tous les produits tensoriels possibles sont isomorphes. Ensuite nous allons montrer qu'un produit tensoriel existe en en construisant un. Voir la proposition 10.114.

Définition 10.109 ([224]).
*Soient deux espaces vectoriels V et W. Un **produit tensoriel** de V et W est un couple (T, h) où T est un espace vectoriel et $h\colon V \oplus W \to T$ est une application*

(1) bilinéaire[32]

(2) surjective

(3) telle que pour tout espace vectoriel U et toute applications bilinéaire $f\colon V \oplus W \to U$, il existe une application linéaire $g\colon T \to U$ telle que $f = g \circ h$.

*La propriété (3) est appelée **propriété universelle** du produit tensoriel.*

Définition 10.110.
*Un **morphisme** entre (T, h) et (T', h') est une application linéaire $\psi\colon T \to T'$ telle que $h' = \psi \circ h$.*

*Nous parlons d'**isomorphisme** si ψ a un inverse qui est également un morphisme.*

Proposition 10.111 ([224]).
Si V et W sont des espaces vectoriels, tous les produits tensoriels entre V et W sont isomorphes entre eux au sens de la définition 10.110.

Plus précisément, si (T, h) et (T', h') sont deux produits tensoriels de V et W, alors

(1) il existe une unique unique application linéaire $g\colon T \to T'$ telle que $h' = g \circ h$,

(2) cette application g est inversible.

En particulier, l'application g est un isomorphisme d'espaces vectoriels.

Démonstration. Soient deux produits tensoriels (T, h) et (T', h').

Existence L'application $h'\colon V \oplus W \to T'$ est bilinéaire, et (T, h) est un produit tensoriel. Donc il existe $g\colon T \to T'$ tel que $h' = g \circ h$. De même, il existe une application $g'\colon T' \to T$ telle que $h = g' \circ h$.

Unicité En ce qui concerne l'unicité, vu que $h\colon V \oplus W \to T$ est surjective, la relation $h' = g \circ h$ prescrit les valeurs de g sur tous les éléments de T.

Inversible Ces deux applications g et g' vérifient $h' = gg'h$ et $h = g'gh$, et de plus $h\colon V \oplus W \to T$ est surjective. Soient $t \in T$ et $x \in V \oplus W$ tel que $t = h(x)$. Nous avons $h(x) = g'gh(x)$. C'est-à-dire $t = (g' \circ g)(t)$. De même dans l'autre sens, il existe $x' \in V \oplus W$ tel que $t = h'(x')$. En appliquant l'égalité $h' = gg'h'$ à x', nous trouvons $t = (g \circ g')(t)$.

Tout cela pour dire que $g' = g^{-1}$. Cette application g est donc un isomorphisme de produits tensoriels entre (T, h) et (T', h').

Au final, l'application $g\colon T \to T'$ étant linéaire et inversible, elle est un isomorphisme d'espaces vectoriels. □

32. Définition 7.182.

Tout cela est fort bien : nous avons unicité à isomorphisme près du produit tensoriel d'espaces vectoriels. Mais nous n'avons pas encore de certitudes à propos de l'existence d'un couple (T, h) vérifiant les propriétés demandées pour être un produit tensoriel.

Nous allons maintenant construire un produit tensoriel.

10.6.3 Le produit tensoriel

C'est le moment pour vous de relire la définition 4.22 d'espace vectoriel librement engendré, et surtout le lemme 4.23 qui en donne une base.

Définition 10.112 ([224]).
Soient deux espaces vectoriels V et W sur le corps commutatif[33] $\mathbb{K}$. *Dans $F_{\mathbb{K}}(V \times W)$ nous considérons les sous-espaces suivants :*

$$A_1 = \{\delta_{(v_1,w)} + \delta_{(v_2,w)} - \delta_{(v_1+v_2,w)} \text{ tel que } v_1, v_2 \in V, w \in W\} \tag{10.242a}$$
$$A_2 = \{\delta_{(v,w_1)} + \delta_{(v,w_2)} - \delta_{(v,w_1+w_2)} \text{ tel que } v \in V, w_1, w_2 \in W\} \tag{10.242b}$$
$$A_3 = \{\lambda\delta_{v,w} - \delta_{(\lambda v,w)} \text{ tel que } v \in V, w \in W, \lambda \in \mathbb{K}\} \tag{10.242c}$$
$$A_4 = \{\lambda\delta_{v,w} - \delta_{(v,\lambda w)} \text{ tel que } v \in V, w \in W, \lambda \in \mathbb{K}\}. \tag{10.242d}$$

Nous considérons alors $N = \mathrm{Span}(A_1, A_2, A_3, A_4)$ et le quotient

$$V \otimes_{\mathbb{K}} W = F_{\mathbb{K}}(V \times W)/N. \tag{10.243}$$

Ce dernier espace vectoriel est le ***produit tensoriel*** *de V par W.*

Remarque 10.113.
Quelque remarques.

(1) Les éléments de $V \otimes W$ ne s'écrivent pas tous sous la forme $v \otimes w$. Certains ont vraiment besoin d'être écrits avec des sommes. En cela, la situation de $V \otimes W$ est réellement différente de celle de $V \times W$. Dans ce dernier, tous les éléments sont des couples.

(2) La classe de l'élément $\delta_{(v,w)} \in F(V \times W)$ sera d'habitude noté $v \otimes w$.

(3) Pour insister sur la notion de classe, nous allons aussi noter $[x]$ la classe de $x \in F(V \times W)$.

(4) L'arithmétique dans $V \otimes W$ est relativement simple. En ajoutant et soustrayant le même élément de A_3 nous avons par exemple

$$(\lambda v) \otimes w = (\lambda v) \otimes w + \lambda(v \otimes w) - (\lambda v) \otimes w. \tag{10.244}$$

Nous obtenons de cette façon

$$\lambda(v \otimes w) = (\lambda v) \otimes w = v \otimes (\lambda w), \tag{10.245}$$

que nous noterons $\lambda v \otimes w$ sans plus de précision.

Proposition 10.114 ([224]).
L'espace vectoriel $V \times W$ muni de

$$\begin{aligned} h\colon V \oplus W &\to V \otimes W \\ (v, w) &\mapsto v \otimes w \end{aligned} \tag{10.246}$$

est un produit tensoriel entre V et W.

Démonstration. Nous devons prouver les conditions de la définition 10.109.

<u>h **est bilinéaire**</u> Ce sont des calculs tels que faits dans la remarque 10.113(4) qui font le travail.

33. À part mention du contraire, tous les corps du Frido sont commutatifs.

h est surjective Un élément de $V \otimes W$ est la classe d'un élément de $F(V \times W)$, c'est-à-dire de la forme

$$\big[\sum_{i\alpha} \delta_{(v_i,w_\alpha)}\big] = \sum_{i\alpha} a_{i\alpha} v_i \otimes w_\alpha. \tag{10.247}$$

Cet élément est dans l'image de h comme le montre le calcul suivant [34] :

$$h\big(\sum_{i\alpha}(v_i, w_\alpha)\big) = \sum_{i\alpha} a_{i\alpha} h(v_i, w_\alpha) = \sum_{i\alpha} v_i \otimes w_\alpha. \tag{10.248}$$

Propriété universelle Soient un espace vectoriel U et une application linéaire $f\colon V \oplus W \to U$. Nous devons trouver une application linéaire $g\colon V \otimes W \to U$ telle que $f = g \circ h$. Pour cela nous commençons par considérer l'application

$$\begin{aligned} g\colon F(V \times W) &\to U \\ \delta_{(v,w)} &\mapsto f(v,w) \end{aligned} \tag{10.249}$$

définie sur tout $F(V \times W)$ par linéarité sans encombres parce que les $\delta_{v,w}$ forment une base par le lemme 4.23.

Nous démontrons que $g(N) = 0$ pour avoir le droit de passer g aux classes et le considérer comme application partant de $V \otimes W$ au lieu de $F(V \times W)$. Prenons par exemple

$$g\big(\delta_{(v_1,w)} + \delta_{(v_2,w)} - \delta_{(v_1+v_2,w)}\big) = g(\delta_{(v_1,w)}) + g(\delta_{(v_2,w)}) - g(\delta_{v_1+v_2,w}) \tag{10.250a}$$

$$= f(v_1, w) + f(v_2, w) - f(v_1 + v_2, w) \tag{10.250b}$$

$$= 0 \tag{10.250c}$$

par la bilinéarité de f. Cela montre que $g(A_1) = 0$. Nous montrons de même que $g(A_2) = g(A_3) = g(A_4) = 0$, et enfin toujours par linéarité que $g(N) = 0$. Pour rappel, les éléments de N sont les combinaisons linéaires finies d'éléments de A_1, A_2, A_3 et A_4.

Par passage aux classes, nous avons une application (que nous notons également g)

$$g\colon F(V \times W)/N \to U \tag{10.251}$$

vérifiant $g(v \otimes w) = f(v, w)$. Mais comme $h(v, w) = v \otimes w$, nous avons $g \circ h\colon V \oplus W \to U$ vérifiant $g \circ h = f$.

L'espace vectoriel $V \otimes W$ est donc un produit tensoriel. □

10.115.
Vu que $V \otimes W$ est un produit tensoriel de V et W, et vu qu'il y a unicité par la proposition 10.111, nous avons bien le droit de dire que $V \otimes W$ est *le* produit tensoriel. Cela justifie le titre.

10.116.
Les prochains lemmes et propositions vont nous dire que l'application

$$\begin{aligned} \varphi\colon V^* \otimes W &\to \mathcal{L}(V, W) \\ \alpha \otimes w &\mapsto \big(v \mapsto \alpha(v)w\big) \end{aligned} \tag{10.252}$$

est un isomorphisme d'espaces vectoriels lorsque V est de dimension finie. Vu que nous aimons les énoncés très explicites, ça va être découpé en plusieurs morceaux, l'énoncé va devenir un peu long ; mais c'est pour la bonne cause.

Lemme 10.117.
Soient deux espaces vectoriels V et W dont W est de dimension finie. Alors l'application définie par

$$\begin{aligned} \varphi\colon F(V^* \times W) &\to \mathcal{L}(V, W) \\ \delta_{(\alpha,w)} &\mapsto \big(v \mapsto \alpha(v)w\big) \end{aligned} \tag{10.253}$$

sur la base « canonique » de $F(V^* \times W)$ passe aux classes.

34. Faites bien la distinction entre $\delta_{v,w}$, (v, w) et $v \otimes w$. Sachez dans quel ensemble se trouvent chacun de ces trois objets.

Démonstration. Avec les notations de la définition 10.112 nous devons prouver que $\varphi(N) = 0$. Nous montrons que $\varphi(A_4) = 0$, et nous vous laissons faire les autres. Pour $\lambda \in \mathbb{K}$, $\alpha \in V^*$ et $w \in W$ en utilisant la linéarité de φ nous avons :

$$\begin{aligned}\varphi\big(\lambda\delta_{(\alpha,w)} - \delta_{(\alpha,\lambda w)}\big)v &= \lambda\varphi(\delta_{(\alpha,w)})(v) - \varphi(\delta_{(\alpha,\lambda w)})(v) && (10.254a)\\ &= \lambda\alpha(v)w - \alpha(v)(\lambda w) && (10.254b)\\ &= 0 && (10.254c)\end{aligned}$$

parce que $\alpha(v)(\lambda w) = \lambda\alpha(v)w$ du fait que $\mathbb{K}$ est commutatif. La commutativité de $\mathbb{K}$ est ce qui permet de permuter le produit $\lambda\alpha(v)$.

Nous laissons à la lectrice le soin de prouver que $\varphi(A_1) = \varphi(A_2) = \varphi(A_3) = 0$. □

Lemme 10.118.
Si W est de dimension finie, alors $\mathcal{L}(V, W)$ muni de

$$\begin{aligned} h'\colon V^* \oplus W &\to \mathcal{L}(V, W)\\ (\alpha, w) &\mapsto \big(v \mapsto \alpha(v)w\big)\end{aligned} \tag{10.255}$$

est un produit tensoriel[35] *de V^* par W.*

Démonstration. Nous devons prouver que

— h est bilinéaire,

— h est surjective

— pour tout espace vectoriel U, et pour toute application bilinéaire $f\colon V^* \oplus W \to U$, il existe une application linéaire $g\colon \mathcal{L}(V, W) \to U$ tel que $f = g \circ h$.

Bilinéaire Le fait que h soit bilinéaire est une simple vérification.

Surjective L'espace W étant de dimension finie, nous pouvons en considérer une base $\{z_i\}_{i\in I}$. Soit $\alpha \in \mathcal{L}(V, W)$. Si $v \in V$, l'élément $\alpha(v)$ peut être décomposé dans la base $\{z_i\}$, ce qui définit des applications linéaires $\alpha_i\colon V \to \mathbb{K}$ par

$$\alpha(v) = \sum_{i\in I} \alpha_i(v)z_i. \tag{10.256}$$

Notons que $\alpha_i \in V^*$. En comparant avec la définition de h', nous voyons que

$$\alpha(v) = \sum_i h(\alpha_i, z_i)(v), \tag{10.257}$$

c'est-à-dire $\alpha = \sum_i h(\alpha_i, w_i) = h\big(\sum_i(\alpha_i, z_i)\big)$. Nous avons donc bien $\alpha \in h(V^* \oplus W)$.

Propriété universelle Soient un espace vectoriel U et une application bilinéaire $f\colon V^* \oplus W \to U$. Pour $\alpha \in \mathcal{L}(V, W)$ nous définissons $g(\alpha)$ comme suit. D'abord nous écrivons α sous la forme

$$\alpha(v) = \sum_i \alpha_i(v)z_i, \tag{10.258}$$

et nous posons

$$g(\alpha) = \sum_i f(\alpha_i, z_i). \tag{10.259}$$

35. Définition 10.109.

Avec cette définition, en posant $w = \sum_i w_i z_i$, nous avons

$$\begin{aligned}
(g \circ h')(\alpha, w) &= g\big(v \mapsto \alpha(v)w\big) && (10.260a)\\
&= g\big(v \mapsto \sum_i \alpha(v) w_i z_i\big) && (10.260b)\\
&= \sum_i f(w_i \alpha, z_i) && (10.260c)\\
&= \sum_i f(\alpha, w_i z_i) && (10.260d)\\
&= f(\alpha, \sum_i w_i z_i) && (10.260e)\\
&= f(\alpha, w). && (10.260f)
\end{aligned}$$

Cela prouve que $g \circ h = f$.

□

Proposition 10.119 ([225]).
Soient deux espaces vectoriels V et W dont V est de dimension finie. Alors l'application

$$\begin{aligned}
\varphi\colon V^* \otimes W &\to \mathcal{L}(V, W)\\
\alpha \otimes w &\mapsto \big(v \mapsto \alpha(v)w\big)
\end{aligned} \tag{10.261}$$

est bien définie[36] *et est un isomorphisme d'espaces vectoriels.*

Démonstration. Le lemme 10.118 donne une structure de produit tensoriel de V^* par W sur $\mathcal{L}(V, W)$. Rappelons les structures :

$$\begin{aligned}
h\colon V^* \oplus W &\to V^* \otimes W\\
(\alpha, w) &\mapsto \alpha \otimes w
\end{aligned} \tag{10.262}$$

et

$$\begin{aligned}
h'\colon V^* \oplus W &\to \mathcal{L}(V, W)\\
(\alpha, w) &\mapsto \big[v \mapsto \alpha(v)w\big].
\end{aligned} \tag{10.263}$$

La proposition 10.111 a déjà fait tout le boulot. La seule chose à faire est de vérifier qu'il existe une application $\varphi\colon V^* \otimes W \to \mathcal{L}(V, W)$ vérifiant simultanément les deux conditions suivantes :

(1) $\varphi(\alpha \otimes w) = \big[v \mapsto \alpha(v)w\big]$
(2) $h' = \varphi\varphi \circ h$.

La seconde condition assure que φ sera un isomorphisme d'espaces vectoriels.

L'existence de φ vérifiant la condition (1) est un effet du lemme 10.117 qui donne une fonction sur $F(V^* \times W)$ dont le φ qui nous concerne est un quotient. Il reste à voir que cette application vérifie $h' = \varphi \circ h$.

En nous rappellant que $\alpha \otimes w = [\delta_{(\alpha,w)}]$ et en écrivant φ à la fois l'application et son passage au quotient,

$$(\varphi \circ h)(\alpha, w) = \varphi(\alpha \otimes w) = \varphi\big([\delta_{(\alpha,w)}]\big) = \varphi(\delta_{(\alpha,w)}). \tag{10.264}$$

En appliquant à $v \in V$ nous avons :

$$(\varphi \circ h)(\alpha, w)v = \varphi(\delta_{(\alpha,w)})v = \alpha(v)w = h'(\alpha, w)v. \tag{10.265}$$

Et voila. Nous avons $\varphi \circ h = h'$. □

Une conséquence de la proposition 10.119 est que

$$\dim(V \otimes W) = \dim(V)\dim(W) \tag{10.266}$$

via le lemme 4.34(2).

36. Au sens où il existe une fonction φ définie sur tout $V^* \otimes W$ qui se réduit à cela pour les éléments de la forme $\alpha \otimes w$.

10.6.4 Bases

Voici un lemme entièrement dédié au principe « dans le Frido, on ne fait pas d'abus de notations, sauf pour la logique formelle et la théorie des ensembles, que nous admettons ».

Lemme 10.120 ([1]).
Si $\tau\colon V_1 \to V_2$ est un isomorphisme d'espaces vectoriels, alors

$$\begin{aligned}\varphi\colon V_1 \otimes W &\to V_2 \otimes W\\ v \otimes w &\mapsto \tau(v) \otimes W\end{aligned} \tag{10.267}$$

est un isomorphisme d'espaces vectoriels.

Démonstration. Comme d'habitude, l'expression (10.267) ne définit pas réellement φ parce que nous ne savons pas du tout si $\{v \otimes w \text{ tel que } v \in V, w \in W\}$ est plus ou moins une base de $V \otimes W$ [37]. Ce que dit réellement ce lemme est qu'il existe une application $V_1 \otimes W \to V_2 \otimes W$ qui est isomorphisme et qui se réduit à l'expression donnée dans le cas d'éléments de $V_1 \otimes W$ de la forme $v \otimes w$.

L'application

$$\begin{aligned}\varphi_0\colon F(V_1 \times W) &\to F(V_2 \times W)\\ \delta(v,w) &\mapsto \delta_{\big(\tau(v),w\big)}\end{aligned} \tag{10.268}$$

est un isomorphisme.

Cette application passe aux classes, mais pas au sens où $x \in [y]$ impliquerait $\varphi_0(x) = \varphi_0(y)$; au sens où si $x \in [y]$, alors $\varphi_0(x) \in [\varphi_0(y)]$. Par exemple

$$\varphi_0\big(\lambda\delta_{(v,w)} - \delta_{(v,\lambda w)}\big) = \lambda\delta_{\big(\tau(v),w\big)} - \delta_{\big(\tau(v),w\big)} \in [0]. \tag{10.269}$$

Nous vous laissons le soin de vérifier les égalités correspondantes pour les autres parties de N.

Le passage au classes de φ_0 signifie que l'on considère l'application

$$\begin{aligned}\varphi\colon V_1 \otimes W &\to V_2 \otimes W\\ [x] &\mapsto [\varphi_0(x)]\end{aligned} \tag{10.270}$$

où vous aurez noté que la prise de classe à gauche n'est pas la même que celle à droite.

Il faut prouver que ce φ est un isomorphisme. En ce qui concerne la linéarité,

$$\begin{aligned}\varphi\big([x]+[y]\big) &= \varphi\big([x+y]\big) && (10.271a)\\ &= [\varphi_0(x+y)] && (10.271b)\\ &= [\varphi_0(x)+\varphi_0(y)] && (10.271c)\\ &= [\varphi_0(x)]+[\varphi_0(y)] && (10.271d)\\ &= \varphi([x])+\varphi([y]). && (10.271e)\end{aligned}$$

Je vous laisse le reste de la linéarité. Et en ce qui concerne le fait que ce soit une bijection, allez-y. □

Proposition 10.121 ([225]).
Soient des espaces vectoriels de dimension finie V et W. Soient une base $\{e_i\}$ de V et une base $\{f_\alpha\}$ de W.

Alors :

(1) La partie $\{e_i \otimes f_\alpha\}$ est une base de $V \otimes W$.

(2) Au niveau des dimensions, $\dim(V \otimes W) = \dim(V)\dim(W)$.

37. Ne lisez pas la proposition 10.121 qui dévoile toute l'intrigue.

Démonstration. Vu que V est de dimension finie, nous avons un isomorphisme d'espaces vectoriels $V^* = V$, et même un isomorphisme d'espaces vectoriels

$$\begin{aligned} \tau\colon V &\to (V^*)^* \\ \tau(v)\alpha &= \alpha(v). \end{aligned} \tag{10.272}$$

Recopions l'isomorphisme de la proposition 10.119 en utilisant V^* au lieu de V :

$$\begin{aligned} \psi_0\colon (V^*)^* \otimes W &\to \mathcal{L}(V^*, W) \\ \tau(v) \otimes w &\mapsto \big(\alpha \mapsto \tau(v)(\alpha)w = \alpha(v)w\big). \end{aligned} \tag{10.273}$$

En écrivant cela, nous avons tenu compte du fait que tout élément de $(V^*)^*$ peut être écrit de façon univoque sous la forme $\tau(v)$ pour un certain $v \in V$.

Vu que τ est un isomorphisme, l'application suivante est encore un isomorphisme[38] :

$$\begin{aligned} \psi\colon V \otimes W &\to \mathcal{L}(V^*, W) \\ v \otimes w &\mapsto \big(\alpha \mapsto \alpha(v)w\big). \end{aligned} \tag{10.274}$$

Nous avançons. Vu que nous avons un isomorphisme, nous pouvons faire passer des bases. Le lemme 4.34 nous donne une base de $\mathcal{L}(V^*, W)$ en les éléments $\beta_{i\alpha}\colon V^* \to W$ définies par

$$\beta_{ij}(\alpha) = \alpha(e_i)f_\alpha. \tag{10.275}$$

Donc $\{\psi^{-1}(\beta_{i\alpha})\}$ est une base de $V \otimes W$.

Pour $a = \sum_i a_i e_i^*$ (base duale, définition 4.130) nous avons :

$$\psi(e_i \otimes f_\alpha)a = a(e_i)f_\alpha = \beta_{i\alpha}(a). \tag{10.276}$$

Cela prouve que $\psi^{-1}(\beta_{i\alpha}) = e_i \otimes f_\alpha$, et donc que ces $e_i \otimes f_\alpha$ est une base de $V \otimes W$.

La formule concernant les dimensions est simplement la définition 4.13 de la dimension : le nombre d'éléments dans une base. □

Exemple 10.122.
Dans le produit tensoriel $\mathbb{R} \otimes \mathbb{R}$, nous avons $x \otimes 1 = 1 \otimes x = x(1 \otimes x)$ pour tout $x \in \mathbb{R}$. Et si $x \geqslant 0$ nous avons aussi $x \otimes 1 = \sqrt{x} \otimes \sqrt{x}$. △

10.6.5 Norme

Nous considérons des espaces vectoriels V et W de dimension finie. L'application (10.274) donne un isomorphisme d'espaces vectoriels

$$\begin{aligned} \psi\colon V \otimes W &\to \mathcal{L}(V^*, W) \\ v \otimes w &\mapsto \big(\alpha \mapsto \alpha(v)w\big). \end{aligned} \tag{10.277}$$

Et ça, c'est très bien, parce que nous connaissons une norme sur $\mathcal{L}(V^*, W)$: la norme opérateur 10.3.

Définition 10.123 ([1]).
Soient deux espaces vectoriels normés de dimension finie V et W. Sur $V \otimes W$ nous définissons, pour $t \in V \otimes W$

$$\|t\| = \|\psi(t)\|_{\mathcal{L}(V^*, W)}. \tag{10.278}$$

Lemme 10.124 ([1]).
La norme sur $V \otimes W$ vérifie

$$\|v \otimes w\| = \|v\|\|w\| \tag{10.279}$$

pour tout $v \in V$ et $w \in W$.

38. Lemme 10.120.

Démonstration. C'est un simple(?) calcul :

$$\|v \otimes w\| = \|\psi(v \otimes w)\| = \|\alpha \mapsto \alpha(v)w\| = \sup_{\|\alpha\|=1} \|\alpha(v)w\| = \sup_{\|\alpha\|=1} |\alpha(v)|\|w\|. \tag{10.280}$$

Étant donné que V est de dimension finie, $\sup_{\|\alpha\|=1} |\alpha(v)| = \|v\|$ [39]. Nous avons donc

$$\|v \otimes w\| = \|v\|\|w\|. \tag{10.281}$$

□

Le lemme suivant montre que $\mathbb{R} \otimes \mathbb{R}$ n'est pas du tout $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. Au contraire, $\mathbb{R} \otimes \mathbb{R}$ est isomorphe à $\mathbb{R}$.

Lemme 10.125 ([1]).
L'application

$$\begin{aligned} \varphi\colon \mathbb{R} \otimes \mathbb{R} &\to \mathbb{R} \\ 1 \otimes 1 &\mapsto 1 \end{aligned} \tag{10.282}$$

prolongée par linéarité est un isomorphisme isométrique.

Démonstration. D'abord une base de $\mathbb{R}$ est $\{1\}$; donc une base de $\mathbb{R} \otimes \mathbb{R}$ est $\{1 \otimes 1\}$ par la proposition 10.121. Donc l'application proposée se prolonge par linéarité à tout $\mathbb{R} \otimes \mathbb{R}$.

Le fait que φ soit une bijection provient du fait que φ transforme une base en une base ; si vous n'y croyez pas, la vérification de l'injectivité et de la surjectivité est facile.

Pour que φ soit isométrique, nous faisons le calcul

$$\|\varphi(x \otimes y)\| = \|xy(1 \otimes 1)\| = |xy|\|1 \otimes 1\| = |xy| = \|x \otimes y\|. \tag{10.283}$$

Nous avons utilisé la propriété 7.113(2) d'une norme ainsi que le lemme 10.124 pour la norme sur $\mathbb{R} \otimes \mathbb{R}$. □

10.6.6 Applications bilinéaires, matrices et produit tensoriel

Soit E, un espace vectoriel de dimension finie. Si α et β sont deux formes linéaires sur un espace vectoriel E, nous définissons $\alpha \otimes \beta$ comme étant la 2-forme donnée par

$$(\alpha \otimes \beta)(u, v) = \alpha(u)\beta(v). \tag{10.284}$$

Si a et b sont des vecteurs de E, ils sont vus comme des formes sur E via le produit scalaire et nous avons

$$(a \otimes b)(u, v) = (a \cdot u)(b \cdot v). \tag{10.285}$$

Cette dernière équation nous incite à pousser un peu plus loin la définition de $a \otimes b$ et de simplement voir cela comme la matrice de composantes

$$(a \otimes b)_{ij} = a_i b_j. \tag{10.286}$$

Cette façon d'écrire a l'avantage de ne pas demander de se souvenir qui est une vecteur ligne, qui est un vecteur colonne et où il faut mettre la transposée. Évidemment $(a \otimes b)$ est soit ab^t soit $a^t b$ suivant que a et b soient ligne ou colonne.

39. Cela est une des raisons pour lesquelles nous sommes en dimension finie : je ne sais pas si cette égalité est vraie en dimension inifinie.

10.6.7 Application d'opérateurs

Lemme 10.126.
Soient $x, y \in E$ et A, B deux opérateurs linéaires sur E vus comme matrices. Alors

$$(Ax \otimes By) = A(x \otimes y)B^t. \tag{10.287}$$

Démonstration. Calculons la composante ij de la matrice $(Ax \otimes By)$. Nous avons

$$\begin{aligned}
(Ax \otimes By)_{ij} &= (Ax)_i(By)_j && (10.288a)\\
&= \sum_{kl} A_{ik}x_kB_{jl}y_l && (10.288b)\\
&= A_{ik}(x \otimes y)_{kl}B_{jl} && (10.288c)\\
&= \big(A(x \otimes y)B^t\big)_{ij}. && (10.288d)
\end{aligned}$$

□

10.7 Calcul différentiel dans un espace vectoriel normé

Quelques motivations pour la notion de différentielle sont données dans 11.20.1.

10.7.1 Définition de la différentielle

Proposition-définition 10.127 ([1]).
Soient deux espaces vectoriels normés E et F ainsi qu'une fonction $f : \mathcal{U} \to F$ où $\mathcal{U}$ est un ouvert de E.

Si il existe une une application linéaire $T \in \mathcal{L}(E, F)$ satisfaisant

$$\lim_{\substack{h \to 0\\ h \in E}} \frac{f(a+h) - f(a) - T(h)}{\|h\|_E} = 0, \tag{10.289}$$

alors il en existe une seule.

*Dans ce cas nous disons que f est **différentiable au point** a et l'application T ainsi définie est appelée **différentielle** de f au point a, et nous la notons df_a.*

Démonstration. Soient deux applications linéaires T_1, T_2 satisfaisant la condition (10.289). Nous avons

$$\frac{\|T_1(h) - T_2(h)\|_F}{\|h\|_E} \leqslant \frac{\|T_1(h) - f(a+h) + f(a)\|}{\|h\|} + \frac{\|f(a+h) - f(a) - T_2(h)\|}{\|h\|} \to 0. \tag{10.290}$$

Nous avons donc

$$\lim_{h \to 0} \frac{\|(T_1 - T_2)(h)\|_F}{\|h\|_E} = 0. \tag{10.291}$$

Soit $\epsilon > 0$. Ce que signifie la limite est qu'il existe un $r > 0$ tel que pour tout $u \in B_E(0, r)$, nous ayons

$$\frac{\|(T_1 - T_2)(u)\|_F}{\|u\|_E} < \epsilon. \tag{10.292}$$

Soit $v \in E$. Nous considérons $\lambda \in \mathbb{R}$ tel que $\lambda v \in B(0, r)$, par exemple $\lambda < r/\|v\|$. Nous avons

$$\epsilon > \frac{\|(T_1 - T_2)(\lambda v)\|_F}{\|\lambda v\|_E} = \frac{\|(T_1 - T_2)(v)\|}{\|v\|}. \tag{10.293}$$

Cela donne

$$\|(T_1 - T_2)(v)\| < \|v\|\epsilon. \tag{10.294}$$

Nous avons donc $\|(T_1 - T_2)(v)\| = 0$, soit $T_1(v) = T_2(v)$. □

L'application différentielle

$$\begin{aligned} df\colon E &\to \mathcal{L}(E,F) \\ a &\mapsto df_a \end{aligned} \tag{10.295}$$

est également très importante.

Définition 10.128.
Une application $f\colon E \to F$ est de ***classe*** C^1 *lorsque l'application différentielle $df\colon E \to \mathcal{L}(E,F)$ est continue. Voir aussi les définitions 11.267 pour les applications de classe C^k.*

Remarque 10.129.
L'application norme étant continue, le critère du théorème 7.105 est en réalité assez général. Par exemple à partir d'une application différentiable[40] $f\colon X \to Y$ nous pouvons considérer la fonction réelle

$$a \mapsto \|df_a\| \tag{10.296}$$

où la norme est la norme opérateur[41]. Si f est de classe C^1 alors cette application est continue et donc bornée sur un compact K de X.

10.7.2 Accroissements finis

Lemme 10.130.
Soit une fonction $f\colon E \to V$ (espaces vectoriels normés) différentiable en $a \in E$. Alors il existe une fonction $\alpha\colon E \to V$ telle que

$$\begin{cases} \lim\limits_{h\to 0} \dfrac{\alpha(h)}{\|h\|} = 0 & \text{(10.297a)} \\ f(a+h) = f(a) + df_a(h) + \alpha(h). & \text{(10.297b)} \end{cases}$$

Démonstration. Il s'agit seulement de poser

$$\alpha(h) = f(a+h) - f(a) - df_a(h). \tag{10.298}$$

Le fait que $\alpha(h)/\|h\| \to 0$ est alors la définition de la différentiabilité de f. □

10.7.3 (non ?) Différentiabilité des applications linéaires

Si E et F sont deux espaces vectoriels nous notons $\mathcal{L}(E,F)$ l'ensemble des applications linéaires de E vers F et $L(E,F)$ l'ensemble des applications linéaires continues de E vers F. Ces espaces seront bien entendu, sauf mention du contraire, toujours munis de la norme opérateur de la définition 10.3.

Lemme 10.131.
Soit une application linéaire f.

(1) Si f est continue, alors elle est différentiable et $df_a(u) = f(u)$ pour tout a et u.

(2) Si f n'est pas continue, alors elle n'est pas différentiable.

Démonstration. La linéarité de f donne :

$$f(a+h) - f(a) - f(h) = 0, \tag{10.299}$$

et donc prendre $T = f$ dans la définition 10.127 fait fonctionner la limite. De plus T est alors continue par hypothèse ; elle est donc bien la différentielle de f.

Supposons que f ne soit pas continue, prenons une application linéaire continue T, et calculons

$$\frac{f(a+h) - f(a) - T(h)}{\|h\|} = \frac{(f-T)(h)}{\|h\|} = (f-T)(e_h) \tag{10.300}$$

40. Définition 10.127.
41. Définition 10.3.

où e_h est le vecteur unitaire dans la direction de h. Vu que f n'est pas continue et que T l'est, l'application $f - T$ n'est pas continue. Elle n'est pas pas bornée par la proposition 10.18. Il existe alors un vecteur h tel que $\|(f-T)(e_h)\| > 1$ (et même plus grand que ce qu'on veut).

Donc la limite de (10.300) pour $h \to 0$ ne peut pas être nulle. □

Lemme 10.132.
Une application linéaire continue est de classe C^∞.

Démonstration. Soit $a \in E$. Étant donné que f est linéaire et continue, elle est différentiable et

$$\begin{aligned} df\colon E &\to L(E,F) \\ a &\mapsto f \end{aligned} \tag{10.301}$$

est une fonction constante et en particulier continue ; nous avons donc $f \in C^1$. Pour la différentielle seconde nous avons $d(df)_a = 0$ parce que $df(a+h) - df(a) = f - f = 0$. Toutes les différentielles suivantes sont nulles. □

10.7.4 Dérivation en chaine et formule de Leibnitz

Proposition 10.133.
Soient $f_i\colon U \to F_i$, des fonctions de classe C^r où U est ouvert dans l'espace vectoriel normé E et les F_i sont des espaces vectoriels normés. Alors l'application

$$\begin{aligned} f = f_1 \times \cdots \times f_n\colon U &\to F_1 \times \cdots \times F_n \\ x &\mapsto \big(f_1(x), \dots, f_n(x)\big) \end{aligned} \tag{10.302}$$

est de classe C^r et

$$d^r f = d^r f_1 \times \dots d^r f_n. \tag{10.303}$$

Démonstration. Soit $x \in U$ et $h \in E$. La différentiabilité des fonctions f_i donne

$$f_i(x+h) = f_i(x) + (df_i)_x(h) + \alpha_i(h) \tag{10.304}$$

avec $\lim_{h\to 0} \alpha_i(h)/\|h\| = 0$. Par conséquent

$$f(x+h) = \big(\dots, f_i(x) + (df_i)_x(h) + \alpha_i(h), \dots\big) \tag{10.305a}$$
$$= \big(\dots, f_i(x), \dots\big) + \big(\dots, (df_i)_x(h), \dots\big) + \big(\dots, \alpha_i(h), \dots\big). \tag{10.305b}$$

Mais la définition 7.230 de la norme dans un espace produit donne

$$\lim_{h\to 0} \frac{\|\big(\alpha_1(h), \dots, \alpha_n(h)\big)\|}{\|h\|} = 0, \tag{10.306}$$

ce qui nous permet de noter $\alpha(h) = \big(\alpha_1(h), \dots, \alpha_n(h)\big)$ et avoir $\lim_{h\to 0} \alpha(h)/\|h\| = 0$. Avec tout ça nous avons bien

$$f(x+h) = f(x) + \big((df_1)_x(h) + \cdots + (df_n)_x(h)\big) + \alpha(h), \tag{10.307}$$

ce qui signifie que f est différentiable et

$$df_x = \big(df_1, \dots, df_n\big). \tag{10.308}$$

□

Théorème 10.134.
Soient des espaces vectoriels normés E, V et W. Nous considérons deux fonctions $f\colon E \to V$ et $g\colon V \to W$. Nous supposons que f est différentiable en $a \in E$ et que g est différentiable en $f(a) \in V$.

Nous supposons de plus que df_a est de norme finie[42].

42. Je ne suis pas totalement certain que cette hypothèse soit nécessaire, mais en tout cas, elle est utilisée.

Alors $g \circ f \colon E \to W$ est différentiable en a et

$$f(g \circ f)_a(u) = df_{f(a)}\big(df_a(u)\big), \tag{10.309}$$

ou encore

$$f(g \circ f)_a = dg_{f(a)} \circ df_a. \tag{10.310}$$

Démonstration. En utilisant le lemme 10.130 pour les fonctions f et g, nous avons

$$f(a+h) = f(a) + df_a(h) + \alpha(h) \tag{10.311}$$

et

$$g\big(f(a)+k\big) = g\big(f(a)\big) + dg_{f(a)}(k) + \beta(k). \tag{10.312}$$

L'application $dg_{f(a)} \circ df_a$ est une application linéaire, et est notre candidat différentielle. En suivant la définition 10.127, nous allons calculer

$$\lim_{h\to 0} \frac{(g \circ f)(a+h) - (g \circ f)(a) - (dg_{f(a)} \circ df_a)(h)}{\|h\|}. \tag{10.313}$$

Si cette limite existe et vaut zéro, alors nous aurons prouvé que le candidat différentielle est correct.

Pour cela, nous emboîtons les formules (10.311) et (10.312) l'une dans l'autre pour avoir :

$$g(a+h) = g\big(f(a)+df_a(h)+\alpha(h)\big) = g\big(f(a)\big) + dg_{f(a)}\big(df_a(h)+\alpha(h)\big) + \beta\big(df_a(h)+\alpha(h)\big). \tag{10.314}$$

Vu que $dg_{f(a)}$ est linéaire, le deuxième terme peut être coupé en deux et après recombinaisons,

$$(g \circ f)(a+h) - (g \circ f)(a) - (df_{f(a)} \circ df_a)(h) = dg_{f(a)}\big(\alpha(h)\big) + \beta\big(df_a(h)+\alpha(h)\big). \tag{10.315}$$

Étant donné que $dg_{f(a)}$ est linéaire,

$$\frac{dg_{f(a)}\big(\alpha(h)\big)}{\|h\|} = dg_{f(a)}\left(\frac{\alpha(h)}{\|h\|}\right) \to 0. \tag{10.316}$$

Il nous reste à voir que

$$\lim_{h\to 0} \frac{\beta\big(df_a(h)+\alpha(h)\big)}{\|h\|} \tag{10.317}$$

existe au vaut zéro. Vu que df_a est linéaire, il existe $M > 0$ tel que[43] $\|df_a(h)\| \leqslant M\|h\|$. D'autre part, vu que $\alpha(h)/\|h\| \to 0$, nous avons $\|\alpha(h)\| \leqslant \|h\|$ pour tout h suffisamment petit.

Donc si h est assez petit, nous avons

$$\|df_a(h) + \alpha(h)\| \leqslant (M+1)\|h\|. \tag{10.318}$$

Soit $\epsilon > 0$. Soit $\delta > 0$ tel que $\|h\| \leqslant \delta$ implique $\beta(h)/\|h\| \leqslant \epsilon$ et (10.318) en même temps. Soit r tel que $(M+1)r < \delta$; et notons que $r < \delta$. Nous considérons alors $h \in B(0,r)$ et nous calculons :

$$\frac{\beta\big(df_a(h)+\alpha(h)\big)}{\|h\|} = \frac{\beta\big(df_a(h)+\alpha(h)\big)}{\|df_a(h)+\alpha(h)\|}\frac{\|df_a(h)+\alpha(h)\|}{\|h\|} \leqslant (M+1)\epsilon. \tag{10.319}$$

La limite (10.317) existe donc et vaut zéro. □

Théorème 10.135 (Différentielle de fonctions composées[226])**.**
Soient E, F et G des espaces vectoriels normés, U ouvert dans E et V ouvert dans F. Soient des applications de classe C^r ($r \geqslant 1$)

$$f \colon U \to V \tag{10.320a}$$

$$g \colon V \to G. \tag{10.320b}$$

Alors l'application $g \circ f \colon V \to G$ est de classe C^r et

$$d(g \circ f)_x = dg_{f(x)} \circ df_x. \tag{10.321}$$

43. Ce M est par exemple la norme opérateur de df_a, comme nous l'assure le lemme 10.10. C'est pour ce passage-ci que nous avons supposé que df_a était de norme finie.

Démonstration. Nous nous fixons $x \in U$. La fonction f est différentiable en $x \in U$ et g en $f(x)$, donc nous pouvons écrire

$$f(x+h) = f(x) + df_x(h) + \alpha(h) \tag{10.322}$$

et

$$g\big(f(x)+u\big) = g\big(f(x)\big) + dg_{f(x)}(u) + \beta(u) \tag{10.323}$$

où la fonction α a la propriété que

$$\lim_{h\to 0} \frac{\|\alpha(h)\|}{\|h\|} = 0; \tag{10.324}$$

et la même chose pour β. La fonction composée en $x+h$ s'écrit donc

$$(g\circ f)(x+h) = g\big(f(x)+df_x(h)+\alpha(h)\big) = g\big(f(x)\big) + dg_{f(x)}\big(df_x(h)+\alpha(h)\big) + \beta\big(df_x(h)+\alpha(h)\big). \tag{10.325}$$

Nous montrons que tous les « petits » termes de cette formule peuvent être groupés. D'abord si h est proche de 0, nous avons

$$\frac{\|df_x(h)+\alpha(h)\|}{\|h\|} \leqslant \frac{\|df_x\|\|h\|}{\|h\|} + \frac{\|\alpha(h)\|}{\|h\|}. \tag{10.326}$$

Si h est petit, le second terme est arbitrairement petit, donc en prenant n'importe que $M > \|df_x\|$ nous avons

$$\frac{\|df_x(h)+\alpha(h)\|}{\|h\|} \leqslant M. \tag{10.327}$$

Par ailleurs, nous avons

$$\frac{\|\beta\big(df_x(h)+\alpha(h)\big)\|}{\|h\|} = \frac{\|\beta\big(df_x(h)+\alpha(h)\big)\|}{\|df_x(h)+\alpha(h)\|}\frac{\|df_x(h)+\alpha(h)\|}{\|h\|} \leqslant M\frac{\|\beta\big(df_x(h)+\alpha(h)\big)\|}{\|df_x(h)+\alpha(h)\|}. \tag{10.328}$$

Vu que la fraction est du type $\frac{\beta(f(h))}{f(h)}$ avec $\lim_{h\to 0} f(h) = 0$, la fraction tend vers zéro lorsque $h \to 0$. En posant

$$\gamma_1(h) = \beta\big(df_x(h)+\alpha(h)\big) \tag{10.329}$$

nous avons $\lim_{h\to 0}\gamma_1(h)/\|h\| = 0$.

L'autre candidat à être un petit terme dans (10.325) est traité en utilisant le lemme 10.13 :

$$\|dg_{f(x)}\big(\alpha(h)\big)\| \leqslant \|dg_{f(x)}\|\|\alpha(h)\|. \tag{10.330}$$

Donc

$$\frac{\|dg_{f(x)}\big(\alpha(h)\big)\|}{\|h\|} \leqslant \|dg_{f(x)}\|\frac{\|\alpha(h)\|}{\|h\|}, \tag{10.331}$$

ce qui nous permet de poser

$$\gamma_2(h) = dg_{f(x)}\big(\alpha(h)\big) \tag{10.332}$$

avec γ_2 qui a la même propriété que γ_1. Avec tout cela, en posant $\gamma = \gamma_1 + \gamma_2$ nous récrivons

$$(g\circ f)(x+h) = g\big(f(x)\big) + dg_{f(x)}\big(df_x(h)\big) + \gamma(h) \tag{10.333}$$

avec $\lim_{h\to 0}\frac{\gamma(h)}{\|h\|} = 0$. Tout cela pour dire que

$$\lim_{h\to 0}\frac{(g\circ f)(x+h) - (g\circ f)(x) - \big(dg_{f(x)}\circ df_x\big)(h)}{\|h\|} = 0, \tag{10.334}$$

ce qui signifie que

$$d(g\circ f)_x = dg_{f(x)}\circ df_x. \tag{10.335}$$

Nous avons donc montré que si f et g sont différentiables, alors $g\circ f$ est différentiable avec différentielle donnée par (10.321).

Nous passons à la régularité. Nous supposons maintenant que f et g sont de classe C^r et nous considérons l'application

$$\begin{aligned}\varphi\colon L(F,G)\times L(E,F)&\to L(E,G)\\(A,B)&\mapsto A\circ B.\end{aligned}\tag{10.336}$$

Montrons que l'application φ est continue en montrant qu'elle est bornée [44]. Pour cela nous écrivons la norme opérateur

$$\|\varphi\|=\sup_{\|(A,B)\|=1}\|\varphi(A,B)\|=\sup_{\|(A,B)\|=1}\|A\circ B\|\leqslant\sup_{\|(A,B)\|=1}\|A\|\|B\|\leqslant 1.\tag{10.337}$$

Justifications : d'une part la norme opérateur est une norme algébrique [45], et d'autre part la définition 7.230 de la norme sur un espace produit pour la dernière majoration. L'application φ est donc continue et donc C^∞ par le lemme 10.132. Nous considérons également l'application

$$\begin{aligned}\psi\colon U&\to L(F,G)\times L(E,F)\\x&\mapsto\big(dg_{f(x)},df_x\big).\end{aligned}\tag{10.338}$$

Vu que f et g sont C^1, l'application ψ est continue. Ces deux applications φ et ψ sont choisies pour avoir

$$(\varphi\circ\psi)(x)=\varphi\big(dg_{f(x)},df_x\big)=dg_{f(x)}\circ df_x,\tag{10.339}$$

c'est-à-dire $\varphi\circ\psi=d(g\circ f)$. Les applications φ et ψ étant continues, l'application $d(g\circ f)$ est continue, ce qui prouve que $g\circ f$ est C^1.

Si f et g sont C^r alors $dg\in C^{r-1}$ et $dg\circ f\in C^{r-1}$ où il ne faut pas se tromper : $dg\colon F\to L(F,G)$ et $f\colon U\to F$; la composée est $dg\circ f\colon x\mapsto dg_{f(x)}\in L(F,G)$.

Pour la récurrence nous supposons que $f,g\in C^{r-1}$ implique $g\circ f\in C^{r-1}$ pour un certain $r\geqslant 2$ (parce que nous venons de prouver cela avec $r=1$ et $r=2$). Soient $f,g\in C^r$ et montrons que $g\circ f\in C^r$. Par la proposition 10.133 nous avons

$$\psi=dg\circ f\times df\in C^{r-1},\tag{10.340}$$

et donc $d(g\circ f)=\varphi\circ\psi\in C^{r-1}$, ce qui signifie que $g\circ f\in C^r$. □

Proposition 10.136 ([1]).
Soit une application $f\colon E\to V$ de classe C^1. Soit une application linéaire $\varphi\colon V\to W$. Alors $\varphi\circ f$ est de classe C^p.

Démonstration. Toute la preuve est un grand jeu de cohérence des espaces en présence, alors soyez attentifs et capable de dire précisément à quel espace appartient chacun de objets entrant en jeu.

Nous posons $V_0=V$ et $V_{k+1}=\mathcal{L}(E,V_k)$. Idem pour les espaces W_k. Ensuite nous posons

$$\begin{aligned}\varphi_1\colon\mathcal{L}(E,V)&\to\mathcal{L}(E,W)\\\alpha&\mapsto\varphi\circ\alpha.\end{aligned}\tag{10.341}$$

et

$$\begin{aligned}\varphi_k\colon\mathcal{L}(E,V_{k-1})&\to\mathcal{L}(E,W_{k-1})\\\alpha&\mapsto\varphi_{k-1}\circ\alpha.\end{aligned}\tag{10.342}$$

Notez la cohérence : si $a\in E$, $\alpha(a)\in V_{k-1}=\mathcal{L}(E,V_{k-2})$, et donc

$$(\varphi_{k-1}\circ\alpha)(a)=\varphi_{k-1}\big(\alpha(a)\big).\tag{10.343}$$

À droite nous avons $\varphi_{k-1}\big(\alpha(a)\big)\in\mathcal{L}(E,W_{k-2})=V_{k-1}$.

De plus, φ est linéaire ; ça se prouve par récurrence en partant de φ_1 et en se basant sur le fait que φ est linéaire.

C'est parti pour une récurrence.

44. Proposition 10.18.
45. Lemme 10.13.

Énoncé Nous allons prouver par récurrence que

$$d^k(\varphi \circ f) = \varphi_k \circ d^k f. \tag{10.344}$$

pour tout $k \leqslant p$.

Initialisation D'abord, f est de classe C^p, donc différentiable et φ est linéaire donc différentiable. Donc la composée est différentiable et le théorème 10.134 nous donne la différentiabilité de $\varphi \circ f$ ainsi que la formule

$$d(\varphi \circ f)_a(u) = d\varphi_{f(a)}\big(df_a(u)\big) = (\varphi \circ df_a)(u) = \varphi_1(df_a)(u). \tag{10.345}$$

Donc $d(\varphi \circ f)_a = \varphi_1(df_a)$, ce qui signifie

$$d(\varphi \circ f) = \varphi_1 \circ df. \tag{10.346}$$

C'est bon pour $k = 1$.

La pas de récurrence Vu que f est de classe C^p, $d^k f$ est encore différentiable. Vu que φ_k est encore linéaire, nous pouvons encore utiliser la règle de différentiation de fonctions composées sur l'application $\varphi_k \circ d^k f$. Nous avons :

$$d^{k+1}(\varphi \circ f)_a(u) = d\big(d^k(\varphi \circ f)\big)_a(u) = d(\varphi_k \circ d^k f)_a(u). \tag{10.347}$$

C'est le moment d'utiliser la formule de différentiation en chaine :

$$d^{k+1}(\varphi \circ f)_a(u) = \big((d\varphi_k)_{d^k f_a} \circ d^{k+1} f_a\big)(u). \tag{10.348}$$

Mais φ_k étant linéaire, $(d\varphi_k)_{d^k f_a} = \varphi_k$, donc

$$d^{k+1}(\varphi \circ f)_a(u) = (\varphi_k \circ d^{k+1} f_a)(u). \tag{10.349}$$

Donc, en oubliant l'application au vecteur u,

$$d^{k+1}(\varphi \circ f)_a = \varphi_k \circ d^{k+1} f_a = \varphi_{k+1}\big(d^{k+1} f_a\big) = (\varphi_{k+1} \circ d^{k+1} f)(a). \tag{10.350}$$

Nous avons donc bien

$$d^{k+1}(\varphi \circ f) = \varphi_{k+1} \circ d^{k+1} f. \tag{10.351}$$

□

Lemme 10.137.
Si $f\colon U \to V$ est un difféomorphisme[46] *alors pour tout $a \in U$, l'application df_a est inversible et*

$$(df_a)^{-1} = (df^{-1})_{f(a)}. \tag{10.352}$$

Démonstration. Il suffit d'apercevoir qu'en vertu de la règle de différentiation en chaine (10.321),

$$(df_a)(df^{-1})_{f(a)} = d(f \circ f^{-1})_{f(a)} = \mathrm{Id}\,. \tag{10.353}$$

□

Proposition 10.138.
Soient des ouverts A de $\mathbb{R}^p$ et B de $\mathbb{R}^m$. Si il existe un difféomorphisme $f\colon A \to B$, alors $p = m$.

Démonstration. Vu que f est un difféomorphisme, le lemme 10.137 fait son travail : l'application linéaire $df_a\colon \mathbb{R}^p \to \mathbb{R}^m$ est inversible d'inverse $df^{-1}_{f(a)}\colon \mathbb{R}^m \to \mathbb{R}^m$.

Or une application linéaire ne peut pas être bijective entre espaces de dimensions différentes (finies). Donc $p = m$. □

46. Définition 10.1

10.7.5 Différentiation de produit

Si nous avons deux application $f\colon E \to V$ et $g\colon E \to W$, alors nous voudrions considérer la fonction

$$\begin{aligned} f \otimes g\colon E &\to V \otimes W \\ a &\mapsto f(a) \otimes g(a). \end{aligned} \tag{10.354}$$

Le problème avec cette notation est que très souvent, les applications f et g sont des éléments d'espaces vectoriels. Si par exemple $f \in \mathcal{L}(E, V)$ et $g \in \mathcal{L}(E, W)$, nous avons $f \otimes g \in \mathcal{L}(E, V) \otimes \mathcal{L}(E, W)$. Dans le Frido nous ne nous permettons pas de dire calmement que $\mathcal{L}(E, V) \otimes \mathcal{L}(E, W) = \mathcal{L}(E, V \otimes W)$. Et je ne vous dit même pas à quel point il n'est pas évident, si $f \in C^\infty(E, V)$ et $g \in C^\infty(E, W)$ que nous aurions $f \otimes g \in C^\infty(E, V) \otimes C^\infty(E, W) = C^\infty(E, V \otimes W)$.

Tout cela pour dire que nous n'allons pas nous lancer dans des abus de notations. Non. Au lieu de cela, nous introduisons une notation. Pour rappel, dans tout le Frido, $\mathrm{Fun}(A, B)$ désigne l'ensemble de toutes les application de A vers B sans suppositions de régularité. Pour les puristes, nous précisions que si $f \in \mathrm{Fun}(A, B)$, nous supposons que f est définie sur tout A. hum ... sauf mention du contraire.

Définition 10.139.
Si $f \in \mathrm{Fun}(E, V)$ et $g \in \mathrm{Fun}(E, W)$, alors nous définissons

$$\begin{aligned} f \tilde{\otimes} g\colon E &\to V \otimes W \\ a &\mapsto f(a) \otimes g(a). \end{aligned} \tag{10.355}$$

Proposition 10.140.
Soient des applications continues $f\colon E \to V$ et $g\colon E \to W$ entre espaces vectoriels de dimension finies. Alors la fonction $f \tilde{\otimes} g\colon E \to V \otimes W$ est continue.

Démonstration. Soient $a \in E$ ainsi qu'une suite $x_k \to a$ dans E. Nous voulons prouver que $f \tilde{\otimes} g(x_k) \xrightarrow{V \otimes W} f(a) \otimes g(a)$. Nous avons :

$$\|f(x_k) \otimes g(x_k) - f(a) \otimes g(a)\| \leqslant \|f(x_k) \otimes g(x_k) - f(x_k) \otimes g(a)\| + \|f(x_k) \otimes g(a) - f(a) \otimes g(a)\|. \tag{10.356}$$

Ensuite en utilisant la classe d'équivalence (10.242b),

$$f(x_k) \otimes g(x_k) - f(x_k) \otimes g(a) = f(x_k) \otimes \big(g(x_k) - g(a)\big), \tag{10.357}$$

et en ce qui concerne les normes,

$$\|f(x_k) \otimes g(x_k) - f(x_k) \otimes g(a)\| = \|f(x_k)\| \| \otimes \big(g(x_k) - g(a)\big)\|. \tag{10.358}$$

Mais par hypothèse, $f(x_k) \to f(a)$ et $g(x_k) \to g(a)$. Donc le tout tend vers zéro lorsque $k \to \infty$.

Le même raisonnement fonctionne avec le second terme de (10.356). □

Lorsque nous parlons de différentielle de produit de fonctions, nous voulons étudier la différentiabilité de $f \tilde{\otimes} g$ sous l'hypothèse de différentiabilité de f et g. Et aussi, si f et g sont de classe C^p, est-ce que $f \tilde{\otimes} g$ est également de classe C^p ?

Nous voudrions avoir une formule du type

$$d(f \tilde{\otimes} g) = df \tilde{\otimes} g + f \tilde{\otimes} dg, \tag{10.359}$$

mais ça ne colle pas au niveau des espaces. En effet, en évaluant cela en $a \in E$, nous avons à gauche $d(f \tilde{\otimes} g)_a \in \mathcal{L}(E, V \otimes W)$, tandis qu'à droite nous avons $df_a \otimes g(a) \in \mathcal{L}(E, V) \otimes W$ et $f(a) \otimes dg_a \in V \otimes \mathcal{L}(E, W)$.

Nous pourrions bien entendu dire que $V \otimes \mathcal{L}(E, W)$ est isomorphe à $\mathcal{L}(E, V \otimes W)$ et hop voila, on n'en parle plus. Ce serait passer sur deux points importants. D'abord est-ce que $V \otimes \mathcal{L}(E, W)$ est vraiment isomorphe à $\mathcal{L}(E, V \otimes W)$? Et ensuite, l'isomorphisme implique une utilisation du théorème 10.134 qui est tout sauf une trivialité.

Bref, fidèle au principe fridesque de ne pas cacher des difficultés techniques sous des abus de notations, nous allons écrire les choses explicitement.

Lemme 10.141.
Si E, V et W sont de dimension finie, les applications

$$\begin{aligned}\psi\colon \mathcal{L}(E,V)\otimes W &\to \mathcal{L}(E,V\otimes W)\\ f\otimes w &\mapsto \Big(u\mapsto f(u)\otimes w\Big)\end{aligned}\tag{10.360}$$

et

$$\begin{aligned}\varphi\colon V\otimes\mathcal{L}(E,W) &\to \mathcal{L}(E,V\otimes W)\\ v\otimes g &\mapsto \big(a\mapsto v\otimes g(a)\big).\end{aligned}\tag{10.361}$$

sont des isomorphismes d'espaces vectoriels.

Dans le meilleur des mondes, ces applications devraient être affublés d'indices V et W.

Démonstration. Nous donnons des détails à propos de ψ. Pour φ c'est la même chose.

Linéaire La formule (10.360) définit ψ en particulier sur une base de $\mathcal{L}(E,V)\otimes W$ par la proposition 10.121(1). Ce que signifie réellement la formule (10.360) est que ψ est ainsi définie sur la base et est prolongée par continuité.

Injective Si pour un f et un w fixé nous avons $\psi(f\otimes w)=0$, alors il y a deux cas : soit $w=0$ soit $w\neq 0$. Dans le premier cas, $f\otimes w=0$, et dans le second cas, nous remarquons que

$$0=\psi(f\otimes w)(a)=f(a)\otimes w\tag{10.362}$$

pour tout $a\in E$. Cela implique $f(a)=0$ pour tout a et donc $f=0$, ce qui signifie que $f\otimes w=0$.

Bijective En utilisant la proposition 10.121 et le lemme 4.34(2), nous avons égalité des dimensions entre $\mathcal{L}(E,V)\otimes W$ et $\mathcal{L}(E,V\otimes W)$.
Une application linéaire injective entre deux espaces vectoriels de même dimension (finie) est une bijection.

□

Proposition 10.142.
Soient des espaces vectoriels normés de dimension finie. Soient $f\colon E\to V$ et $g\colon E\to W$ des fonctions de classe C^1. Alors $f\tilde{\otimes}g\colon E\to V\otimes W$ est de classe C^1 nous avons les formules

$$d(f\tilde{\otimes}g)_a(u)=df_a(u)\otimes g(a)+f(a)\otimes dg_a(u)\tag{10.363}$$

ainsi que

$$d(f\tilde{\otimes}g)=\psi\circ(df\tilde{\otimes}g)+\varphi\circ(f\tilde{\otimes}dg).\tag{10.364}$$

Démonstration. Nous commençons par prouver que $f\tilde{\otimes}g$ est différentiable en injectant le candidat (10.363) dans la définition. Au numérateur nous avons :

$$(f\tilde{\otimes}g)(a+h)-(f\tilde{\otimes}g)(a)-df_a(h)\otimes g(a)-f(a)\otimes dg_a(h).\tag{10.365}$$

Le lemme 10.130 assure qu'il existe une fonction $\alpha\colon E\to V$ telle que $\lim_{h\to 0}\alpha(h)/\|h\|$ et $f(a+h)+f(a)+df_a(h)+\alpha(h)$. Même chose pour g. Nous avons donc

$$(f\tilde{\otimes}g)(a+h)=f(a+h)\otimes g(a+h)=\big(f(a)+df_a(h)+\alpha(h)\big)\otimes\big(g(a)+dg_a(h)+\beta(h)\big)\tag{10.366}$$

qui se développe en 9 termes. En effectuant les différences dans (10.365), nous nous retrouvons avec un numérateur qui vaut

$$f(a)\otimes\beta(h)+df_a(h)\otimes dg_a(h)+df_a(h)\otimes\beta(h)+\alpha(h)\otimes g(a)+\alpha(h)\otimes dg_a(h)+\alpha(h)\otimes\beta(h).\tag{10.367}$$

Nous pouvons prouver terme à terme qu'en divisant par $\|h\|$ nous avons une limite qui vaut zéro. Par exemple,

$$\lim_{h\to 0}\frac{f(a)\otimes\beta(h)}{\|h\|}\tag{10.368}$$

se calcule en prenant la norme du numérateur et en utilisant le lemme 10.124 :

$$\frac{\|f(a)\otimes\beta(h)\|}{\|h\|}=\frac{\|f(a)\|\|\beta(h)\|}{\|h\|}\to 0. \tag{10.369}$$

Tous les termes contenant $\alpha(h)$ ou $\beta(h)$ se traitent de la même manière. Le dernier terme à traiter est

$$\lim_{h\to 0}\frac{df_a(h)\otimes dg_a(h)}{\|h\|}. \tag{10.370}$$

En prenant la norme du numérateur, en utilisant encore le lemme 10.124 et en utilisant le lemme 10.10, nous avons

$$\|df_a(h)\otimes dg_a(h)\|=\|df_a(h)\|\|dg_a(h)\|\leqslant\|df_a\|\|dg_a\|\|h\|^2, \tag{10.371}$$

donc

$$\lim_{h\to 0}\frac{df_a(h)\otimes dg_a(h)}{\|h\|}=0. \tag{10.372}$$

Notons que l'utilisation du lemme 10.10 requière que df_a soit continue, ce qui n'est pas évident en dimension infinie : une application linéaire n'est pas spécialement continue. C'est donc ici que nous utilisons le fait que E, V et W sont de dimension finie [47].

Ceci prouve que $f\tilde{\otimes}g$ est différentiable et nous donne la formule (10.363) pour appliquer sa différentielle à un élément de E. La formule (10.364) est un corolaire : elle se vérifie en l'appliquant à a puis à u.

Pour terminer nous devons prouver que $d(f\tilde{\otimes}g)$ est continue. Vu que f et g sont de classe C^1, les applications f, g, df et dg sont continues. Les applications ψ et φ sont également continues parce que linéaires sur des espaces de dimensioy finie. La proposition 10.140 appliquée à df et g montre que $df\tilde{\otimes}g$ est continue. La composition avec ψ qui est linéaire conserve la continuité.

Dons le membre de droite de (10.364) est continu et $f\tilde{\otimes}g$ est a une différentielle continue. Elle est donc de classe C^1. □

Il est temps de démontrer le truc difficile, à savoir que si f et g sont de classe C^p, alors $f\tilde{\otimes}g$ est également de classe C^p.

Proposition 10.143.
Nous applellons P_k la propriété suivante :

> *Pour tout $p\geqslant k$, pour tout espaces vectoriels normés E, V, W de dimension finies et pour toutes applications $f\colon E\to V$ et $g\colon E\to W$ de classe C^k, la fonction $f\tilde{\otimes}g$ est de classe C^k.*

(1) La propriété P_k est vraie pour tout k.

(2) Si $f\colon E\to V$ et $g\colon E\to W$ sont de classe C^p, alors $f\tilde{\otimes}g\colon E\to V\otimes W$ est de classe C^p.

Démonstration. Le gros de la preuve est le point (1). Le point (2) est alors une utilisation de la propriété P_p avec $p=k$.

Pour $k=0$. Si f et g sont de classe C^p avec $p\geqslant k$, alors f et g sont a fortiori continues. La proposition 10.140 montre alors que $f\tilde{\otimes}g$ est continue.

Bien que ce ne soit pas tout à fait nécessaire, nous prouvons que P_1 est également vraie avant de passer à la récurrence. Si f et g sont de classe C^p avec $p\geqslant 1$, alors elles sont de classe C^1 et la proposition 10.142 s'applique : $f\tilde{\otimes}g$ est de classe C^1.

Nous faisons la récurrence en supposant que P_k est vraie, et en prouvant que P_{k+1} est vraie. Soit $p\geqslant k+1$ ainsi que des applications $f\colon E\to V$ et $g\colon E\to W$ de classe C^{k+1}. La proposition 10.142 dit que $f\tilde{\otimes}g$ est de classe C^1 et que

$$d(f\tilde{\otimes}g)=\psi\circ(df\tilde{\otimes}g)+\varphi\circ(f\tilde{\otimes}dg). \tag{10.373}$$

47. Il y a surement moyen de paufiner, et d'affaiblir cette hypothèse, mais je ne me lance pas là-dedans.

À droite, df et g sont de classe C^k parce que f et g sont de classe C^{k+1}. Donc $df\tilde{\otimes}g$ est de classe C^k par l'hypothèse de récurrence appliquée aux espaces $\mathcal{L}(E,V)$ et W. La proposition 10.136 nous assure alors que $\psi\circ(df\tilde{\otimes}g)$ est de classe C^k également.

Nous avons prouvé que $d(f\tilde{\otimes}g)$ est de classe C^k, donc $f\tilde{\otimes}g$ est de classe C^{k+1}. Cela nous fait la récurrence. □

10.7.6 Formule des accroissements finis

Proposition 10.144.
Soient $a<b$ dans $\mathbb{R}$ et deux fonctions

$$f\colon [a,b]\to E \tag{10.374a}$$
$$g\colon [a,b]\to \mathbb{R} \tag{10.374b}$$

continues sur $[a,b]$ et dérivables sur $]a,b[$. Si pour tout $t\in]a,b[$ nous avons $\|f'(t)\|\leqslant g'(t)$ alors

$$\|f(b)-f(a)\|\leqslant g(b)-g(a). \tag{10.375}$$

Démonstration. Soit $\epsilon>0$ et la fonction

$$\begin{aligned}\varphi_\epsilon\colon [a,b]&\to\mathbb{R}\\ t&\mapsto \|f(t)-f(a)\|-g(t)-\epsilon t.\end{aligned} \tag{10.376}$$

Cela est une fonction continue réelle à variable réelle. En particulier pour tout $u\in]a,b[$ la fonction φ_ϵ est continue sur le compact $[u,b]$ et donc y atteint son minimum en un certain point $c\in[u,b]$; c'est le bon vieux théorème de Weierstrass 7.105. Nous commençons par montrer que pour tout u, ledit minimum ne peut être que b. Pour cela nous allons montrer que si $t\in[u,b[$, alors $\varphi_\epsilon(s)<\varphi_\epsilon(t)$ pour un certain $s>t$. Par continuité si s est proche de t nous avons

$$\left\|\frac{f(s)-f(t)}{s-t}\right\|-\frac{\epsilon}{2}<\|f'(t)\|<g'(t)+\frac{\epsilon}{2}=\frac{g(s)-g(t)}{s-t}+\frac{\epsilon}{2}. \tag{10.377}$$

Ces inégalités proviennent de la limite

$$\lim_{s\to t}\frac{f(s)-f(t)}{s-t}=f'(t), \tag{10.378}$$

donc si s et t sont proches,

$$\left\|\frac{f(s)-f(t)}{s-t}-f'(t)\right\| \tag{10.379}$$

est petit. Si $s>t$ nous pouvons oublier des valeurs absolues et transformer l'inégalité en

$$\|f(s)-f(t)\|<g(s)-g(t)+\epsilon(s-t). \tag{10.380}$$

Utilisant cela et l'inégalité triangulaire,

$$\varphi_\epsilon(s)\leqslant\|f(s)-f(t)\|+\|f(t)-f(a)\|-g(s)-\epsilon s \tag{10.381a}$$
$$\leqslant g(s)-g(t)+\epsilon s-\epsilon t+\|f(t)-f(a)\|-g(s)-\epsilon s \tag{10.381b}$$
$$=\varphi_\epsilon(t). \tag{10.381c}$$

Donc nous avons bien $\varphi_\epsilon(s)<\varphi_\epsilon(t)$ avec l'inégalité stricte. Par conséquent pour tout $u\in]a,b[$ nous avons $\varphi_\epsilon(b)<\varphi_\epsilon(u)$ et en prenant la limite $u\to a$ nous avons

$$\varphi_\epsilon(b)\leqslant\varphi_\epsilon(a). \tag{10.382}$$

Cette inégalité donne immédiatement

$$\|f(b)-f(a)\|\leqslant g(b)-g(a)+\epsilon(b-a) \tag{10.383}$$

pour tout $\epsilon>0$ et donc

$$\|f(b)-f(a)\|\leqslant g(b)-g(a). \tag{10.384}$$

□

Théorème 10.145 (Théorème des accroissements finis)**.**
Soient E et F des espaces vectoriels normés, U ouvert dans E et une application différentiable $f\colon U \to F$. Pour tout segment $[a,b] \subset U$ nous avons

$$\|f(b) - f(a)\| \leqslant \left(\sup_{x\in[a,b]} \|df_x\| \right) \|b-a\|. \tag{10.385}$$

Démonstration. Nous prenons les applications

$$\begin{aligned} k\colon [0,1] &\to E \\ t &\mapsto f\big((1-t)a + tb\big) \end{aligned} \tag{10.386}$$

et

$$\begin{aligned} g\colon [0,1] &\to \mathbb{R} \\ t &\mapsto t \sup_{x\in[a,b]} \|df_x\| \|b-a\|. \end{aligned} \tag{10.387}$$

Pour tout t nous avons $g'(t) = M\|b-a\|$ où il n'est besoin de dire ce qu'est M. D'un autre côté nous avons aussi

$$\begin{aligned} k'(t) &= \lim_{\epsilon\to 0} \frac{f\big((1-t-\epsilon)a + (t+\epsilon)b\big) - f\big((1-t)a+tb\big)}{\epsilon} \\ &= \frac{d}{d\epsilon}\Big[f\big((1-t)a + tb + \epsilon(b-a)\big)\Big]_{\epsilon=0} \\ &= df_{(1-t)a+tb}(b-a) \end{aligned} \tag{10.388}$$

où nous avons utilisé l'hypothèse de différentiabilité de f sur $[a,b]$ et donc en $(1-t)a+tb$. Nous avons donc

$$\|k'(t)\| \leqslant \|b-a\| \|df_{(1-t)a+tb}\| \leqslant M\|b-a\| = g'(t) \tag{10.389}$$

La proposition 10.144 est donc utilisable et

$$\|k(1) - k(0)\| = g(1) - g(0), \tag{10.390}$$

c'est-à-dire

$$\|f(b) - f(a)\| = M\|b-a\| \tag{10.391}$$

comme il se doit. □

Proposition 10.146.
Soient E et F des espaces vectoriels normés, U ouvert dans E et une application $f\colon U \to F$. Soient $a,b \in U$ tels que $[a,b] \subset U$. Nous posons $u = (b-a)/\|b-a\|$ et nous supposons que pour tout $x \in [a,b]$, la dérivée directionnelle

$$\frac{\partial f}{\partial u}(x) = \frac{d}{dt}\Big[f(x+tu)\Big]_{t=0} \tag{10.392}$$

existe. Nous supposons de plus que $\frac{\partial f}{\partial u}(x)$ est continue en $x = a$. Alors

$$\|f(b) - f(a)\| \leqslant \left(\sup_{x\in[a,b]} \|\frac{\partial f}{\partial u}(x)\| \right) \|b-a\|. \tag{10.393}$$

Démonstration. Nous posons évidemment

$$M = \sup_{x\in[a,b]} \|\frac{\partial f}{\partial u}(x)\| \tag{10.394}$$

et nous considérons les fonctions

$$k(t) = f\big((1-t)a + tb\big) \tag{10.395}$$

et

$$g(t) = tM\|b-a\|. \tag{10.396}$$

Pour alléger les notations nous posons $x = (1-t)a+tb$ et nous calculons avec un petit changement de variables dans la limite :

$$k'(t) = \frac{d}{d\epsilon}\Big[f\big(x+\epsilon(b-a)\big)\Big]_{\epsilon=0} = \|b-a\|\frac{d}{d\epsilon}\Big[f\big(x+\frac{\epsilon}{\|b-a\|}(b-a)\big)\Big]_{\epsilon=0} = \|b-a\|\frac{\partial f}{\partial u}(x), \tag{10.397}$$

et donc encore une fois nous avons

$$\|k'(t)\| \leqslant g'(t), \tag{10.398}$$

ce qui donne

$$\|k(1)-k(0)\| = g(1)-g(0), \tag{10.399}$$

c'est-à-dire

$$\|f(b)-f(a)\| \leqslant \sup_{x\in[a,b]} \|\frac{\partial f}{\partial u}(x)\|\|b-a\|. \tag{10.400}$$

□

Théorème 10.147.
Soient E, V deux espaces vectoriels normés, une application $f\colon E \to V$, un point $a \in E$ tel que pour tout $u \in E$, la dérivée

$$\frac{d}{dt}\Big[f(x+tu)\Big]_{t=0} \tag{10.401}$$

existe pour tout $x \in B(a,r)$ et est continue (par rapport à x) en $x = a$. Nous supposons de plus que

$$\frac{\partial f}{\partial u}(a) = 0 \tag{10.402}$$

pour tout $u \in E$. Alors f est différentiable en a et

$$df_a = 0 \tag{10.403}$$

Démonstration. Soit $\epsilon > 0$. Pourvu que $\|h\|$ soit assez petit pour que $a+h \in B(a,r)$, la proposition 10.146 nous donne

$$\|f(a+h)-f(a)\| \leqslant \sup_{x\in[a,a+h]} \|\frac{\partial f}{\partial u}(x)\||h| \tag{10.404}$$

où $u = h/\|h\|$. Par continuité de $\partial_u f(x)$ en $x = a$ et par le fait que cela vaut 0 en $x = a$, il existe un $\delta > 0$ tel que si $\|h\| < \delta$ alors

$$\|\frac{\partial f}{\partial u}(a+h)\| \leqslant \epsilon. \tag{10.405}$$

Pour de tels h nous avons

$$\|f(a+h)-f(a)\| \leqslant \epsilon\|h\|, \tag{10.406}$$

ce qui prouve que l'application linéaire $T(u) = 0$ convient parfaitement pour faire fonctionner la définition 10.127. □

10.7.7 Applications multilinéaires

Nous avons déjà parlé d'applications multilinéaires dans la définition 10.43.

Lemme 10.148 (Leibnitz pour les formes bilinéaires[226])**.**
Si $B\colon E\times F \to G$ est bilinéaire et continue, elle est C^∞ et

$$dB_{(x,y)}(u,v) = B(x,v)+B(u,y). \tag{10.407}$$

Démonstration. D'abord le membre de droite de (10.407) est une application linéaire et continue, donc c'est un bon candidat à être différentielle. Nous allons prouver que ça l'est, ce qui prouvera la différentiabilité de B. Avec ce candidat, le numérateur de la définition (10.127) s'écrit dans notre cas

$$B\big((x,y)+(u,v)\big)-B(x,y)-B(x,v)-B(u,y)=B(u,v). \tag{10.408}$$

Il reste à voir que

$$\lim_{(u,v)\to(0,0)}\frac{B(u,v)}{\|(u,v)\|}=0 \tag{10.409}$$

Par l'équation (10.89) nous avons

$$\frac{\|B(u,v)\|}{\|(u,v)\|}\leqslant\frac{\|B\|\|u\|\|v\|}{\|u\|}=\|B\|\|v\| \tag{10.410}$$

parce que $\|(u,v)\|\geqslant\|u\|$. À partir de là il est maintenant clair que

$$\lim_{(u,v)\to(0,0)}\frac{\|B(u,v)\|}{\|(u,v)\|}=0, \tag{10.411}$$

ce qu'il fallait. □

Proposition 10.149 (Règle de Leibnitz[226])**.**
Soient E,F_1,F_2 des espaces vectoriels normés, U ouvert dans E et des applications de classe C^r ($r\geqslant 1$)

$$f_1\colon U\to F_1 \tag{10.412a}$$
$$f_2\colon U\to F_2 \tag{10.412b}$$
$$\tag{10.412c}$$

et $B\in L(F_1\times F_2,G)$. Alors l'application

$$\begin{aligned}\varphi\colon U&\to G\\ x&\mapsto B\big(f_1(x),f_2(x)\big)\end{aligned} \tag{10.413}$$

est de classe C^r et

$$d\varphi_x(u)=\varphi\big((df_1)_x(u),f_2(x)\big)+\varphi\big(f_1(x),(df_2)_x(u)\big). \tag{10.414}$$

Démonstration. Par hypothèse B est continue (c'est la définition de l'espace L), et donc C^∞ par le lemme 10.148. Par ailleurs la fonction $f_1\times f_2$ est de classe C^r parce que f_1 et f_2 le sont et parce que la proposition 10.133 le dit. L'application composée $B\circ(f_1\times f_2)$ est donc également de classe C^r par le théorème 10.135.

Il ne nous reste donc qu'à prouver la formule 10.414. En utilisant la différentielle du produit cartésien [48] nous avons

$$f\big(B\circ(f_1\times f_2)\big)_x(h)=dB_{(f_1\times f_2)(x)}\big((df_1)_x(h),(df_2)_x(h)\big). \tag{10.415}$$

Nous développons cela en utilisant le lemme 10.148 :

$$d\big(B\circ(f_1\times f_2)\big)_x(h)=dB_{\big(f_1(x),f_2(x)\big)}\big((df_1)_x(h),(df_2)_x(h)\big) \tag{10.416a}$$
$$=B\big(f_1(x),(df_2)_x(h)\big)+B\big((df_1)_x(h),f_2(x)\big), \tag{10.416b}$$

comme souhaité. □

48. Proposition 10.133.

10.7.8 Différentielle partielle

Définition 10.150 (Différentielle partielle)**.**
*Soient E, F et G des espaces vectoriels normés et une fonction $f\colon E\times F\to G$. Nous définissons sa **différentielle partielle** sur l'espace E par*

$$\begin{aligned} d_1f_{(x_0,y_0)}\colon E&\to G\\ u&\mapsto \frac{d}{dt}\Big[f(x_0+tu,y_0\Big]_{t=0}. \end{aligned}\tag{10.417}$$

La différentielle d_2 se définit de la même façon.

Proposition 10.151 ([226])**.**
Soient E_1, E_2 et F des espaces vectoriels normés, soit un ouvert $U\subset E_1\times E_2$ et une fonction $f\colon U\to F$.

(1) Si f est différentiable alors les différentielles partielles existent et

$$d_1f_{(x_0,y_0)}(u)=df_{(x_0,y_0)}(u,0)\tag{10.418a}$$
$$d_2f_{(x_0,y_0)}(v)=df_{(x_0,y_0)}(0,v)\tag{10.418b}$$

où $u\in E_1$ et $v\in E_2$.

(2) Si f est différentiable alors

$$df_{(x_0,y_0)}(u,v)=d_1f_{(x,y_0)}(u)+d_2f_{(x_0,y_0)}(v).\tag{10.419}$$

Démonstration. Nous posons $\alpha=(x_0,y_0)\in U$ et

$$\begin{aligned} j_\alpha^{(1)}\colon E_1&\to E_1\times E_2\\ x&\mapsto (x,y_0). \end{aligned}\tag{10.420}$$

C'est une fonction de classe C^∞ et

$$(dj_\alpha^{(1)})_{x_0}(u)=\frac{d}{dt}\Big[j_\alpha^{(1)}(x_0+tu)\Big]_{t=0}=\frac{d}{dt}\Big[(x_0+tu,y_0)\Big]_{t=0}=(u,0).\tag{10.421}$$

D'autre part

$$(d_1f)_\alpha(u)=\frac{d}{dt}\Big[f(x_0+tu,y_0)\Big]_{t=0}\tag{10.422a}$$
$$=\frac{d}{dt}\Big[(f\circ j_\alpha^{(1)})(x_0+tu)\Big]_{t=0}\tag{10.422b}$$
$$=\big(d(f\circ j_\alpha^{(1)})\big)_{x_0}(u).\tag{10.422c}$$

À ce moment nous utilisons la règle des différentielles composées 10.135 pour dire que

$$(d_1f)_\alpha(u)=df_{j_\alpha^{(1)}(x_0)}\circ(dj_\alpha^{(1)})_{x_0}(u)=df_\alpha(u,0).\tag{10.423}$$

Voila qui prouve déjà le point (1).

Pour la suite nous considérons les fonctions

$$\begin{aligned} P_1(x,y)&=x, & J_1(u)&=(u,0),\\ P_2(x,y)&=y, & J_2(v)&=(0,v) \end{aligned}\tag{10.424}$$

et nous avons l'égalité évidente

$$J_1\circ P_1+J_2\circ P_2=\mathbb{1}\tag{10.425}$$

sur $E_1\times E_2$. En appliquant df_α à cette dernière égalité, en appliquant à (u,v) et en utilisant la linéarité de df_α nous trouvons

$$df_\alpha(u,v)=df_\alpha\big((J_1\circ P_1)(u,v)\big)+df_\alpha\big((J_2\circ P_2)(u,v)\big)\tag{10.426a}$$
$$=df_\alpha(u,0)+df_\alpha(0,v)\tag{10.426b}$$
$$=(d_1f)_\alpha(u)+(d_2f)_\alpha(v)\tag{10.426c}$$

où nous avons utilisé le point (1) pour la dernière égalité. □

10.7.9 L'inverse, sa différentielle

Si E est un espace de Banach, nous sommes intéressés à l'espace $\mathrm{GL}(E)$ des endomorphismes inversibles de E sur E. Cet ensemble est métrique par la formule usuelle

$$\|T\| = \sup_{\|x\|=1} \|T(x)\|_E. \tag{10.427}$$

Proposition 10.152 (Thème 44)**.**
Soit E un espace de Banach (espace vectoriel normé complet). Si A est un endomorphisme de E satisfaisant $\|A\| < 1$ pour la norme opérateur, alors $(\mathbb{1} - A)$ est inversible et son inverse est donné par

$$(\mathbb{1} - A)^{-1} = \sum_{k=0}^{\infty} A^k. \tag{10.428}$$

Démonstration. Étant donné que la norme opérateur est une norme algébrique (lemme 10.13), nous avons $\|A^k\| \leqslant \|A\|^k$. Par conséquent la série $\|A^k\|$ est majorée par la série géométrique qui converge[49]. Par conséquent $\sum_k A^k$ est une série absolument convergente et donc convergente par la proposition 10.54 et le fait que $\mathcal{L}(E)$ est complet (proposition 10.57).

Montrons à présent que la somme est l'inverse de $\mathbb{1} - A$ en utilisant le produit terme à terme autorisé par la proposition 10.34 :

$$\sum_{k=0}^{n} A^k(\mathbb{1} - A) = \sum_{k=0}^{n} (A^k - A^{k+1}) = \mathbb{1} - A^{n+1}. \tag{10.429}$$

Par conséquent

$$\|\mathbb{1} - \sum_{k=0}^{n} A^k(\mathbb{1} - A)\| = \|A^{n+1}\| \leqslant \|A\|^{n+1} \to 0. \tag{10.430}$$

□

Théorème 10.153 (Inverse dans $\mathrm{GL}(E)$[227, 226])**.**
Soient E et F des espaces vectoriels normés.

(1) L'ensemble $\mathrm{GL}(E)$ est ouvert dans $\mathrm{End}(E)$.

(2) L'application inverse

$$\begin{aligned} i\colon \mathrm{GL}(E,F) &\to \mathrm{GL}(F,E) \\ u &\mapsto u^{-1} \end{aligned} \tag{10.431}$$

est de classe C^∞ et

$$di_{u_0}(h) = -u_0^{-1} \circ h \circ u_0^{-1} \tag{10.432}$$

pour tout $h \in \mathrm{End}(E)$

Démonstration. Nous supposons que $\mathrm{GL}(E,F)$ n'est pas vide, sinon ce n'est pas du jeu.

Cas de dimension finie Si la dimension de E et F est finie, elles doivent être égales, sinon il n'y a pas de fonctions inversibles $E \to F$. L'ensemble $\mathrm{GL}(E,F)$ est donc naturellement $\mathrm{GL}(n,\mathbb{R})$. Un élément de $\mathbb{M}(n,\mathbb{R})$ est dans $\mathrm{GL}(n,\mathbb{R})$ si et seulement si son déterminant est non nul. Le déterminant étant une fonction continue (polynomiale) en les entrées de la matrice, l'ensemble $\mathrm{GL}(n,\mathbb{R})$ est ouvert dans $\mathbb{M}(n,\mathbb{R})$.

Même idée pour la régularité de la fonction $i\colon \mathrm{GL}(n,\mathbb{R}) \to \mathrm{GL}(n,\mathbb{R})$, $X \mapsto X^{-1}$. Les entrées de X^{-1} sont les cofacteurs de X divisé par $\det(X)$, et donc des polynômes en les entrées de X divisés par un polynôme qui ne s'annule pas sur $\mathrm{GL}(n,\mathbb{R})$, et donc sur un ouvert autour de X et de X^{-1}. Bref, tout est C^∞.

Le reste de la preuve parle de la dimension infinie.

49. Voir l'exemple 10.67.

Ouvert autour de l'identité Nous commençons par prouver que $B(\mathbb{1},1)\subset \mathrm{GL}(E)$. Pour cela il suffit de remarquer que si $\|u\|<1$ alors le lemme 10.152 nous donne un inverse de $(1+u)$ en la personne de $\sum_{k=0}^{\infty}(-u)^k$.

Ouvert en général Soit maintenant $u_0\in\mathrm{GL}(E)$. Si $\|u\|<\frac{1}{\|u_0^{-1}\|}$ alors $\|u_0^{-1}u\|<1$, ce qui signifie que

$$\mathbb{1}+u_0^{-1}u \tag{10.433}$$

est inversible. Mais $u_0+u=u_0(\mathbb{1}+u_0^{-1}u)$, donc $u_0+u\in\mathrm{GL}(E)$ ce qui signifie que

$$B\left(u_0,\frac{1}{\|u_0^{-1}\|}\right)\subset\mathrm{GL}(E). \tag{10.434}$$

Différentielle en l'identité Nous commençons par prouver que $di_{\mathbb{1}}(u)=-u$. Pour cela nous posons

$$\alpha(h)=\sum_{k=2}^{\infty}(-1)^k h^k \tag{10.435}$$

et nous calculons

$$di_{\mathbb{1}}(u)=\frac{d}{dt}\Big[i(\mathbb{1}+tu)\Big]_{t=0}=\frac{d}{dt}\Big[\mathbb{1}-tu+\alpha(tu)\Big]_{t=0}. \tag{10.436}$$

Il suffit de prouver que $\frac{d}{dt}\Big[\alpha(tu)\Big]_{t=0}=0$ pour conclure que $di_{\mathbb{1}}(u)=-u$. Pour cela, nous remarquons que $\alpha(0)=0$ et donc que

$$\frac{d}{dt}\Big[\alpha(tu)\Big]_{t=0}=\lim_{t\to 0}\frac{\alpha(tu)-\alpha(0)}{t} \tag{10.437a}$$

$$=\lim_{t\to 0}\sum_{k=2}^{\infty}(-1)^k\frac{(tu)^k}{t} \tag{10.437b}$$

$$=-\lim_{t\to 0}u\sum_{k=1}^{\infty}(-1)^k t^k u^k. \tag{10.437c}$$

La norme de ce qui est dans la limite est majorée par

$$\|u\|\sum_{k=1}^{\infty}\|tu\|^k=\|u\|\left(\frac{1}{1-\|tu\|}-1\right), \tag{10.438}$$

et cela tend vers zéro lorsque $t\to\infty$. Nous avons utilisé la somme 10.126 de la série géométrique. Nous avons bien prouvé que $di_{\mathbb{1}}(u)=-u$.

Différentielle en général Soit maintenant $u_0\in\mathrm{GL}(E)$ et $h\in\mathrm{End}(E)$ tel que $u_0+h\in\mathrm{GL}(E)$; par le premier point, il suffit de prendre $\|h\|$ suffisamment petit. Vu que $u_0+h=u_0(\mathbb{1}+u_0^{-1}h)$ nous avons

$$(u_0+h)^{-1}=(\mathbb{1}+u_0^{-1}h)^{-1}u_0^{-1}. \tag{10.439}$$

Nous pouvons donc calculer

$$(u_0+h)^{-1}=\big(\mathbb{1}-u_0^{-1}h+\alpha(u_0^{-1}h)\big)u_0^{-1}=u_0^{-1}-u_0^{-1}hu_0^{-1}+\alpha(u_0^{-1}h)u_0^{-1}, \tag{10.440}$$

et ensuite

$$di_{u_0}(h)=\frac{d}{dt}\Big[i(u_0+th)\Big]_{t=0}=\frac{d}{dt}\Big[u_0^{-1}-tu_0^{-1}hu_0^{-1}+\alpha(tu_0^{-1}h)u_0^{-1}\Big]_{t=0}, \tag{10.441}$$

mais nous avons déjà vu que

$$\frac{d}{dt}\Big[\alpha(th)\Big]_{t=0}=0, \tag{10.442}$$

donc

$$di_{u_0}(h)=-u_0^{-1}hu_0^{-1} \tag{10.443}$$

Cela donne la différentielle de l'application inverse.

Continuité de l'inverse L'application i est continue parce que différentiable.

L'inverse est C^∞ Nous allons écrire la fonction inverse comme une composée. Soient les applications

$$\begin{aligned} B\colon L(F,E)\times L(F,E) &\to L\big(L(E,F),L(F,E)\big)\\ B(\psi_1,\psi_2)(A) &= -\psi_1\circ A\circ\psi_2 \end{aligned} \tag{10.444}$$

et

$$\begin{aligned} \Delta\colon L(F,E) &\to L(F,E)\times L(F,E)\\ \varphi &\mapsto (\varphi,\varphi) \end{aligned} \tag{10.445}$$

Nous avons alors

$$di = B\circ\Delta\circ i. \tag{10.446}$$

L'application Δ est de classe C^∞. Nous devons voir que B l'est aussi. Pour le voir nous commençons par prouver qu'elle est bornée :

$$\begin{aligned} \|B\| &= \sup_{\|\psi_1\|,\|\psi_2\|=1} \|B(\psi_1,\psi_2)\|_{\mathcal{L}\left(L(E,F),L(F,E)\right)}\\ &= \sup_{\|\psi_1\|,\|\psi_2\|=1}\ \sup_{\|A\|=1} \|\psi_1\circ A\circ\psi_2\|_{L(F,E)}\\ &\leqslant \sup_{\|\psi_1\|,\|\psi_2\|=1}\ \sup_{\|A\|=1} \|\psi_1\|\|A\|\|\psi_2\|\\ &\leqslant 1. \end{aligned} \tag{10.447}$$

Donc B est bien bornée et par conséquent continue. Une application bilinéaire continue est C^∞ par le lemme 10.148. La décomposition $di = B\circ\Delta\circ i$ nous donne donc que $i\in C^\infty$ dès que i est continue, ce que nous avions déjà montré.

□

10.8 Exponentielle de matrice

Proposition 10.154.
Soit V un espace vectoriel de dimension finie et $A\in\operatorname{End}(V)$. La série

$$\exp(A) = \mathbb{1} + A + \frac{A^2}{2} + \frac{A^3}{3} + \ldots = \sum_{k=1}^{\infty}\frac{A^k}{k!}. \tag{10.448}$$

*converge normalement dans $\big(\operatorname{End}(V),\|.\|_{op}\big)$. L'**exponentielle** de la matrice A est cette matrice.*

Démonstration. Vu que la norme opérateur est une norme d'algèbre par le lemme 10.13, nous avons pour tout k la majoration $\|A^k\|\leqslant\|A\|^k$. Nous avons donc

$$\sum_{k=0}^{\infty}\frac{\|A^k\|}{k!}\leqslant\sum_k\frac{\|A\|^k}{k!}. \tag{10.449}$$

La dernière somme converge en vertu de la convergence de la série exponentielle donnée en exemple 10.69. □

Étant donné que c'est une limite, il y a une question de convergence et donc de topologie. C'est pour cela que nous ne pouvions pas introduire l'exponentielle de matrice avant d'avoir introduit la norme des matrices. La convergence de la série pour toute matrice sera prouvée au passage dans la proposition 10.155.

La fonction exponentielle $x\mapsto e^x$ n'est pas un polynôme en x, mais nous avons le résultat marrant suivant.

Proposition 10.155.
Si u est un endomorphisme, alors $\exp(u)$ est un polynôme en u[50].

50. Nan, mais j'te jure : exp n'est pas un polynôme, mais $\exp(u)$ est un polynôme de u.

Démonstration. Nous considérons l'application

$$\begin{aligned}\varphi_u\colon \mathbb{K}[X] &\to \mathrm{End}(E)\\ P &\mapsto P(u)\end{aligned} \tag{10.450}$$

Étant donné que l'image de φ_u est un fermé dans $\mathrm{End}(E)$, il suffit de montrer que la série

$$\sum_{k=0}^{\infty} \frac{\varphi_u(X)^k}{k!} \tag{10.451}$$

converge dans $\mathrm{End}(E)$ pour qu'elle converge dans $\mathrm{Image}(\varphi_u)$. Pour ce faire nous nous rappelons de la norme opérateur [51] et de la propriété fondamentale $\|A^k\| \leqslant \|A\|^k$. En notant $A = \varphi_u(X)$,

$$\left\| \sum_{k=n}^{m} \frac{A^k}{k!} \right\| \leqslant \sum_{k=n}^{m} \frac{\|A^k\|}{k!} \leqslant \sum_{k=n}^{m} \frac{\|A\|^k}{k!}, \tag{10.452}$$

ce qui est une morceau du développement de $e^{\|A\|}$. La limite $n \to \infty$ est donc zéro par la convergence de l'exponentielle réelle. La suite des sommes partielles de e^A est donc de Cauchy. La série converge donc parce que nous sommes dans un espace vectoriel réel de dimension finie ($\mathrm{End}(E)$). □

10.156.
Pourquoi $\exp(u)$ est-il un polynôme d'endomorphisme alors que exp n'est pas un polynôme ? Lorsque nous disons que la fonction $x \mapsto \exp(x)$ n'est pas un polynôme, nous sommes en train de localiser la fonction exp à l'intérieur de l'espace de toutes les fonctions $\mathbb{R} \to \mathbb{R}$, c'est-à-dire à l'intérieur d'un espace de dimension infinie. Au contraire lorsqu'on parle de $\exp(u)$ et qu'on le compare aux endomorphismes $P(u)$, nous sommes en train de repérer $\exp(u)$ à l'intérieur de l'espace des matrices qui est de dimension finie. Il n'est donc pas étonnant que l'on parvienne moins à faire la distinction.

Si par contre nous considérons exp en tant qu'application $\exp\colon \mathrm{End}(E) \to \mathrm{End}(E)$, ce n'est pas un polynôme.

Si u et v sont des endomorphismes, nous aurons des polynômes P et Q tels que $e^u = P(u)$ et $e^v = Q(v)$; mais nous n'aurons en général évidemment pas $P = Q$. En cela, exp n'est pas un polynôme.

10.9 Espace dual

Définition 10.157.
*Soit un espace vectoriel normé $(V, \|.\|)$ sur le corps $\mathbb{C}$ ou $\mathbb{R}$ (que nous nommons $\mathbb{K}$). Son **dual topologique**, noté V' est l'ensemble des applications linéaires continues $V \to \mathbb{K}$.*

10.9.1 Topologies

Il est possible de mettre sur V' (au moins) deux topologies distinctes. La première est la topologie de la norme opérateur ; rien de nouveau pour elle. La seconde est la topologie $*$-faible dont nous avons déjà un peu parlé dans la définition 7.333.

En termes de notations, nous allons noter les semi-normes de la topologie faible par

$$p_x(\varphi) = |\varphi(x)| \tag{10.453}$$

pour $x \in V$ et $\varphi \in V'$. À droite, les barres dénotent soit la valeur absolue (si $\mathbb{K} = \mathbb{R}$), soit le module (si $\mathbb{K} = \mathbb{C}$).

51. Définition 10.3.

Lemme 10.158.
Soit $\varphi \in V'$ et $x \in V$. Alors

$$p_x(\varphi) \leqslant \frac{\|\varphi\|}{\|x\|}. \tag{10.454}$$

Si $\varphi_0 \in V'$, si $r > 0$ et si $x \in V$ nous avons aussi :

$$B(\varphi_0, r) \subset B_x(\varphi_0, \frac{r}{\|x\|}). \tag{10.455}$$

Démonstration. En posant $x' = x/\|x\|$ nous avons

$$p_x(\varphi) = |\varphi(x)| = \frac{1}{\|x\|}|\varphi(x')| \leqslant \frac{1}{\|x\|}\|\varphi\|. \tag{10.456}$$

En ce qui concerne la seconde affirmation, si $\varphi \in B(\varphi_0, r)$ alors en notant $x' = x/\|x\|$ nous avons :

$$p_x(\varphi_0 - \varphi) = |\varphi_0(x) - \varphi(x)| = \frac{1}{\|x\|}|\varphi_0(x') - \varphi(x')| \leqslant \frac{1}{\|x\|}\|\varphi_0 - \varphi\| \leqslant \frac{r}{\|x\|}. \tag{10.457}$$

Donc $\varphi \in B_x\big(\varphi_0, \frac{r}{\|x\|}\big)$. □

Proposition 10.159.
En ce qui concerne la convergence d'une suite (φ_k) dans V' mais si elle vérifie

$$\varphi_k \xrightarrow{\|.\|} \varphi \tag{10.458}$$

alors

$$\varphi_k \xrightarrow{*} \varphi. \tag{10.459}$$

Démonstration. Soit une suite (φ_k) dans V', convergente vers φ pour la topologie de la norme. Soit $x \in V$, et $x' = x/\|x\|$. Nous avons

$$p_x(\varphi_k - \varphi) = \frac{1}{\|x\|}|\varphi_k(x') - \varphi(x)| \leqslant \frac{1}{\|x\|}\|\varphi_k - \varphi\| \to 0. \tag{10.460}$$

□

Lemme 10.160.
La translation dans V' est une opération continue pour la topologie de la norme opérateur et pour celle de la topologie $*$.

Démonstration. Soit une suite φ_k tendant vers 0 ; nous devons prouver que $\tau_\sigma(\varphi_k) \to \tau_\sigma(0) = \sigma$. Et ce, pour chacune des deux topologies.

Norme opérateur L'hypothèse $\varphi_k \xrightarrow{\|.\|} 0$ signifie que $\|\varphi_k\| \to 0$, c'est-à-dire que

$$\sup_{\|v\|=1} |\varphi_k(v)| \to 0. \tag{10.461}$$

Nous avons alors

$$\|\tau_\sigma(\varphi_k) - \sigma\| = \sup_{\|v\|=1} |\tau_\sigma(\varphi_k)v - \sigma(v)| = \sup_{\|v\|=1} |\varphi_k(v)| \to 0. \tag{10.462}$$

Donc d'accord pour $\tau_\sigma(\varphi) \to \sigma$.

Topologie $*$ Nous supposons maintenant que $\varphi_k \xrightarrow{*} 0$. Pour tout $v \in V$ nous avons

$$p_v\big(\tau_\sigma(\varphi_k) - \sigma\big) = \big|\tau_\sigma(\varphi_k)v - \sigma(v)\big| = |\varphi_k(v)| = p_v(\varphi_k). \tag{10.463}$$

Mais par hypothèse, $p_v(\varphi_k) \to 0$.

□

Pour la suite, nous allons préfixer par N les concepts liés à la topologie de V' associée à la norme opérateur et par $*$, les concepts de la topologie $*$.

Proposition 10.161.
Soit un espace vectoriel normé V. Un $*$-ouvert et toujours un N-ouvert.

Démonstration. Soit un $*$-ouvert $\mathcal{O}$ de V'. Il existe donc $x \in V$ et $r > 0$ tels que $B_x(\varphi, r) \subset \mathcal{O}$. Nous avons alors, en utilisant le lemme 10.158,

$$B(\varphi, r\|x\|) \subset B_x(\varphi, r) \subset \mathcal{O}. \tag{10.464}$$

Donc $\mathcal{O}$ est un N-ouvert. □

Corolaire 10.162.
Soit un espace topologique X. Si $f\colon (V', *) \to X$ est continue, alors $f\colon (V', \|.\|) \to X$ est continue.

Démonstration. Soit un ouvert $\mathcal{O}$ de X. Vu que f est $*$-continue, la partie $f^{-1}(\mathcal{O})$ est un $*$-ouvert de V'. Il est onc un N-ouvert de V' par la proposition 10.161. □

10.9.2 Réflexivité

Pour la suite nous notons V'' le dual de $(V', \|.\|)$. Certes en tant qu'ensembles, $(V', *)$ et $(V', \|.\|)$ sont identiques, mais comme ils n'ont pas la même topologie, les duaux ne sont pas les mêmes.

Bref, V'' est l'ensemble des applications linéaires continues $(V', \|.\|) \to \mathbb{C}$. Et lorsque nous disons $\mathbb{C}$ ici, ça peut aussi bien être $\mathbb{R}$ selon le contexte.

De plus nous considérons que V'' la norme opérateur qui dérive de la norme de V', laquelle dérive de la norme vectorielle sur V.

Proposition-définition 10.163.
Soit un espace vectoriel normé V sur $\mathbb{R}$ ou $\mathbb{C}$. Nous considérons l'application

$$\begin{aligned} J\colon V &\to V'' \\ J(x)\varphi &= \varphi(x). \end{aligned} \tag{10.465}$$

(1) L'application J est bien définie : $J(x)$ est continue.
(2) L'application J est continue.
(3) Elle est injective.

*Lorsque J est bijective, l'espace V est dit **réflexif**.*

Démonstration. Point par point.

(1) Nous commençons par montrer que $J(x)\colon (V', \|.\|) \to \mathbb{C}$ est continue pour chaque $x \in V$. Soit une suite $\varphi_k \xrightarrow{\|.\|} 0$. Nous avons :

$$J(x)\varphi_k = \varphi_k(x) \leqslant \|\varphi_k\|\|x\| \to 0 \tag{10.466}$$

où vous aurez noté l'utilisation du lemme 10.10. Cela prouve que $J(x)$ est continue et donc que J est bien à valeurs dans V''.

(2) Soit une suite $x_k \xrightarrow{V} 0$, et étudions $\|J(x_k)\|$ pour la norme dans V''. Nous posons $x'_k = x_k/\|x_k\|$ et nous calculons (encore une fois, nous écrivons « $\mathbb{C}$ », mais ça pourrait être $\mathbb{R}$)

$$\|J(x_k)\| = \sup_{\|\varphi\|=1} |J(x_k)\varphi|_{\mathbb{C}} = \sup_{\|\varphi\|=1} |\varphi(x_k)| = \|x_k\| \sup_{\|\varphi\|=1} |\varphi(x'_k)| \leqslant \|x_k\| \to 0. \tag{10.467}$$

La dernière inégalité pourrait être sans doute une égalité [52], mais nous n'en avons pas besoin ici.

□

52. Écrivez-moi si vous en êtes certain.

10.9.3 Module de continuité

Définition 10.164.
Soient deux espaces topologiques normés X et Y, ainsi qu'une application $f\colon X\to Y$. Le ***module de continuité*** *de f est la fonction*

$$\begin{aligned}\omega_f\colon \mathbb{R}^+ &\to \mathbb{R}^+\cup\{\infty\}\\ h &\mapsto \sup_{\substack{x,y\in X\\ d_X(x,y)<h}} d_Y\big(f(x),f(y)\big).\end{aligned}\tag{10.468}$$

Nous définissons aussi $\omega_f(h)=0$ pour $h\leqslant 0$.

Notons que le module de continuité est une fonction croissante.

Lemme 10.165.
Soit $f\in C^0\big([0,1],\mathbb{C}\big)$ et ω son module de continuité. Si λ et h sont strictement positifs avec $\lambda h\in[0,1]$ alors

$$\phi(\lambda h)\leqslant(\lambda+1)\omega(h).\tag{10.469}$$

Démonstration. La fonction ω est décroissante, et pour $h,k>0$ nous avons $\omega(h+k)\leqslant\omega(h)+\omega(k)$. Par récurrence pour tout $k\in\mathbb{N}$ nous avons

$$\omega(kh)\leqslant k\omega(h).\tag{10.470}$$

En écrivant cela pour $k=\lceil\lambda\rceil$, nous avons

$$\omega(\lambda h)\leqslant\omega(kh)\leqslant k\omega(h)\leqslant(\lambda+1)\omega(h).\tag{10.471}$$

□

Lemme 10.166.
Une fonction est uniformément continue[53] *si et seulement si son module de continuité est continu en zéro*[54].

Démonstration. Nous commençons par supposer que f est uniformément continue. Soit $\epsilon>0$. Par uniforme continuité, il existe $\delta>0$ tel que $d\big(f(x),f(y)\big)\leqslant\epsilon$ dès que $d(x,y)\leqslant\delta$. Si $h\in B(0,\delta)$, alors

$$\omega_f(h)\leqslant\omega_f(\delta)=\sup_{\substack{x,y\in X\\ d(x,y)\leqslant\delta}} d\big(f(x),f(y)\big)\leqslant\epsilon.\tag{10.472}$$

Cela prouve que $\lim_{h\to 0}\omega_f(h)=0$.

Dans l'autre sens, si $\epsilon>0$ est fixé, il suffit de prendre δ tel que $\omega_f(h)\leqslant\epsilon$ pour tout $h\leqslant\delta$ pour faire fonctionner la définition de l'uniforme continuité. □

Lemme 10.167 ([228])**.**
Soient des espaces métriques E et E' et une suite de fonctions $(f_i)_{i\geqslant 0}$ qui converge uniformément vers f. Alors pour chaque $\delta>0$ nous avons

$$\limsup_{i\to\infty}\omega_{f_i}(\delta)\leqslant\omega_f(\delta).\tag{10.473}$$

Démonstration. Soient $\delta>0$ ainsi que $x,y\in E$ tels que $\|x-y\|\leqslant\delta$. Pour chaque i nous avons

$$|f_i(x)-f_i(y)|\leqslant|f_i(x)-f(x)|+|f(x)-f(y)|+|f(y)-f_i(y)|\tag{10.474a}$$
$$\leqslant|f(x)-f(y)|+2\|f_i-f\|_\infty\tag{10.474b}$$
$$\leqslant\omega_f(\delta)+2\|f_i-f\|_\infty.\tag{10.474c}$$

53. Définition 7.323.
54. Dans ce lemme, nous avons deux espaces métriques, mais nous allons noter d la distance des deux côtés.

Nous prenons le supremum de cela sur $\{x, y \in E$ tel que $\|x - y\| \leqslant \delta\}$ pour obtenir :

$$\omega_{f_i}(\delta) \leqslant \omega_f(\delta) + 2\|f_i - f\|_\infty. \tag{10.475}$$

La tentation est grande à ce point de prendre la limite des deux côtés pour $i \to \infty$. Cependant, rien ne nous permet de dire que la suite $i \mapsto \omega_{f_i}(\delta)$ ait une limite. Nous pouvons cependant prendre la limite supérieures[55] et obtenir

$$\limsup_{i\to\infty} \omega_{f_i}(\delta) \leqslant \omega_f(\delta). \tag{10.476}$$

□

10.10 Mini introduction aux nombres p-adiques

10.10.1 La flèche d'Achille

C'est un grand classique que je donne ici juste comme introduction pour montrer que des séries infinies peuvent donner des nombres finis de manière tout à fait intuitive.

Achille tire une flèche vers un arbre situé à 10 m de lui. Disons que la flèche avance à une vitesse constante de 1 m/s. Il est clair que la flèche mettra 10 s pour toucher l'arbre. En 5 s, elle aura parcouru la moitié de son chemin. On le note :

$$\text{temps} = 5s + \dots$$

Reste 5 m à faire. En 2.5 s, elle aura fait la moitié de ce chemin chemin, soit $2.5m = \frac{10}{4}m$. On le note :

$$\text{temps} = \frac{10}{2}s + \frac{10}{4}s +$$

Reste $2.5m$ à faire. La moitié de ce trajet, soit $\frac{10}{8}m$, est parcouru en $\frac{10}{8}s$; on le note encore, mais c'est la dernière fois !

$$\text{temps} = \frac{10}{2}s + \frac{10}{4}s + \frac{10}{8}s +$$

En continuant ainsi à regarder la flèche qui parcours des demi-trajets puis des moitiés de demi-trajets et encore des moitiés de moitiés de demi-trajets, et en sachant que le temps total est $10s$, on trouve :

$$10\left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots\right) = 10.$$

On doit donc croire que la somme jusqu'à l'infini des inverses des puissances de deux vaut 1 :

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = 1.$$

Cela peut être démontré à la loyale.

10.10.2 La tortue et Achille

Maintenant qu'on est convaincu que des sommes infinies peuvent représenter des nombres tout à fait normaux, passons à un truc plus marrant.

Achille, qui marche peinard à 10 m/h, part avec $1m$ d'avance sur une tortue qui avance à 1 m/h. Le temps que la tortue arrive au point de départ d'Achille, Achille aura parcouru $10m$, et le temps que la tortue mettra pour arriver à ce point, eh bien, Achille ne sera déjà plus là : il sera à $100m$. Si la tortue tient bon pendant un temps infini, et si l'on est confiant en le genre de raisonnements faits à la section 10.10.1, elle rattrapera Achille dans

$$1m + 10m + 100m + 1000m + \dots$$

55. Définition 7.145.

Autant dire que ça ne risque pas d'arriver. Et pourtant, mettons en équations :

$$\begin{cases} x_{\text{Achile}}(t) = 1 + 10t & (10.477a) \\ x_{\text{tortue}}(t) = t. & (10.477b) \end{cases}$$

La tortue rejoint Achille au temps t tel que $x_{\text{Achille}(t)} = x_{\text{tortue}}(t)$. Un mini calcul donne $t = -1/9$. Physiquement, c'est une situation logique. Peut-on en déduire une égalité mathématique du style de

$$1 + 10 + 100 + 1000 + \ldots = -\frac{1}{9} \text{ ???}$$

Là où les choses deviennent jolies, c'est quand on cherche à voir ce que peut bien être la valeur d'un hypothétique $x = 1 + 10 + 100 + 1000 + \ldots$. En effet, logiquement on devrait avoir

$$\begin{aligned} \frac{x}{10} &= \frac{1}{10} + 1 + 10 + 100 + \ldots \\ &= \frac{1}{10} + x. \end{aligned}$$

Reste à résoudre l'équation du premier degré : $\frac{x}{10} = x + \frac{1}{10}$. Ai-je besoin de donner la solution ?

10.10.3 Dans les nombres p-adiques, c'est vrai

Nous nous proposons d'apprendre sur les nombres p-adiques juste ce qu'il faut pour montrer que l'égalité

$$\sum_{k=0}^{\infty} 10^k = -\frac{1}{9} \tag{10.478}$$

est vraie dans les nombres 5-adiques. Tout ce qu'il faut est sur wikipedia.

Soit $a \in \mathbb{N}$ et p, un nombre premier. La **valuation** p-adique de a est l'exposant de p dans la décomposition de a en nombres premiers. On la note $v_p(a)$. Pour un rationnel on définit

$$v_p\left(\frac{a}{b}\right) = v_p(a) - v_p(b) \tag{10.479}$$

La **valeur absolue** p-adique de $r \in \mathbb{Q}$ est

$$|r|_p = p^{-v_p(r)}. \tag{10.480}$$

Nous posons $|0|_p = 0$. De là nous considérons la distance

$$d_p(x, y) = |x - y|_p. \tag{10.481}$$

Lemme 10.168.
L'espace $(\mathbb{Q}, d_p)$ est un espace métrique[56].

Nous considérons maintenant $p = 5$. Étant donné que $a = 5 \cdot 2$ nous avons $v_5(10) = 1$ et

$$v_5\left(\frac{1}{9}\right) = v_5(1) - v_5(9) = 0. \tag{10.482}$$

Nous avons

$$\sum_{k=0}^{N} 10^k + \frac{1}{9} = \frac{10^{N+1}}{9} \tag{10.483}$$

mais

$$v_p\left(\frac{10^{N+1}}{9}\right) = v_5(10^{N+1}) - v_5(9) = N + 1. \tag{10.484}$$

56. Définition 7.93

Par conséquent

$$d_5\big(\sum_{k=0}^{N} 10^k, -\frac{1}{9}\big) = |\frac{10^{N+1}}{9}|_p = p^{-(N+1)}. \tag{10.485}$$

En passant à la limite,

$$\lim_{N\to\infty} d_5\big(\sum_{k=0}^{N} 10^k, -\frac{1}{9}\big) = 0, \tag{10.486}$$

ce qui signifie que[57]

$$\sum_{k=0}^{\infty} 10^k = -\frac{1}{9}. \tag{10.487}$$

57. Voir la définition 10.48 de la convergence d'une série dans un espace métrique.

Chapitre 11

Analyse réelle

11.1 Intervalles

Définition 11.1 (Intervalle)**.**
Une partie I de $\mathbb{R}$ est un ***intervalle*** *si pour tout $a, b \in I$ nous avons $t \in I$ dès que $a \leqslant t \leqslant b$.*

Un intervalle est ***ouvert*** *s'il est de la forme $]a, b[$ avec éventuellement $a = -\infty$ ou $b = +\infty$. Un intervalle est* ***fermé*** *s'il est de la forme $[a, b]$ ou $]-\infty, b]$ ou $[a, +\infty[$ avec $a, b \in \mathbb{R}$.*

Remarque 11.2.
L'ensemble $\mathbb{R}$ ne contient pas $-\infty$ et $-\infty$. L'intervalle $[-\infty, 5]$ par exemple, n'est pas une partie de $\mathbb{R}$.

Exemple 11.3. (1) Les ensembles $]3, 7[$ et $]-\infty, \pi[$ sont des intervalles ouverts.

(2) Les ensembles $[10, 15]$ et $[-1, +\infty[$ sont des intervalles fermés.

(3) L'ensemble $]-4, -2[\cup]2, 9[$ n'est pas un intervalle (il y a un « trou » entre -2 et 2).

(4) L'ensemble $\mathbb{R}$ lui-même est un intervalle ; par convention, il est à la fois ouvert et fermé.

Un intervalle peut n'être ni ouvert ni fermé ; par exemple $]4, 8]$. Cet intervalle est « ouvert en 4 et fermé en 8 » . △

Définition 11.4 (Fonction, domaine, image, graphe)**.**
Soient X et Y deux ensembles. Une ***fonction*** *f définie sur X et à valeurs dans Y est une correspondence qui associe à chaque élément x dans X* **au plus** *un élément y dans Y. On écrit $y = f(x)$.*

— *La partie de X qui contient tous les x sur lesquels f peut opérer est dite* ***domaine*** *de f. Le domaine de f est indiqué par* $\operatorname{Domaine} f$.

— *L'élément de $y \in Y$ associé par f à un élément $x \in \operatorname{Domaine} f$ (c'est-à-dire $f(x) = y$) est appellé* ***image*** *de x par f. L'****image*** *de la fonction f est la partie de Y qui contient les images de tous les éléments de* $\operatorname{Domaine} f$. *L'image de f est indiquée par $\Im f$.*

— *Le* ***graphe*** *de f est l'ensemble de toutes les couples $(x, f(x))$ pour $x \in \operatorname{Domaine} f$. Le graphe de f est une partie de l'ensemble noté $X \times Y$ et il est indiqué par* $\operatorname{Graph} f$. *Dans ce cours $X = \mathbb{R}$ et $Y = \mathbb{R}$, donc le graphe de f est contenu dans le plan cartésien.*

Définition 11.5 (Fonction croissante, décroissante et monotone)**.**
Soit $f \colon \mathbb{R} \to \mathbb{R}$ une fonction définie sur un intervalle $I \subset \mathbb{R}$.

(1) Le fonction f est ***croissante*** *sur I si pour tout $x < y$ dans I nous avons $f(x) \leqslant f(y)$. Elle est* strictement *croissante si $f(x) < f(y)$ dès que $x < y$.*

(2) Le fonction f est ***décroissante*** *sur I si pour tout $x < y$ dans I nous avons $f(x) \geqslant f(y)$. Elle est* strictement *décroissante si $f(x) > f(y)$ dès que $x < y$.*

(3) La fonction f est dite ***monotone*** *sur I si elle est soit croissante soit décroissante sur I.*

Exemple 11.6.
La fonction $x \mapsto x^2$ est décroissante sur l'intervalle $]-\infty, 0]$ et croissante sur l'intervalle $[0, \infty[$. Elle n'est par contre ni croissante ni décroissante sur l'intervalle $[-4, 3]$. △

11.2 Application réciproque

11.2.1 Définitions

Les définitions d'injection, surjection, bijection et d'application réciproque sont les définitions 7.81 et 7.82.

Exemple 11.7. (1) La fonction $x \mapsto x^2$ n'est pas une bijection de $\mathbb{R}$ vers $\mathbb{R}$ parce qu'il n'existe aucun x tel que $x^2 = -1$.

(2) La fonction

$$\begin{aligned} f\colon [0, +\infty[&\to [0, +\infty[\\ x &\mapsto x^2 \end{aligned} \tag{11.1}$$

est une bijection. Notez que c'est la même fonction que celle de l'exemple précédent. Seul l'intervalle sur laquelle nous nous plaçons a changé.

(3) La fonction

$$\begin{aligned} f\colon \mathbb{R} &\to [0, \infty[\\ x &\mapsto x^2 \end{aligned} \tag{11.2}$$

n'est pas une bijections parce qu'il existe plusieurs x pour lesquels $f(x) = 4$.

(4) Nous verrons un peu plus tard (11.324) que l'application

$$\begin{aligned} f\colon [0, \infty[&\to [0, \infty[\\ x &\mapsto x^2 \end{aligned} \tag{11.3}$$

est une bijection.

En conclusion : il est très important de préciser les domaines des fonctions considérées. △

Remarque 11.8.
Dire que la fonction $f\colon I \to J$ est bijective, c'est dire que l'équation $f(x) = y$ d'inconnue x peut être résolue de façon univoque pour tout $y \in J$.

Remarque 11.9.
Toute fonction strictement monotone sur un intervalle I est injective.

Exemple 11.10.
Trouvons la fonction réciproque de la fonction affine $f\colon \mathbb{R} \to \mathbb{R}$, $x \mapsto 3x - 2$. Si $y \in \mathbb{R}$ le nombre $f^{-1}(y)$ est la valeur de x pour laquelle $f(x) = y$. Il s'agit donc de résoudre

$$3x - 2 = y \tag{11.4}$$

par rapport à x. La solution est $x = \frac{y+2}{3}$ et donc nous écrivons

$$f^{-1}(y) = \frac{y+2}{3}. \tag{11.5}$$

Notons que dans les calculs, il est plus simple d'écrire « y » que « x » la variable de la fonction réciproque. Il est néanmoins (très) recommandé de nommer « x » la variable dans la réponse finale. Dans notre cas nous concluons donc

$$f^{-1}(x) = \frac{x+2}{3}. \tag{11.6}$$

△

11.2.2 Graphe de la fonction réciproque

Par définition le graphe de la fonction f est l'ensemble des points de la forme (x, y) vérifiant $y = f(x)$. Afin de déterminer le graphe de la bijection réciproque nous pouvons faire le raisonnement suivant.

Le point (x_0, y_0) est sur le graphe de f

$\Leftrightarrow$

La relation $f(x_0) = y_0$ est vérifiée

$\Leftrightarrow$

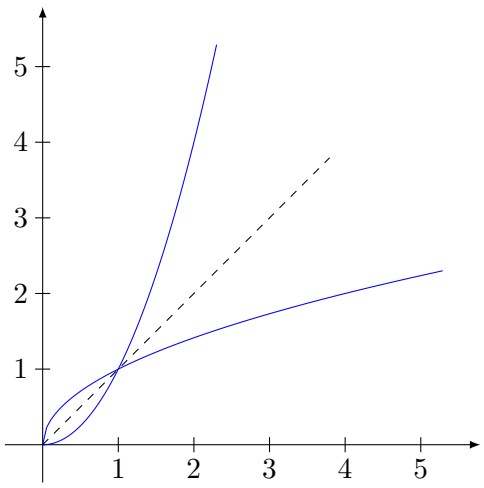
La relation $x_0 = f^{-1}(y_0)$ est vérifiée

$\Leftrightarrow$

Le point (y_0, x_0) est sur le graphe de f^{-1}.

À retenir 11.11
Dans un repère orthonormal, le graphe de la bijection réciproque est obtenu à partir du graphe de f en effectuant une symétrie par rapport à la droite d'équation $y = x$.

Le dessin suivant montre le cas de la courbe de la fonction carré comparé à celle de la racine carrée.



11.3 Limite de fonctions

11.3.1 Définition

La définition générale de la limite est 7.67. Dans le cas de fonctions $\mathbb{R} \to \mathbb{R}$, elle peut s'écrire de façon plus efficace. La proposition suivante montre comment fonctionne la limite pour une fonction définie sur tout $\mathbb{R}$.

Proposition 11.12 (Caractérisation de la limite)**.**
Soit une fonction $f : \mathbb{R} \to \mathbb{R}$ définie sur $\mathbb{R}$ et $a \in \mathbb{R}$. La fonction f admet la limite ℓ pour $x \to a$ si et seulement si il existe un réel ℓ tel que pour tout $\epsilon > 0$, il existe un $\delta > 0$ tel que

$$0 < |x - a| < \delta \Rightarrow |f(x) - \ell| < \varepsilon. \tag{11.7}$$

Démonstration. Il s'agit de montrer l'équivalence avec la définition 7.67. Nous allons faire un usage intensif de la remarque 7.95(3).

<u>**Sens direct**</u> Soient $\epsilon > 0$ et $V = B(\ell, \epsilon)$. Alors il existe un voisinage W de a dans $\mathbb{R}$ tel que

$$f\big(W \backslash \{a\}\big) \subset V. \tag{11.8}$$

Soit δ tel que $B(a,\delta)\subset W$. Nous avons encore

$$f\big(B(a,\delta)\backslash\{a\}\big)\subset V. \tag{11.9}$$

Soit maintenant $x\in\mathbb{R}$ tel que $0<|x-a|<\delta$. Cela signifie $x\in B(a,\delta)\backslash\{a\}$. Pour un tel x nous avons donc $f(x)\in B(\ell,\epsilon)$, c'est-à-dire $|f(x)-\ell|<\epsilon$.

Dans l'autre sens Soient un voisinage V de ℓ et $\epsilon>0$ tel que $B(\ell,\epsilon)\subset V$. Nous considérons δ tel que $0<|x-a|<\delta$ implique $|f(x)-\ell|<\epsilon$.

Avec tout cela nous posons $W=B(x,\delta)$, et nous avons

$$f\big(W\backslash\{a\}\big)\subset B(\ell,\alpha)\subset V. \tag{11.10}$$

□

Si aucun nombre ℓ ne vérifie la condition de la définition, alors on dit que la fonction n'admet pas de limite en a. Lorsque f possède la limite ℓ en a, nous notons

$$\lim_{x\to a} f(x)=\ell. \tag{11.11}$$

La proposition suivante a déjà été démontrée dans la proposition 7.70. Nous en donnons ici une démonstration adaptée au cas $\mathbb{R}\to\mathbb{R}$.

Proposition 11.13.
Soit une fonction $f\colon D\to\mathbb{R}$. Si a est un point d'accumulation de D et s'il existe une limite de f en a, alors il en existe une seule.

Démonstration. Nous prouvons qu'il ne peut pas exister deux nombres $\ell\neq\ell'$ vérifiant tout les deux la condition (11.7).

Soient ℓ et ℓ' deux limites de f au point a. Par définition, pour tout ε nous avons des nombres δ et δ' tels que

$$\begin{aligned}|x-a|<\delta &\Rightarrow \big|f(x)-\ell\big|<\varepsilon\\ |x-a|<\delta' &\Rightarrow \big|f(x)-\ell'\big|<\varepsilon\end{aligned} \tag{11.12}$$

Pour fixer les idées, supposons que $\delta<\delta'$ (le cas $\delta\geqslant\delta'$ se traite de la même manière).

Étant donné que a est un point d'accumulation du domaine D de f, il existe un $x\in D$ tel que $|x-a|<\delta$. Évidemment, nous avons aussi $|x-a|<\delta'$. Les conditions (11.12) signifient alors que ce x vérifie en même temps

$$|f(x)-\ell|<\varepsilon, \tag{11.13}$$

et

$$|f(x)-\ell'|<\varepsilon. \tag{11.14}$$

Afin de prouver que $\ell=\ell'$, nous allons maintenant calculer $|\ell-\ell'|$ et montrer que cette distance est plus petite que tout nombre. Nous avons (voir remarque 11.14)

$$|\ell-\ell'|=|\ell-f(x)+f(x)-\ell'|\leqslant|\ell-f(x)|+|f(x)-\ell'|<\varepsilon+\varepsilon. \tag{11.15}$$

En résumé, pour tout $\varepsilon>0$ nous avons

$$|\ell-\ell'|<2\varepsilon, \tag{11.16}$$

et donc $|\ell-\ell'|=0$, ce qui signifie que $\ell=\ell'$. □

Remarque 11.14.
Les inégalités (11.15) utilisent deux techniques très classiques en analyse qu'il convient d'avoir bien compris. La première est de faire

$$|A-B|=|A-C+C-B|. \tag{11.17}$$

Il s'agit d'ajouter $-C+C$ dans la norme. Évidemment, cela ne change rien.

La seconde technique est l'inégalité

$$|A+B|\leqslant|A|+|B|. \tag{11.18}$$

Exemple 11.15.
Considérons la fonction $f(x) = 2x$, et calculons la limite $\lim_{x\to 3} f(x)$. Vu que $f(3) = 6$, nous nous attendons à avoir $\ell = 6$. C'est ce que nous allons prouver maintenant. Pour chaque $\varepsilon > 0$ nous devons trouver un $\delta > 0$ tel que $|x - 3| < \delta$ implique $|f(x) - 6| < \varepsilon$. En remplaçant $f(x)$ par sa valeur en fonction de x et avec quelques manipulations nous trouvons :

$$\begin{aligned} |f(x) - 6| &< \varepsilon \\ |2x - 6| &< \varepsilon \\ 2|x - 3| &< \varepsilon \\ |x - 3| &< \frac{\varepsilon}{2} \end{aligned} \tag{11.19}$$

Donc dès que $|x - 3| < \frac{\varepsilon}{2}$, nous avons $|f(x) - 6| < \varepsilon$. Nous posons donc $\delta = \frac{\varepsilon}{2}$.

Plus généralement, nous avons $\lim_{x\to a} f(x) = 2a$, et cela se prouve en étudiant $|f(x) - 2a|$ exactement de la même manière. △

11.3.2 Quelques règles de calcul

Proposition 11.16.
La limite est une opération linéaire, c'est-à-dire que si f et g sont des fonctions qui admettent des limites en a et si λ est un nombre réel,

(1) $\lim_{x\to a}(\lambda f)(x) = \lambda \lim_{x\to a} f(x)$,

(2) $\lim_{x\to a}(f + g)(x) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$.

En combinant les deux propriétés de la proposition 11.16, nous pouvons écrire

$$\lim_{x\to a}(\lambda f + \mu g)(x) = \lambda \lim_{x\to a} f(x) + \mu \lim_{x\to a} g(x). \tag{11.20}$$

pour toutes fonctions f et g admettant une limite en a et pour tout réels λ et μ.

En plus d'être linéaire, la limite possède les deux propriétés suivantes.

Proposition 11.17.
Si f et g sont deux fonctions qui admettent une limite en a, alors

$$\lim_{x\to a}(fg)(x) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x). \tag{11.21}$$

Si de plus $\lim_{x\to a} g(x) \neq 0$, *alors*

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}. \tag{11.22}$$

Théorème 11.18.
Si

$$\lim_{x\to a} f(x) = b, \tag{11.23}$$

alors

$$\lim_{x\to a}(\lambda f)(x) = \lambda b \tag{11.24}$$

pour n'importe quel $\lambda \in \mathbb{R}$.

Démonstration. Soit $\epsilon > 0$. Afin de prouver la propriété (11.24), il faut trouver un δ tel que pour tout x dans $[a - \delta, a + \delta]$, on ait $|(\lambda f)(x) - \lambda b| \leqslant \epsilon$. Cette dernière inégalité est équivalente à $|\lambda||f(x) - b| \leqslant \epsilon$. Nous devons donc trouver un δ tel que

$$|f(x) - b| \leqslant \frac{\epsilon}{|\lambda|}. \tag{11.25}$$

soit vraie pour tout x dans $[a - \delta, a + \delta]$. Mais l'hypothèse (11.23) dit précisément qu'il existe un δ tel que pour tout x dans $[a - \delta, a + \delta]$ on ait cette inégalité. □

Théorème 11.19.
Si

$$\lim_{x\to a} f(x) = b_1 \tag{11.26a}$$
$$\lim_{x\to a} g(x) = b_2, \tag{11.26b}$$

alors

$$\lim_{x\to a}(f+g)(x) = b_1 + b_2. \tag{11.27}$$

Démonstration. Soit $\epsilon > 0$. Par hypothèse, il existe δ_1 tel que

$$|f(x) - b_1| \leqslant \frac{\epsilon}{2} \tag{11.28}$$

dès que $|x-a| \leqslant \delta_1$. Il existe aussi δ_2 tel que

$$|g(x) - b_2| \leqslant \frac{\epsilon}{2}. \tag{11.29}$$

dès que $|x-a| \leqslant \delta_2$. Tu notes l'astuce de prendre $\epsilon/2$ dans la définition de limite pour f et g. Maintenant, ce qu'on voudrait c'est un δ tel que l'on ait $|(f+g)(x)-(b_1+b_2)| \leqslant \epsilon$ dès que $|x-a| \leqslant \delta$. Moi je dit que $\delta = \min\{\delta_1, \delta_2\}$ fonctionne. En effet, en utilisant l'inégalité $|a+b| \leqslant |a| + |b|$, nous trouvons :

$$|(f+g)(x) - (b_1+b_2)| = |(f(x)-b_1) + (g(x)-b_2)| \leqslant |f(x)-b_1| + |g(x)-b_2|. \tag{11.30}$$

Comme on suppose que $|x-a| \leqslant \delta$, on a évidemment $|x-a| \leqslant \delta_1$, et donc l'équation (11.28) tient. Mais si $|x-a| \leqslant \delta$, on a aussi $|x-a| \leqslant \delta_2$, et donc l'équation (11.28) tient également. Chacun des deux termes de (11.30) est donc plus petits que $\epsilon/2$, et donc le tout est plus petit que ϵ, ce qu'il fallait montrer.

□

Voici une formule qui résume la linéarité de la limite :

$$\lim_{x\to a}[\alpha f(x) + \beta g(x)] = \alpha \lim_{x\to a} f(x) + \beta \lim_{x\to a} g(x). \tag{11.31}$$

Proposition 11.20 ([147]).
Soient des fonctions $f, g\colon \mathbb{R} \to \mathbb{R}$ telles que $\lim_{x\to a} f(x) = \ell$ et $\lim_{x\to a} g(x) = \ell' \neq 0$. Alors

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\ell}{\ell'}. \tag{11.32}$$

Démonstration. Nous avons :

$$\left|\frac{f(x)}{g(x)} - \frac{\ell}{\ell'}\right| = \frac{|\ell' f(x) - g(x)\ell|}{|g(x)\ell|}. \tag{11.33}$$

Soit s, un minorant de $|g(x)|$ sur un voisinage de a ; vu que la limite en a est $\ell' \neq 0$, nous pouvons prendre par exemple $s = \ell'/2$: $|g(x)| > \ell'/2$ sur $B(a,\delta)$ dès que δ est assez petit. Nous considérons $x \in B(a,\delta)$. Avec cela nous avons :

$$\left|\frac{f(x)}{g(x)} - \frac{\ell}{\ell'}\right| = \frac{|\ell' f(x) - g(x)\ell|}{|g(x)\ell|} \tag{11.34a}$$
$$\leqslant \frac{2}{|\ell'|}\left(\frac{|\ell' f(x) - g(x)\ell|}{|\ell'|}\right) \tag{11.34b}$$
$$\leqslant \frac{2}{|\ell'|^2}\big(|\ell' f(x) - \ell\ell'| + |\ell\ell' - g(x)\ell|\big) \tag{11.34c}$$
$$= \frac{2}{|\ell'|^2}\big(|\ell'||f(x)-\ell| + |\ell||\ell' - g(x)|\big). \tag{11.34d}$$

Soient $\epsilon > 0$ et δ tel que $|f(x) - \ell| < \epsilon$ et $|g(x) - \ell'| < \epsilon$ pour tout $x \in B(a, \delta)$. Avec cela nous avons

$$\left|\frac{f(x)}{g(x)} - \frac{\ell}{\ell'}\right| \leqslant \frac{2}{|\ell'|^2\big(|\ell'| + |\ell|\big)}\epsilon. \tag{11.35}$$

D'où la limite attendue. □

Lemme 11.21.
Si $\lim_{x\to a} f(x) = b$ *avec* a, $b \in \mathbb{R}$, *alors il existe un* $\delta > 0$ *et un* $M > 0$ *tels que*

$$(|x - a| \leqslant \delta) \Rightarrow |f(x)| \leqslant M.$$

Ce que signifie ce lemme, c'est que quand la fonction f admet une limite finie en un point, alors il est possible de majorer la fonction sur un intervalle autour du point.

Démonstration. Cela va être démontré par l'absurde. Supposons qu'il n'existe pas de δ ni de M qui vérifient la condition. Dans ce cas, pour tout δ et pour tout M, il existe un x tel que $|x-a| \leqslant \delta$ et $|f(x)| > M$. Cela est valable pour tout M, donc prenons par exemple $b + 1000$. Donc

$$\forall \delta > 0, \exists x \text{ tel que } |x - a| \leqslant \delta \text{ et } |f(x)| > b + 1000. \tag{11.36}$$

Cela signifie qu'aucun δ ne peut convenir dans la définition de $\lim_{x\to a} f(x) = b$, ce qui contredit les hypothèses. □

Dans le même ordre d'idée, on peut prouver que si la limite de la fonction en un point est positive, alors elle est positive autour ce ce point. Plus précisément, nous avons la

Proposition 11.22.
Si f *est une fonction telle que* $\lim_{x\to a} f(x) > 0$, *alors il existe un voisinage de* a *sur lequel* f *est positive.*

Démonstration. Supposons que $\lim_{x\to a} f(x) = y_0$. Par la définition de la limite fait que si pour tout x dans un voisinage autour de a, on ait $|f(x) - a| < \epsilon$. Cela est valable pour tout ϵ, pourvu que le voisinage soit assez petit. Si je choisis un voisinage pour lequel $|f(x) - a| < \frac{y_0}{2}$, alors sur ce voisinage, f est positive. □

Théorème 11.23.
Si

$$\lim_{x\to a} f(x) = b_1 \qquad \text{et} \qquad \lim_{x\to a} g(x) = b_2, \tag{11.37}$$

alors

$$\lim_{x\to a} (fg)(x) = b_1 b_2. \tag{11.38}$$

Démonstration. Soit $\epsilon > 0$, et tentons de trouver un δ tel que $|f(x)g(x) - b_1b_2| \leqslant \epsilon$ dès que $|x - a| \leqslant \delta$. Nous avons

$$\begin{aligned}
|f(x)g(x) - b_1b_2| &= |f(x)g(x) - b_1b_2 + f(x)b_2 - f(x)b_2| \\
&= \big|f(x)\big(g(x) - b_2\big) + b_2\big(f(x) - b_1\big)\big| \\
&\leqslant \big|f(x)\big(g(x) - b_2\big)\big| + \big|b_2\big(f(x) - b_1\big)\big| \\
&= |f(x)||g(x) - b_2| + |b_2||f(x) - b_1|.
\end{aligned} \tag{11.39}$$

À la première ligne se trouve la subtilité de la démonstration : on ajoute et on enlève[1] $f(x)b_2$. Maintenant nous savons par le lemme 11.21 que pour un certain δ_1, la quantité $|f(x)|$ peut être majoré par un certain M dès que $|x-a| \leqslant \delta_1$. Prenons donc un tel δ_1 et supposons que $|x-a| \leqslant \delta_1$. Nous savons aussi que pour n'importe quel choix de ϵ_2 et ϵ_3, il existe des nombres δ_2 et δ_3 tels que

1. Comme exercice, tu peux essayer de refaire la démonstration en ajoutant et enlevant $g(x)b_1$ à la place.

$|f(x) - b_1| \leqslant \epsilon_2$ et $|g(x) - b_1| \leqslant \epsilon_3$ dès que $|x - a| \leqslant \delta_2$ et $|x - a| \leqslant \delta_3$. Dans ces conditions, la dernière expression (11.39) se réduit à

$$|f(x)g(x) - b_1b_2| \leqslant M\epsilon_2 + |b_2|\epsilon_3. \tag{11.40}$$

Pour terminer la preuve, il suffit de choisir ϵ_2 et ϵ_3 tels que $M\epsilon_2 + |b_2|\epsilon_3 \leqslant \epsilon$, et puis prendre $\delta = \min\{\delta_1, \delta_2, \delta_3\}$.

Remettons les choses dans l'ordre. L'on se donne ϵ au départ. La première chose est de trouver un δ_1 qui permet de majorer $|f(x)|$ par M selon le lemme 11.21, et puis choisissons ϵ_2 et ϵ_3 tels que $M\epsilon_2 + |b_2|\epsilon_3 \leqslant \epsilon$. Ensuite nous prenons, en vertu des hypothèses de limites pour f et g, les nombres δ_2 et δ_3 tels que $|f(x) - b_1| \leqslant \epsilon_2$ et $|g(x) - b_2| \leqslant \epsilon_3$ dès que $|x - a| \leqslant \delta_2$ et $|x - a| \leqslant \delta_3$.

Si avec tout ça on prend $\delta = \min\{\delta_1, \delta_2, \delta_3\}$, alors la majoration et les deux inégalités sont valables en même temps et au final

$$|f(x)g(x) - b_1b_2| \leqslant M\epsilon_2 + b_2\epsilon_3 \leqslant \epsilon,$$

ce qu'il fallait prouver.

□

À l'aide de ces petits résultats, nous pouvons déjà calculer pas mal de limites. Nous pouvons déjà par exemple calculer les limites de tous les polynômes en tous les nombres réels. En effet, nous savons la limite de la fonction $f(x) = x$. La fonction $x \mapsto x^2$ n'est rien d'autre que le produit de f par elle-même. Donc

$$\lim_{x\to a} x^2 = \big(\lim_{x\to a} x\big) \cdot \big(\lim_{x\to a} x\big) = a^2.$$

De la même façon, nous trouvons facilement que

$$\lim_{x\to a} x^n = a^n. \tag{11.41}$$

11.3.3 Limite en l'infini

Non, sur $\mathbb{R}$ nous n'allons pas ajouter ∞ avec la topologie d'Alexandrov de la définition 7.66. Nous n'allons pas considérer $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$.

Définition 11.24 (Droite réelle achevée[229])**.**
Nous considérons l'ensemble

$$\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\} \tag{11.42}$$

où $+\infty$ et $-\infty$ ne sont pas des éléments de $\mathbb{R}$.

Nous mettons sur $\bar{\mathbb{R}}$ la relation d'ordre en prenant celle de $\mathbb{R}$ à laquelle nous ajoutons les règles

(1) $-\infty < x$ pour tout $x \in \mathbb{R} \cup \{+\infty\}$

(2) $+\infty > x$ pour tout $x \in \mathbb{R} \cup \{-\infty\}$.

Nous mettons une topologie sur $\bar{\mathbb{R}}$ en donnant la base[2] suivante :

— $]a, b[$,

— $]a, +\infty]$,

— $[-\infty, b[$

pour tout réels a et b.

11.25.
En principe, la notation « ∞ » est réservée à l'infini du compactifié d'Alexandrov, et pour les infinis de la droite réelle achevée, il faudrait bien écrire « $+\infty$ » et « $-\infty$ ». Cependant, nous allons souvent écrire $\lim_{x\to\infty}$ au lieu de $\lim_{x\to+\infty}$.

2. Base de topologie, définition 7.59.

Lemme 11.26.
La topologie sur $\mathbb{R}$ induite de celle sur $\bar{\mathbb{R}}$ est la topologie usuelle.

Lemme 11.27.
Si $n \geqslant 1$ nous avons

$$\lim_{x\to+\infty} x^n = +\infty \tag{11.43}$$

et

$$\lim_{x\to+\infty} \frac{1}{x^n} = 0. \tag{11.44}$$

Démonstration. Si V est un voisinage de $+\infty$, alors nous devons montrer qu'il existe un voisinage W de $+\infty$ tel que $x^n \in V$ pour tout $x \in W$.

Un ouvert est une union d'éléments de la base de topologie [3] ; en regardant la liste donnée dans la définition 7.60, nous voyons que V contient au moins une partie de la forme $]R, +\infty]$. Nous supposons que $R > 1$.

Si $x > R > 1$, alors nous avons $x^n > x$ et donc

$$x^n > x > R, \tag{11.45}$$

ce qui signifie $x \in V$.

En prenant $W =]R, +\infty]$, nous avons bien $W^n \subset V$. Cela prouve (11.43).

En ce qui concerne la seconde limite, la démonstration est du même type. Remarquez seulement que vous n'avez pas formellement le droit d'utiliser la proposition 11.20 en invoquant $\frac{1}{+\infty} = 0$. □

11.3.4 Limite en des nombres

Nous posons la définition suivante.

Définition 11.28.
Lorsque $a \in \mathbb{R}$, on dit que la fonction f ***tend vers l'infini quand*** *x* ***tend vers*** *a si*

$$\forall M \in \mathbb{R}, \exists \delta \text{ tel que } (|x-a| \leqslant \delta) \Rightarrow f(x) \geqslant M \text{ quand } x \in \operatorname{dom} f.$$

Cela signifie que l'on demande que dès que x est assez proche de a (c'est-à-dire dès que $|x-a| \leqslant \delta$), alors $f(x)$ est plus grand que M, et que l'on peut trouver un δ qui fait ça pour n'importe quel M. Une autre façon de le dire est que pour toute hauteur M, on peut trouver un intervalle de largeur δ autour de a [4] tel que sur cet intervalle, la fonction f est toujours plus grande que M.

Montrons sur un dessin pourquoi je disais que la fonction $x \to 1/x$ n'est pas de ce type.

Le problème est qu'il n'existe par exemple aucun intervalle autour de 0 sur lequel f serait toujours plus grande que 10. En effet n'importe quel intervalle autour de 0 contient au moins un nombre négatif. Or quand x est négatif, f n'est certainement pas plus grande que 10. Nous y reviendrons.

Pour l'instant, montrons que la fonction $f(x) = 1/x^2$ est une fonction qui vérifie la définition 11.28. Avant de prendre n'importe quel M, prenons par exemple 100. Nous avons besoin d'un intervalle autour de zéro sur lequel f est toujours plus grande que 100. C'est vite vu que $f(0.1) = f(-0.1) = 100$, donc l'intervalle $[-\frac{1}{10}, \frac{1}{10}]$ est le bon. Partout dans cet intervalle, f est plus grande que 100. Partout ? Ben non : en $x = 0$, la fonction n'est même pas définie, donc c'est un peu dur de dire qu'elle est plus grande que 100. C'est pour cela que nous avons ajouté la condition « quand $x \in \operatorname{dom} f$ » dans la définition de la limite.

Prenons maintenant un $M \in \mathbb{R}$ arbitraire, et trouvons un intervalle autour de 0 sur lequel f est toujours plus grande que M. La réponse est évidemment l'intervalle de largeur $1/\sqrt{M}$, c'est-à-dire

$$\left[-\frac{1}{\sqrt{M}}, \frac{1}{\sqrt{M}}\right].$$

3. C'est la proposition 7.60.
4. C'est-à-dire un intervalle de la forme $[a-\delta, a+\delta]$.

11.3.5 Limites quand tout va bien

D'abord définissons ce qu'on entend par la limite d'une fonction en un point quand il n'y a aucun infini en jeu.

Définition 11.29.
On dit que la fonction f ***tend vers*** *b* ***quand*** *x* ***tend vers*** *a si*

$$\forall \epsilon > 0, \exists \delta \text{ tel que } (|x-a| \leqslant \delta) \Rightarrow |f(x)-b| \leqslant \epsilon \text{ quand } x \in \operatorname{dom} f.$$

Dans ce cas, nous notons

$$\lim_{x \to a} f(x) = b. \tag{11.46}$$

Commençons par un exemple très simple : prouvons que $\lim_{x\to 0} x = 0$. C'est donc $a = b = 0$ dans la définition. Prenons $\epsilon > 0$, et trouvons un intervalle autour de zéro tel que partout dans l'intervalle, $x \leqslant \epsilon$. Bon ben c'est clair que $\delta = \epsilon$ fonctionne.

Plus compliqué maintenant, mais toujours sans surprises.

Proposition 11.30.

$$\lim_{x \to 0} x^2 = 0.$$

Démonstration. Soit $\epsilon > 0$. On veut un intervalle de largeur δ autour de zéro tel que x^2 soit plus petit que ϵ sur cet intervalle. Cette fois-ci, le δ qui fonctionne est $\delta = \sqrt{\epsilon}$. En effet un élément de l'intervalle $[-\delta, \delta]$ est un r de valeur absolue plus petite ou égale à δ :

$$|r| \leqslant \delta = \sqrt{\epsilon}.$$

En prenant le carré de cette inégalité on a :

$$r^2 \leqslant \epsilon,$$

ce qu'il fallait prouver. □

Calculer et prouver des valeurs de limites, mêmes très simples, devient vite de l'arrachage de cheveux à essayer de trouver le bon δ en fonction de ϵ si on n'a pas quelques théorèmes généraux. Heureusement nous en avons déjà quelques uns : 11.16, 11.17, 11.18, 11.19, 11.20.

Proposition 11.31 ([1])**.**
Soit $f\colon \mathbb{R}^2 \to \mathbb{R}$ une application continue dont la variable y varie dans un compact I de $\mathbb{R}$. Alors la fonction

$$\begin{aligned} d\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \sup_{y \in I} f(x,y) \end{aligned} \tag{11.47}$$

est continue.

Démonstration. Soit x_0 fixé. Prouvons que d est continue en x_0. Nous notons y_0 la valeur de y qui réalise le maximum (par le théorème 7.154 et le fait que les fonctions projection soient continues, lemme 7.88). Soit aussi $\epsilon > 0$ tellement fixé que même avec un tourne vis hydraulique, il ne bougerait pas. Nous considérons δ tel que si $\|(x,y) - (x_0,y_0)\| \leqslant \delta$ alors $\|f(x,y) - f(x_0,y_0)\| < \epsilon$.

Si $|x - x_0| < \delta$ alors pour y assez proche de y_0 nous avons $\|(x,y) - (x_0,y_0)\| \leqslant \delta$, et donc $\|f(x,y) - f(x_0,y_0)\| \leqslant \epsilon$. Cela montre qu'il existe δ tel que $|x - x_0| \leqslant \delta$ implique $d(x) \geqslant d(x_0) - \epsilon$.

Nous devons encore trouver un δ tel que si $|x - x_0| \leqslant \delta$ alors $d(x) \leqslant d(x_0) + \epsilon$. Supposons que non. Alors pour tout δ il existe un x tel que $|x - x_0| \leqslant \delta$ et $d(x) > d(x_0) + \epsilon$. Cela nous donne une suite $x_i \to x_0$.

Pour chaque x_i nous notons y_i la valeur de y qui réalise le supremum correspondant. La suite (y_i) étant contenue dans un compact nous supposons prendre une sous-suite de (x_i) telle que la

suite (y_i) converge. Nous nommons a la limite (et non y_0 parce que nous ne savons pas si $y_i \to y_0$). Pour chaque i nous avons

$$f(x_i, y_i) > \sup_{y \in I} f(x_0, y) + \epsilon. \tag{11.48}$$

En prenant la limite et en utilisant la continuité de f,

$$f(x_0, a) > \sup_{y \in I} f(x_0, y) + \epsilon, \tag{11.49}$$

ce qui est impossible. □

11.3.6 Limites de fonctions

Définition 11.32.
Soit $f\colon D \subset \mathbb{R}^m \to \mathbb{R}$ une fonction et a un point d'accumulation de D. On dit que f possède une ***limite*** *s'il existe un élément $\ell \in \mathbb{R}$ tel que*

$$\forall \varepsilon > 0, \exists \delta > 0 \text{ tel que } 0 < \|x - a\| < \delta \Rightarrow |f(x) - \ell| < \varepsilon. \tag{11.50}$$

Pour une fonction $f\colon D \subset \mathbb{R}^m \to \mathbb{R}^n$, la définition est la même, sauf que nous remplaçons la valeur absolue par la norme dans $\mathbb{R}^n$. Nous disons donc que ℓ est la limite de f lorsque x tend vers a, et nous notons $\lim_{x \to a} f(x) = \ell$ lorsque pour tout $\varepsilon > 0$, il existe un $\delta > 0$ tel que

$$0 < \|x - a\|_{\mathbb{R}^m} < \delta \Rightarrow \|f(x) - \ell\|_{\mathbb{R}^n} < \varepsilon. \tag{11.51}$$

Remarque 11.33.
Dans l'équation (11.51), nous avons explicitement écrit les normes $\|.\|_{\mathbb{R}^m}$ et $\|.\|_{\mathbb{R}^n}$. Dans la suite nous allons le plus souvent noter $\|.\|$ sans plus de précision. Il est important de faire l'exercice de bien comprendre à chaque fois de quelle norme nous parlons.

Remarque 11.34.
Il est important de remarquer à quel point les définitions 11.32, et les caractérisations 9.190, 11.12 sont analogues. En réalité, la définition fondamentale est la définition de la limite dans les espaces vectoriels normés ; les deux autres sont des cas particuliers, adaptés à $\mathbb{R}$ et $\mathbb{R}^m$. Il en sera de même pour les définitions de fonctions continues : il y aura une définition pour la continuité de fonctions entre espaces vectoriels normés, et ensuite une définition pour les fonctions de $\mathbb{R}^m$ dans $\mathbb{R}^n$ qui en sera un cas particulier.

Tentons de comprendre ce que signifie qu'un nombre ℓ *ne soit pas* la limite de f lorsque $x \to a$. Il s'agit d'inverser la condition (11.50). Le nombre ℓ n'est pas une limite de f pour $x \to a$ lorsque

$$\exists \varepsilon > 0 \text{ tel que } \forall \delta > 0, \exists x \text{ tel que } 0 < \|x - a\| < \delta \text{ et } \|f(x) - \ell\| > \varepsilon, \tag{11.52}$$

c'est-à-dire qu'il existe un certain seuil ε tel qu'on a beau s'approcher aussi proche qu'on veut de a (distance δ), on trouvera toujours un x tel que $f(x)$ n'est pas ε-proche de ℓ.

Lemme 11.35 (Unicité de la limite)**.**
Si ℓ et ℓ' sont deux limites de $f(x)$ lorsque x tend vers a, alors $\ell = \ell'$.

Démonstration. Soit $\varepsilon > 0$. Nous considérons δ tel que $\|f(x) - \ell\| < \varepsilon$ pour tout x tel que $\|x - a\| < \delta$. De la même manière, nous prenons δ' tel que $\|x - a\| < \delta'$ implique $\|f(x) - \ell'\| < \varepsilon$. Pour les x tels que $\|x - a\|$ est plus petit que δ et δ' en même temps, nous avons

$$\|\ell - \ell'\| = \|\ell - f(x) + f(x) - \ell'\| \leqslant \|\ell - f(x)\| + \|f(x) - \ell'\| < 2\varepsilon, \tag{11.53}$$

et donc $\|\ell - \ell'\| = 0$ parce que c'est plus petit que 2ε pour tout ε. □

11.3.7 Limite à gauche et à droite

Si a est à l'intérieur du domaine de f, nous savons ce que signifie $\lim_{x\to a} f(x)$. Nous donnons également une définition des limites à gauche et à droite.

Définition 11.36.
*Soient $D \subset \mathbb{R}$ et une fonction $f\colon D \to \mathbb{R}$. Si $a \in \mathrm{Adh}(D)$ nous définissons la **limite à droite** de f en a par*

$$\lim_{x\to a^+} f(x) = \lim_{x\to a} \tilde{f}(x) \tag{11.54}$$

où $\tilde{f}$ est la fonction f restreinte à $D \cap \{x \text{ tel que } x > a\}$. La limite (11.54) est souvent écrite sous la forme condensée

$$\lim_{\substack{x\to a\\ x>a}} f(x). \tag{11.55}$$

Pour la limite à gauche c'est un peu la même chose :

$$\lim_{x\to a^-} f(x) = \lim_{\substack{x\to a\\ x<a}} f(x). \tag{11.56}$$

Proposition 11.37 ([230])**.**
Soit une fonction $f\colon D \to \mathbb{R}$ où D est une partie de $\mathbb{R}$. Si $a \in \mathrm{Adh}(D)$ alors la limite $\lim_{x\to a} f(x)$ existe si et seulement si les limites à gauche et à droite existent et sont égales. Dans ce cas nous avons égalité :

$$\lim_{x\to a} f(x) = \lim_{x\to a^+} f(x) = \lim_{x\to a^-} f(x). \tag{11.57}$$

11.38.
Quelques remarques à propos de la proposition 11.37.

(1) Cette proposition ne se généralise pas du tout aux dimensions supérieures. Dans $\mathbb{R}^2$ par exemple, il ne faudrait pas croire que si les limites suivant toutes les directions existent alors la limite existe.

(2) Cette proposition est souvent utilisée pour calculer des limites dans lesquelles arrivent des valeurs absolues. Par exemple durant la démonstration de la proposition 11.288.

11.4 Limite en compactifié d'Alexandrov

Nous considérons l'espace topologique localement compact $\mathbb{R}$, et son compactifié d'Alexandrov défini en 7.66. Nous avons donc un point supplémentaire noté ∞. Ce point n'est ni du côté des grands nombres positifs, ni du côté des grands nombres négatifs. Il n'est ni $+\infty$ ni $-\infty$.

Proposition 11.39.
Dans cet espace topologique $\hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$,

$$\lim_{x\to 0} \frac{1}{x} = \infty. \tag{11.58}$$

Démonstration. Soit un voisinage V de ∞ dans $\hat{\mathbb{R}}$. Il s'écrit $V = K^c \cup \{\infty\}$ pour un certain compact de $\mathbb{R}$. Le théorème 7.127 nous assure que K est borné. Donc il existe $R > 0$ tel que $K \subset B(0, R)$. Pour $x \in B(0, 1/R)$ nous avons

$$|\frac{1}{x}| > R, \tag{11.59}$$

et donc $1/x \in K^c$. Donc aussi $\frac{1}{x} \in V$. □

De la même façon, dans $\mathbb{C} \cup \{\infty\}$ nous avons

$$\lim_{z\to 0} \frac{1}{z} = \infty. \tag{11.60}$$

11.40.
Je vous laisse deviner la topologie à considérer sur $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$. Dans cet espace topologique la limite $\lim_{x\to 0} \frac{1}{x}$ n'existe pas.

11.5 Continuité

11.41.
Nous allons considérer trois approches différentes de la continuité. La première sera de définir la continuité de fonctions de $\mathbb{R}$ vers $\mathbb{R}$ au moyen du critère usuel. Ensuite, nous définiront la continuité des applications entre n'importe quels espaces métriques, et nous montrerons que les deux définitions sont équivalentes dans le cas des fonctions sur $\mathbb{R}$ à valeurs réelles.

Enfin, un peu plus tard nous verrons que la continuité peut également être vue en termes de limites. Encore une fois nous verrons que dans le cas de fonctions de $\mathbb{R}$ vers $\mathbb{R}$ cette troisième approche est équivalentes aux deux premières.

La définition de fonction continue est la définition 7.71. Dans le cas d'une fonction $f\colon \mathbb{R} \to \mathbb{R}$, elle devient ceci.

Proposition 11.42.
*La fonction $f\colon \mathbb{R} \to \mathbb{R}$ est **continue en** a si et seulement si*

$$\forall \epsilon > 0, \exists \delta \text{ tel que } \big(|x-a| \leqslant \delta\big) \Rightarrow |f(x) - f(a)| \leqslant \epsilon. \tag{11.61}$$

Nous allons maintenant étudier quelques conséquences de la continuité sur $\mathbb{R}$.

(1) D'abord on voit que la continuité n'a été définie qu'en un point. On peut dire que la fonction f est continue *en tel point donné*, mais nous n'avons pas dit ce qu'est une fonction continue *dans son ensemble*.

(2) Le théorème 7.74 nous précise que si I est un intervalle de $\mathbb{R}$, la fonction f est continue sur I si et seulement si elle est continue en chaque point de I.

(3) Comme la définition de f continue en a fait intervenir $f(x)$ pour tous les x pas trop loin de a, il faut au moins déjà que f soit définie sur ces x. En d'autres termes, dire que f est continue en a demande que f existe sur un intervalle autour de a.

Ceci couplé à la définition précédente laisse penser qu'il est surtout intéressant d'étudier les fonctions qui sont continues sur un intervalle.

(4) L'intuition comme quoi une fonction continue doit pouvoir être tracée sans lever la main correspond aux fonctions continues sur des intervalles. Au moins sur l'intervalle où elle est continue, elle est traçable en un morceau.

Exemple 11.43.
Il est très possible d'être continue en un seul point. Par exemple la fonction

$$f(x) = x(1 - \mathbb{1}_{\mathbb{Q}}(x)) \tag{11.62}$$

où $\mathbb{1}_{\mathbb{Q}}$ est la fonction indicatrice de $\mathbb{Q}$ dans $\mathbb{R}$. △

Proposition 11.44.
Si $f\colon \mathbb{R} \to \mathbb{R}$ est continue au point $a \in \mathbb{R}$ et si $f(a) \neq 0$, alors il existe un voisinage de a sur lequel f ne s'annule pas.

Démonstration. Si f s'annulait sur tout voisinage de a (mais pas ne a lui-même), nous aurions, pour tout n un réel

$$x_n \in B\big(a, \frac{1}{n}\big) \backslash \{a\} \tag{11.63}$$

tel que $f(x_n) = 0$. Cela donnerait une suite $x_n \to a$ avec $f(x_n) \to 0$, ce qui contredit la continuité de f en a en vertu de la proposition 7.807.264 sur la continuité séquentielle en un point. □

Notons que ce résultat se généralise beaucoup : si f est continue et pas égale à r en a, alors elle continue à n'être pas égale à r dans un voisinage de a.

11.5.1 Opération sur la continuité

Nous allons démontrer maintenant une série de petits résultats qui permettent de simplifier la démonstration de la continuité de fonctions.

Théorème 11.45.
Si la fonction f est continue au point a, alors la fonction λf est également continue en a.

Démonstration. Soit $\epsilon > 0$. Nous avons besoin d'un $\delta > 0$ tel que pour chaque x à moins de δ de a, la fonction λf soit à moins de ϵ de $(\lambda f)(a) = \lambda f(a)$. Étant donné que la fonction f est continue en a, on sait déjà qu'il existe un δ_1 (nous notons δ_1 afin de ne pas confondre ce nombre dont on est sûr de l'existence avec le δ que nous sommes en train de chercher) tel que

$$(|x-a| \leqslant \delta_1) \Rightarrow |f(x) - f(a)| \leqslant \epsilon_1.$$

Hélas, ce δ_1 n'est pas celui qu'il faut faut parce que nous travaillons avec λf au lieu de f, ce qui fait qu'au lieu d'avoir $|f(x) - f(a)|$, nous avons $|\lambda f(x) - \lambda f(a)| = |\lambda| \cdot |f(x) - f(a)|$. Ce que δ_1 fait avec (λf), c'est

$$(|x-a| \leqslant \delta_1) \Rightarrow |(\lambda f)(x) - (\lambda f)(a)| \leqslant |\lambda|\epsilon_1.$$

Ce que nous apprend la continuité de f, c'est que pour chaque choix de ϵ_1, on a un δ_1 qui fait cette implication. Comme cela est vrai pour chaque choix de ϵ_1, essayons avec $\epsilon_1 = \epsilon/|\lambda|$ pour voir ce que ça donne. Nous avons donc un δ_1 qui fait

$$(|x-a| \leqslant \delta_1) \Rightarrow |(\lambda f)(x) - (\lambda f)(a)| \leqslant |\lambda|\epsilon_1 = \epsilon.$$

Ce δ_1 est celui qu'on cherchait. □

Théorème 11.46.
Si f et g sont deux fonctions continues en a, alors la fonction $f + g$ est également continue en a.

Démonstration. La continuité des fonctions f et g au point a fait en sorte que pour tout choix de ϵ_1 et ϵ_2, il existe δ_1 et δ_2 tels que

$$(|x-a| \leqslant \delta_1) \Rightarrow |f(x) - f(a)| \leqslant \epsilon_1.$$

et

$$(|x-a| \leqslant \delta_2) \Rightarrow |g(x) - g(a)| \leqslant \epsilon_2.$$

La quantité que nous souhaitons analyser est $|f(x) + g(x) - f(a) - g(a)|$. Tout le jeu de la démonstration de la continuité est de triturer cette expression pour en tirer quelque chose en termes de ϵ_1 et ϵ_2. Si nous supposons avoir pris $|x-a|$ plus petit en même temps que δ_1 et que δ_2, nous avons

$$|f(x) + g(x) - f(a) - g(a)| \leqslant |f(x) - g(x)| + |g(x) - g(a)| \leqslant \epsilon_1 + \epsilon_2$$

en utilisant la formule générale $|a+b| \leqslant |a|+|b|$. Maintenant si on choisit ϵ_1 et ϵ_2 tels que $\epsilon_1+\epsilon_2 < \epsilon$, et les δ_1, δ_2 correspondants, on a

$$|f(x) + g(x) - f(a) - g(a)| \leqslant \epsilon,$$

pourvu que $|x-a|$ soit plus petit que δ_1 et δ_2. Le bon δ à prendre est donc le minimum de δ_1 et δ_2 qui eux-mêmes sont donnés par un choix de ϵ_1 et ϵ_2 tels que $\epsilon_1 + \epsilon_2 \leqslant \epsilon$. □

Pour résumer ces deux théorèmes, on dit que si f et g sont continues en a, alors la fonction $\alpha f + \beta g$ est également continue en a pour tout α, $\beta \in \mathbb{R}$.

Parmi les propriétés immédiates de la continuité d'une fonction, nous avons ceci qui est souvent bien utile.

Corolaire 11.47.
Si la fonction f est continue en a et si $f(a) > 0$, alors f est positive sur un intervalle autour de a.

Démonstration. Prenons $\epsilon < f(a)$ et voyons[5] ce que la continuité de f en a nous offre : il existe un δ tel que

$$(|x-a| \leqslant \delta) \Rightarrow |f(x)-f(a)| \leqslant \epsilon < f(a).$$

Nous en retenons que sur un intervalle (de largeur δ), nous avons $|f(x)-f(a)| \leqslant f(a)$. Par hypothèse, $f(a) > 0$, donc si $f(x) < 0$, alors la différence $f(x)-f(a)$ donne un nombre encore plus négatif que $-f(a)$, c'est-à-dire que $|f(x)-f(a)| > f(a)$, ce qui est contraire à ce que nous venons de démontrer. D'où la conclusion que $f(x) > 0$. □

11.5.2 La fonction la moins continue du monde

Parmi les exemples un peu sales de fonctions non continues, il y a celle-ci :

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 \text{ si } x \in \mathbb{Q} \\ 0 \text{ sinon.} \end{cases}$$

Par exemple, $\chi_{\mathbb{Q}}(0) = 1$, et[6] $\chi_{\mathbb{Q}}(\pi) = \chi_{\mathbb{Q}}(\sqrt{2}) = 0$. Malgré que $\chi_{\mathbb{Q}}(0) = 1$, il n'existe *aucun* voisinage de 1 sur lequel la fonction reste proche de 1, parce que tout voisinage va contenir au moins un irrationnel. À chaque millimètre, cette fonction fait une infinité de bonds !

Cette fonction n'est donc continue nulle part.

À partir de là, nous pouvons construire la fonction suivante qui n'est continue qu'en un point :

$$f(x) = x\chi_{\mathbb{Q}}(x) = \begin{cases} x \text{ si } x \in \mathbb{Q} \\ 0 \text{ sinon.} \end{cases}$$

Cette fonction est continue en zéro. En effet, prenons $\delta > 0$; il nous faut un ϵ tel que $|x| \leqslant \epsilon$ implique $f(x) \leqslant \delta$ parce que $f(0) = 0$. Bon ben prendre simplement $\epsilon = \delta$ nous contente. Cette fonction est donc très facilement continue en zéro.

Et pourtant, dès que l'on s'écarte un tant soit peu de zéro, elle fait des bons une infinité de fois par millionième de millimètre ! Cette fonction est donc la plus discontinue du monde en tous les points saut un (zéro) où elle est une fonction continue !

11.5.3 Approche topologique

Nous avons vu que sur tout ensemble métrique, nous pouvons définir ce qu'est un ouvert : c'est un ensemble qui contient une boule ouverte autour de chacun de ses points. Quand on est dans un ensemble ouvert, on peut toujours un peu se déplacer sans sortir de l'ensemble.

Le théorème suivant est une très importante caractérisation des fonctions continues (de $\mathbb{R}$ dans $\mathbb{R}$) en termes de topologie, c'est-à-dire en termes d'ouverts.

Théorème 11.48.
Si I est un intervalle ouvert contenu dans $\operatorname{dom} f$*, alors f est continue sur I si et seulement si pour tout ouvert $\mathcal{O}$ dans $\mathbb{R}$, l'image inverse $f|_I^{-1}(\mathcal{O})$ est ouvert.*

Par abus de langage, nous exprimons souvent cette condition par « une fonction est continue si et seulement si l'image inverse de tout ouvert est un ouvert ».

Démonstration. Dans un premier temps, nous allons transformer le critère de continuité en termes de boules ouvertes, et ensuite, nous passerons à la démonstration proprement dite. Le critère de continuité de f au point x dit que

$$\forall \delta > 0, \exists\, \epsilon > 0 \text{ tel que } \big(|x-a| < \epsilon\big) \Rightarrow |f(x)-f(a)| < \delta. \tag{11.64}$$

5. ici, nous insistons sur le fait que nous prenons ϵ *strictement* plus petit que $f(a)$.

6. Pour prouver que $\sqrt{2}$ n'est pas rationnel, c'est pas trop compliqué, mais pour prouver que π ne l'est pas non plus, il faudra encore manger de la soupe.

Cette condition peut être exprimée sous la forme suivante :

$$\forall \delta > 0, \exists \epsilon \text{ tel que } a \in B(x,\epsilon) \Rightarrow f(a) \in B\big(f(x),\delta\big),$$

ou encore

$$\forall \delta > 0, \exists \epsilon \text{ tel que } f\big(B(x,\epsilon)\big) \subset B\big(f(x),\delta\big). \tag{11.65}$$

Jusque ici, nous n'avons fait que du jeu de notations. Nous avons exprimé en termes de topologie des inégalités analytiques. La condition (11.65) est le plus souvent utilisée comme définition de la continuité d'une fonction en x, lorsque le contexte ne demande pas de définitions plus générales. Si tel est le choix, il faut pouvoir retrouver (11.64) à partir de (11.65).

Passons maintenant à la démonstration proprement dite du théorème.

D'abord, supposons que f est continue sur I, et prenons $\mathcal{O}$, un ouvert quelconque. Le but est de prouver que $f|_I^{-1}(\mathcal{O})$ est ouvert. Pour cela, nous prenons un point $x_0 \in f|_I^{-1}(\mathcal{O})$ et nous allons trouver un ouvert autour ce point contenu dans $f|_I^{-1}(\mathcal{O})$. Nous écrivons $y_0 = f(x_0)$. évidemment, $y_0 \in \mathcal{O}$, donc on a une boule autour de y_0 qui est contenue dans $\mathcal{O}$, soit donc $\delta > 0$ tel que

$$B(y_0,\delta) \subset \mathcal{O}.$$

Par hypothèse, f est continue en x_0, et nous pouvons donc y appliquer le critère (11.65). Il existe donc $\epsilon > 0$ tel que

$$f\big(B(x_0,\epsilon)\big) \subset B\big(f(x_0),\delta\big) \subset \mathcal{O}.$$

Cela prouve que $B(x_0,\epsilon) \subset f|_I^{-1}(\mathcal{O})$.

Dans l'autre sens, maintenant. Nous prenons $x_0 \in I$ et nous voulons prouver que f est continue en x_0, c'est-à-dire que pour tout δ nous cherchons un ϵ tel que $f\big(B(x_0,\epsilon)\big) \subset B\big(f(x_0),\delta\big)$. Oui, mais $B\big(f(x_0),\delta\big)$ est ouverte, donc par hypothèse, $f|_I^{-1}\Big(B\big(f(x_0),\delta\big)\Big)$ est ouvert, inclus dans I et contient x_0. Donc il existe un ϵ tel que

$$B(x_0,\epsilon) \subset f|_I^{-1}\Big(B\big(f(x_0),\delta\big)\Big),$$

et donc tel que

$$f\big(B(x_0,\epsilon)\big) \subset B\big(f(x_0),\delta\big),$$

ce qu'il fallait prouver. □

Lemme 11.49.
L'image d'un ensemble connexe par une fonction continue est connexe.

Démonstration. Nous allons encore faire la contraposée. Soit A une partie de $\mathbb{R}$ telle que $f(A)$ ne soit pas connexe. Nous allons prouver que A elle-même n'est pas connexe. Dire que $f(A)$ n'est pas connexe, c'est dire qu'il existe $\mathcal{O}_1$ et $\mathcal{O}_2$, deux ouverts disjoints qui recouvrent $f(A)$. Je prétends que $f^{-1}(\mathcal{O}_1)$ et $f^{-1}(\mathcal{O}_2)$ sont ouverts, disjoints et qu'ils recouvrent A.

— Ces deux ensembles sont ouverts parce qu'ils sont images inverses d'ouverts par une fonction continue (théorème 11.48).

— Si $x \in f^{-1}(\mathcal{O}_1) \cap f^{-1}(\mathcal{O}_2)$, alors $f(x) \in \mathcal{O}_1 \cap \mathcal{O}_2$, ce qui contredirait le fait que $\mathcal{O}_1$ et $\mathcal{O}_2$ sont disjoints. Il n'y a donc pas d'éléments dans l'intersection de $f^{-1}(\mathcal{O}_1)$ et de $f^{-1}(\mathcal{O}_2)$.

— Si $f^{-1}(\mathcal{O}_1)$ et $f^{-1}(\mathcal{O}_2)$ ne recouvrent pas A, il existe un x dans A qui n'est dans aucun des deux. Dans ce cas, $f(x)$ est dans $f(A)$, mais n'est ni dans $\mathcal{O}_1$, ni dans $\mathcal{O}_2$, ce qui contredirait le fait que ces deux derniers recouvrent $f(A)$.

Nous déduisons que A n'est pas connexe. Et donc le lemme. □

Théorème 11.50 (Théorème des valeurs intermédiaires)**.**
Soit f, une fonction continue sur $[a,b]$, et supposons que $f(a) < f(b)$. Alors pour tout y tel que $f(a) \leqslant y \leqslant f(b)$, il existe un $x \in [a,b]$ tel que $f(x) = y$.

Démonstration. Nous savons que $[a,b]$ est connexe parce que c'est un intervalle (proposition 7.153). Donc $f\big([a,b]\big)$ est connexe (lemme 11.49) et donc est un intervalle (à nouveau la proposition 7.153). Étant donné que $f\big([a,b]\big)$ est un intervalle, il contient toutes les valeurs intermédiaires entre n'importe quels deux de ses éléments. En particulier toutes les valeurs intermédiaires entre $f(a)$ et $f(b)$. □

Corolaire 11.51.
L'image d'un intervalle par une fonction continue est un intervalle.

Démonstration. Soient I un intervalle, $\alpha < \beta \in f(I)$ et $\gamma \in]\alpha,\beta[$. Nous considérons $a,b \in I$ tels que $\alpha = f(a)$ et $\beta = f(b)$. Par le théorème des valeurs intermédiaires 11.50, il existe $t \in]a,b[$ tel que $f(t) = \gamma$. Par conséquent $\gamma \in f(I)$. □

Corolaire-définition 11.52 (Existence de la racine carrée)**.**
Si $x \geqslant 0$ alors il existe un unique $y \geqslant 0$ tel que $y^2 = x$. Ce nombre est noté $\sqrt{x}$ et est nommé ***racine carrée*** *de x.*

Démonstration. La fonction $f\colon t \mapsto t^2$ est continue et strictement croissante. Nous avons $f(0) = 0$ et[7] $f(x+1) > x$. Donc le théorème des valeurs intermédiaires 11.50 nous assure qu'il existe un unique $y \in [0, x+1]$ tel que $f(y) = x$. □

11.5.4 Continuité de la racine carrée, invitation à la topologie induite

Pourquoi nous intéresser particulièrement à cette fonction ? Parce qu'elle a une sale condition d'existence : son domaine de définition n'est pas ouvert. Or dans tous les théorèmes de continuité d'approche topologique que nous avons vus, nous avons donné des conditions *pour tout ouvert*. Nous nous attendons donc a avoir des difficultés avec la continuité de $\sqrt{x}$ en zéro.

Prenons I, n'importe quel intervalle ouvert dans $\mathbb{R}^+$, et voyons que la fonction

$$\begin{aligned} f\colon \mathbb{R}^+ &\to \mathbb{R}^+ \\ x &\mapsto \sqrt{x} \end{aligned} \tag{11.66}$$

est continue sur I. Remarque déjà que si I est un ouvert dans $\mathbb{R}^+$, il ne peut pas contenir zéro. Avant de nous lancer dans notre propos, nous prouvons un lemme qui fera tout le travail[8].

Lemme 11.53.
Soit $\mathcal{O}$, un ouvert dans $\mathbb{R}^+$. Alors $\mathcal{O}^2 = \{x^2 \text{ tel que } x \in \mathcal{O}\}$ est également ouvert .

Démonstration. Un élément de $\mathcal{O}^2$ s'écrit sous la forme x^2 pour un certain $x \in \mathcal{O}$. Le but est de trouver un ouvert autour de x^2 qui soit contenu dans $\mathcal{O}^2$. Étant donné que $\mathcal{O}$ est ouvert, on a une boule centrée en x contenue dans $\mathcal{O}$. Nous appelons δ le rayon de cette boule :

$$B(x,\delta) \subset \mathcal{O}.$$

Étant donné que cet ensemble est connexe, nous savons par le lemme 11.49 que $B(x,\delta)^2$ est également connexe (parce que la fonction $x \mapsto x^2$ est continue). Son plus grand élément est $(x+\delta)^2 = x^2 + \delta^2 + 2x\delta > x^2 + \delta^2$, et son plus petit élément est $(x-\delta)^2 = x^2 + \delta^2 - 2x\delta$.

Ce qui serait pas mal, c'est que ces deux bornes entourent x^2 ; de cette façon elles définiraient un ouvert autour de x^2 qui soit dans $\mathcal{O}^2$. Hélas, c'est pas gagné que $x^2 + \delta^2 - 2x\delta$ soit plus petit que x^2.

Heureusement, en fait c'est vrai parce que d'une part, du fait que $\mathcal{O} \subset \mathbb{R}^+$, on a $x > 0$, et d'autre part, pour que $\mathcal{O}$ soit positif, il faut que $\delta < x$. Donc on a évidemment que $\delta < 2x$, et donc que

$$x^2 + \delta^2 - 2x\delta = x^2 + \delta\underbrace{(\delta - 2x)}_{<0} < x^2.$$

7. Faites deux cas suivant $x \geqslant 1$ ou non si vous le voulez, moi je prends $x+1$.
8. C'est toujours ingrat d'être un lemme : on fait tout le travail et c'est toujours le théorème qui est nommé.

Donc nous avons fini : l'ensemble

$$B(x,\delta)^2 =]x^2+\delta^2-2x\delta, x^2+\delta^2+2x\delta[\subset \mathcal{O}^2$$

est un intervalle qui contient x^2, et donc qui contient une boule ouverte centrée en x^2.

□

Maintenant nous pouvons nous attaquer à la continuité de la racine carrée sur tout ouvert positif en utilisant le théorème 11.48. Soit $\mathcal{O}$ n'importe quel ouvert de $\mathbb{R}$, et prouvons que $f|_I^{-1}(\mathcal{O})$ est ouvert. Par définition,

$$f|_I^{-1}(\mathcal{O}) = \{x \in I \text{ tel que } \sqrt{x} \in \mathcal{O}\}. \tag{11.67}$$

Maintenant c'est un tout petit effort que de remarquer que $f|_I^{-1}(\mathcal{O}) = \mathcal{O}^2 \cap I$. De là, on a gagné parce que $\mathcal{O}^2$ et I sont des ouverts. Or l'intersection de deux ouverts est ouvert.

Nous n'en avons pas fini avec la fonction $\sqrt{x}$. Nous avons la continuité de la racine carrée pour tous les réels strictement positifs. Il reste à pouvoir dire que la fonction est continue en zéro malgré qu'elle ne soit pas définie sur un ouvert autour de zéro.

Il est possible de dire que la racine carrée est continue en 0, malgré qu'elle ne soit pas définie sur un ouvert autour de 0... en tout cas pas un ouvert au sens que tu as en tête. Nous allons rentabiliser un bon coup notre travail sur les espaces métriques.

Nous pouvons définir la notion de boule ouverte sur n'importe quel espace métrique A en disant que

$$B(x,r) = \{y \in A \text{ tel que } d(x,y) < r\}.$$

Définition 11.54.
*Soit $f\colon A \to B$, une application entre deux espaces métriques. Nous disons que f est **continue** au point $a \in A$ si $\forall \delta > 0$, $\exists \epsilon > 0$ tel que*

$$f\big(B(a,\epsilon)\big) \subset B\big(f(a),\delta\big). \tag{11.68}$$

Tu reconnais évidemment la condition (11.65). Nous l'avons juste recopiée. Tu remarqueras cependant que cette définition généralise immensément la continuité que l'on avait travaillé à propos des fonctions de $\mathbb{R}$ vers $\mathbb{R}$. Maintenant tu peux prendre n'importe quel espace métrique et c'est bon.

Nous n'allons pas faire un tour complet des conséquences et exemples de cette définition. Au lieu de cela, nous allons juste montrer en quoi cette définition règle le problème de la continuité de la racine carrée en zéro.

La fonction que nous regardons est

$$\begin{aligned} f\colon \mathbb{R}^+ &\to \mathbb{R}^+ \\ x &\mapsto \sqrt{x}. \end{aligned} \tag{11.69}$$

Mais cette fois, nous ne la voyons pas comme étant une fonction dont le domaine est une partie de $\mathbb{R}$, mais comme fonction dont le domaine est $\mathbb{R}^+$ vu comme un espace métrique en soi. Quelles sont les boules ouvertes dans $\mathbb{R}^+$ autour de zéro ? Réponse : la boule ouverte de rayon r autour de zéro dans $\mathbb{R}^+$ est :

$$B(0,r)_{\mathbb{R}^+} = \{x \in \mathbb{R}^+ \text{ tel que } d(x,0) < r\} = [0,r[.$$

Cet intervalle est un ouvert. Aussi incroyable que cela puisse paraitre !

Testons la continuité de la racine carrée en zéro dans ce contexte. Il s'agit de prendre $A = \mathbb{R}^+$, $B = \mathbb{R}^+$ et $a = 0$ dans la définition 11.54. Nous avons que $B(\sqrt{0},\delta) = B(0,\delta) = [0,\delta[$ pour la topologie de $\mathbb{R}^+$.

Il s'agit maintenant de trouver un ϵ tel que $f\big(B(0,\epsilon)\big) \subset [0,\delta[$. Par définition, nous avons que

$$f\big(B(0,\epsilon)\big) = [0,\sqrt{\epsilon}[,$$

le problème revient dont à trouver ϵ tel que $\sqrt{\epsilon} \leqslant \delta$. Prendre $\epsilon < \delta^2$ fait l'affaire.

Donc voilà. Au sens de la topologie propre à $\mathbb{R}^+$, nous pouvons dire que la fonction racine carrée est partout continue.

11.5.5 Prolongement par continuité

11.5.5.1 Discussion avec mon ordinateur

Voici un extrait de ce peut donner Sage. Nous lui donnons la fonction

$$f(x) = \frac{x+4}{3x^2+10x-8}. \tag{11.70}$$

Cette fonction est faite exprès pour que le dénominateur s'annule en -4. En fait $3x^2+10x-8 = (x+4)(3x-2)$, et la fraction peut se simplifier en

$$f(x) = \frac{1}{3x-2}. \tag{11.71}$$

Et avec cela nous écririons $f(-4) = -\frac{1}{14}$. Voyons comment cela passe dans Sage.

```
----------------------------------------------------------------------
| Sage Version 5.2, Release Date: 2012-07-25                         |
| Type "notebook()" for the browser-based notebook interface.        |
| Type "help()" for help.                                            |
----------------------------------------------------------------------
sage: f(x)=(x+4)/(3*x**2+10*x-8)
sage: f(-4)
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
ValueError: power::eval(): division by zero
```

Il produit donc une erreur de division par zéro. Cela n'est pas étonnant. Pourtant si on lui demande, il est capable de simplifier. En effet :

```
sage: f.simplify_full()
x |--> 1/(3*x - 2)
sage: f.simplify_full()(-4)
-1/14
```

Nous considérons la question suivante : étant donné une fonction f définie sur $I\backslash\{x_0\}$, est-il possible de définir f en x_0 de telles façon à ce qu'elle soit continue ?

Exemple 11.55.
La fonction

$$\begin{aligned} f\colon \mathbb{R}\backslash\{0\} &\to \mathbb{R} \\ x &\mapsto \frac{1}{x} \end{aligned} \tag{11.72}$$

n'est pas définie pour $x = 0$ et il n'y a pas moyen de définir $f(0)$ de telle sorte que f soit continue parce que $\lim_{x\to 0} \frac{1}{x}$ n'existe pas. △

11.5.5.2 Limite et prolongement

Reprenons l'exemple de la fonction (11.70) que mon ordinateur refusait de calculer en zéro :

$$f(x) = \frac{x+4}{3x^2+10x-8} = \frac{x+4}{(x+4)\left(x-\frac{2}{3}\right)}. \tag{11.73}$$

Cette fonction a une condition d'existence en $x = -4$. Et pourtant, tant que $x \neq 4$, cela a un sens de simplifier les $(x+4)$ et d'écrire

$$f(x) = \frac{1}{x-\frac{2}{3}} = \frac{3}{3x-2}.$$

Étant donné que pour toute valeur de x différente de -4, la fonction f s'exprime de cette façon, nous avons que

$$\lim_{x\to -4} f(x) = \lim_{x\to -4}\left(\frac{3}{3x-2}\right).$$

Oui, mais la fonction[9] $g(x) = 3/(3x-2)$ est continue en -4 et donc sa limite vaut sa valeur. Nous en déduisons que

$$\lim_{x\to -4} f(x) = -\frac{3}{14}.$$

Que dire maintenant de la fonction ainsi définie ?

$$\tilde{f}(x) = \begin{cases} f(x) & \text{si } x \neq -4 \\ -3/14 & \text{si } x = -4. \end{cases} \tag{11.74}$$

Cette fonction est continue en -4 parce qu'elle y est égale à sa limite. Les étapes suivies pour obtenir ce résultat sont :

— Repérer un point où la fonction n'existe pas,

— calculer la limite de la fonction en ce point, et en particulier vérifier que cette limite existe, ce qui n'est pas toujours le cas,

— définir une nouvelle fonction qui vaut partout la même chose que la fonction originale, sauf au point considéré où l'on met la valeur de la limite.

C'est ce qu'on appelle **prolonger la fonction par continuité** parce que la fonction résultante est continue. La prolongation de f par continuité est donc en général définie par

$$\tilde{f}(x) = \begin{cases} f(x) & \text{si } f(x) \\ \lim_{y\to x} f(y) & \text{si } f(x) \end{cases} \tag{11.75}$$

Dans le cas que nous regardions,

$$f(x) = \frac{x+4}{3x^2+10x-8},$$

le prolongement par continuité est donné par

$$\tilde{f} = \frac{3}{3x-2}. \tag{11.76}$$

Remarquons que cette fonction n'est toujours pas définie en $x = 2/3$.

11.5.6 Prolongement par continuité

Proposition-définition 11.56 (Prolongement par continuité)**.**
Soit $f\colon I\backslash\{x_0\} \to \mathbb{R}$ telle que $\lim_{x\to x_0} f(x) = \ell \in \mathbb{R}$. La fonction

$$\begin{aligned} &\tilde{f}\colon I \to \mathbb{R} \\ &\tilde{f}(x) = \begin{cases} f(x) & \textit{si } x \neq x_0 \\ \ell & \textit{si } x = x_0 \end{cases} \end{aligned} \tag{11.77}$$

*est une fonction continue sur I et est appelée le **prolongement par continuité** de f en x_0.*

Vous noterez que dans cet énoncé nous demandons $\ell \in \mathbb{R}$. Les cas $\ell = \pm\infty$ sont donc exclus.

11.57.
Le lemme 11.72 donnera un autre gros morceau de prolongement par continuité. Là, ce ne sera pas juste une valeur qui manquera, mais carrément la majorité des valeurs ; mais par contre, ce ne sera pas vraiment de la prolongation par continuité, mais de la prolongation par Cauchy-continuité.

9. Cette fonction g n'est pas f parce que g a en plus l'avantage d'être définie en -4.

Exemple 11.58.
La fonction

$$\begin{aligned} f\colon \mathbb{R}\backslash\{-3,2\} &\to \mathbb{R} \\ x &\mapsto \frac{x^2+2x-3}{(x+3)(x-2)} \end{aligned} \tag{11.78}$$

admet pour limite $\lim_{x\to -3} f(x) = \frac{4}{5}$. Son prolongement par continuité en $x=-3$ est donné par

$$\tilde{f}(x) = \frac{x-1}{x-2}. \tag{11.79}$$

Notons que les fonctions f et $\tilde{f}$ ne sont pas identiques : l'une est définie pour $x=-3$ et l'autre pas. Lorsqu'on fait le calcul

$$\frac{x^2+2x-3}{(x+3)(x-2)} = \frac{(x-1)(x+3)}{(x+3)(x-2)} = \frac{x-1}{x-2}, \tag{11.80}$$

la simplification n'est pas du tout un acte anodin. Le dernier signe « = » est discutable parce que les deux dernières expressions ne sont pas égales pour tout x ; elles ne sont égales « que » pour les x pour lesquels les deux expressions existent. △

Les fonctions trigonométriques donneront quelques exemples intéressants de prolongements par continuité. Voir l'exemple 17.207. Et une avec la fonction logarithme dans l'exemple 14.83.

11.5.7 Théorème de la bijection

Proposition 11.59.
Une fonction monotone et surjective d'un intervalle I sur un autre intervalle J est continue sur I.

Proposition 11.60.
Soient $f\colon I\to J$ une bijection et $f^{-1}\colon J\to I$ sa réciproque. Alors pour tout $x_0\in I$ nous avons

$$f^{-1}\big(f(x_0)\big) = x_0 \tag{11.81}$$

et pour tout $y_0\in J$ nous avons

$$f\big(f^{-1}(y_0)\big) = y_0. \tag{11.82}$$

Démonstration. Nous prouvons la relation (11.81) et nous laissons (11.82) comme exercice au lecteur.

Soit $x_0\in I$. Posons $y_0=f(x_0)$. La définition de l'application réciproque est que pour $y\in J$, $f^{-1}(y)$ est l'unique élément x de I tel que $f(x)=y$. Donc $f^{-1}(y_0)$ est l'unique élément de I dont l'image est y_0. C'est donc x_0 et nous avons $f^{-1}(y_0)=x_0$, c'est-à-dire

$$f^{-1}\big(f(x_0)\big) = x_0. \tag{11.83}$$

□

Théorème 11.61 (Théorème de la bijection)**.**
Soit I un intervalle et f une fonction continue et strictement monotone de I dans $\mathbb{R}$. Nous avons alors :

(1) $f(I)$ est un intervalle de $\mathbb{R}$;
(2) La fonction $f\colon I\to f(I)$ est bijective
(3) La fonction $f^{-1}\colon f(I)\to I$ est strictement monotone de même sens que f ;
(4) La fonction $f\colon I\to f(I)$ est un homéomorphisme, c'est-à-dire que $f^{-1}\colon f(I)\to I$ est continue.

Démonstration. Prouvons les choses point par point.

(1) Supposons pour fixer les idées que f est monotone croissante[10].
Soient $a < b$ dans $f(I)$. Par définition il existe $x_1, x_2 \in I$ tels que $a = f(x_1)$ et $b = f(x_2)$. La fonction f est continue sur l'intervalle $[x_1, x_2]$ et vérifie $f(x_1) < f(x_2)$. Donc le théorème des valeurs intermédiaires 11.50 nous dit que pour tout t dans $[f(x_2), f(x_2)]$, il existe un $x_0 \in [x_1, x_2]$ tel que $f(x_0) = t$. Cela montre que toutes les valeurs intermédiaires entre a et b sont atteintes par f et donc que $f(I)$ est un intervalle.

(2) Nous prouvons maintenant que f est bijective en prouvant séparément qu'elle est surjective et injective.

f est surjective Une fonction est toujours surjective depuis un intervalle I vers l'ensemble $\Im f$.

f est injective Soit $x \neq y$ dans I ; pour fixer les idées nous supposons que $x < y$. La stricte monotonie de f implique que $f(x) < f(y)$ ou que $f(x) > f(y)$. Dans tous les cas $f(x) \neq f(y)$.

La fonction f est donc bijective.

(3) Comme d'accoutumée nous supposons que f est croissante. Soient $y_1 < y_2$ dans $f(I)$; nous devons prouver que $f^{-1}(y_1) \leqslant f^{-1}(y_2)$. Pour cela nous considérons les nombres $x_1, x_2 \in I$ tels que $f(x_1) = y_1$ et $f(x_2) = y_2$. Nous allons en prouver la contraposée en supposant que $f^{-1}(y_1) > f^{-1}(y_2)$. En appliquant f (qui est croissante) à cette dernière inégalité il vient

$$f\big(f^{-1}(y_1)\big) \geqslant f\big(f^{-1}(y_2)\big), \tag{11.84}$$

ce qui signifie

$$y_1 \geqslant y_2 \tag{11.85}$$

par l'équation (11.82).

(4) La fonction $f^{-1}\colon f(I) \to I$ est une fonction monotone et surjective, donc continue par la proposition 11.59.

□

Exemple 11.62.
La fonction

$$\begin{aligned} f\colon [2,3] &\to [4,9] \\ x &\mapsto x^2 \end{aligned} \tag{11.86}$$

est une bijection. Sa réciproque est la fonction

$$\begin{aligned} f^{-1}\colon [4,9] &\to [2,3] \\ x &\mapsto \sqrt{x}. \end{aligned} \tag{11.87}$$

△

11.6 Limite et continuité

Voir les remarques dans l'index thématique 17 pour comprendre la place et la portée de ce qui va venir à propos de limite et de continuité.

Théorème 11.63 (Limite et continuité)**.**
La fonction f est continue au point a si et seulement si $\lim_{x\to a} f(x) = f(a)$.

Démonstration. Nous commençons par supposer que f est continue en a, et nous prouvons que $\lim_{x\to a} f(x) = a$. Soit $\epsilon > 0$; ce qu'il nous faut c'est un δ tel que $|x-a| \leqslant \delta$ implique $|f(x)-f(a)| \leqslant \epsilon$. La caractérisation 11.42 de la continuité donne l'existence d'un δ comme il nous faut.

Dans l'autre sens, c'est-à-dire prouver que f est continue au point a sous l'hypothèse que $\lim_{x\to a} f(x) = f(a)$, la preuve se fait de la même façon. □

10. Traitez en tant qu'exercice le cas où f est décroissante.

Nous en déduisons que si nous voulons gagner quelque chose à parler de limites, il faut prendre des fonctions non continues. En effet, si une fonction est continue en un point, la limite ne donne aucune nouvelle information que la valeur de la fonction elle-même en ce point.

Prenons une fonction qui fait un saut. Pour se fixer les idées, prenons celle-ci :

$$f(x) = \begin{cases} 2x & \text{si } x \in]\infty, 2[\\ x/2 & \text{si } x \in [2, \infty[\end{cases} \tag{11.88}$$

Essayons de trouver la limite de cette fonction lorsque x tend vers 2. Étant donné que f n'est pas continue en 2, nous savons déjà que $\lim_{x\to 2} f(x) \neq f(2)$. Donc ce n'est pas 1. Cette limite ne peut pas valoir 4 non plus parce que si je prends n'importe quel ϵ, la valeur de $f(2+\epsilon)$ est très proche de 2, et donc ne peut pas s'approcher de 4. En fait, tu peux facilement vérifier que *aucun nombre ne vérifie la condition de limite pour f en* 2. Nous disons que la limite n'existe pas.

Il ne faudrait pas en déduire trop vite que si une fonction n'est pas continue en a, alors la limite $x \to a$ n'existe pas. Ce que dit le théorème 11.63 est que si une fonction n'est pas continue en a, alors sa limite (si elle existe) ne vaut pas $f(a)$.

Exemple 11.64 (Un exemple de continuité Thème 17)**.**
Soit la fonction

$$f(x) = \begin{cases} x & \text{si } x \neq 0 \\ 4 & \text{si } x = 0. \end{cases} \tag{11.89}$$

Cette fonction n'est pas continue en $x = 0$, et pourtant la limite existe : $\lim_{x\to 0} f(x) = 0$. Faisons cela en détail pour nous assurer de ce qu'il se passe.

Considérons l'ouvert $]3, 5[$. L'image réciproque de cet ouvert par f est la partie $]3, 5[\cup \{0\}$ qui n'est pas ouvert. Donc la fonction f n'est pas continue comme fonction $\mathbb{R} \to \mathbb{R}$.

Considérons pour comprendre la restriction $f\colon [-1, 1] \to \mathbb{R}$. L'image inverse de $]3, 5[$ par cette fonction est $\{0\}$ qui n'est pas un ouvert.

Plus généralement tant qu'on considère des restrictions de f sur des parties contenant un voisinage de 0, la fonction ne peut pas être continue [11].

Voyons ce qui en est de la continuité ponctuelle de f en $x = 0$. La définition 7.71 est celle de la continuité en un point ; elle dit que f sera continue en 0 si $f(0) = 4$ est une limite de f. Nous voila parti vers la définition 7.67.

Soit le voisinage $V =]3, 5[$ de $f(0)$. Quel que soit le voisinage W de 0 dans $\mathbb{R}$, il existe un $\epsilon > 0$ tel que $W \subset B(0, \epsilon)$. Nous avons alors

$$f\big(W\backslash\{a\}\big) \subset f\big(B(0, \epsilon)\backslash\{0\}\big). \tag{11.90}$$

Mais le nombre $\epsilon/2$ fait partie de $f\big(B(0, \epsilon)\backslash\{0\}\big)$ et n'est pas dans V. Donc $f(0)$ n'est pas une limite de f en zéro. Cette fonction n'est donc pas continue en zéro. △

Exemple 11.65 (Même exemple, limite)**.**
Nous avons vu que, pour la fonction (11.89), le nombre 4 n'est pas une limite de f en zéro. Nous montrons à présent que 0 est une limite (et même la seule par la proposition 7.70 que nous ne rappellerons plus à chaque fois) de f.

Montrons que 0 est une limite de f en zéro, c'est-à-dire que $\lim_{x\to 0} f(x) = 0$.

Nous suivant la définition 7.67. Soit un voisinage V de 0 dans $\mathbb{R}$. Il existe δ tel que $B(0, \delta) \subset V$. En posant $\epsilon = \delta$ et en définissant $W = B(0, \epsilon)$ nous avons

$$f\big(B(0, \epsilon)\backslash\{0\}\big) = B(0, \epsilon)\backslash\{0\} \subset B(0, \delta) \subset V. \tag{11.91}$$

Donc 0 est une limite de f en zéro. △

Nous avons déjà vu par le corolaire 7.156 qu'une suite croissante et bornée était convergente. Il en va de même pour les fonctions.

11. Les plus acharnés se demanderont ce qu'il se passe pour la restriction de f à la partie $\{0\}$ munie de la topologie induite de $\mathbb{R}$.

Proposition 11.66 ([1]).
Si la fonction réelle $f\colon I = [a, b[\to \mathbb{R}$ est croissante et bornée, alors la limite

$$\lim_{x\to b} f(x) \tag{11.92}$$

existe et est finie.

Démonstration. Commençons par prouver que si (x_n) est une suite dans I convergent vers b, alors $f(x_n)$ est une suite convergente. Dans un second temps nous allons prouver que si (x_n) et (x'_n) sont deux suites qui convergent vers b, alors les suites convergentes $f(x_n)$ et $f(x'_n)$ convergent vers la même limite. Alors le critère séquentiel de la limite d'une fonction conclura (proposition 7.263).

Nous pouvons extraire de x_n une sous-suite croissante $(x_{\alpha(n)})$. Alors la suite $f\big(x_{\alpha(n)}\big)$ est une suite croissante et majorée, donc convergente par le corolaire 7.156 [12]. Nommons ℓ la limite et montrons qu'elle est aussi limite de f sur la suite originale.

Pour tout $\epsilon > 0$, il existe K tel que si $n > K$ alors $\big|f\big(x_{\alpha(n)}\big) - \ell\big| < \epsilon$. Soit K' tel que pour tout $n > K'$ nous ayons $x_n > x_{\alpha(K')}$. Cela est possible parce que la suite est bornée par b et converge vers b : il suffit de prendre K' de telle sorte que $|x_n - b| \leqslant |x_{\alpha(n)} - b|$. Si $n > K'$ alors $x_n > x_{\alpha(K)}$ et

$$f(x_n) \geqslant f(x_{\alpha(n)}) \geqslant \ell - \epsilon; \tag{11.93}$$

en résumé si $n > K$ alors $|f(x_n) - \ell| < \epsilon$. Cela prouve que $f(x_n) \to \ell$.

Soit maintenant une autre suite (x'_n) qui converge également vers b. Comme nous venons de le voir la suite $f(x'_n)$ est convergente et nous nommons ℓ' la limite. Si nous considérons (x''_n) la suite « alternée » $(x_1, x'_1, x_2, x'_2, \cdots)$ alors nous avons encore une suite qui converge vers b et donc $f(x''_n) \to \ell'$.

Mais étant donné que $f(x_n)$ et $f(x'_n)$ sont des sous-suites, elles doivent converger vers la même valeur. Donc $\ell = \ell' = \ell''$. □

11.6.1 Règles simples de calcul

Les opérations simples passent à la limite, sauf la division pour laquelle il faut faire attention au dénominateur.

Proposition 11.67.
Soient f et g deux fonctions telles que $\lim_{x\to a} f(x) = \alpha$ et $\lim_{x\to a} g(x) = \beta$. Alors

(1) $\lim_{x\to a} f(x) + g(x) = \alpha + \beta$,

(2) $\lim_{x\to a} f(x)g(x) = \alpha\beta$,

(3) s'il existe un voisinage de a sur lequel g ne s'annule pas, alors $\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\alpha}{\beta}$.

Le résultat suivant est pratique pour le calcul des limites.

Proposition 11.68.
Quand la limite existe, nous avons

$$\lim_{x\to a} f(x) = \lim_{\epsilon\to 0} f(a + \epsilon),$$

ce qui correspond à un « changement de variables » dans la limite.

Démonstration. Si $A = \lim_{x\to a} f(x)$, par définition,

$$\forall \epsilon' > 0,\ \exists \delta \text{ tel que } |x - a| \leqslant \delta \Rightarrow |f(x) - A| \leqslant \epsilon'. \tag{11.94}$$

La seule subtilité de la démonstration est de remarquer que si $|x - a| \leqslant \delta$, alors x peut être écrit sous la forme $x = a + \epsilon$ pour un certain $|\epsilon| \leqslant \delta$. En remplaçant x par $a + \epsilon$ dans la condition 11.94, nous trouvons

$$\forall \epsilon' > 0,\ \exists \delta \text{ tel que } |\epsilon| \leqslant \delta \Rightarrow |f(x + \epsilon) - A| \leqslant \epsilon', \tag{11.95}$$

12. En gros nous sommes en train de dire que toute la théorie des fonctions convexes est un vulgaire corolaire de Bolzano-Weierstrass.

ce qui signifie exactement que $\lim_{\epsilon\to 0} f(x+\epsilon) = A$. □

Il y a une petite différence de point de vue entre $\lim_{x\to a} f(x)$ et $\lim_{\epsilon\to 0} f(a+\epsilon)$. Dans le premier cas, on considère $f(x)$, et on regarde ce qu'il se passe quand x se rapproche de a, tandis que dans le second, on considère $f(a)$, et on regarde ce qu'il se passe quand on s'éloigne un tout petit peu de a. Dans un cas, on s'approche très près de a, et dans l'autre on s'en éloigne un tout petit peu. Le contenu de la proposition 11.68 est de dire que ces deux points de vue sont équivalents.

Proposition 11.69 ([1]).
L'ensemble $\mathbb{R}\cup\{\infty\}$ muni de la topologie de la compactification en un point[13] *est connexe par arcs.*

Démonstration. Nous allons montrer que le chemin

$$\begin{aligned}\gamma\colon [0,1] &\to [0,\infty]\\ x &\mapsto \begin{cases}\frac{1}{x} & \text{si } x\neq 0\\ \infty & \text{si } x=0\end{cases}\end{aligned} \tag{11.96}$$

est continu au sens de la définition 7.71(2) (qui est le seul sens possible au mot « continu »).

Soit un ouvert $\mathcal{O}$ dans $\mathbb{R}\cup\{\infty\}$. Si cet ouvert ne contient pas ∞, alors $\gamma^{-1}(\mathcal{O})$ est ouvert dans $\mathbb{R}$ parce que la fonction $x\mapsto 1/x$ est continue[14].

Si $\infty\in\mathcal{O}$, alors $\mathcal{O}=\{\infty\}\cup\mathcal{O}'$ où $\mathcal{O}'$ est un ouvert de $\mathbb{R}$ ayant la propriété que $\mathbb{R}\backslash\mathcal{O}'$ est compact.

Nous avons $\gamma^{-1}(\mathcal{O})=\{0\}\cup\gamma^{-1}(\mathcal{O}')$. Le fait est que $\gamma^{-1}(\mathcal{O}')$ est ouvert de $\mathbb{R}$ contenu dans $[0,1]$.

Vu que le complémentaire de $\mathcal{O}'$ est compact, il existe $a\in\mathbb{R}$ tel que $]a,\infty[\subset\mathcal{O}'$. Donc $\mathcal{O}'=]a,\infty[\cup\mathcal{O}''$ où $\mathcal{O}''$ est un ouvert.

Nous avons :

$$\gamma^{-1}(\mathcal{O}) = \{\gamma^{-1}(\infty)\}\cup\gamma^{-1}\big(]a,\infty[\big)\cup\gamma^{-1}(\mathcal{O}'') \tag{11.97a}$$

$$= \{0\}\cup\,]0,\frac{1}{a}[\,\cup\gamma^{-1}(\mathcal{O}'') \tag{11.97b}$$

$$= [0,\frac{1}{a}[\,\cup\gamma^{-1}(\mathcal{O}''). \tag{11.97c}$$

Vous noterez le point essentiel où la topologie de la compactification agit : vu que $\{0\}$ n'est pas un ouvert de $[0,1]$, nous devons nous assurer que la partie $\gamma^{-1}(\mathcal{O}')$ vienne se coller à $\{0\}$ pour compléter en un ouvert.

L'ensemble $\gamma^{-1}(\mathcal{O}'')$ est un ouvert de $\mathbb{R}$ contenu dans $[0,1]$. Nous avons donc

$$\gamma^{-1}(\mathcal{O}) = \Big(]-1,\frac{1}{a}[\,\cup\gamma^{-1}(\mathcal{O}'')\Big)\cap[0,1]. \tag{11.98}$$

Cela est un ouvert de $[0,1]$ par définition de la topologie induite[15]. □

11.6.2 Prolongement des rationnels vers les réels

Si $f\colon\mathbb{Q}\to\mathbb{R}$ est une fonction continue pour la topologie induite, est-ce qu'on peut la prolonger en une fonction continue sur $\mathbb{R}$? La réponse est hélas non.

13. Définition 7.66.
14. Voir par exemple la proposition 11.67.
15. Définition 7.15.

Exemple 11.70 ([231]).
Vu que $\sqrt{2}$ est irrationnel[16], ceci définit bien une fonction sur $\mathbb{Q}$:

$$\begin{aligned} f\colon \mathbb{Q} &\to \mathbb{R} \\ q &\mapsto \begin{cases} 0 & \text{si } q < \sqrt{2} \\ 1 & \text{si } q > \sqrt{2}. \end{cases} \end{aligned} \tag{11.99}$$

Cela est une fonction continue sur $\mathbb{Q}$. En effet, soient $q \in \mathbb{Q}$ et $\epsilon > 0$. Nous prenons $\delta > 0$ tel que $\sqrt{2}$ ne soit pas dans $B(q, \delta)$. Alors si $p \in B_{\mathbb{Q}}(q, \delta)$ nous avons $f(q) = f(p)$ et donc

$$|f(p) - f(q)| < \epsilon. \tag{11.100}$$

Il n'est cependant pas possible de la prolonger en une fonction continue sur $\mathbb{R}$. △

Pour qu'une fonction sur $\mathbb{Q}$ puisse être prolongée en une fonction continue sur $\mathbb{R}$, il faut un peu plus que la continuité. Il fait la Cauchy-continuité que nous définissons pas plus tard qu'immédiatement.

Définition 11.71 ([231]).
*Soient X et Y deux espaces métriques. Une application $f\colon X \to Y$ est dite **Cauchy-continue** si pour toute suite de Cauchy (x_n) dans X, la suite $\big(f(x_n)\big)$ est de Cauchy dans Y.*

En terme de prolongement continu, nous avons ce lemme qui demande à une fonction d'être Cauchy continue. Vous pouvez comparer avec le principe de prolongement analytique 16.128 qui donne un énoncé similaire pour un prolongement analytique.

Lemme 11.72 ([1, 232, 233]).
Soit une fonction Cauchy continue[17] $f\colon \mathbb{Q} \to \mathbb{R}$.

(1) La limite $\lim_{q\to x} f(q)$ existe pour tout $x \in \mathbb{R}$.

(2) Il existe un unique prolongement continu $\tilde{f}\colon \mathbb{R} \to \mathbb{R}$.

(3) Ce prolongement est donné par

$$\tilde{f}(x) = \begin{cases} f(x) & \text{si } x \in \mathbb{Q} \\ \lim_{q\to x} f(q) & \text{sinon} \end{cases} \tag{11.101}$$

Démonstration. Imprégniez vous bien de la la définition 7.67 de la limite avant de commencer.

Unicité Prouvons rapidement l'unicité avant l'existence parce que c'est facile.

L'unicité du prolongement est la proposition 7.269 à propos de fonctions continues égales sur une partie denses. La densité de $\mathbb{Q}$ dans $\mathbb{R}$, si vous la cherchez est la proposition 7.121.

Candidat limite Soit $x \in \mathbb{R}$. Vu que $x \in \bar{\mathbb{Q}}$, nous pouvons chercher à savoir si $\lim_{q\to x} f(q)$ existe. Si elle existe, elle sera unique.

Soit une suite (q_i) d'éléments de $\mathbb{Q}$ qui converge vers x dans $\mathbb{R}$ (i.e. pour la topologie de $\mathbb{R}$). Les nombres réels $f(q_i)$ forment une suite dans $\mathbb{R}$. La suite (q_i) étant convergente, elle est de Cauchy[18].

Vu que f est supposée Cauchy-continue, la suite $\big(f(q_i)\big)$ est de Cauchy dans $\mathbb{R}$, et elle est donc convergente.

C'est bien la limite Nous prouvons à présent que le nombre réel $\lim_{i\to\infty} f(q_i)$ (dont l'existence vient d'être prouvée) vérifie bien la définition de la limite $\lim_{q\to x} f(q)$.

Soit un voisinage V de $\lim f(q_i)$ dans $\mathbb{R}$. Nous devons trouver un voisinage W de x dans $\mathbb{R}$ tel que

$$f\big(W \cap \mathbb{Q}\backslash\{x\}\big) \subset V. \tag{11.102}$$

16. Proposition 1.123. Le fait que $\sqrt{2}$ existe dans $\mathbb{R}$ est la proposition 1.166.
17. Définition 11.71 ; nous en avons discuté dans l'exemple 11.70.
18. Théorème 7.294(2).

Pour cela nous considérons $\epsilon > 0$ tel que $B\big(\lim f(q_i), \epsilon\big) \subset V$. Vu que f est continue sur $\mathbb{Q}$, il existe δ tel que

$$|p - q| < 2\delta \Rightarrow |f(p) - f(q)| < \epsilon. \tag{11.103}$$

Nous posons $W = B(x, \delta)$.

Soit $q \in W \cap \mathbb{Q}\backslash\{x\}$. Nous nous proposons de majorer la quantité $|f(q) - \lim f(q_i)|$ par un multiple de ϵ.

Pour cela nous considérons k suffisamment grand pour que $|f(q_k) - \lim f(q_i)| < \epsilon$. Et de plus, vu que $q_i \to x$ nous considérons k suffisamment grand pour que $|q_k - x| < \delta$. L'indice k est choisi pour vérifier les deux conditions en même temps.

Nous écrivons alors la majoration suivante :

$$|f(q) - \lim f(q_i)| \leqslant |f(q) - f(q_k)| + |f(q_k) + \lim f(q_i)|. \tag{11.104}$$

Le second terme est majoré par ϵ. Pour le premier terme, $q \in B(x, \delta)$ et $q_k \in B(x, \delta)$, donc $|q - q_k| \leqslant 2\delta$, ce qui implique $|f(q) - f(q_k)| < \epsilon$.

Au final, $|f(q) - \lim f(q_i)| \leqslant 2\epsilon$. En reprenant tout le travail avec $\epsilon/2$ au lieu de ϵ nous trouvons $f(q) \in B\big(\lim f(q_i), \epsilon\big) \subset V$.

Intermède Jusqu'à présent, nous avons prouvé que

$$\lim_{q \to x} f(q) \tag{11.105}$$

existe et vaut

$$\lim f(q_i) \tag{11.106}$$

lorsque (q_i) est une suite quelconque de rationnels qui converge vers x. Nous l'écrivons pour la référentier plus tard :

$$\lim_{q \to x} f(q) = \lim f(q_i). \tag{11.107}$$

La limite (11.105) est une limit de fonction définie sur $\mathbb{Q} \subset \mathbb{R}$ en un pour adhérent à l'ensemble de définition de f. La limite (11.106) est une limite usuelle d'une suite dans $\mathbb{R}$.

Le prolongement Nous posons

$$\tilde{f}(x) = \begin{cases} f(x) & \text{si } x \in \mathbb{Q} \\ \lim_{q \to x} f(q) & \text{sinon} \end{cases} \tag{11.108}$$

et nous allons prouver que $\tilde{f}$ est une fonction continue sur $\mathbb{R}$.

Continuité Soit $a \in \mathbb{R}$; nous allons montrer la continuité de $\tilde{f}$ en a. Nous fixonx bien entendu $\epsilon > 0$, et nous nous acharnons à majorer la quantité $|\tilde{f}(x) - \tilde{f}(a)|$.

Vu que f est continue sur $\mathbb{Q}$ nous considérons δ' tel que (dans $\mathbb{Q}$) $0 < |q - q'| < \delta'$ implique $|f(q) - f(q')| < \epsilon$.

$a \in \mathbb{Q}$, $x \in \mathbb{Q}$ Alors $\tilde{f}(x) = f(x)$ et $\tilde{f}(a) = f(a)$. Par la continuité de f sur $\mathbb{Q}$, il existe un δ tel que $0 < |x - a| < \delta$ implique $|f(x) - f(a)| < \epsilon$.

$a \in \mathbb{Q}$, x irrationnel Nous considérons une suite de rationnels $q_k \to x$ (vous penserez à l'utilisation du lemme 1.150). Nous avons la majoration

$$|\tilde{f}(x) - \tilde{f}(a)| = |\lim_{q \to x} f(q) - f(a)| \leqslant |\lim_{q \to x} f(q) - f(q_k)| + |f(q_k) - f(a)|. \tag{11.109}$$

Nous considérons $\delta < \delta'$ et k suffisament grand pour que $|q_k - x| < \delta' - \delta$. Avec ces choix,

$$|q_k - a| \leqslant |q_k - x| + |x - a| \leqslant \delta'. \tag{11.110}$$

Enfin nous prenons également k suffisament grand pour avoir $|\lim_{q \to x} f(q) - f(q_k)| \leqslant \epsilon$. Les inégalités (11.109) peuvent alors être prolongées pour avoir

$$|\tilde{f}(x) - \tilde{f}(a)| \leqslant 2\epsilon. \tag{11.111}$$

a irrationnel, $x \in \mathbb{Q}$ Nous faisons encore la majoration

$$|\tilde{f}(x) - \tilde{f}(a)| = |f(x) - \lim_{q\to a} f(q)| \leqslant |f(x) - f(q_k)| + |f(q_k) - \lim_{q\to a} f(a)|. \tag{11.112}$$

Nous prenons $\delta < \delta'/2$ et nous choisissons k assez grand pour que $|q_k - a| < \delta'/2$. De ces choix il ressort que

$$|q_k - x| \leqslant |q_k - a| + |a - x| \leqslant \frac{\delta'}{2} + \frac{\delta'}{2} \leqslant \delta'. \tag{11.113}$$

Donc $|f(x) - f(q_k)| < \epsilon$. De plus, pour k assez grand, $|f(q_k) - \lim_{q\to a} f(q)| \leqslant \epsilon$.

a et x irrationnels Nous avons

$$|\tilde{f}(x) - \tilde{f}(a)| = |\lim_{q\to x} f(q) - \lim_{r\to a} f(r)|, \tag{11.114}$$

et nous considérons des suites de rationnels $q_k \to x$ et $r_i \to a$. De plus nous considérons $\delta < \delta'/4$, et k, i suffisament grands pour avoir $|q_k - x| \leqslant \delta'/4$ et $|r_i - a| < \delta'/4$. Avec tout cela nous avons

$$|q_k - r_i| \leqslant |q_k - x| + |x - a| + |a - r_i| \leqslant 3\delta'/4 < \delta'. \tag{11.115}$$

Enfin, en choisissant i et k de telle sorte à avoir $|\lim_{q\to x} f(q) - f(q_k)| \leqslant \epsilon$ et $|f(r_i) - \lim_{r\to a} f(r)| < \epsilon$ nous avons les majorations

$$|\tilde{f}(x) - \tilde{f}(a)| = |\lim_{q\to x} f(q) - \lim_{r\to a} f(r)| \tag{11.116a}$$

$$\leqslant |\lim_{q\to x} f(x) - f(q_k)| + |f(q_k) - f(r_i)| + |f(r_i) - \lim_{r\to q} f(r)| \tag{11.116b}$$

$$\leqslant 3\epsilon. \tag{11.116c}$$

□

Proposition 11.73.
Soient des fonctions continues $f, g\colon \mathbb{R} \to \mathbb{R}$. Si f et g sont égales sur $\mathbb{Q}$, alors elles sont égales sur $\mathbb{R}$.

Démonstration. Sinon nous pouvons utiliser les propriétés fondamentales des réels et de la continuité. Soit $x \in \mathbb{R}$; nous voulons montrer que $f(x) = g(x)$. En prenant par exemple le lemme 1.150, il existe une suite q_i de rationnels telle que $q_i \xrightarrow{\mathbb{R}} x$.

Par ailleurs, f et g sont continues sur $\mathbb{R}$ et donc en chaque point de $\mathbb{R}$ (théorème 7.74). Par la caractérisation séquentielle 7.80 de la continuité, nous avons

$$f(x) = \lim_{i\to\infty} f(q_i) = \lim_{i\to\infty} g(q_i) = g(x). \tag{11.117}$$

□

Proposition 11.74 ([1]).
Soit une fonction strictement croissante $f\colon \mathbb{Q} \to \mathbb{R}$. Alors la prolongation continue $\tilde{f}\colon \mathbb{R} \to \mathbb{R}$ est également strictement croissante.

Démonstration. Soient $x, y \in \mathbb{R}$ avec $x < y$. Notons $d = y - x$. Nous considérons des suites de rationnels $x_k \to x$ et $y_l \to y$ telles que pour tout k, $x_k \in B(x, d/3)$ et $y_k \in B(y, d/3)$. En particulier, $x_k < y_l$ pour tout k et l.

Soient des rationnels q et q' tels que pour tout k,

$$x_k < q < q' < y_k. \tag{11.118}$$

Pour trouver de tels rationnels, il suffit de les chercher dans $]x + \frac{d}{3}, y - \frac{d}{3}[$. Cet intervalle étant de longueur $d/3$, il contient des rationnels.

Vue la croissance de f sur $\mathbb{Q}$, nous avons, pour tout k :

$$f(x_k) < f(q) < f(q') < f(y_k), \tag{11.119}$$

et à la limite :

$$\tilde{f}(x) \leqslant f(q) < f(q') \leqslant \tilde{f}(y). \tag{11.120}$$

Notez que les inégalités strictes se changent en inégalités larges au passage à la limite. D'où l'utilisé de prendre *deux* rationnels entre x_k et y_k pour maintenir une inégalité stricte entre $\tilde{f}(x)$ et $\tilde{f}(y)$. □

11.7 Espace des fonctions continues

Définition 11.75.
*Soit I, un intervalle de $\mathbb{R}$. L'**oscillation** sur I est le nombre*

$$\omega_f(I) = \sup_{x\in I} f(x) - \inf_{x\in I} f(x). \tag{11.121}$$

Pour chaque x fixé, la fonction

$$x \mapsto \omega_f\big(B(x,\delta)\big) \tag{11.122}$$

est une fonction positive, croissante et a donc une limite (pour $\delta \to 0$). Nous notons $\omega_f(x)$ cette limite qui est l'**oscillation** de f en ce point. Une propriété immédiate est que f est continue en x_0 si et seulement si $\omega_f(x_0) = 0$.

Lemme 11.76.
L'ensemble des points de discontinuité d'une fonction $f\colon \mathbb{R} \to \mathbb{R}$ est une réunion dénombrable de fermés.

Démonstration. Soit D l'ensemble des points de discontinuité de f. Nous avons

$$D = \bigcup_{n=1}^{\infty} \{x \text{ tel que } \omega_f(x) \geqslant \frac{1}{n}\}. \tag{11.123}$$

Il nous suffit donc de montrer que pour tout ϵ, l'ensemble

$$\{x \text{ tel que } \omega_f(x) < \epsilon\} \tag{11.124}$$

est ouvert. Soit en effet x_0 dans cet ensemble. Il existe δ tel que $\omega_f\big(B(x_0,\delta)\big) < \epsilon$. Si $x \in B(x_0,\delta)$, alors si on choisit δ' tel que $B(x,\delta') \subset B(x_0,\delta)$, nous avons $\omega_f\big(B(x,\delta')\big) < \epsilon$, ce qui justifie que $\omega_f(x) < \epsilon$ et donc que x est également dans l'ensemble considéré. □

Théorème 11.77.
L'ensemble des points de discontinuité d'une limite simple de fonctions continues est de première catégorie.

Démonstration. Soit (f_n) une suite de fonctions qui converge simplement vers f. Nous devons écrire l'ensemble des points de discontinuité de f comme une union dénombrable d'ensembles tels que sur tout intervalle I, aucun de ces ensembles n'est dense. Nous savons déjà par le lemme 11.76 que l'ensemble des points de discontinuité de f est donné par

$$D = \bigcup_{n=1}^{\infty} \{x \text{ tel que } \omega_f(x) \geqslant \frac{1}{n}\}. \tag{11.125}$$

Nous essayons donc de prouver que pour tout ϵ, l'ensemble

$$F = \{x \text{ tel que } \omega_f(x) \geqslant \epsilon\} \tag{11.126}$$

est nulle part dense. Soit

$$E_n = \bigcap_{i,j>n} \{x \text{ tel que } |f_i(x) - f_j(x)| < \epsilon\}. \tag{11.127}$$

Nous montrons que cet ensemble est fermé en étudiant le complémentaire. Soit $x \notin E_n$; alors il existe un couple (i, j) tel que

$$|f_i(x) - f_j(x)| > \epsilon. \tag{11.128}$$

Par continuité, cette inégalité reste valide dans un voisinage de x. Donc il existe un voisinage de x contenu dans $\complement E_n$ et E_n est donc fermé.

De plus nous avons $E_n \subset E_{n+1}$ et $\bigcup_n E_n = \mathbb{R}$. Ce dernier point est dû au fait que pour tout x, il existe N tel que $i, j > N$ implique $|f_i(x) - f_j(x)| \leqslant \epsilon$. Cela est l'expression du fait que la suite $\big(f_n(x)\big)_{n\in\mathbb{N}}$ est de Cauchy.

Soit I, un intervalle fermé de $\mathbb{R}$. Nous voulons trouver un intervalle $J \subset I$ sur lequel f est continue. Nous écrivons I sous la forme

$$I = \bigcup_{n=1}^{\infty} (E_n \cap I). \tag{11.129}$$

Tous les ensembles $J_n = E_n \cap I$ ne peuvent être nulle part dense en même temps (à cause du théorème de Baire 7.325). Il existe donc un n tel que J_n contienne un ouvert J. Le but est de montrer que f est continue sur J. Pour ce faire, nous n'allons pas simplement majorer $|f(x)-f(x_0)|$ par ϵ lorsque $|x - x_0|$ est petit. Nous allons majorer l'oscillation de f sur $B(x_0, \delta)$ lorsque δ est petit. Pour cela nous prenons x_0 et x dans J et nous écrivons

$$|f(x) - f(x_0)| \leqslant |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)|. \tag{11.130}$$

À ce niveau nous rappelons que n est fixé par le choix de J, dans lequel ϵ est déjà inclus. Nous choisissons évidemment $|x - x_0| \leqslant \delta$ de telle sorte que le second terme soit plus petit que ϵ en vertu de la continuité de f_n. Pour le premier terme, pour tout $i, j \geqslant n$ nous avons

$$|f_i(x) - f_j(x)| < \epsilon. \tag{11.131}$$

Si nous posons $j = n$ et $i \to \infty$, en tenant compte du fait que $f_i \to f$ simplement,

$$|f(x) - f_n(x)| \leqslant \epsilon. \tag{11.132}$$

Nous avons donc obtenu $|f(x)-f_n(x_0)| \leqslant 2\epsilon$. Cela signifie que dans un voisinage de rayon δ autour de x_0, les valeurs extrêmes prises par $f(x)$ sont $f_n(x_0) \pm 4\epsilon$. Nous avons donc prouvé que pour tout ϵ, il existe δ tel que

$$\omega_f\big([x_0 - \delta, x_0 + \delta]\big) \leqslant 4\epsilon. \tag{11.133}$$

De là nous concluons que

$$\lim_{\delta\to 0} \omega_f\big([x_0 - \delta, x_0 + \delta]\big) = 0, \tag{11.134}$$

ce qui signifie que f est continue en x_0. □

Exemple 11.78.
Une fonction discontinue sur $\mathbb{Q}$ et continue ailleurs. La fonction

$$f(x) = \begin{cases} 0 & \text{si } x \notin \mathbb{Q} \\ \frac{1}{q} & \text{si } x = p/q \end{cases} \tag{11.135}$$

où par « $x = p/q$ » nous entendons que p/q est la fraction irréductible.

Cette fonction est discontinue sur $\mathbb{Q}$ parce que si $q \in \mathbb{Q}$ alors $f(q) \neq 0$ alors que dans tous voisinage de q il existe un irrationnel sur qui la fonction vaudra zéro.

Montrons que f est continue sur les irrationnels. Si $x_0 \notin \mathbb{Q}$ alors $f(x_0) = 0$. Mais si on prend un voisinage suffisamment petit de x_0, nous pouvons nous arranger pour que tous les rationnels aient un dénominateur arbitrairement grand. En effet si nous nous fixons un premier rayon $r_0 > 0$ alors il existe un nombre fini de fractions de la forme $1, \frac{k}{2}, \frac{k}{3}, \ldots, \frac{k}{N}$ dans $B(x_0, r_0)$. Il suffit maintenant de choisir $0 < r \leqslant r_0$ tel que ces fractions soient toutes hors de $B(x_0, r)$. Dans cette boule nous avons $f < \frac{1}{N}$. Du coup f est continue en x_0. △

Définition 11.79 (Point périodique[234])**.**
Soit $f\colon I \to I$ une application d'un ensemble I dans lui-même. Si $x \in I$ vérifie $f^n(x) = x$ et $f^k(x) \neq x$ pour $k = 1, \ldots, n-1$ alors on dit que x est un point n-périodique.

Lemme 11.80.
Soit I un segment[19] *de $\mathbb{R}$ et une fonction continue $f\colon I \to I$. Si K est un segment fermé avec $K \subset f(I)$ alors il existe un segment fermé $L \subset I$ tel que $K = f(L)$.*

Démonstration. Mentionnons immédiatement que f est continue sur I qui est compact[20]. Par conséquent tous les nombres dont nous allons parler sont finis parce que f est bornée par le théorème 7.154.

Soit $K = [\alpha, \beta]$. Si $\alpha = \beta$ alors le segment $L = \{a\}$ convient. Nous supposons donc que $\alpha \neq \beta$ et nous considérons $a, b \in I$ tels que $\alpha = f(a)$ et $\beta = f(b)$. Vu que $a \neq b$ nous supposons $a < b$ (le cas $a > b$ se traite de façon similaire).

Nous posons

$$A = \{x \in [a, b] \text{ tel que } f(x) = \alpha\}. \tag{11.136}$$

C'est un ensemble borné par a et b. De plus il est fermé ; ce dernier point n'est pas tout à fait évident parce que f n'est pas définit sur $\mathbb{R}$ mais sur I qui est fermé, le corolaire 11.47 n'est donc pas immédiatement utilisable. Prouvons donc que $Z = \{x \in \mathbb{R} \text{ tel que } f(x) = \alpha\}$ est fermé. Si x_0 est hors de Z alors soit x_0 est dans I soit il est hors de I. Dans ce second cas, le complémentaire de I étant ouvert, on a un voisinage de x_0 hors de I et par conséquent hors de Z. Si au contraire $x_0 \in I$ alors il y a (encore) deux cas : soit $x_0 \in \mathrm{Int}(I)$ soit x_0 est sur le bord de I. Dans le premier cas, le théorème des valeurs intermédiaires[21] fonctionne. Pour le second cas, nous supposons $x_0 = \max(I)$ (le cas $x_0 = \min(I)$ est similaire). Le théorème des valeurs intermédiaires dit que sur $[x_0 - \epsilon, x_0]$, $f \neq \alpha$ et en même temps, sur $]x_0, x_0 + \epsilon]$, nous sommes en dehors du domaine. Au final $\{f(x) = \alpha\}$ est fermé et A est alors fermé en tant que intersection de deux fermés.

L'ensemble A étant non vide ($a \in A$), il possède donc un maximum que nous nommons u :

$$u = \max(A). \tag{11.137}$$

Nous posons aussi

$$B = \{x \in [u, b] \text{ tel que } f(x) = \beta\} \tag{11.138}$$

qui est encore fermé, borné et non vide. Nous pouvons donc définir

$$v = \min(B). \tag{11.139}$$

Nous prouvons maintenant que $f\big([u, v]\big) = [\alpha, \beta]$. D'abord $f\big([u, v]\big)$ est un intervalle compact[22] contenant $f(u) = \alpha$ et $f(v) = \beta$. Par conséquent $[\alpha, \beta] \subset f\big([u, v]\big)$. Pour l'inclusion inverse supposons $t \in [u, v]$ tel que $f(t) > \beta$. Vu que $f(a) = \alpha$ et $\alpha < \beta$ le théorème des valeurs intermédiaires il existe $t_0 \in [a, t]$ tel que $f(t_0) = \beta$. Cela donne $t_0 < v$ et donc contredit la minimalité de v dans B. Nous en déduisons que $f\big([u, v]\big)$ ne contient aucun élément plus grand que β. Même jeu pour montrer que ça ne contient aucun élément plus petit que α.

En définitive, le segment $L = [u, v]$ fonctionne. □

19. définition 7.152. Un segment est un intervalle fermé borné.
20. Par le lemme 7.125.
21. Théorème 11.50.
22. Corolaire 11.51 et théorème 7.92.

Lorsque $I_2 \subset f(I_1)$ nous notons $I_1 \to I_2$ ou, si une ambiguïté est à craindre, $I_1 \xrightarrow{f} I_2$. Cette flèche se lit « recouvre ».

Lemme 11.81 ([202, 234]).
Soient les segments $I_0, \dots, I_{n-1}$ tels que nous ayons le cycle

$$I_0 \to I_1 \to \dots \to I_{n-1} \to I_0. \tag{11.140}$$

Alors f^n admet un point fixe $x_0 \in I_0$ tel que $f^k(x_0) \in I_k$ pour tout $k = 0, \dots, n-1$.

Démonstration. Nous prouvons les cas $n = 1$ et $n = 2$ séparément.

$\underline{n = 1}$ Nous avons $I_0 \to I_0$, c'est-à-dire que $I_0 \subset f(I_0)$. Si $I_0 = [a, b]$ alors nous posons $a = f(\alpha)$ et $b = f(\beta)$ pour certains $\alpha, \beta \in I_0$. Nous posons ensuite $g(x) = f(x) - x$.

Dans un premier temps, $g(\alpha) = a - \alpha \leqslant 0$ parce que $a =\in (I_0)$ et $\alpha \in I_0$. Pour la même raison, $g(\beta) = b - \beta \geqslant 0$. Le théorème des valeurs intermédiaires donne alors $t_0 \in [\alpha, \beta] \subset I_0$ tel que $g(t_0) = 0$. Nous avons donc $f(t_0) = t_0$.

$\underline{n = 2}$ Nous avons $I_0 \to I_1 \to I_0$. Vu que $I_1 \subset f(I_0)$, le lemme 11.80 donne un segment $J_1 \subset I_0$ tel que $f(J_1) = I_1$. Mézalors

$$J_1 \subset I_0 \subset f(I_1) = f^2(J_1). \tag{11.141}$$

Nous avons donc $J_1 \xrightarrow{f^2} J_1$ et par le cas $n = 1$ traité plus haut, la fonction f^2 a un point fixe x_0 dans J_1. De plus

$$f(x_0) \in f(J_1) = I_1, \tag{11.142}$$

le point x_0 est donc bien celui que nous cherchions.

Cas général. Nous avons

$$I_0 \to I_1 \to \dots \to I_{n-1} \to I_0. \tag{11.143}$$

Vu que $I_1 \subset f(I_0)$, il existe $J_1 \subset I_0$ tel que $f(J_1) = I_1$. Mais

$$I_2 \subset f(I_1) = f^2(J_1), \tag{11.144}$$

donc il existe $J_2 \subset J_1$ tel que $I_2 = f^2(J_2)$. En procédant encore longtemps ainsi nous construisons les ensembles $J_1, \dots, J_{n-1}$ tels que

$$J_{n-1} \subset J_{n-2} \subset \dots \subset J_1 \subset J_0 \tag{11.145}$$

tels que $I_k = f^k(J_k)$ pour tout $k = 1, \dots, n-1$. La dernière de ces inclusions est $I_{n-1} = f^{n-1}(J_{n-1})$, mais $I_{n-1} \to I_0$, c'est-à-dire que

$$I_0 \subset f(I_{n-1}) = f^n(J_{n-1}), \tag{11.146}$$

et il existe $J_n \subset J_{n-1}$ tel que $I_0 \subset f^n(J_n)$. Mais comme $J_n \subset J_0$ nous avons en particulier $J_n \subset f^n(J_n)$.

Cela donne un point fixe $x_0 \in J_n$ pour f^n. Par construction nous avons $J_n \subset J_{n-1} \subset \dots \subset J_1 \subset J_0$ et donc $x_0 \in J_k$ pour tout k. En particulier

$$f^k(x_0) \in f^k(J_k) = I_k \tag{11.147}$$

pour tout k.

□

Théorème 11.82 (Théorème de Sarkowski[202, 234]).
Soit I, un segment de $\mathbb{R}$ et une application continue $f\colon I \to I$. Si f admet un point 3-périodique, alors f admet des points n-périodiques pour tout $n \geqslant 1$.

Démonstration. Soit $a \in I$ un point 3-périodique pour f et notons $b = f(a)$, $c = f(b)$. Les points b et c sont également des points 3-périodiques. Quitte à renommer, nous pouvons supposer que a est le plus petit des trois. Il reste deux possibilités : $a < b < c$ et $a < c < b$. Nous traitons d'abord le premier cas.

Supposons $a < b < c$. Nous posons $I_0 = [a, b]$ et $I_1 = [b, c]$. Nous avons immédiatement $I_1 \subset f(I_0)$ et comme $f(b) = c$ et $f(c) = a$, $f(I_1)$ recouvre $[a, c]$ et donc recouvre en même temps I_1 et I_2. Nous avons donc $I_0 \to I_1$, $I_1 \to I_0$ et $I_1 \to I_1$.

Un point 1-périodique Nous avons $I_1 \to I_1$ qui prouve que f a un point fixe dans I_1. C'est le cas $n = 1$ du lemme 11.81. Voilà un point 1-périodique.

Un point 2-périodique Nous avons $I_0 \to I_1 \to I_0$. Par conséquent, le lemme 11.81 dit que f^2 a un point fixe $x_0 \in I_0$ tel que $f(x_0) \in I_1$. Montrons que $f(x_0) \neq x_0$. Pour avoir $x_0 = f(x_0)$, il faudrait $x_0 \in I_0 \cap I_1 = \{b\}$. Mais b est un point 3-périodique, donc ne vérifiant certainement pas $f^2(b) = b$. Nous en déduisons que $f(x_0) \neq x_0$ et donc que x_0 est 2-périodique.

Un point 3-périodique On en a par hypothèse.

Un point n-périodique pour $n \geqslant 4$ Nous avons le cyle

$$I_0 \to \underbrace{I_1 \to I_1 \to \ldots \to I_1}_{\text{n-1}fois} \to I_0. \tag{11.148}$$

Le lemme donne alors un point fixe $x \in I_0$ pour f^n tel que $f^k(x) \in I_1$ pour $k = 1, \ldots, n-1$. Est-ce possible que $x = b$? Non parce que $f^2(b) = a \in I_0$ alors que $f^2(x) \in I_1$. Mais $I_0 \cap I_1 = \{b\}$.

Par conséquent la relation $f^k(x) \in I_1$ exclu d'avoir $f^k(x) = x$, et le point x est bien n-périodique.

Passons au cas $a < c < b$. Alors nous posons $I_0 = [a, c]$ et $I_1 = [c, b]$. Encore une fois $f(I_0)$ contient a et b, donc $I_0 \to I_0$ et $I_0 \to I_1$. Mais en même temps $f(I_1)$ contient a et c, donc $I_1 \to I_0$.

Nous pouvons donc refaire comme dans le premier cas, en inversant les rôles de I_0 et I_1. En particulier nous pouvons considérer le cycle

$$I_1 \to I_0 \to I_0 \to \ldots \to I_0 \to I_1. \tag{11.149}$$

□

11.8 Uniforme continuité

Définition 11.83.
*Une partie $A \subset \mathbb{R}^m$ est dite **bornée** s'il existe un $M > 0$ tel que $A \subset B(0, M)$. Le **diamètre** de la partie A est le nombre*

$$\mathrm{Diam}(A) = \sup_{x,y \in A} \|x - y\| \in [0, \infty]. \tag{11.150}$$

Lorsque A est borné, il existe un M tel que $\|x\| \leqslant M$ pour tout $x \in A$.

Lemme 11.84.
Si A est une partie non vide de $\mathbb{R}^m$, alors $\mathrm{Diam}(A) = \mathrm{Diam}(\bar{A})$.

Nous n'allons pas donner de démonstrations de ce lemme.

Si (x_n) est une suite et I est un sous-ensemble infini de $\mathbb{N}$, nous désignons par x_I la suite des éléments x_n tels que $n \in I$. Par exemple la suite $x_{\mathbb{N}}$ est la suite elle-même, la suite $x_{2\mathbb{N}}$ est la suite obtenue en ne prenant que les éléments d'indice pair.

Les suites x_I ainsi construites sont dites des **sous-suites** de la suite (x_n).

Pour une fonction $f \colon D \subset \mathbb{R}^m \to \mathbb{R}$, la continuité au point a signifie que pour tout $\varepsilon > 0$,

$$\exists \delta > 0 \text{ tel que } 0 < \|x - a\| < \delta \Rightarrow |f(x) - f(a)| < \varepsilon. \tag{11.151}$$

Le δ qu'il faut choisir dépend évidemment de ε, mais il dépend en général aussi du point a où l'on veut tester la continuité. C'est-à-dire que, étant donné un $\varepsilon > 0$, nous pouvons trouver un δ qui fonctionne pour certains points, mais qui ne fonctionne pas pour d'autres points.

Il peut cependant également arriver qu'un même δ fonctionne pour tous les points du domaine. Dans ce cas, nous disons que la fonction est uniformément continue sur le domaine.

Définition 11.85.
Une fonction $f\colon D \subset \mathbb{R}^m \to \mathbb{R}$ est dite ***uniformément continue*** *sur D si*

$$\forall \varepsilon > 0,\ \exists \delta > 0 \text{ tel que } \forall x, y \in D,\ \|x - y\| \leqslant \delta \Rightarrow |f(x) - f(a)| < \varepsilon. \tag{11.152}$$

Il est intéressant de voir ce que signifie le fait de *ne pas* être uniformément continue sur un domaine D. Il s'agit essentiellement de retourner tous les quantificateurs de la condition (11.152) :

$$\exists \varepsilon > 0 \text{ tel que } \forall \delta > 0,\ \exists x, y \in D \text{ tel que } \|x - y\| < \delta \text{ et } \big|f(x) - f(y)\big| > \varepsilon. \tag{11.153}$$

Dans cette condition, les points x et y peuvent être fonction du δ. L'important est que pour tout δ, on puisse trouver deux points δ-proches dont les images par f ne soient pas ε-proches.

Exemple 11.86.
Prenons la fonction $f(x) = \frac{1}{x}$, et demandons nous pour quel δ nous sommes sûr d'avoir

$$|f(a + \delta) - f(a)| = \left|\frac{1}{a + \delta} - \frac{1}{a}\right| < \varepsilon. \tag{11.154}$$

Pour simplifier, nous supposons que $a > 0$. Nous calculons

$$\begin{aligned}
\frac{1}{a} - \frac{1}{a + \delta} &< \varepsilon \\
\frac{\delta}{a(a + \delta)} &< \varepsilon \\
\delta &< \varepsilon a^2 + \varepsilon a \delta \\
\delta(1 - \varepsilon a) &< \varepsilon a^2 \\
\delta &< \frac{\varepsilon a^2}{1 - \varepsilon a}.
\end{aligned} \tag{11.155}$$

Notons que, à ε fixé, plus a est petit, plus il faut choisir δ petit. La fonction $x \mapsto \frac{1}{x}$ n'est donc pas uniformément continue. Cela correspond au fait que, proche de zéro, la fonction monte très vite. Une fonction uniformément continue sera une fonction qui ne montera jamais très vite. △

Proposition 11.87.
Quelques propriétés des fonctions uniformément continues.

(1) Toute application uniformément continue est continue ;

(2) la composée de deux fonctions uniformément continues est uniformément continue ;

Nous verrons qu'une application lipschitzienne est uniformément continue (proposition 11.264).

Une fonction peut être uniformément continue sur un domaine et pas sur un autre. Le théorème suivant donne une importante indication à ce sujet.

Théorème 11.88 (Heine)**.**
Soit K un compact de $\mathbb{R}^n$. Une fonction continue $f\colon \mathbb{R}^n \to \mathbb{R}^m$ est uniformément continue sur K.

La démonstration qui suit est valable pour une fonction $f\colon \mathbb{R}^n \to \mathbb{R}^m$ et utilise le fait que le produit cartésien de compacts est compact. Dans le cas de fonctions sur $\mathbb{R}$, nous pouvons modifier la démonstration pour ne pas utiliser ce résultat ; voir plus bas.

Démonstration. Nous allons prouver ce théorème par l'absurde. Nous commençons par écrire la condition (11.153) qui exprime que f n'est pas uniformément continue sur le compact K :

$$\exists \varepsilon > 0 \text{ tel que } \forall \delta > 0, \exists x, y \in K \text{ tels que } \|x - y\| < \delta \text{ et } \big|f(x) - f(y)\big| > \varepsilon. \tag{11.156}$$

En particulier (en prenant $\delta = \frac{1}{n}$ pour tout n), pour chaque n nous pouvons trouver x_n et y_n dans K qui vérifient simultanément les deux conditions suivantes :

$$\begin{cases} \|x_n - y_n\| < \dfrac{1}{n} & \text{(11.157a)} \\ \big|f(x_n) - f(y_n)\big| > \varepsilon. & \text{(11.157b)} \end{cases}$$

Nous insistons que c'est le même ε pour chaque n. L'ensemble K étant compact, l'ensemble $K \times K$ est compact (théorème 7.318) et nous pouvons trouver une sous-suite convergente *du couple* (x_n, y_n) dans $K \times K$. Quitte à passer à ces sous-suites, nous nous supposons que (x_n, y_n) converge dans $K \times K$ et en particulier que les suites (x_n) et (y_n) sont convergentes. Étant donné que pour chaque n elles vérifient $\|x_n - y_n\| < \frac{1}{n}$, les limites sont égales :

$$\lim x_n = \lim y_n = x. \tag{11.158}$$

L'ensemble K étant fermé, la limite x est dans K. Par continuité de f, nous avons finalement

$$\lim f(x_n) = \lim f(y_n) = f(x), \tag{11.159}$$

mais alors

$$\lim_{n\to\infty} \big|f(x_n) - f(y_n)\big| = 0, \tag{11.160}$$

ce qui est en contradiction avec le choix (11.157b).

Tout ceci prouve que $f(K)$ est bornée supérieurement et que f atteint son supremum (qui est donc un maximum). Le fait que $f(K)$ soit borné inférieurement se prouve en considérant la fonction $-f$ au lieu de f.

□

Remarque 11.89.
Nous pouvons ne pas utiliser le fait que le produit de compacts est compact. Cela est particulièrement commode lorsqu'on considère des fonctions de $\mathbb{R}$ dans $\mathbb{R}$ parce que dans ce cadre nous ne pouvons pas supposer connue la notion de produit d'espace topologiques.

Pour choisir les sous-suites (x_n) et (y_n), il suffit de prendre une sous-suite convergente de (x_n) et d'invoquer le fait que $\|x_n - y_n\| \leqslant \frac{1}{n}$. Les suites (x_n) et (y_n) étant adjacentes[23], la convergence de (x_n) implique la convergence de (y_n) vers la même limite.

Il est donc un peu superflus de parler de la convergence du couple (x_n, y_n).

Proposition 11.90 (Heine[235])**.**
Toute application continue d'un espace métrique compact dans un espace métrique quelconque est uniformément continue.

11.9 Fonctions sur un compact

Par le théorème des valeurs intermédiaires 11.50, l'image d'un intervalle par une fonction continue est un intervalle, et nous avons l'importante propriété suivante des fonctions continues sur un compact.

Le théorème suivant est un cas particulier du théorème 7.154.

Théorème 11.91.
Si f est une fonction continue sur l'intervalle compact $[a, b]$. Alors f est bornée sur $[a, b]$ et elle atteint ses bornes.

23. Définition 7.142.

Démonstration. Étant donné que $[a, b]$ est un intervalle compact, son image est également un intervalle compact, et donc est de la forme $[m, M]$. Ceci découle du théorème 7.92 et le corolaire 11.51. Le maximum de f sur $[a, b]$ est la borne M qui est bien dans l'image (parce que $[m, M]$ est fermé). Idem pour le minimum m. □

11.10 Polynômes

L'algèbre des polynômes sur un anneau est définie en 3.86. Si $P \in A[X]$ et si $\alpha \in A$ nous avons également défini l'évaluation de P en α ; c'est la définition 3.87. Dans le cadre de l'analyse, lorsque nous considérons des polynômes, nous allons complètement confondre le polynôme avec la fonction qu'il définit.

11.10.1 Polynômes sur les réels

Proposition 11.92.
Tout polynôme à coefficients réels de degré impair possède une racine réelle.

Démonstration. Nous mettons le plus haut degré en facteur :

$$P(x) = \sum_{k=0}^{n} a_k x^k = x^n \sum_{k=0}^{n} \frac{a_k}{x^{n-k}}. \tag{11.161}$$

Le terme $k = 0$ vaut $a_n x^n$ tandis que les autres sont de la forme (à coefficient près) $\frac{1}{x^l}$ pour un $l \geqslant 1$. Lorsque $x \to \infty$, chacun de ces termes s'annule (lemme 11.27). Nous avons donc

$$\lim_{x\to\infty} P(x) = +\infty, \tag{11.162}$$

et de même, n étant impair, $\lim_{x\to-\infty} P(x) = -\infty$. Le théorème des valeurs intermédiaires 11.50 nous donne alors l'existence d'un réel sur lequel P s'annule. □

11.10.2 Polynômes sur les complexes

Nous allons parler de comportement asymptotique de polynômes définis sur $\mathbb{C}$. La topologique que nous considérons est celle de la compactification en un point décrite en 7.66.

Le lemme suivant donne une caractérisation de la limite en l'infini dans le compactifié $\hat{\mathbb{C}}$. Dans beaucoup de cas, cette caractérisation est prise comme la définition de la limite. Hélas, dans le Frido nous sommes des extrémistes et nous ne parvenons pas à dire le mot « limite » si il n'y a pas une topologie.

Lemme 11.93 ([1]).
Nous considérons la compatification en un point d'Alexandrov[24]*. Soit une fonction $f\colon \mathbb{C} \to \mathbb{C}$. Nous avons $\lim_{z\to\infty} f(z) = \infty$ si et seulement si pour tout $M > 0$, il existe $R > 0$ tel que $|z| > R$ implique $|f(z)| > M$.*

Démonstration. Souvenons-nous que, en général[25], nous avons

$$\lim_{x\to a} f(x) = b \tag{11.163}$$

si pour tout voisinage V de b, il existe un voisinage W de a tel que $z \in W\backslash\{a\}$ implique $f(z) \in V$.

Précisons encore un point de notation. Si K est une partie de $\mathbb{C}$, nous notons K^c son complémentaire dans $\mathbb{C}$, pas dans $\hat{\mathbb{C}}$.

Ceci étant dit, nous passons à la preuve.

24. Définition 7.66.
25. Définition 7.67.

Sens direct Nous supposons que $\lim_{z\to\infty} f(z) = \infty$. Soit $M > 0$; nous considérons le voisinage $V = \overline{B(0,M)}^c \cup \{\infty\}$. Par définition de la limite, il existe un voisinage W de ∞ tel que $z \in W \Rightarrow f(z) \in V\backslash\{\infty\} = \overline{B(0,M)}^c$. Ce voisinage est de la forme $K^c \cup \{\infty\}$. Vu que K est compact, il est borné et il existe $R > 0$ tel que $K \subset B(0,R)$.

Avec tout cela nous avons la chaine suivante d'implications :

$$|z| > R \Rightarrow z \in K^c \Rightarrow z \in W \Rightarrow f(z) \in V\backslash\{\infty\} = \overline{B(0,M)}^c \Rightarrow |f(z)| > M. \tag{11.164}$$

C'est bien la propriété que nous voulions.

Sens réciproque Soit un voisinage V de ∞. Nous avons $V = K^c \cup \{\infty\}$ où K est compact dans $\mathbb{C}$. Il existe $M > 0$ tel que $K \subset B(0,M)$.

Par hypothèse, il existe R tel que $|z| > R \Rightarrow |f(z)| > M$. Soit $W = \overline{B(0,R)}^c \cup \{\infty\}$. Nous avons la chaine

$$z \in W \Rightarrow |z| > R \Rightarrow |f(z)| > M \Rightarrow f(z) \in K^c \Rightarrow f(z) \in V. \tag{11.165}$$

□

Proposition 11.94 ([1]).
Soit le polynôme

$$\begin{aligned} P\colon \mathbb{C} &\to \mathbb{C} \\ z &\mapsto \sum_{i=0}^{n} a_i z^i \end{aligned} \tag{11.166}$$

où nous sous-entendons que $a_n \neq 0$. La fonction $z \mapsto |P(z)|$ est équivalente[26] *en l'infini à la fonction*

$$\begin{aligned} w\colon \mathbb{C} &\to \mathbb{R}^+ \\ z &\mapsto |a_n z^n|. \end{aligned} \tag{11.167}$$

Démonstration. Nous voudrions prouver qu'il existe une fonction $\alpha\colon \mathbb{C} \to \mathbb{R}$ telle que

$$\left\{ |\sum_{i=0}^{n} a_i z^i| = \big(1 + \alpha(z)\big)|a_n z^n|. \lim_{z\to\infty} \alpha(z) = 0. \right. \tag{11.168a}$$

Nous trouvons un candidat pour être une telle fonction en isolant simplement $\alpha(z)$ de cette égalité. Nous trouvons

$$\alpha(z) = |\sum_{i=0}^{n} \frac{a_i}{a_n} z^{i-n}| - 1. \tag{11.169}$$

Elle vérifie immédiatement (11.168). Le point qui fait intervenir la topologie de est de vérifier que $\lim_{z\to\infty} \alpha(z) = 0$. Le terme $i = 0$ de la somme vaut 1. Il suffit donc de montrer que pour $i \neq 0$ nous avons

$$\lim_{z\to\infty} \frac{1}{z^{n-i}} = 0. \tag{11.170}$$

Soit $\epsilon > 0$. Nous devons prouver qu'il existe un voisinage V de ∞ dans $\hat{\mathbb{C}}$ tel que

$$|\frac{1}{z^{n-i}} - 0| \leqslant \epsilon \tag{11.171}$$

pour tout $z \in V$.

En utilisant la proposition 1.167 nous avons déjà

$$|\frac{1}{z^{n-i}}| = \frac{1}{|z^{n-i}|} = \frac{1}{|z|^{n-i}}. \tag{11.172}$$

26. Définition 7.40.

Soit $R > 0$ tel que $\frac{1}{R} < \epsilon$. Nous considérons le voisinage $\{|z| > R\} \cup \{\infty\}$ de ∞. Dans ce voisinage, nous avons

$$\frac{1}{|z|^{n-i}} \leqslant \frac{1}{|z|} \leqslant \frac{1}{R} < \epsilon. \tag{11.173}$$

Et voila. □

Le lemme suivant parle de polynôme sur $\mathbb{C}$. Vous pouvez l'adapter à $\hat{\mathbb{R}}$ et $\bar{\mathbb{R}}$.

Lemme 11.95.
Si $P\colon \mathbb{C} \to \mathbb{C}$ est un polynôme, alors $|P|$ atteint une borne inférieure globale.

Démonstration. Nous savons, par l'équivalence de fonctions prouvée dans la proposition 11.94 que $\lim_{z\to\infty} P(z) = \infty$. Soit $a > 0$ dans $\mathbb{R}$. Par le lemme 11.93 il existe un $R > a$ tel que $|z| > R \Rightarrow |f(z)| > |f(a)|$.

La fonction $|P|$ est continue sur le compact $\overline{B(0,R)}$. Soit z_0 le point de minimum [27] de $|P|$ sur $\overline{B(0,R)}$.

Nous devons prouver que z_0 donne même un minimum global. Vu que $a \in \overline{B(0,R)}$ nous avons

$$|f(z_0)| \leqslant |f(a)|. \tag{11.174}$$

Si $z \in \overline{B(0,R)}^c$, nous avons

$$|f(z)| > |f(a)| \geqslant |f(z_0)|. \tag{11.175}$$

Donc ce z_0 est un minimum sur $B(0,R)$ et sur $\overline{B(0,R)}^c$. Bref, un minimum global. □

Lemme 11.96.
Soit le polynôme

$$\begin{aligned} P\colon \mathbb{C} &\to \mathbb{C} \\ z &\mapsto \sum_{i=0}^{n} a_i z^i. \end{aligned} \tag{11.176}$$

La fonction P est équivalente à $a_0 + a_1 z$ en $z = 0$.

Démonstration. En posant $g(z) = a_0 + a_1 z$, nous devons trouver une fonction α telle que

$$P(z) = \big(1 + \alpha(z)\big)g(z). \tag{11.177}$$

Si $a_0 \neq 0$, il existe un voisinage de $z = 0$ sur lequel la fonction

$$\alpha(z) = \frac{z^2 \sum_{i=2}^{n} a_i z^{i-2}}{a_0 + a_1 z} \tag{11.178}$$

existe. Il n'y a aucun problème à ce que $\alpha(z) \to 0$ pour $z \to 0$ [28], et un simple calcul [29] donne (11.178).

Si par contre $a_0 = 0$, nous faisons le calcul intermédiaire suivant :

$$\alpha(z)g(z) = P(z) - g(z) = z^2 \sum_{i=2}^{n} a_i z^{i-2}, \tag{11.179}$$

et donc, en isolant $\alpha(z)$ et en simplifiant par z, nous voyons que la fonction α définie par

$$\alpha(z) = \frac{z}{a_1} \sum_{i=2}^{n} a_i z^{i-2} \tag{11.180}$$

fonctionne. □

27. Théorème de Weierstrass 7.105.
28. En remarquant toutefois que c'est une limite à deux dimensions. Sachez la définir.
29. En fait, la formule (11.178) est obtenue en isolant $\alpha(z)$ dans (11.177).

Proposition 11.97 ([236, 1]).
Soient $a, b \in \mathbb{R}$.

(1) L'équation $z^2 = a + bi$ a une solution dans $\mathbb{C}$.

(2) Pour tout l, l'équation $z^{2^l} = a + bi$ a une solution dans $\mathbb{C}$.

Nous ne disons pas que ces solutions sont uniques[30].

Démonstration. Pour prouver (1), l'équation $z^2 = a + bi$ a pour solution $\pm\xi$ où

$$\xi = \sqrt{\frac{1}{2}a + \frac{1}{2}\sqrt{a^2 + b^2}} + i\,\mathrm{sgn}(b)\sqrt{-\frac{1}{2}a + \frac{1}{2}\sqrt{a^2 + b^2}}. \tag{11.181}$$

Nous n'avons en fait pas besoin de montrer que $\pm\xi$ sont toutes deux des solutions, ni que ce sont les seules. Un calcul direct montre que $\xi^2 = a + bi$ et nous sommes content.

Pour (2), nous faisons une récurrence sur l. Nous savons que

$$z^{2^{k+1}} = (z^{2^k})^2. \tag{11.182}$$

Soit $\xi \in \mathbb{C}$ tel que $\xi^{2^k} = a + bi$; un tel ξ existe par hypothèse de récurrence. Alors si z est tel que $z^2 = \xi$, nous avons

$$z^{2^{k+1}} = a + bi. \tag{11.183}$$

□

Le théorème de d'Alembert possède de nombreuses démonstrations. En voici une qui à ma connaissance est celle demandant le moins d'analyse ; une démonstration à base de théorie de Galois peut être trouvée dans [51, 237]. Si vous lisez ces lignes pour savoir qu'un polynôme de degré n possède au *maximum* n racines, ce n'est pas ici qu'il faut regarder, mais le corolaire 3.195.

Théorème 11.98 (d'Alembert[236]).
Tout polynôme non constant à coefficients complexes admet au moins une racine complexe.

Démonstration. Nous effectuons une preuve tout à la fois par l'absurde et par récurrence en supposant que le polynôme

$$\begin{aligned} f\colon \mathbb{C} &\to \mathbb{C} \\ z &\mapsto z^n + a_1 z^{n-1} + \ldots + a_n \end{aligned} \tag{11.184}$$

n'a pas de racines dans $\mathbb{C}$, et que n soit le plus petit entier pour lequel un tel polynôme existe. Nous notons

$$n = 2^k m \tag{11.185}$$

où m est impair.

Le lemme 11.95 donne un point z_0 qui réalise le minimum global de $|f|$ sur $\mathbb{C}$. Nous posons $g(z) = f(z_0 + z)$ et nous définissons ses coefficients A_i par

$$g(z) = \sum_{i=0}^{n} A_i z^i. \tag{11.186}$$

Nous avons $A_n = 1$ et $|A_0| = |f(z_0)|$. Soit A_r le premier à être non nul parmi les A_1, A_2, ...

Si $r < n$ Par hypothèse de récurrence, il existe $\xi \in \mathbb{C}$ tel que $\xi^r = -A_1/A_r$. Nous avons

$$g(t\xi) = A_0 + \frac{-A_r t^r A_0}{A_r} + t^{r+1} \sum_{i=r+1}^{n} A_i \xi^i t^{i-r-1}. \tag{11.187}$$

En notant $P(t)$ le dernier polynôme, nous pouvons écrire cela sous forme compacte :

$$g(t\xi) = A_0 - t^r A_0 + t^{r+1} P(t). \tag{11.188}$$

30. Comme vous en conviendrez en pensant à $z^2 = 1$ qui a déjà les solutions 1 et -1.

Vu que

$$\lim_{t\to 0}\frac{t^{r+1}P(t)}{t^r|A_0|}=\lim_{t\to 0}tP(t)=0, \tag{11.189}$$

il existe $t_0>0$ tel que

$$|t_0^{r+1}P(t_0)|<|A_0t_0r|. \tag{11.190}$$

Nous choisissons de plus $t_0<1$, de telle sorte que $1-t^r>0$. Avec cela nous avons

$$|g(t\xi)|\leqslant|A_0|(1-t^r)+|t^{r+1}P(t)|=|A_0|\underbrace{-t^r|A_0|+|t^{r+1}P(t)|}_{<0}<|A_0|. \tag{11.191}$$

Or $|A_0|$ était un minimum global de $|g|$. Contradiction.

Si $r=n$ Dans ce cas,

$$g(z)=f(z_0+z)=A_0+z^n, \tag{11.192}$$

et nous rappelons que $n=2^km$ où m est impair. Nous allons trouver une contradiction dans les quatre cas $\operatorname{Re}A_0>0$, $\operatorname{Re}(A_0)<0$, $\operatorname{Im}(A_0)>0$ et $\operatorname{Im}(A_0)<0$. Bien entendu ces cas se recouvrent largement, mais en toute généralité, nous avons besoin des quatre.

Si $\operatorname{Re}(A_0)>0$ La proposition 11.97 nous permet de considérer $v\in\mathbb{C}$ tel que $v^{2^k}=-1$. Nous avons alors

$$g(tv)=A_0+(tv)^n=A_0+t^n(v^{2^k})^m=A_0+t^n(-1)^m=A_0-t^n \tag{11.193}$$

parce que m est impair. Nous avons $\operatorname{Im}\big(g(tv)\big)=\operatorname{Im}(A_0)$. Si t est assez petit pour que $t^n<|\operatorname{Re}(A_0)|$ nous avons aussi $|\operatorname{Re}\big(g(tv)\big)|<|\operatorname{Re}(A_0)|$. Donc

$$|g(tv)|^2=|\operatorname{Re}\big(g(tv)\big)|^2+|\operatorname{Re}\big(g(tv)\big)|^2<|\operatorname{Re}(A_0)|^2+|\operatorname{Im}(A_0)|^2=|A_0|^2. \tag{11.194}$$

Donc $|g(tv)|<|A_0|$. Contradiction.

Si $\operatorname{Re}(A_0)<0$ Nous prenons $v=1$, et même histoire.

Si $\operatorname{Im}(A_0)<0$ Nous prenons $w\in\mathbb{C}$ tel que

$$w^{2^k}=i(-1)^{\frac{1}{2}(m-1)}. \tag{11.195}$$

Là, il y a un peu d'arrachage de cheveux pour bien voir les cas. La difficulté est que les puissances de i alternent entre 1, -1, i et $-i$. Vu que m est impair, nous avons un l tel que $m=2l+1$. Nous subdivisons les cas l pair et l impair.

Si l est pair Alors d'une part $\frac{1}{2}(m-1)=l$ est pair et donc

$$(-1)^{\frac{1}{2}(m-1)}=1. \tag{11.196}$$

Et d'autre part, $i^{2l+1}=(-1)^li=i$. En tout,

$$i^m(-1)^{\frac{1}{2}(m-1)}=i. \tag{11.197}$$

Si l est impair Alors $\frac{1}{2}(m-1)=l$ et $(-1)^{\frac{1}{2}(m-1)}=-1$. Mais en même temps, $i^{2l+1}=-i$, ce qui donne encore une fois

$$i^m(-1)^{\frac{1}{2}(m-1)}=i. \tag{11.198}$$

Bref, que l soit pair ou impair, nous avons $i^m(-1)^{\frac{1}{2}(m-1)}=i$.

Nous avons donc $\operatorname{Re}\big(g(tw)\big)=\operatorname{Re}(A_0)$ et $\operatorname{Im}\big(g(tw)\big)<\operatorname{Im}(A_0)$. Encore contradiction.

Si $\operatorname{Im}(A_0)=0$ Même chose que ce que nous venons de faire, mais avec

$$w^{2^k}=-i(-1)^{\frac{1}{2}(m-1)}. \tag{11.199}$$

□

Corolaire 11.99.
Le corps $\mathbb{C}$ est algébriquement clos.

Corolaire 11.100 ([1])**.**
Tout polynôme de degré 3 à coefficients réels possède au moins une racine réelle.

Démonstration. Soient les racines λ_1, λ_2 et λ_3 du polynôme en question. Toutes trois sont dans $\mathbb{C}$. Supposons que λ_1 ne soit pas réelle. Alors λ_2 ou λ_3 doit être égale à $\bar{\lambda}_1$. Disons λ_2. Nous avons donc les racines λ_1, $\bar{\lambda}_1$ et λ_3. Le polynôme se factorise alors en

$$a(X-\lambda_1)(X-\bar{\lambda}_1)(X-\lambda_3). \tag{11.200}$$

Le coefficient a doit être réel parce qu'il est le coefficient du terme en X^3 (réel par hypothèse). Si λ_3 n'est pas réel, alors ce polynôme ne peut pas avoir des coefficients réels. Entre autres parce que terme indépendant est $a|\lambda_1|^2\lambda_3$, qui est réel si et seulement si λ_3 est réel[31]. □

Tant que vous y êtes, vous pouvez voir que le polynôme (11.200) est à coefficient réels si et seulement si $a \in \mathbb{R}$ et $\lambda_3 \in \mathbb{R}$.

Exemple 11.101.
Toute application linéaire $\mathbb{R}^3 \to \mathbb{R}^3$ a un vecteur propre. En effet si $R\colon \mathbb{R}^3 \to \mathbb{R}^3$ est linéaire, son polynôme caractéristique χ_R est de degré 3. Le corolaire 11.100 indique qu'un tel polynôme possède au moins une racine réelle. Une telle racine est une valeur propre de R par le théorème 9.111. △

Définition 11.102.
Si $\lambda \in \mathbb{K}$ est une racine de χ_u, l'ordre de l'annulation est la ***multiplicité algébrique*** *de la valeur propre λ de u. À ne pas confondre avec la* ***multiplicité géométrique*** *qui sera la dimension de l'espace propre.*

Proposition 11.103.
Un polynôme irréductible à coefficients réels est soit de degré un soit de degré 2 avec un discriminant négatif.

Démonstration. Soit un polynôme P à coefficients réels de degré plus grand que 1. Alors le théorème de d'Alembert-Gauss (théorème 11.98) implique l'existence d'une racine $\alpha \in \mathbb{C}$. Si α est un réel, P est réductible. Si α n'est pas réel, alors conjugué complexe $\bar{\alpha}$ est également une racine. Par conséquent les polynômes $(X-\alpha)$ et $(X-\bar{\alpha})$ divisent P dans $\mathbb{C}[X]$..

Ces deux polynômes sont premiers entre eux parce que

$$a(X-\alpha)+b(X-\bar{\alpha})=0 \tag{11.201}$$

implique $a=b=0$. Par conséquent le produit

$$X^2-(\alpha+\bar{\alpha})X+\alpha\bar{\alpha} \tag{11.202}$$

divise également P. Ce dernier est un polynôme à coefficients réels de degré 2. Donc tout polynôme de degré 3 ou plus est réductible. □

Proposition 11.104.
Si E est un espace vectoriel sur $\mathbb{C}$, tout endomorphisme possède au moins une valeur propre.

Démonstration. Soit un endomorphisme u sur E. Le théorème 9.111 dit que $\lambda \in \mathbb{C}$ est une valeur propre si et seulement si λ est une racine du polynôme caractéristique χ_u. Or ce polynôme possède au moins une racine dans $\mathbb{C}$ par le théorème de d'Alembert 11.98. □

31. Notez l'utilisation du lemme 1.169.

11.11 Dérivée : exemples introductifs

11.11.1 La vitesse

Lorsqu'un mobile se déplace à une vitesse variable, nous obtenons la *vitesse instantanée* en calculant une vitesse moyenne sur des intervalles de plus en plus petits. Si le mobile a un mouvement donné par $x(t)$, la vitesse moyenne entre $t = 2$ et $t = 5$ sera

$$v_{\text{moy}}(2 \to 5) = \frac{x(5) - x(2)}{5 - 2}.$$

Plus généralement, la vitesse moyenne entre 2 et $2 + \Delta t$ est donnée par

$$v_{\text{moy}}(2 \to 2 + \Delta t) = \frac{x(2 + \Delta t) - x(2)}{\Delta t}.$$

Cela est une fonction de Δt. Oui, mais je te rappelle qu'on a dans l'idée de calculer une vitesse instantanée, c'est-à-dire de voir ce que vaut la vitesse moyenne sur un intervalle très très très très petit. La notion de limite semble toute indiquée pour décrire mathématiquement l'idée physique de vitesse instantanée.

Nous allons dire que la vitesse instantanée d'un mobile est la limite quand Δt tend vers zéro de sa vitesse moyenne sur l'intervalle de temps Δt, ou en formule :

$$v(t_0) = \lim_{\Delta t \to 0} \frac{x(t_0) - x(t_0 + \Delta t)}{\Delta t}. \tag{11.203}$$

11.11.2 La tangente à une courbe

Passons maintenant à tout autre chose, mais toujours dans l'utilisation de la notion de limite pour résoudre des problèmes intéressants. Comment trouver l'équation de la tangente à la courbe $y = f(x)$ au point $(x_0, f(x_0))$?

Essayons de trouver la tangente au point P donné de la courbe donnée à la figure 11.1.

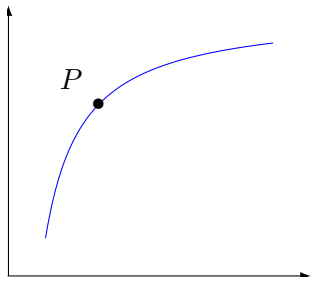


FIGURE 11.1 – Comment trouver la tangente à la courbe au point P ?

La tangente est la droite qui touche la courbe en un seul point sans la traverser. Afin de la construire, nous allons dessiner des droites qui touchent la courbe en P et un autre point Q, et nous allons voir ce qu'il se passe quand Q est très proche de P. Cela donnera une droite qui, certes, touchera la courbe en deux points, mais en deux points *tellement proches que c'est comme si c'étaient les mêmes*. Tu sens que la notion de limite va encore venir.

Nous avons placé le point, sur la figure 11.2, le point P en a et le point Q un peu plus loin x. En d'autres termes leurs coordonnées sont

$$P = \big(a, f(a)\big) \qquad\qquad Q = \big(x, f(x)\big). \tag{11.204}$$

En regardant par exemple la figure 11.2, le coefficient directeur de la droite qui passe par ces deux points est donné par

$$\frac{f(x) - f(a)}{x - a}, \tag{11.205}$$

et bang ! Encore le même rapport que celui qu'on avait trouvé à l'équation (11.203) en parlant de vitesses. Si tu regardes la figure 11.3, tu verras que réellement en faisant tendre x vers a on obtient la tangente.

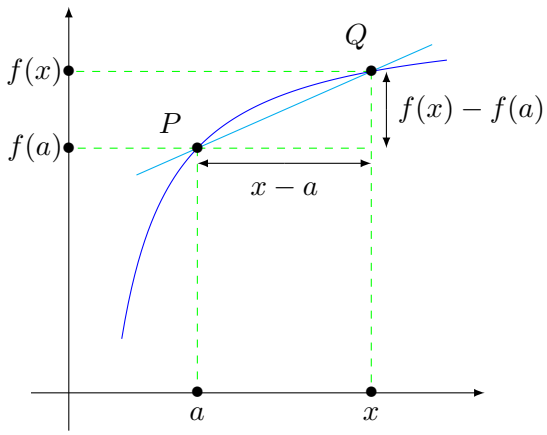
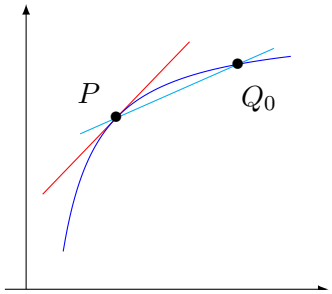
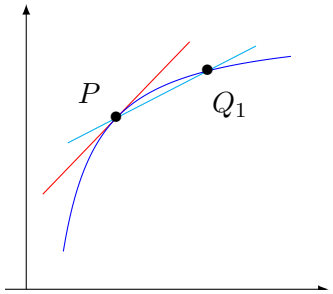


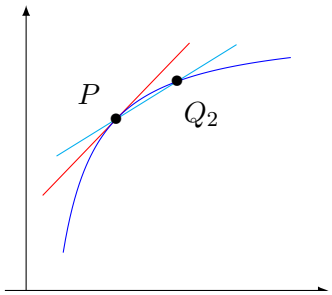
FIGURE 11.2 – Traçons d'abord une corde entre le point P et un point Q un peu plus loin.



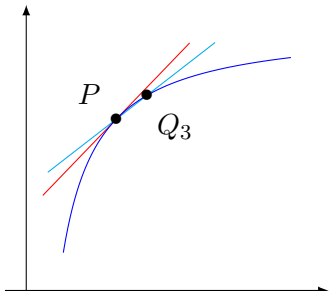
(a) Pas très bon ...



(b) ...de mieux en mieux ...



(c) ...de mieux en mieux ...



(d) ...presque parfait

FIGURE 11.3 – Recherche de la tangente par approximations successives.

11.11.3 L'aire en dessous d'une courbe

Encore un exemple. Nous voudrions bien pouvoir calculer l'aire en dessous d'une courbe. Nous notons $S_f(x)$ l'aire en dessous de la fonction f entre l'abscisse 0 et x, c'est-à-dire l'aire bleue de la figure 11.4.

Si la fonction f est continue et que Δx est assez petit, la fonction ne varie pas beaucoup entre x et $x + \Delta x$. L'augmentation de surface entre x et $x + \Delta x$ peut donc être approximée par le rectangle de surface $f(x)\Delta x$. Ce que nous avons donc, c'est que quand Δx est très petit,

$$S_f(x + \Delta x) - S_f(x) = f(x)\Delta x, \tag{11.206}$$

c'est-à-dire

$$f(x) = \lim_{\Delta x \to 0} \frac{S_f(x + \Delta x) - S_f(x)}{\Delta x}. \tag{11.207}$$

Donc, la fonction f est la dérivée de la fonction qui représente l'aire en dessous de f. Calculer des surfaces revient donc au travail inverse de calculer des dérivées.

Nous avons déjà vu que calculer la dérivée d'une fonction n'est pas très compliqué. Aussi étonnant que cela puisse paraitre, il se fait que le processus inverse est très compliqué : il est en

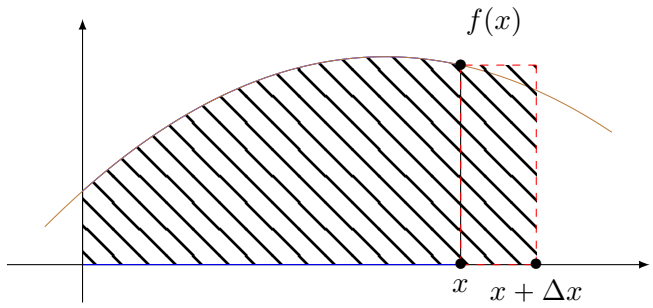


FIGURE 11.4 – L'aire en dessous d'une courbe. Le rectangle rouge d'aire $f(x)\Delta x$ approxime l'augmentation de l'aire lorsqu'on passe de x à $x + \Delta x$.

général extrêmement difficile (et même souvent impossible) de trouver une fonction dont la dérivée est une fonction donnée.

Une fonction dont la dérivée est la fonction f s'appelle une **primitive** de f, et la fonction qui donne l'aire en dessous de la fonction f entre l'abscisse 0 et x est notée

$$S_f(x) = \int_0^x f(t)dt. \tag{11.208}$$

Nous pouvons nous demander si, pour une fonction f donnée, il existe une ou plusieurs primitives, c'est-à-dire s'il existe une ou plusieurs fonctions F telles que $F' = f$. La réponse viendra...

11.12 Dérivation de fonctions réelles

On considère dans la suite une fonction $f : A \to \mathbb{R}$, où $a \in A \subset \mathbb{R}$; cependant, les notions de continuité et de dérivabilité se généralisent immédiatement au cas de fonctions à valeurs vectorielles ; la notion de continuité se généralise au cas des fonctions à plusieurs variables (la notion de dérivabilité est remplacée par celle de différentiabilité dans ce cadre).

Définition 11.105.
*La fonction f est **dérivable** en a si $a \in \operatorname{int} A$ et si*

$$\lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

*existe. On note alors cette quantité $f'(a)$, c'est le nombre dérivé de f en a. La **fonction dérivée** de f est*

$$\begin{aligned} f' \colon A' &\to \mathbb{R} \\ a &\mapsto f(a) \end{aligned} \tag{11.209}$$

définie sur l'ensemble noté A' des points a où f est dérivable.

Exemple 11.106.
Montrons que la fonction $f : \mathbb{R} \to \mathbb{R} : x \mapsto x$ est continue et dérivable. Exceptionnellement (bien qu'on sache que la dérivabilité implique la continuité), montrons ces deux assertions séparément.

Continuité Pour prouver la continuité au point $a \in \mathbb{R}$ nous devons montrer que

$$\lim_{x \to a} x = a \tag{11.210}$$

c'est-à-dire

$$\forall \epsilon > 0, \exists \delta > 0 : \forall x \in \mathbb{R}\, |x - a| < \delta \Rightarrow |x - a| < \epsilon \tag{11.211}$$

ce qui est clair en prenant $\delta = \epsilon$.

Dérivabilité Soit $a \in \mathbb{R}$. Calculons la limite du quotient différentiel

$$\lim_{\substack{x \to a \\ x \neq a}} \frac{x-a}{x-a} = \lim_{\substack{x \to a \\ x \neq a}} 1 = 1 \tag{11.212}$$

ce qui prouve que f est dérivable et que sa dérivée vaut 1 en tout point a de $\mathbb{R}$.

On a donc montré que la fonction $x \mapsto x$ est continue, dérivable, et que sa dérivée vaut 1 en tout point a de son domaine.

△

Proposition 11.107.
Une fonction dérivable sur un intervalle est continue sur cet intervalle.

Démonstration. Soit I un intervalle sur lequel la fonction f est dérivable, et soit $x_0 \in I$. Nous allons prouver la continuité de f en x_0. Le fait que la limite

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0+h) - f(x_0)}{h} \tag{11.213}$$

existe implique a fortiori que

$$\lim_{h \to 0} f(x_0+h) - f(x_0) = 0. \tag{11.214}$$

Cela signifie la continuité de f en vertu du critère 11.63. □

Théorème 11.108.
Toute fonction dérivable en un point est continue en ce point.

Démonstration. Soient $f\colon \mathbb{R} \to \mathbb{R}$ et $a \in \mathbb{R}$. Nous supposons que f n'est pas continue en a et nous allons en déduire qu'elle n'est pas non plus dérivable en a. Pour cela nous considérons le lien entre limite et continuité donné dans le théorème 11.63. Nier que f est continue en a revient à dire qu'il existe un voisinage V de $f(a)$ tel que

$$\forall r > 0, \exists \epsilon < r \text{ tel que } f(a+\epsilon) \notin V. \tag{11.215}$$

Si $B\big(f(a), R\big) \subset V$ [32], et si $r = 1/n$, nous construisons une suite $\epsilon_n \to 0$ telle que

$$|f(a+\epsilon_n) - f(a)| > R. \tag{11.216}$$

Avec cela nous avons

$$\frac{|f(a+\epsilon_n) - f(a)|}{\epsilon_n} > \frac{R}{\epsilon_n} \to \infty. \tag{11.217}$$

Donc la fonction f ne peut pas être dérivable en a. □

Remarque 11.109.
La réciproque du théorème précédent n'est pas vraie : il existent bien des fonctions qui sont continues à un point x_0 mais qui ne sont pas dérivables en x_0. La fonction valeur absolue, $x \mapsto |x|$, par exemple est continue sur tout $\mathbb{R}$ mais elle n'est pas dérivable en 0.

Si f est une fonction dérivable, il peut arriver que la fonction dérivée f' soit elle-même dérivable. Dans ce cas nous notons f'' ou $f^{(2)}$ la dérivée de la fonction f'. Cette fonction f'' est la **dérivée seconde** de f. Elle peut encore être dérivable ; dans ce cas nous notons $f^{(3)}$ sa dérivée, et ainsi de suite. Nous définissons $f^{(n)} = (f^{(n-1)})'$ la dérivée n^{e}de f. Nous posons évidemment $f^{(0)} = f$.

32. Existence par la définition de la topologie métrique 7.94.

11.12.1 Exemples

Exemple 11.110.
Commençons par la fonction $f(x) = x$. Dans ce cas nous avons

$$\frac{f(x) - f(a)}{x - a} = \frac{x - a}{x - a} = 1. \tag{11.218}$$

La dérivée est donc 1. △

Proposition 11.111.
La dérivé de la fonction $x \mapsto x$ vaut 1, en notations compactes : $(x)' = 1$.

Démonstration. D'après la définition de la dérivée, si $f(x) = x$, nous avons

$$f(x) = \lim_{\epsilon \to 0} \frac{(x + \epsilon) - x}{\epsilon} = \lim_{\epsilon \to 0} \frac{\epsilon}{\epsilon} = 1, \tag{11.219}$$

et c'est déjà fini. □

11.12.1.1 La fonction carré

Prenons ensuite $f(x) = x^2$. En utilisant le produit remarquable $(x^2 - a^2) = (x - a)(x + a)$ nous trouvons

$$\frac{f(x) - f(a)}{x - a} = x + a. \tag{11.220}$$

Lorsque $x \to a$, cela devient $2a$. Nous avons par conséquent

$$f'(x) = 2x. \tag{11.221}$$

Lemme 11.112.
Si $f(x) = x^2$, alors $f'(x) = 2x$.

Démonstration. Utilisons la définition, et remplaçons f par sa valeur :

$$\begin{aligned}
f'(x) &= \lim_{\epsilon \to 0} \frac{f(x + \epsilon) - f(x)}{\epsilon} && \text{(11.222a)}\\
&= \lim_{\epsilon \to 0} \frac{(x + \epsilon)^2 - x^2}{\epsilon} && \text{(11.222b)}\\
&= \lim_{\epsilon \to 0} \frac{x^2 + 2x\epsilon + \epsilon^2 - x^2}{\epsilon} && \text{(11.222c)}\\
&= \lim_{\epsilon \to 0} \frac{\epsilon(2x + \epsilon)}{\epsilon} && \text{(11.222d)}\\
&= \lim_{\epsilon \to 0} (2x + \epsilon) && \text{(11.222e)}\\
&= 2x, && \text{(11.222f)}
\end{aligned}$$

ce qu'il fallait prouver. □

11.12.1.2 La fonction racine carré

Considérons maintenant la fonction $f(x) = \sqrt{x}$. Nous avons

$$\begin{aligned}
\frac{f(x) - f(a)}{x - a} &= \frac{\sqrt{x} - \sqrt{a}}{x - a}\\
&= \frac{(\sqrt{x} - \sqrt{a})(\sqrt{x} + \sqrt{x})}{(x - a)(\sqrt{x} + \sqrt{x})}\\
&= \frac{1}{\sqrt{x} + \sqrt{x}}.
\end{aligned} \tag{11.223}$$

Lorsque $x \to 0$, nous obtenons

$$f'(a) = \frac{1}{2\sqrt{a}}. \tag{11.224}$$

Notons que la dérivée de $f(x) = \sqrt{x}$ n'existe pas en $x = 0$. En effet elle serait donnée par le quotient

$$f'(0) = \lim_{x\to 0} \frac{\sqrt{x} - \sqrt{0}}{x} = \lim_{x\to 0} \frac{\sqrt{x}}{x} = \lim_{x\to 0} \frac{1}{\sqrt{x}}. \tag{11.225}$$

Mais si x devient très petit, la dernière fraction tend vers l'infini.

11.12.2 Interprétation géométrique de la dérivée : tangente

Considérons le **graphe** de la fonction f sur I, c'est-à-dire l'ensemble

$$\big\{\big(x, f(x)\big) \text{ tel que } x \in I\big\}. \tag{11.226}$$

Le nombre

$$\frac{f(x) - f(a)}{x - a} \tag{11.227}$$

est la pente de la droite qui joint les points $\big(x, f(x)\big)$ et $\big(a, f(a)\big)$, voir la figure 11.5.

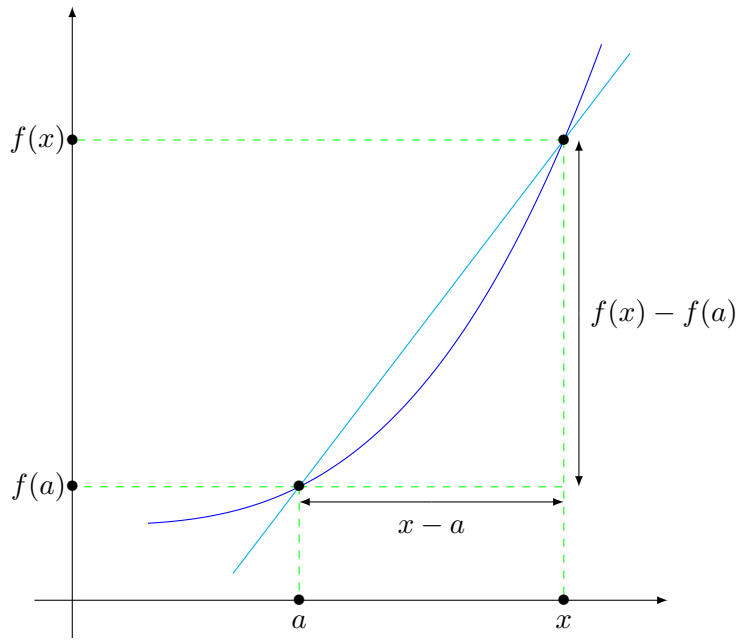


FIGURE 11.5 – Le coefficient directeur de la corde entre a et x.

Étant donné que $f'(a)$ est le coefficient directeur de la tangente au point $\big(a, f(a)\big)$, l'équation de la tangente est

$$y - f(a) = f'(a)(x - a). \tag{11.228}$$

11.12.3 Interprétation géométrique de la dérivée : approximation affine

Le fait que la fonction f soit dérivable au point $a \in I$ signifie que

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a} = \ell \tag{11.229}$$

pour un certain nombre ℓ. Cela peut être récrit sous la forme

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a} - \ell = 0, \tag{11.230}$$

ou encore

$$\lim_{x\to a} \frac{f(x)-f(a)-\ell(x-a)}{x-a} = 0. \tag{11.231}$$

Introduisons la fonction

$$\alpha(t) = \frac{f(a+t)-f(a)-t\ell}{t}. \tag{11.232}$$

Cette fonction est faite exprès pour que

$$\alpha(x-a) = \frac{f(x)-f(a)-\ell(x-a)}{x-a}; \tag{11.233}$$

par conséquent $\lim_{x\to a} \alpha(x-a) = 0$. Nous récrivons l'équation (11.233) sous la forme

$$f(x)-f(a)-\ell(x-a) = (x-a)\alpha(x-a). \tag{11.234}$$

Le second membre tend vers zéro lorsque x tend vers a avec une « vitesse au carré » : c'est le produit de deux facteurs tous deux tendant vers zéro. Si x n'est pas très loin de a, il n'est donc pas une mauvaise approximation de dire

$$f(x)-f(a)-\ell(x-a) \simeq 0, \tag{11.235}$$

c'est-à-dire

$$f(x) \simeq f(a)+f'(a)(x-a). \tag{11.236}$$

Nous avons retrouvé l'équation (11.228). La manipulation que nous venons de faire revient donc à dire que la fonction f, au voisinage de a, est bien approximée par sa tangente.

L'équation (11.236) peut être aussi écrite sous la forme

$$f(x+\Delta x) \simeq f(x)+f'(x)\Delta x \tag{11.237}$$

qui est une approximation d'autant meilleure que Δx est petit.

11.12.4 Développement limité au premier ordre

Si une fonction est dérivable en a alors elle peut être approximée « au premier ordre » par une formule simple qui sera généralisé pour des dérivées d'ordre supérieurs avec les séries de Taylor, théorème 11.375.

Proposition 11.113 (Développement limité au premier ordre)**.**
Si $f\colon \mathbb{R}\to\mathbb{R}$ est une fonction dérivable, alors is existe une fonction $\alpha\colon \mathbb{R}\to\mathbb{R}$ telle que

$$f(a+h) = f(a)+hf'(a)+\alpha(h) \tag{11.238}$$

et

$$\lim_{h\to 0} \frac{\alpha(h)}{h} = 0. \tag{11.239}$$

Démonstration. La fonction f étant dérivable en a nous avons l'existence de la limite suivante :

$$f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}, \tag{11.240}$$

ce qui revient à dire qu'en définissant la fonction β par

$$f'(a) = \frac{f(a+h)-f(a)}{h} + \beta(h) \tag{11.241}$$

alors $\beta(h)\to 0$ lorsque $h\to 0$. En multipliant par h et en nommant $\alpha(h) = h\beta(h)$ nous trouvons le résultat :

$$f(a+h) = f(a)+hf'(a)+\alpha(h) \tag{11.242}$$

avec

$$\lim_{h\to 0} \frac{\alpha(h)}{h} = \lim_{h\to 0} \beta(h) = 0. \tag{11.243}$$

□

11.13 Règles de calcul

D'abord une dérivée facile, qui sera utile pour démontrer la formule de dérivation d'un quotient.

Lemme 11.114.
Nous avons :

$$\left(\frac{1}{x}\right)' = -\frac{1}{x^2}. \tag{11.244}$$

Démonstration. En posant $f(x) = 1/x$, nous avons le calcul

$$\frac{f(x+\epsilon) - f(x)}{\epsilon} = \frac{\frac{1}{x+\epsilon} - \frac{1}{x}}{\epsilon} = \frac{x - (x+\epsilon)}{\epsilon x(x+\epsilon)} = \frac{-1}{x(x+\epsilon)}. \tag{11.245}$$

Nous trouvons le résultat en passant à la limite et en tenant compte de la proposition 11.20 sur la limite d'un quotient. □

Proposition 11.115 ([238, 239, 240])**.**
Nous avons les règles suivantes.

(1) Si $f, g\colon \mathbb{R} \to \mathbb{R}$ sont dérivables en $a \in \mathbb{R}$, alors $f + g$ est dérivable en a et

$$(f+g)'(a) = f'(a) + g'(a). \tag{11.246}$$

(2) Si $f\colon \mathbb{R} \to \mathbb{R}$ est dérivable en $a \in \mathbb{R}$ et si $\lambda \in \mathbb{R}$, alors (λf) est dérivable en a et

$$(\lambda f)'(a) = \lambda f'(a). \tag{11.247}$$

(3) Si $f, g\colon \mathbb{R} \to \mathbb{R}$ sont dérivables en $a \in \mathbb{R}$, alors fg est dérivable en a et

$$(fg)'(a) = f'(a)g(a) + f(a)g'(a). \tag{11.248}$$

*Cette formule est appelée **règle de Leibnitz**.*

(4) Soient deux intervalles I, J dans $\mathbb{R}$. Soient des fonctions $f\colon I \to J$ et $g\colon J \to \mathbb{R}$. Soit encore $a \in I$. SI f est dérivable en a et si g est dérivable en $f(a)$, alors $g \circ f$ est dérivable en a et

$$(g \circ f)'(a) = g'\big(f(a)\big)f'(a). \tag{11.249}$$

(5) Soient $f, g\colon I \to \mathbb{R}$ des fonction sur un intervalle ouvert I. Soit $a \in I$; supposons que $g(a) \neq 0$. Alors la fonction $\frac{f}{g}$ est dérivable en a et

$$\left(\frac{f}{g}\right)'(a) = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}. \tag{11.250}$$

En particulier, la dérivation est une opération linéaire sur l'espace des fonctions infiniement dérivables.

Démonstration. Point par point.

Pour (1)

Pour (2) Écrivons la définition de la dérivée avec (λf) au lieu de f, et calculons un petit peu :

$$(\lambda f)'(x) = \lim_{\epsilon \to 0} \frac{(\lambda f)(x+\epsilon) - (\lambda f)(x)}{\epsilon} \tag{11.251a}$$

$$= \lim_{\epsilon \to 0} \frac{\lambda\big(f(x+\epsilon)\big) - \lambda f(x)}{\epsilon} \tag{11.251b}$$

$$= \lim_{\epsilon \to 0} \lambda \frac{f(x+\epsilon) - f(x)}{\epsilon} \tag{11.251c}$$

$$= \lambda \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon} \tag{11.251d}$$

$$= \lambda f'(x). \tag{11.251e}$$

Pour (3), règle de Leibnitz La définition de la dérivée dit que

$$(fg)'(x) = \lim_{\epsilon\to 0} \frac{f(x+\epsilon)g(x+\epsilon) - f(x)g(x)}{\epsilon}. \tag{11.252}$$

La subtilité est d'ajouter au numérateur la quantité $-f(x)g(x+\epsilon) + f(x)g(x+\epsilon)$, ce qui est permis parce que cette quantité est nulle [33]. Le numérateur de (11.252) devient donc

$$\begin{aligned} f(x+\epsilon)g(x+\epsilon) - f(x)g(x+\epsilon) + f(x)g(x+\epsilon) - f(x)g(x) \\ = g(x+\epsilon)\big(f(x+\epsilon) - f(x)\big) + f(x)\big(g(x+\epsilon) - g(x)\big), \end{aligned} \tag{11.253}$$

où nous avons effectué deux mises en évidence. Étant donné que nous avons deux termes, nous pouvons couper la limite en deux :

$$\begin{aligned} (fg)'(x) &= \lim_{\epsilon\to 0} g(x+\epsilon)\frac{f(x+\epsilon)-f(x)}{\epsilon} &&+ \lim_{\epsilon\to 0} f(x)\frac{g(x+\epsilon)-g(x)}{\epsilon} \\ &= \lim_{\epsilon\to 0} g(x+\epsilon) \lim_{\epsilon\to 0} \frac{f(x+\epsilon)-f(x)}{\epsilon} &&+ f(x) \lim_{\epsilon\to 0} \frac{g(x+\epsilon)-g(x)}{\epsilon}, \end{aligned} \tag{11.254}$$

où nous avons utilisé le théorème 11.23 pour scinder la première limite en deux, ainsi que la propriété (11.24) pour sortir le $f(x)$ de la limite dans le second terme. Maintenant, dans le premier terme, nous avons évidemment [34] $\lim_{\epsilon\to 0} g(x+\epsilon) = g(x)$. Les limites qui restent sont les définitions classiques des dérivées de f et g au point x :

$$(fg)'(x) = g(x)f'(x) - f(x)g'(x), \tag{11.255}$$

ce qu'il fallait démontrer.

Pour (4) Nous posons $b = f(a)$ et nous considérons la fonction suivante :

$$\begin{aligned} u\colon J &\to \mathbb{R} \\ y &\mapsto u(y) = \begin{cases} \frac{g(y)-g(b)}{y-b} & \text{si } y \neq b \\ g'(b) & \text{si } y = b. \end{cases} \end{aligned} \tag{11.256}$$

Vu que g est dérivable en b, la seconde ligne existe et u est continue en $y = b = f(a)$. C'est la définition de la dérivée.

Mais f est continue en a, donc $u \circ f$ est également continue en a, et nous avons

$$\lim_{x\to a}(u\circ f)(x) = u\big(f(a)\big) = u(b) = g'(b). \tag{11.257}$$

En récrivant la définition de u en $f(x)$, l'expression suivante est une fonction continue de x :

$$u\big(f(x)\big) = \begin{cases} \frac{g\big(f(x)\big)-g(b)}{f(x)-b} & \text{si } f(x) \neq b \\ g'(b) & \text{si } y = b. \end{cases} \tag{11.258}$$

Si $f(x) \neq b$ nous avons :

$$g\big(f(x)\big) - g(b) = u\big(f(x)\big)\big(f(x) - b\big). \tag{11.259}$$

Si par contre $f(x) = b$, en réalité, l'égalité (11.259) est encore valable parce qu'elle se résume à $0 = 0$. Nous divisons par $x - a$ et nous avons l'égalité

$$\frac{g\big(f(x)\big) - f\big(f(a)\big)}{x-a} = u\big(f(x)\big)\frac{f(x)-f(a)}{x-a} \tag{11.260}$$

33. Nous avons déjà faut le coup d'ajouter et enlever la même chose durant la démonstration du théorème 11.23. C'est une technique assez courante en analyse.

34. Pas tout à fait évidemment : selon le théorème 11.63, *limite et continuité*, il faut que g soit continue.

qui est valable sur $I\backslash\{a\}$.

Il ne s'agit pas maintenant de prendre la limite $x \to a$ des deux côtés, parce que la limite du membre de gauche est précisément ce que ce théorème s'efforce de prouver exister. Nous montrons que la limite du membre de gauche existe en montrant que celle de droite existe.

D'une part, $u \circ f$ est continue et

$$\lim_{x\to a} u\big(f(x)\big) = u\big(f(a)\big) = u(b) = g'(b). \tag{11.261}$$

D'autre par, f est dérivable en a, donc

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a} = f'(a). \tag{11.262}$$

Tout cela pour dire qu'à droite, la limite existe et vaut $g'(b)f'(a)$. Donc nous avons l'existence de la limite que nous définissant $(g \circ f)'(a)$, et la valeur

$$\lim_{x\to a} \frac{g\big(f(x)\big) - f\big(f(a)\big)}{x - a} = g'\big(f(a)\big)f'(a). \tag{11.263}$$

Le résultat est prouvé.

Pour (5) Nous considérons la fonction

$$\begin{aligned} i\colon \mathbb{R}\backslash\{0\} &\to \mathbb{R} \\ x &\mapsto \frac{1}{x}. \end{aligned} \tag{11.264}$$

La fonction g est dérivable en a, la fonction i est dérivable en $g(a)$. Donc par le théorème de dérivation des fonctions composées[35], la fonction $i \circ f$ est dérivable en a et

$$(i \circ g)'(a) = i'\big(g(a)\big)g'(a) = -\frac{g'(a)}{g(a)^2}. \tag{11.265}$$

Pour le quotient, nous utilisons la formule de la dérivée du produit sur $\frac{f}{g}(x) = f(x)\frac{1}{g(x)}$ pour dire que f/g est dérivable en a et

$$\left(\frac{f}{g}\right)'(a) = f'(a)\frac{1}{g(a)} + f(a)\left(\frac{1}{g}\right)'(a) = \frac{f'(a)}{g(a)} - \frac{f(a)g'(a)}{g(a)^2} = \frac{f'(a)g(a) - f(a)g'(a)}{g(a)^2}, \tag{11.266}$$

ce qu'il fallait démontrer.

□

Remarque 11.116.
Nous ne pouvons pas dire que la dérivée est une opération linéaire sur l'espace des fonctions dérivables. Certes la proposition 11.115 implique entre autres que l'ensemble des fonctions dérivables est un espace vectoriel. Mais la dérivée d'une fonction dérivable n'est pas spécialement dérivable.

Remarque 11.117.
La formule $(1/u)' = -u'/u^2$ ne peut pas être vue comme un cas particulier de $(u^\alpha)' = \alpha u^{\alpha-1}$ (proposition 11.363) parce que cette formule est utilisée dans la démonstration de la formule générale.

Pour les fonctions à valeurs dant $\mathbb{R}^n$, nous posons la définition suivante.

Définition 11.118.
Soit une fonction $f\colon \mathbb{R} \to \mathbb{R}^n$ dont les composantes $f_i\colon \mathbb{R} \to \mathbb{R}$ sont dérivables. Nous définissons la fonction f' par

$$f'(x) = \sum_i f_i'(x)e_i, \tag{11.267}$$

c'est-à-dire une dérivation composante par composante.

35. Proposition 11.115(4).

Cette définition est celle pour une fonction $\mathbb{R} \to \mathbb{R}^n$, et elles est facile. Très différente est la situation d'une fonction $\mathbb{R}^n \to \mathbb{R}$ dans laquelle il faudra introduire la notion de différentielle [36].

Par rapport à la dérivation, les produits scalaire et vectoriel vérifient une règle de Leibnitz.

Proposition 11.119.
Soit I un intervalle de $\mathbb{R}$. Si u et u sont dans $C^1(I, \mathbb{R}^3)$, alors

$$\begin{aligned} \frac{d}{dt}\big(u(t) \cdot v(t)\big) &= \big(u'(t) \cdot v(t)\big) + \big(u(t) \cdot v'(t)\big) \\ \frac{d}{dt}\big(u(t) \times v(t)\big) &= \big(u'(t) \times v(t)\big) + \big(u(t) \times v'(t)\big). \end{aligned} \tag{11.268}$$

Nous faisons la preuve pour le produit scalaire ; sans doute que le produit vectoriel sera la même chose.

Démonstration. Nous considérons des fonctions dérivables $f, g \colon \mathbb{R} \to \mathbb{R}^3$, et nous posons $\varphi(t) = f(t) \cdot g(t)$. En ce qui concerne la dérivée de la fonction $f \cdot g \colon \mathbb{R} \to \mathbb{R}$, nous devons étudier la limite

$$\lim_{\epsilon \to 0} \frac{\varphi(t+\epsilon) - \varphi(t)}{\epsilon} = \lim_{\epsilon \to 0} \frac{f(t+\epsilon) \cdot g(t+\epsilon) - f(t) \cdot g(t)}{\epsilon}. \tag{11.269}$$

La fonction f étant dérivable, la proposition 11.113 nous donne une fonction $\alpha \colon \mathbb{R} \to \mathbb{R}^3$ telle que

$$f(t+\epsilon) = f(t) + \epsilon f'(t) + \epsilon\alpha(\epsilon) \tag{11.270}$$

et $\lim_{\epsilon \to 0} \alpha(\epsilon) = 0$. En substituant cela dans le numérateur de (11.269) nous calculons un peu :

$$\begin{aligned} f(t+\epsilon) \cdot g(t+\epsilon) - f(t) \cdot g(t) &= \big(f(t) + \epsilon f'(t) + \epsilon\alpha(\epsilon)\big) \cdot \big(g(t) + \epsilon g'(t) + \epsilon\beta(\epsilon)\big) && \text{(11.271a)} \\ &\quad - f(t) \cdot g(t) && \text{(11.271b)} \\ &= \epsilon f(t) \cdot g'(t) + \epsilon^2 f'(t) \cdot \beta(\epsilon) && \text{(11.271c)} \\ &\quad + \epsilon\alpha(\epsilon) \cdot g(t) + \alpha(\epsilon)\epsilon^2 \cdot g'(t) + \epsilon^2\alpha(\epsilon) \cdot \beta(\epsilon). && \text{(11.271d)} \end{aligned}$$

En divisant cela par ϵ et en prenant la limite $\epsilon \to 0$, et nous reste

$$f(t) \cdot g'(t) + f'(t) \cdot g(t). \tag{11.272}$$

□

11.13.1 Dérivée de la réciproque

Proposition 11.120 ([241]).
Soit $f \colon I \to J = f(I)$ une fonction bijective, continue et dérivable [37]. Soient $x_0 \in I$ et $y_0 = f(x_0)$. Si $f'(x_0) \neq 0$ alors la fonction réciproque f^{-1} est dérivable en y_0 et sa dérivée est donnée par

$$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}. \tag{11.273}$$

Démonstration. Pour rappel, une fonction dérivable est toujours continue (proposition 11.107).

Prouvons que f^{-1} est dérivable au point $b = f(a) \in J$. Étant donné que f est dérivable en a, nous avons

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}. \tag{11.274}$$

Par ailleurs, étant donnée la continuité de f^{-1} donnée par la proposition 11.61(4), nous avons

$$\lim_{\epsilon \to 0} f^{-1}(b+\epsilon) = f^{-1}(b) = a. \tag{11.275}$$

36. Ce sera pour la définition 10.127.
37. Définition 11.105.

Nous pouvons donc remplacer dans (11.274) tous les x par $f^{-1}(b+\epsilon)$ et prendre la limite $\epsilon \to 0$ au lieu de $x \to a$:

$$\begin{aligned} f'(a) &= \lim_{\epsilon\to 0} \frac{f\big(f^{-1}(b+\epsilon)\big) - f(a)}{f^{-1}(b+\epsilon) - a} \\ &= \lim_{\epsilon\to 0} \frac{b+\epsilon - f(a)}{f^{-1}(b+\epsilon) - f^{-1}(b)} \\ &= \lim_{\epsilon\to 0} \frac{\epsilon}{f^{-1}(b+\epsilon) - f^{-1}(b)} \\ &= \frac{1}{\lim_{\epsilon\to 0} \frac{f^{-1}(b+\epsilon)-f^{-1}(b)}{\epsilon}} \\ &= \frac{1}{(f^{-1})'(b)}. \end{aligned} \tag{11.276}$$

Nous avons utilisé le fait que $f(a) = b$ et $a = f^{-1}(b)$. □

Proposition 11.121 ([242])**.**
Soit f une fonction dérivable et strictement monotone de l'intervalle I sur l'intervalle J (f est alors une bijection de I vers J). Si f' ne s'annule par sur I alors

(1) la fonction f est une bijection de I vers J,

(2) la fonction f^{-1} est dérivable sur J,

(3) et nous avons la formule

$$(f^{-1})' = \frac{1}{f' \circ f^{-1}}. \tag{11.277}$$

11.122.
Très souvent on préfère retenir la formule

$$(f^{-1})'(y_0) = \frac{1}{f'\left((f^{-1})(y_0)\right)} \tag{11.278}$$

Elle est très simple à retrouver : il suffit d'écrire

$$f^{-1}\big(f(x)\big) = x \tag{11.279}$$

puis de dériver les deux côtés par rapport à x en utilisant la règle de dérivation des fonctions composées :

$$(f^{-1})'\big(f(x)\big)f'(x) = 1. \tag{11.280}$$

Exemple 11.123 (difféomorphisme entre $\mathbb{R}$ et un ouvert borné)**.**
Nous cherchons à construire une application dérivable et d'inverse dérivable entre $\mathbb{R}$ (en entier) et un ouvert borné de $\mathbb{R}$. Il serait tentant de prendre l'application arc tangente

$$\begin{aligned} \arctan\colon \mathbb{R} &\to \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[, \\ x &\mapsto \arctan(x) \end{aligned} \tag{11.281}$$

mais elle ne sera définie que dans le théorème 17.30.

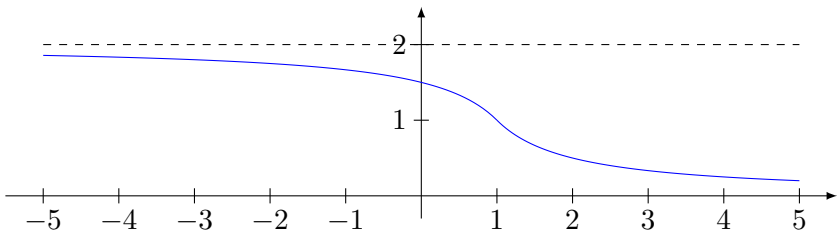
Nous posons

$$f(x) = \begin{cases} 2 + \frac{1}{x-2} & \text{si } x \leqslant 1 \\ \frac{1}{x} & \text{si } x > 1. \end{cases} \tag{11.282}$$

Cela est continue en $x = 1$: il suffit de calculer les deux valeurs. En ce qui concerne la dérivabilité en $x = 1$, nous devons faire

$$\lim_{\epsilon\to 0} \frac{f(1+\epsilon) - f(1)}{\epsilon}. \tag{11.283}$$

La limite à gauche est égale à la dérivée de $x \mapsto 2 + \frac{1}{x-2}$ en $x = 1$ et la limite à droite est égale à la dérivée de $x \mapsto 1/x$ en $x = 1$. Dans les deux cas nous trouvons -1.



Nous voyons vite que cette fonction est strictement décroissante ; et un calcul de limite nous dit qu'il s'agit d'une bijection dérivable

$$f\colon \mathbb{R} \to]0,2[. \tag{11.284}$$

La proposition 11.120 s'applique et la bijection réciproque est également dérivable (donc continue aussi). △

Problèmes et choses à faire

Si vous connaissez un autre exemple, plus simple, de difféomorphisme $f\colon \mathbb{R} \to]a,b[$, faites-le moi savoir. Ne pas utiliser d'exponentielle (vous pensiez à bricoler quelque chose à partir de la primitive de $x \mapsto e^{-x^2}$?) ni de fonctions trigonométriques.

Exemple 11.124.
Nous aimerions donner le logarithme comme exemple, mais l'exponentielle ne sera définie que dans longtemps à partir des séries entières. Allez voir l'exemple 14.64 pour le logarithme comme inverse de l'exponentielle. △

11.14 Dérivation et croissance

Supposons une fonction dont la dérivée est positive. Étant donné que la courbe est « collée » à ses tangentes, tant que les tangentes montent, la fonction monte. Or, une tangente qui monte correspond à une dérivée positive, parce que la dérivée est le coefficient directeur de la tangente.

Ce résultat très intuitif peut être prouvé rigoureusement. C'est la tache à laquelle nous allons nous atteler maintenant.

Proposition 11.125.
Si f et f' sont des fonctions continues sur l'intervalle $[a,b]$ et si f' est strictement positive sur $[a,b]$, alors f est croissante sur $[a,b]$.

De la même manière, si f' est strictement négative sur $[a,b]$, alors f est décroissante sur $[a,b]$.

Démonstration. Nous n'allons prouver que la première partie. La seconde partie se prouve en considérant $-f$ et en invoquant alors la première[38]. Prenons x_1 et x_2 dans $[a,b]$ tels que $x_1 < x_2$. Par hypothèse, pour tout x dans $[x_1,x_2]$, nous avons

$$f'(x) = \lim_{\epsilon\to 0} \frac{f(x+\epsilon)-f(x)}{\epsilon} > 0. \tag{11.285}$$

Maintenant, la proposition 11.22 dit que quand une limite est positive, alors la fonction dans la limite est positive sur un voisinage. En appliquant cette proposition à la fonction

$$r(\epsilon) = \frac{f(x+\epsilon)-f(x)}{\epsilon}, \tag{11.286}$$

dont la limite en zéro est positive, nous trouvons que $r(\epsilon) > 0$ pour tout ϵ pas trop éloigné de zéro. En particulier, il existe un $\delta > 0$ tel que $\epsilon < \delta$ implique $r(\epsilon) > 0$; pour un tel ϵ, nous avons donc

$$r(\epsilon) = \frac{f(x+\epsilon)-f(x)}{\epsilon} > 0. \tag{11.287}$$

38. Méditer cela.

Étant donné que $\epsilon > 0$, nous avons que $f(x+\epsilon) - f(x) > 0$, c'est-à-dire que f est strictement croissante entre x et $x+\epsilon$.

Jusqu'ici, nous avons prouvé que la fonction f était strictement croissante dans un voisinage autour de chaque point de $[a,b]$. Cela n'est cependant pas encore tout à fait suffisant pour conclure. Ce que nous voudrions faire, c'est de dire, c'est prendre un voisinage $]a,m_1[$ autour de a sur lequel f est croissante. Donc, $f(m_1) > f(a)$. Ensuite, on prend un voisinage $]m_1,m_2[$ de m_1 sur lequel f est croissante. De ce fait, $f(m_2) > f(m_1) > f(a)$. Et ainsi de suite, nous voulons construire des m_3, m_4,... jusqu'à arriver en b. Hélas, rien ne dit que ce processus va fonctionner. Il faut trouver une subtilité. Le problème est que les voisinages sur lesquels la fonction est croissante sont peut-être de plus en plus petits, de telle sorte à ce qu'il faille une infinité d'étapes avant d'arriver à bon port (en b).

Heureusement, nous pouvons drastiquement réduire le nombre d'étapes en nous souvenant du théorème de Borel-Lebesgue 7.124. Nous notons par $\mathcal{O}_x$, un ouvert autour de x tel que f soit strictement croissante sur $\mathcal{O}_x$. Un tel voisinage existe. Cela fait une infinité d'ouverts tels que

$$[a,b] \subseteq \bigcup_{x\in[a,b]} \mathcal{O}_x. \tag{11.288}$$

Ce que le théorème dit, c'est qu'on peut en choisir un nombre fini qui recouvre encore $[a,b]$. Soient $\{\mathcal{O}_{x_1},\dots,\mathcal{O}_{x_n}\}$, les heureux élus, que nous supposons pris dans l'ordre : $x_1 < x_2 < \dots < x_n$. Nous avons

$$[a,b] \subseteq \bigcup_{i=1}^{n} \mathcal{O}_i. \tag{11.289}$$

Quitte à les rajouter à la collection, nous supposons que $x_1 = a$ et que $x_n = b$. Maintenant nous allons choisir encore un sous-ensemble de cette collection d'ouverts. On pose $\mathcal{A}_1 = \mathcal{O}_{x_1}$. Nous savons que $\mathcal{A}_1$ intersecte au moins un des autres $\mathcal{O}_{x_i}$. Cette affirmation vient du fait que $[a,b]$ est connexe (proposition 7.153), et que si $\mathcal{O}_{x_1}$ n'intersectait personne, alors

$$\mathcal{O}_{x_1} \quad \text{et} \quad \bigcup_{i=2}^{n} \mathcal{O}_{x_i} \tag{11.290}$$

forment une partition de $[a,b]$ en deux ouverts disjoints, ce qui n'est pas possible parce que $[a,b]$ est connexe. Nous nommons $\mathcal{A}_2$, un des ouverts $\mathcal{O}_{x_i}$ qui intersecte $\mathcal{A}_1$. Disons que c'est $\mathcal{O}_k$. Notons que $\mathcal{A}_1 \cup \mathcal{A}_2$ est un intervalle sur lequel f est strictement croissante. En effet, si y_{12} est dans l'intersection, $f(a) < f(y_{12})$ parce que f est strictement croissante sur $\mathcal{A}_1$, et pour tout $x > y_{12}$ dans $\mathcal{A}_2$, $f(x) > f(y_{12})$ parce que f est strictement croissante dans $\mathcal{A}_2$.

Maintenant, nous éliminons de la liste des $\mathcal{O}_{x_i}$ tous ceux qui sont inclus dans $\mathcal{A}_1 \cup \mathcal{A}_2$. Dans ce qu'il reste, il y en a automatiquement un qui intersecte $\mathcal{A}_1 \cup \mathcal{A}_2$, pour la même raison de connexité que celle invoquée plus haut. Nous appelons cet ouvert $\mathcal{A}_3$, et pour la même raison qu'avant, f est strictement croissante sur $\mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$.

En recommençant suffisamment de fois, nous finissons par devoir prendre un des $\mathcal{O}_{x_i}$ qui contient b, parce qu'au moins un des $\mathcal{O}_{x_i}$ contient b. À ce moment, nous avons fini la démonstration. □

Il est intéressant de noter que ce théorème concerne la croissance d'une fonction sous l'hypothèse que la dérivée est positive. Il nous a fallu très peu de temps, en utilisant la positivité de la dérivée, pour conclure qu'autour de tout point, la fonction était strictement croissante. À partir de là, c'était pour ainsi dire gagné. Mais il a fallu un réel travail de topologie très fine [39] pour conclure. Étonnant qu'une telle quantité de topologie soit nécessaire pour démontrer un résultat essentiellement analytique dont l'hypothèse est qu'une limite est positive, n'est-ce pas ?

Une petite facile, maintenant.

39. et je te rappelle que nous avons utilisé la proposition 7.153, qui elle même était déjà un très gros boulot !

Proposition 11.126.
Si f est croissante sur un intervalle, alors $f' \geqslant 0$ à l'intérieur de cet intervalle, et si f est décroissante sur l'intervalle, alors $f' \leqslant 0$ à l'intérieur de l'intervalle.

Note qu'ici, nous demandons juste la croissance de f, et non sa *stricte* croissance.

Démonstration. Soit f, une fonction croissante sur l'intervalle I, et x un point intérieur de I. La dérivée de f en x vaut

$$f'(x) = \lim_{\epsilon \to 0} \frac{f(x+\epsilon) - f(x)}{\epsilon}, \tag{11.291}$$

mais, comme f est croissante sur I, nous avons toujours que $f(x+\epsilon) - f(x) \geqslant 0$ quand $\epsilon > 0$, et $f(x+\epsilon) - f(x) \leqslant 0$ quand $\epsilon < 0$, donc cette limite est une limite de nombres positifs ou nuls, qui est donc positive ou nulle. Cela prouve que $f'(x) \geqslant 0$. □

11.14.1 Théorèmes de Rolle et des accroissements finis

Théorème 11.127 (Théorème de Rolle[243, 244])**.**
Soit f, une fonction continue sur $[a,b]$ et dérivable sur $]a,b[$. Si $f(a) = f(b)$, alors il existe un point $c \in]a,b[$ tel que $f'(c) = 0$.

Démonstration. Étant donné que $[a,b]$ est un intervalle compact, l'image de $[a,b]$ par f est un intervalle compact, soit $[m,M]$ (théorème 7.92). Si $m = M$, alors le théorème est évident : c'est que la fonction est constante, et la dérivée est par conséquent nulle. Supposons que $M > f(a)$ (il se peut que $M = f(a)$, mais alors si f n'est pas constante, il faut avoir $m < f(a)$ et le reste de la preuve peut être adaptée).

Comme M est dans l'image de $[a,b]$ par f, il existe $c \in]a,b[$ tel que $f(c) = M$. Considérons maintenant la fonction

$$\tau(x) = \frac{f(c+x) - f(c)}{x}. \tag{11.292}$$

Par définition, $\lim_{x\to 0} \tau(x) = f'(c)$. Par hypothèse, si $u < c$,

$$\tau(u-c) = \frac{f(u) - f(c)}{u-c} > 0 \tag{11.293}$$

parce que $u - c < 0$ et $f(u) - f(c) < 0$. Par conséquent, $\lim_{x\to 0} \tau(x) \geqslant 0$. Nous avons aussi, pour $v > c$,

$$\tau(v-c) = \frac{f(v) - f(c)}{v-c} < 0 \tag{11.294}$$

parce que $v - c > 0$ et $f(v) - f(c) < 0$. Par conséquent, $\lim_{x\to 0} \tau(x) \leqslant 0$. Mettant les deux ensemble, nous avons $f'(c) = \lim_{x\to 0} \tau(x) = 0$, et c est le point que nous cherchions. □

Voici une généralisation du théorème de Rolle, dans le cas où nous n'aurions pas deux points sur lesquels la fonction est identique, mais deux points en lesquels la limite de la fonction est identique. Typiquement, lorsque les points en question sont $\pm\infty$.

Théorème 11.128 (Généralisation de Rolle[243])**.**
Soient $-\infty \leqslant a < b \leqslant +\infty$. Soit une fonction dérivable $f\colon]a,b[\to \mathbb{R}$ telle que

$$\lim_{x\to a} f(x) = \lim_{x\to b} f(x) = \ell \tag{11.295}$$

avec $\ell \in \bar{\mathbb{R}}$. Alors il existe $x \in]a,b[$ tel que $f'(x) = 0$.

Démonstration. Soit un difféomorphisme[40] strictement croissant $\varphi\colon \mathbb{R} \to]\alpha,\beta[$. Pour cela vous pouvez bricoler à partir de l'exemple 11.123. Mais n'utilisez pas la fonction arc tangente, parce qu'elle n'est définie qu'au théorème 17.30.

40. Définition 10.1.

Nous posons $a' = \varphi(a)$, $b' = \varphi(b)$ et

$$g = \varphi \circ f \circ \varphi^{-1} \colon]a', b'[\to]\alpha, \beta[. \tag{11.296}$$

Cela est une fonction dérivable et continue sur $[a', b']$ en posant $g(a') = g(b') = \varphi(\ell)$.

Donc il existe $c' \in]a', b'[$ tel que $g'(c') = 0$. En posant $c = \varphi^{-1}(c')$ nous avons $c \in]a, b[$ et, en utilisant de nombreuses fois la règle de dérivation des fonctions composées 11.115(4),

$$\begin{aligned}
f'(c) &= f'\big(\varphi^{-1}(c')\big) && \text{(11.297a)}\\
&= (\varphi^{-1})'\Big((g \circ \varphi)\big(\varphi^{-1}(c')\big)\Big)(g \circ \varphi)'\big(\varphi^{-1}(c')\big) && \text{(11.297b)}\\
&= (\varphi^{-1})'\big(g(c')\big)\underbrace{g'(c')}_{=0}\varphi'\big(\varphi^{-1}(c')\big) && \text{(11.297c)}\\
&= 0. && \text{(11.297d)}
\end{aligned}$$

□

Une autre généralisation de Rolle, avec des dérivées d'ordre supérieur.

Proposition 11.129 ([245]).
Soit un intervalle ouvert $I \subset \mathbb{R}$ contenant a, b ($a \neq b$). Soit une fonction $f \in C^{k+1}(I, \mathbb{R})$. Si $f(a) = f(b)$ et si $f^{(j)}(a) = 0$ pour $j = 1, \dots, n$, alors il existe $x \in]a, b[$ tel que $f^{(n+1)}(c) = 0$.

Démonstration. Le théorème de Rolle 11.127 nous dit qu'il existe $c_1 \in]a, b[$ tel que $f'(c_1) = 0$. Mais alors $f'(a) = f'(x_1) = 0$, et le théorème de Rolls appliqué à f' donne $c_2 \in]a, c_1[$ tel que $f''(c_2) = 0$. Continuant ainsi n fois, il existe $c \in]a, b[$ tel que $f^{(n+1)}(c) = 0$. □

Le théorème suivant est le théorème des **accroissements finis**.

Théorème 11.130 (Accroissements finis).
Soit f, une fonction continue sur $[a, b]$ et dérivable sur $]a, b[$.

(1) Il existe au moins un réel $c \in]a, b[$ tel que

$$f'(c) = \frac{f(b) - f(a)}{b - a} \tag{11.298}$$

Autrement dit, la tangente en c est parallèle à la corde entre a et b.

(2) Nous avons la majoration

$$\Big|\frac{f(b) - f(a)}{b - a}\Big| \leqslant \sup_{x \in [a,b]} |f'(x)||b - a|. \tag{11.299}$$

Démonstration. Considérons la fonction

$$\tau(x) = f(x) - \Big(\frac{f(b) - f(a)}{b - a}x + f(a) - a\frac{f(b) - f(a)}{b - a}\Big), \tag{11.300}$$

c'est-à-dire la fonction qui donne la distance entre f et le segment de droite qui lie $(a, f(a))$ à $(b, f(b))$. Par construction, $\tau(a) - \tau(b) = 0$, donc le théorème de Rolle s'applique à τ pour laquelle il existe donc un $c \in]a, b[$ tel que $\tau'(c) = 0$.

En utilisant les règles de dérivation, nous trouvons que la dérivée de τ vaut

$$\tau'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}, \tag{11.301}$$

donc dire que $\tau'(c) = 0$ revient à dire que $f(b) - f(a) = (b - a)f'(c)$, ce qu'il fallait démontrer.

La majoration est une conséquence immédiate, parce que le supremum de $|f'(x)|$ est forcément plus grand que $|f'(c)|$. □

Une généralisation pour une fonction sur un intervalle $]a,b[$ où a et b peuvent être infinis.

Théorème 11.131 (Généralisation des accroissements finis)**.**
Soient $-\infty \leqslant a < b \leqslant +\infty$ et f,g des fonctions continues sur $[a,b]$ et dérivables sur $]a,b[$.
Si $a=-\infty$:

— *Nous demandons la continuité sur $]-\infty,b]$ et la dérivabilité sur $]-\infty,b[$.*

— *Nous notons $f(a)$ la limite $\lim_{x\to-\infty} f(x)$, et nous supposons qu'elle est finie.*

Mêmes conventions si $b=+\infty$.
Alors il existe $c\in\,]a,b[$ tel que

$$\big(f(b)-f(a)\big)g'(c)=\big(g(b)-g(a)\big)f'(c). \tag{11.302}$$

Démonstration. Nous posons

$$h(t)=\big(g(b)-g(a)\big)f(t)-\big(f(b)-f(a)\big)g(t). \tag{11.303}$$

Nous avons $\lim_{t\to a} h(t)=\lim_{t\to b} h(t)$, de telle sorte que le théorème de Rolle généralisé 11.128 s'applique et il existe $c\in\,]a,b[$ tel que $h'(c)=0$. Pour ce c nous avons

$$0=h'(c)=\big(f(b)-f(a)\big)g'(c)-\big(g(b)-g(a)\big)f'(c), \tag{11.304}$$

et donc

$$\big(f(b)-f(a)\big)g'(c)=\big(g(b)-g(a)\big)f'(c). \tag{11.305}$$

□

11.14.2 Règle de l'Hospital

Proposition 11.132 (Règle de l'Hospital pour $\frac{0}{0}$[246])**.**
Soient des fonctions f,g dérivables sur $]a,b[$ et dont la limite en a est nulle. Si g' ne s'annule pas sur $]a,b[$ et si

$$\lim_{x\to a^+}\frac{f'(x)}{g'(x)}=\ell \tag{11.306}$$

alors

$$\lim_{x\to a^+}\frac{f(x)}{g(x)}=\ell. \tag{11.307}$$

Ici $\ell\in\bar{\mathbb{R}}$, et les hypothèses garantissent l'existence de la limite (11.307).

Démonstration. Soit $x\in\,]a,b[$. Les fonctions f et g sont dérivables sur $]a,x[$ et continues sur $[a,x]$, de telle sorte que le théorème 11.131 s'applique et nous avons $c_x\in\,]a,x[$ tel que

$$\big(f(x)-f(a)\big)g'(c_x)=\big(g(x)-g(a)\big)f'(c_x). \tag{11.308}$$

Nous nous souvenons de ce que signifient les notations dans le théorème : les notations $f(a)$, $f(x)$, $g(a)$ et $g(x)$ désignent en réalité les limites. Donc dans (11.308), nous avons $f(a)=g(a)=0$.

D'autre part nous avons $g(x)\neq g(a)$, sinon le théorème de Rolle 11.128 annulerait g' quelque part dans $]a,x[$. Nous pouvons donc récrire (11.308) sous la forme

$$\frac{f(x)}{g(x)}=\frac{f'(c_x)}{g'(c_x)}. \tag{11.309}$$

Mais $\lim_{x\to a^+} c_x=a$ parce que $c_x\in\,]a,x[$. Donc la limite du membre de droite de (11.309) lorsque $x\to a^+$ existe et vaut ℓ. La même limite à gauche doit alors exister et valoir la même valeur. □

Proposition 11.133 (L'Hospital pur $\frac{\infty}{\infty}$)**.**
Soit f et g deux fonctions

(1) dérivables sur $]a,b[$,

(2) dont les limites en a sont toutes deux ∞,

(3) $g' \neq 0$ sur $]a,b[$.

(4)

$$\lim_{x\to a^+} \frac{f'(x)}{g'(x)} = \ell \in \bar{\mathbb{R}}. \tag{11.310}$$

Alors

$$\lim_{x\to a^+} \frac{f(x)}{g(x)} = \ell. \tag{11.311}$$

Cette dernière égalité signifie « la limite existe et vaut ℓ ».

Démonstration. Soit un intervalle $]x,y[$ strictement inclus dans $]a,b[$ avec $x,y \in \mathbb{R}$. Par le théorème des accroissements finis généralisés 11.131 il existe $c \in]x,y[$ tel que

$$\frac{f(x)-f(y)}{g(x)-g(y)} = \frac{f'(c)}{g'(c)}. \tag{11.312}$$

Notons que le dénominateur à gauche n'est pas nul à cause de Rolle et de l'hypothèse que g' ne s'annule pas sur $[x,y]$. Nous isolons $f(x)$:

$$f(x) = \frac{f'(c)}{g'(c)}\Big(g(x)-g(y)\Big) + f(y). \tag{11.313}$$

Avant de diviser par $g(x)$ nous devons prendre quelques précautions. Soit V, un voisinage de ℓ [41]. Vu la limite (11.310), il existe $y \in]a,b[$ tel que

$$\frac{f'(t)}{g'(t)} \in V \tag{11.314}$$

pour tout $t \in]a,y[$. Nous utilisons ici avec subtilité le fait que ces intervalles sont une base de la topologie autour de ∞. Maintenant $f(y)$ et $g(y)$ sont fixés et sont des nombres réels. Vu que $\lim_{x\to a} g(x) = 0$ nous pouvons choisir $r < y$ tel que nous ayons simultanément

(1) $g(x) \neq 0$ sur $]a,r[$,

(2)

$$|\frac{g(y)}{g(x)}| \leqslant \epsilon \tag{11.315}$$

et

$$|\frac{f(y)}{g(x)}| \leqslant \epsilon \tag{11.316}$$

pour tout $x \in]a,r[$.

Nous sommes maintenant armés de y et r satisfaisant tout cela et nous pouvons traiter avec la formule (11.313) en ne la considérant que pour $x \in]a,r[$. Soit $x \in]a,r[$; il existe $c_x \in]a,x[$ tel que

$$\frac{f(x)}{g(x)} = \frac{f'(c_x)}{g'(c_x)}\left(1-\frac{g(y)}{g(x)}\right) + \frac{f(y)}{g(x)}. \tag{11.317}$$

Nous avons :

(1) $\lim_{x\to a^+} c_x = a$,

(2)

$$\lim_{x\to a^+} \frac{f'(c_x)}{g'(c_x)} = \lim_{x\to a^+} \frac{f'(x)}{g'(x)} = \ell, \tag{11.318}$$

(3)

$$\lim_{x\to a^+} \left(1-\frac{g(y)}{g(x)}\right) = 1, \tag{11.319}$$

41. Vous savez ce que signifie un « voisinage de ∞ » ? Allez voir la sous-section 7.7.15.

(4) $\lim_{x\to a^+} \frac{f(y)}{g(x)} = 0$.

Donc chaque partie du membre de droite de (11.317) a une limite bien déterminée pour $x \to a^+$. Les règles de calcul s'appliquent et nous avons

$$\lim_{x\to a^+} \frac{f(x)}{g(x)} = \ell \times 1 + 0 = \ell. \tag{11.320}$$

□

11.14.3 Dérivée et primitive

Corolaire 11.134.
Soit f une fonction dérivable sur $[a, b]$ telle que $f'(x) = 0$ pour tout $x \in [a, b]$. Alors f est constante sur $[a, b]$.

Démonstration. Si f n'était pas constante sur $[a, b]$, il existerait un $x_1 \in]a, b[$ tel que $f(a) \neq f(x_1)$, et dans ce cas, il existerait, par le théorème des accroissements finis 11.130, un $c \in]a, x_1[$ tel que

$$f'(c) = \frac{f(x_1) - f(a)}{x_1 - a} \neq 0, \tag{11.321}$$

ce qui contredirait les hypothèses. □

Corolaire 11.135.
Soient f et g, deux fonctions dérivables sur $[a, b]$ telles que

$$f'(x) = g'(x) \tag{11.322}$$

pour tout $x \in [a, b]$. Alors existe un réel C tel que $f(x) = g(x) + C$ pour tout $x \in [a, b]$.

Démonstration. Considérons la fonction $h(x) = f(x) - g(x)$, dont la dérivée est, par hypothèse, nulle. L'annulation de la dérivée entraine par le corolaire 11.135 que h est constante. Si $h(x) = C$, alors $f(x) = g(x) + C$, ce qu'il fallait prouver. □

Définition 11.136.
*Soit I un intervalle ouvert de $\mathbb{R}$ et une fonction $f\colon I \to \mathbb{R}$. La fonction $F\colon I \to \mathbb{R}$ est une **primitive** de f si F est dérivable sur I et si $F'(x) = f(x)$ pour tout x dans I.*

Exprimé en termes des primitives, le corolaire 11.135 signifie que

Corolaire 11.137.
Si F et G sont deux primitives de la même fonction f sur un intervalle, alors il existe une constante C pour laquelle $F(x) = G(x) + C$.

Cela signifie qu'il n'y a, en réalité, pas des milliards de primitives différentes à une fonction. Il y en a essentiellement une seule, et puis les autres, ce sont juste les mêmes, mais décalées d'une constante.

Remarque 11.138.
L'hypothèse de se limiter à un intervalle est importante parce que si on considère la fonction sur deux intervalles disjoints, nous pouvons choisir la constante indépendamment dans l'un et dans l'autre. Par exemple la fonction

$$F(x) = \begin{cases} \ln(x) + 1 & \text{si } x > 0 \\ \ln(x) - 7 & \text{si } x < 0 \end{cases} \tag{11.323}$$

est une primitive de $\frac{1}{x}$ sur l'ensemble $\mathbb{R}\backslash\{0\}$.

Certains ne s'en privent pas. Le logiciel Sage par exemple fait ceci :

```
sage: f(x)=1/x
sage: F=f.integrate(x)
sage: A=F(x)-F(-x)
sage: A.full_simplify()
I*pi
```

En réalité lorsque $x > 0$, Sage définit $\ln(-x) = \ln(x) + i\pi$. Cela a une certaine logique parce que $\ln(-1) = i\pi$ (du fait que $e^{i\pi} = -1$), mais si on ne le sait pas, ça peut étonner.

11.139.
Il existe plusieurs primitives à une fonction donnée. En physique, la constante arbitraire est souvent fixée par une condition initiale, comme nous le verrons dans la section 41.1.

11.15 Fonctions de plusieurs variables

La physique, et les sciences en général, regorgent de fonctions à plusieurs variables.

Accélération centripète [42] Si une masse m tourne sur un cercle, elle subira une accélération dirigée vers l'intérieur égale à

$$F(v, r) = \frac{mv^2}{r} \tag{11.324}$$

où r est le rayon du cercle et v est la vitesse.

Pression dans un gaz Si on a n moles de gaz dans un volume V a une température T, alors la pression sera donnée par la fonction de trois variables

$$p = \frac{nRT}{V} \tag{11.325}$$

où R est la constante des gaz parfaits.

En mathématique, on peut inventer de nombreuses fonctions de plusieurs variables. La fonction

$$f(x, y) = x^2 + xy\cos(x^2 + y^3) \tag{11.326}$$

est définie sur $\mathbb{R}^2$. La fonction

$$f(x, y, z) = \frac{x + y - 2z}{1 - x^2 - y^2 - z^2} \tag{11.327}$$

est définie sur $\mathbb{R}^3$ moins la sphère unité $\{x^2 + y^2 + z^2 = 1\}$.

La plus grande partie de ce cours est consacrée à l'étude des fonctions de plusieurs variables. Nous allons maintenant donner quelques indications sur comment «dessiner» une telle fonction. Vous connaissez déjà la définition de graphe pour une fonction f d'une seule variable à valeurs dans $\mathbb{R}$: c'est l'ensemble des points du plan de la forme $(x, f(x))$. Vous voyez que cet ensemble n'est pas vraiment un gros morceau de $\mathbb{R}^2$ parce que son intérieur est vide : il y a une seule valeur de f qui correspond au point x, donc une boule de $\mathbb{R}^2$ centrée en $(x, f(x))$ de n'importe quel rayon contient toujours des points qui ne font pas partie du graphe de f.

Nous voulons donner une définition assez générale pour le graphe d'une fonction

Définition 11.140.
*Soit f une fonction de $\mathbb{R}^m$ dans $\mathbb{R}^n$. Le **graphe** de f est la partie de $\mathbb{R}^m \times \mathbb{R}^n$ de la forme*

$$\operatorname{Graph} f = \{(x, y) \in \mathbb{R}^m \times \mathbb{R}^n \mid y = f(x)\}. \tag{11.328}$$

42. Appelez la « centrifuge » si vous voulez ; ça ne me fait ni chaud ni froid.

Cette définition se spécialise de la façon suivante dans les cas communs. Soit f une fonction de $\mathbb{R}^m$ dans $\mathbb{R}$. Le graphe de f est la partie de $\mathbb{R}^m \times \mathbb{R}$ donné par

$$\operatorname{Graph} f = \{(x, y) \in \mathbb{R}^m \times \mathbb{R} \,|\, y = f(x)\}. \tag{11.329}$$

Et pour les fonctions $\mathbb{R}^2 \to \mathbb{R}$:

$$\operatorname{Graph} f = \{(x, y, z) \in \mathbb{R}^2 \text{ tel que } z = f(x, y)\}. \tag{11.330}$$

C'est cette définition qu'il faut garder à l'esprit lorsqu'on travaille sur des dessins en trois dimensions.

Si f est une fonction de deux variables indépendantes x et y à valeurs dans $\mathbb{R}$, alors un point dans le graphe de f est un point $(x, y, z) \in \mathbb{R}^3$ tel que

$$z = f(x, y), \tag{11.331}$$

ou encore, un point de la forme

$$\big(x, y, f(x, y)\big). \tag{11.332}$$

Nous avons parfois besoin de donner des représentations graphiques d'une fonction. Nous pouvons, par exemple, penser à la fonction que associe à un point de la Terre son altitude. Lorsqu'on part pour une promenade en montagne on a envie de connaitre le graphe de cette fonction qui correspond en fait à la surface de la montagne. Bien sur nous ne voulons pas amener avec nous un modèle en 3D de la montagne donc il nous faut une méthode efficace pour projeter le graphe de f sur le plan x-y tout en gardant les informations fondamentales. Pour cela nous avons besoin de deux définitions (à ne pas confondre !)

Définition 11.141.
*Soit f une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ et soit c dans $\mathbb{R}$. La **z-section** de* $\operatorname{Graph} f$ *à la hauteur c est donné par*

$$S_c^z = \{(x, y, c) \in \mathbb{R}^3 \,|\, f(x, y) = c\}.$$

Définition 11.142.
*Soit f une fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ et soit c dans $\mathbb{R}$. La **courbe de niveau** de f à la hauteur c est l'ensemble*

$$N_c = \{(x1, \dots, x_n) \in \mathbb{R}^n \,|\, f(x1, \dots, x_n) = c\}. \tag{11.333}$$

On peut représenter la fonction f d'une façon très précise en traçant quelques-unes de ses courbes de niveau. Dans la suite on pourra considérer aussi les x-sections et les y-sections du graphe d'une fonction de deux variables. La x-section de $\operatorname{Graph} f$ à la hauteur a est

$$S_a^x = \{(a, y, z) \in \mathbb{R}^3 \,|\, f(a, y) = z\}.$$

Comme vous avez peut être déjà compris, S_a^x est le graphe de la fonction de y qu'on obtient de f en fixant $x = a$. Cette fonction est appelée x-section de f pour $x = a$.

Certaines surfaces dans $\mathbb{R}^3$ sont le graphe d'une fonction.

Exemple 11.143.
Quelques graphes importants.

Un plan non vertical Tout plan dans $\mathbb{R}^3$ peut être décrit par une équation de la forme

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = r,$$

où, (x_0, y_0, z_0) est vecteur dans $\mathbb{R}^3$, et a, b, c et r sont des nombres réels. Si $c \neq 0$ alors le plan n'est pas vertical et on peut dire que il est le graphe de la fonction

$$P(x, y) = \frac{r + cz_0 - a(x - x_0) - b(y - y_0)}{c},$$

quitte à choisir des nouvelles constantes s, t, q,

$$P(x, y) = sx + ty + q.$$

Un paraboloïde elliptique Pour tous α et β dans $\mathbb{R}$ les graphes des fonctions

$$PE_1(x,y) = \frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2}$$

ou de la fonction

$$PE_2(x,y) = -\frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2}$$

sont des paraboloïdes elliptiques. Le premier est contenu dans le demi-espace $z \geqslant 0$, l'autre dans $z \leqslant 0$. Le nom de cette surface vient de la forme de ses sections. En fait toutes sections S_c^z sont des ellipses, alors que les sections S_a^x et S_b^y sont des paraboles.

Un paraboloïde hyperbolique (selle) Pour tous α et β dans $\mathbb{R}$ les graphes des fonctions

$$PH_1(x,y) = \frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2}$$

ou de la fonction

$$PH_2(x,y) = -\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2}$$

sont des paraboloïdes hyperboliques. Remarquez que les sections S_c^z de ce graphe sont des hyperboles, alors que les sections S_a^x et S_b^y sont des paraboles.

Une demi-sphère La fonction $S^+ = \sqrt{R^2 - x^2 - y^2}$ a pour graphe la demi-sphère supérieure centrée en l'origine et de rayon R. Le dernier de ces exemples nous signale une chose très importante : une sphère entière n'est pas le graphe d'une fonction de x et y. Par contre, une demi-sphère est bien le graphe de la fonction $f(x,y) = \sqrt{1 - x^2 - y^2}$.

L'équation que nous utilisons pour d'écrire une sphère de rayon R centrée en l'origine est

$$x^2 + y^2 + z^2 = R^2$$

Donc, à chaque point (x,y) dans le disque $x^2 + y^2 \leqslant R^2$ (notez que ce disque est contenu dans la section S_0^z), on peut associer deux valeurs de z : $z_1 = \sqrt{R^2 - x^2 - y^2}$ et $z_2 = -\sqrt{R^2 - x^2 - y^2}$. Par définition, une fonction n'associe qu'un seul valeur à chaque point de son domaine, d'où l'impossibilité de décrire cette sphère comme le graphe d'une fonction de x et y.

△

Considérons la fonction $Sp : \mathbb{R}^3 \to \mathbb{R}$ qui associe à (x,y,z) la valeur $x^2 + y^2 + z^2$. La sphère de rayon R centrée en l'origine est l'ensemble de niveau N_{R^2} de Sp. L'ensemble de niveau N_0 de Sp est l'origine, et tous les ensembles de niveau de hauteur négative sont vides. La même chose est vraie pour les ellipsoïdes centrées en l'origine avec les axes x, y et z comme axes principaux et comme longueurs de demi-axes a, b et c. Voici la fonction dont il sont les ensembles de niveau

$$El(x,y,z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}.$$

Exemple 11.144.
Des ensembles de niveau importants.

Tout graphe Le graphe de toute fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ peut être considéré comme l'ensemble de niveau zéro de la fonction $F(x,y,z) = z - f(x,y)$.

Hyperboloïdes Les hyperboloïdes, comme les ellipsoïdes, sont une famille d'ensemble de niveau. En particulier, nous considérons des hyperboloïdes dont l'axe de symétrie est l'axe des z et qui sont symétriques par rapport un plan x-y. Une fois que les paramètres a, b et c sont fixés la fonction que nous intéresse est

$$Hyp(x,y,z) = \frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2}.$$

Les ensembles de niveau N_d pour $d > 0$ sont connexes, on les appelle *hyperboloïdes à une feuille*. L'ensemble de niveau N_0 est *cône (elliptique)*, les deux moitiés du cône se touchent en l'origine. Enfin, les ensembles de niveau N_d pour $d < 0$ ne sont pas connexes et pour cette raison on les appelle *hyperboloïdes à deux feuilles*.

△

11.15.1 Graphes de fonctions à plusieurs variables

Le **graphe** d'une fonction de deux variables $f\colon D \subset \mathbb{R}^2 \to \mathbb{R}$ est l'ensemble

$$\Big\{\big(x, y, f(x, y)\big) \text{ tel que } (x, y) \in D\Big\} \subset \mathbb{R}^3. \tag{11.334}$$

Ce graphe est une surface dans $\mathbb{R}^3$.

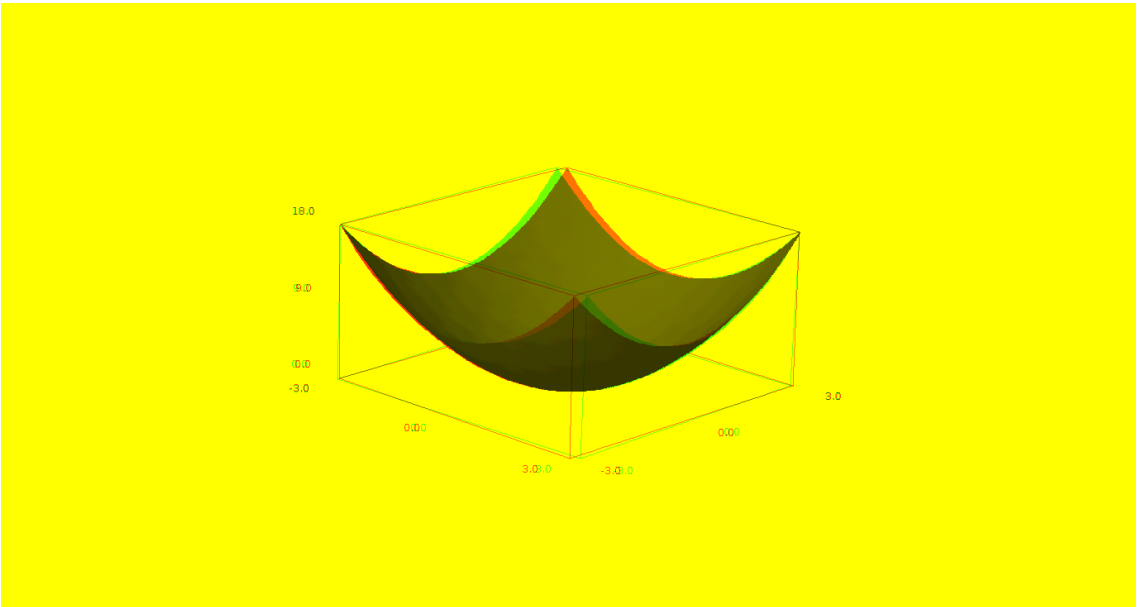
Exemple 11.145.
Tracer le graphe de la fonction

$$(x, y) \mapsto x^2 + y^2. \tag{11.335}$$

Le plus simple est de demander à Sage de nous fournir une représentation 3D

```
----------------------------------------------------------------------
| Sage Version 4.6.1, Release Date: 2011-01-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: f(x,y)=x**2+y**2
sage: plot3d(f,(x,-3,3),(y,-3,3))
```

Voici ce que cela donne[43] : (à regarder avec des lunettes bleues et rouges) :



À part que l'ordinateur l'a dit, est-ce qu'on peut comprendre pourquoi le graphe de la fonction $x^2 + y^2$ ressemble à un bol ? En coordonnées cylindriques, le graphe s'écrit

$$z = r^2. \tag{11.336}$$

Donc il se fait que plus on s'éloigne du point $(0, 0)$ dans le plan XY, plus le graphe va monter. Et il monte à quelle vitesse ? Il monte à la vitesse r^2. Il s'agit donc de dessiner la fonction $z = r^2$ dans le plan et de la « faire tourner ».

△

43. En vrai, ce que Sage donne est un objet qu'on peut même faire bouger.

11.15.2 Courbes de niveau

Une technique utile pour se faire une idée de la forme d'une fonction en trois dimensions est de tracer les **courbes de niveau**. La courbe de niveau de hauteur h est la courbe dans le plan donnée par l'équation

$$f(x, y) = h. \tag{11.337}$$

Exemple 11.146.
Dessinons par exemple les courbes de niveau de la fonction

$$f(x, y) = x + y + 2. \tag{11.338}$$

La courbe de niveau h est donnée par l'équation $x + y + 2 = h$, c'est-à-dire

$$y(x) = -x + h - 2. \tag{11.339}$$

Par conséquent la courbe de niveau de hauteur 0 est $y = -x - 2$, celle de hauteur 5 est $y = -x + 3$, etc.

Nous pouvons également nous aider de Sage pour ce faire :

```
----------------------------------------------------------------------
| Sage Version 4.6.1, Release Date: 2011-01-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: f(x,y)=x+y+2
sage: var('h')
h
sage: niveau(h,x)=solve(f(x,y)==h,y)[0].rhs()
sage: g1(x)=niveau(1,x)
sage: g1
x |--> -x - 1
```

Ici la fonction `g1` est la courbe de niveau 1.

Si on veut faire tracer une courbe de niveau, Sage peut le faire :

```
sage: implicit_plot(f(x,y)==1,(x,-3,3),(y,-4,4))
```

Cela tracera la courbe de niveau $h = 1$ dans la partie du plan $x \in [-3, 3]$ et $y \in [-4, 4,]$.

△

Il est bien entendu possible de créer automatiquement 50 courbes de niveau et de demander de les tracer toutes sur le même graphe.

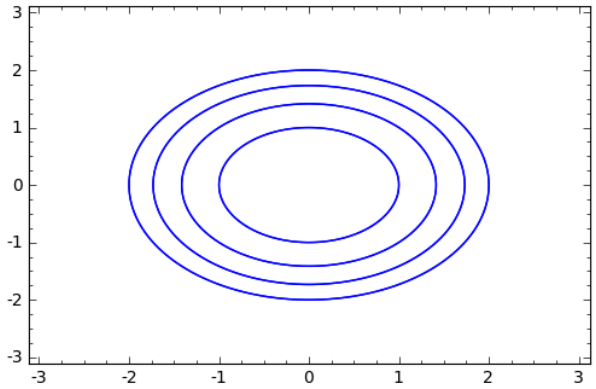
```
#! /usr/bin/sage -python
# -*- coding: utf8 -*-

from sage.all import *

var('x,y')
f=x**2+y**2
G=Graphics()
a=3
for i in range(0,5):
  G=G+implicit_plot(f==i,(x,-a,a),(y,-a,a))
show(G)
```

tex/frido/courbeNiveau.py

Le résultat est :



Notez que les courbes sont censées être des cercles : les axes X et Y n'ont pas la même échelle.

Exemple 11.147.
Un exemple plus riche en enseignements est celui de la fonction

$$f(x, y) = x^2 - y^2. \tag{11.340}$$

La courbe de niveau h est donnée par l'équation $x^2 - y^2 = h$.

Commençons par $h = 0$. Dans ce cas nous avons $(x + y)(x - y) = 0$ et par conséquent les courbes de niveau de hauteur zéro sont les deux droites $x + y = 0$ et $x - y = 0$.

Voyons ensuite la courbe de niveau $h = 1$. Cela est l'équation $x^2 - y^2 = 1$, c'est-à-dire

$$y(x) = \pm\sqrt{x^2 - 1}. \tag{11.341}$$

C'est une fonction qui n'est définie que pour $|x| \geqslant 1$. Avec $x = 1$ nous avons $y = 1$. Ensuite, lorsque x grandit, y grandit également, mais la courbe ne peut pas croiser la courbe de niveau $h = 0$. Donc, suivant les notations de la figure 11.6, la courbe de niveau « part » de P et doit monter sans croiser les diagonales.

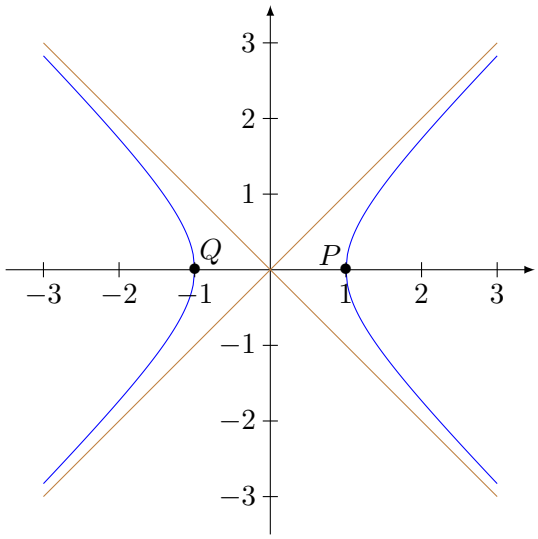


FIGURE 11.6 – La courbe de niveau $h = 1$ de $x^2 - y^2$. Notez qu'elle est en deux morceaux.

En ce qui concerne la courbe de niveau $h = -1$, elle correspond à la courbe $y = \pm\sqrt{1 + x^2}$ qui est définie pour tous les $x \in \mathbb{R}$. Le même raisonnement que précédemment nous amène à la figure 11.7.

△

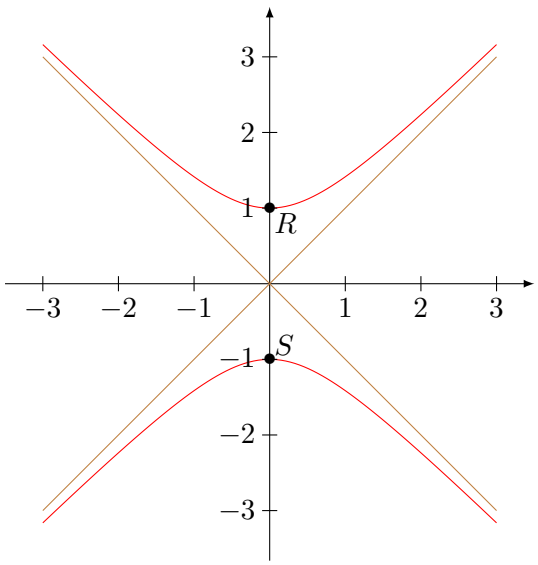


FIGURE 11.7 – La courbe de niveau $x^2 - y^2 = -1$.

Une autre façon de voir les courbes de niveau est de dire que la courbe de niveau de hauteur h est la projection dans le plan XY de la section du graphe de f par le plan $z = h$.

On peut également définir le graphe de fonctions de trois (ou plus) variables. Le graphe de la fonction $f\colon D \subset \mathbb{R}^3 \to \mathbb{R}$ est l'ensemble

$$\big\{\big(x, y, z, f(x, y, z)\big) \text{ tel que } (x, y, z) \in D\big\} \subset \mathbb{R}^4. \tag{11.342}$$

De tels graphes ne peuvent pas être représentés sur une feuille de papier. Il est toutefois possible de définir les ensembles de niveaux :

$$E_h = \big\{(x, y, z) \in D \text{ tel que } f(x, y, z) = h\big\}. \tag{11.343}$$

Ce sont des surfaces dans $\mathbb{R}^3$ que l'on peut dessiner.

Exemple 11.148.
Les surfaces de niveau de la fonction $f(x, y, z) = x^2 + y^2 + z^2$ sont des sphères. Il n'y a pas de surfaces de niveau pour les « hauteurs » négatives. △

Exemple 11.149.
Considérons la fonction $f(x, y, z) = x^2 + y^2 - z^2$. En coordonnées cylindriques, cette fonction s'écrit

$$f(r, \theta, z) = r^2 - z^2. \tag{11.344}$$

La surface de niveau 0 est donnée par l'équation $r = |z|$. Cela fait un cercle à chaque hauteur, dont le rayon grandit linéairement avec la hauteur ; le tout est donc un cône. C'est d'ailleurs le cône obtenu par rotation de la courbe de niveau $h = 0$ que nous avions obtenu pour la fonction $x^2 - y^2$.

En ce qui concerne les ensembles de niveau positifs, ils sont donnés par

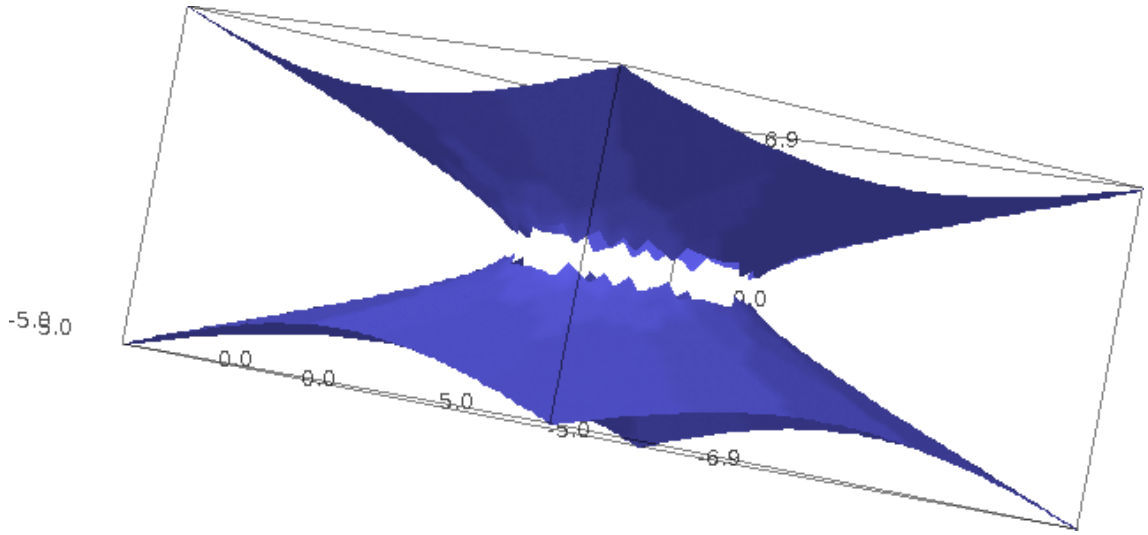
$$z = \pm\sqrt{x^2 + y^2 - h}. \tag{11.345}$$

Notez qu'ils ne sont pas définis pour $r \geqslant h$. Cela pose un petit problème quand on veut le tracer à l'ordinateur :

```
----------------------------------------------------------------------
| Sage Version 4.6.1, Release Date: 2011-01-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: var('x,y')
```

```
(x, y)
sage: f(x,y)=sqrt(x**2+y**2-3)
sage: F=plot3d(f(x,y),(x,-5,5),(y,-5,5))
sage: G=plot3d(-f(x,y),(x,-5,5),(y,-5,5))
sage: F+G
```

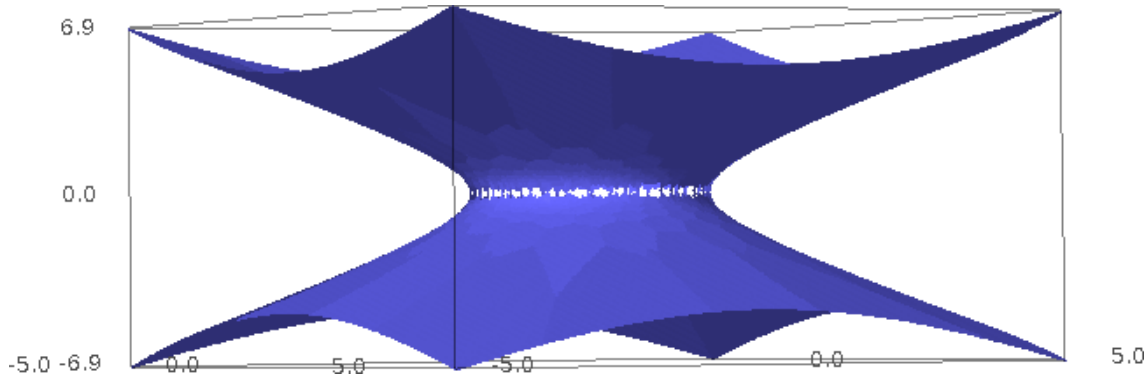
Le résultat est [44] :



On voit qu'il y a un grand trou au centre correspondant aux z proches de zéro. Or d'après l'équation, il n'en est rien : en $z = 0$ il y a bel et bien tout un cercle. Afin d'obtenir une meilleur image, il faut demander de tracer avec un maillage plus fin :

```
----------------------------------------------------------------------
| Sage Version 4.6.1, Release Date: 2011-01-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: var('x,y')
(x, y)
sage: f(x,y)=sqrt(x**2+y**2-3)
sage: F=plot3d(f(x,y),(x,-5,5),(y,-5,5),plot_points=300)
sage: G=plot3d(-f(x,y),(x,-5,5),(y,-5,5),plot_points=300)
sage: F+G
```

Le temps de calcul est un peu plus long, mais le résultat est meilleur :



△

44. Encore une fois : ça donne mieux à l'écran, et vous pouvez le faire bouger ; je vous encourage à le faire !

11.16 Limites à plusieurs variables

Prenons une fonction $f\colon \mathbb{R}^n \to \mathbb{R}$. Nous disons que

$$\lim_{x\to x_0} f(x) = l \in \mathbb{R} \tag{11.346}$$

lorsque $\forall \epsilon > 0$, $\exists \delta$ tel que $\|x - x_0\| \leqslant \delta$ implique $|f(x) - l| \leqslant \epsilon$.

Remarquez qu'ici, $x \in \mathbb{R}^n$, et sachez distinguer $\|.\|$, la norme dans $\mathbb{R}^n$ de $|.|$ qui est la valeur absolue dans $\mathbb{R}$. Une autre façon d'exprimer cette définition est que l'ensemble des valeurs atteintes par f dans une boule de rayon δ autour de x_0 n'est pas très loin de l. Nous définissons donc

$$E_\delta = \{f(x) \text{ tel que } x \in B(x_0, \delta)\}. \tag{11.347}$$

Notez que si f n'est pas définie en x_0, il n'y a pas de valeurs correspondantes au centre de la boule dans E_δ. Ceci est évidemment la situation générique lorsqu'il y a une indétermination à lever dans le calcul de la limite. Nous avons alors que

$$\lim_{x\to x_0} f(x) = l \tag{11.348}$$

lorsque $\forall \epsilon > 0$, $\exists \delta$ tel que

$$\sup\{|v - l| \text{ tel que } v \in E_\delta\} \leqslant \epsilon. \tag{11.349}$$

Une façon classique de montrer qu'une limite n'existe pas, est de prouver que, pour tout δ, l'ensemble E_δ contient deux valeurs constantes. Si par exemple $0 \in E_\delta$ et $1 \in E_\delta$ pour tout δ, alors aucune valeur de l (même pas $l = \pm\infty$) ne peut satisfaire à la condition (11.349) pour toute valeur de ϵ.

Nous laissons à la sagacité de l'étudiant le soin d'adapter tout ceci pour le cas $\lim_{x\to x_0} f(x) = \pm\infty$.

La proposition suivante semble évidente, mais nous sera tellement utile qu'il est préférable de l'expliciter :

Proposition 11.150.
Soient $f : D \to \mathbb{R}$ une fonction de domaine D, $a \in \mathrm{Adh}(D)$ et un voisinage V de a. Nous supposons que $V \cap D$ s'écrive comme une intersection finie :

$$V \cap D = \bigcup_{i=1}^{k} A_i$$

telle que $a \in \mathrm{Adh}\, A_i$ pour tout $i \leqslant k$. Alors, la limite

$$\lim_{x\to a} f(x) \tag{11.350}$$

existe et vaut $b \in \mathbb{R}$ si et seulement si chacune des limites

$$\lim_{\substack{x\to a\\ x\in A_i}} f(x) \tag{11.351}$$

existe et vaut b.

Démonstration. On sait déjà que si la limite de $f : D \to \mathbb{R}$ existe, alors toute restriction à A_i admet la même limite[45]. Il suffit donc de prouver la réciproque.

Fixons provisoirement un entier i entre 1 et k ainsi que $\epsilon > 0$. Vu que $\lim_{\substack{x\to a\\ x\in A_i}} f(x) = b$, il existe $\delta_i > 0$ tel que si $x \in A_i$ et si $0 < |x - a| < \delta_i$, alors

$$|f(x) - f(a)| < \epsilon. \tag{11.352}$$

Quitte à prendre δ_i un peu plus petit, nous supposons que $V \subset B(a, \delta_i)$.

Nous posons $\delta = \min\{\delta_i\}_{i=1,\dots,k}$, et nous considérons $x \in D$ tel que $0 < |x - a| < \delta$. Nous avons alors

45. C'est une conséquence de la caractérisation séquentielle de la continuité 7.79.

(1) $x \in V \cap D$,

(2) il existe i tel que $x \in A_i$.

Ce x est donc un élément de A_i vérifiant $0 < |x-a| < \delta \leqslant \delta_i$. Il vérifie donc (11.352) : $|f(x)-f(a)| < \epsilon$.

Cela prouve la limite (11.350). □

Exemple 11.151. (1) Pour qu'une fonction $f : \mathbb{R} \to \mathbb{R}$ admette une limite en $a \in \mathbb{R}$, il faut et il suffit qu'elle y admette une limite à droite et une limite à gauche qui soient égales.

Cela est une application de la proposition 11.150 avec $\mathbb{R} =]-\infty, a[\cup]a, \infty[$.

(2) Une suite (x_k) admet une limite si et seulement si les sous-suites (x_{2k}) et (x_{2k+1}) convergent vers la même limite. Ceci n'est pas une application directe de la proposition, mais la teneur est la même.

△

Lemme 11.152 ([1]).
Soient deux espaces vectoriels E et F. Soit une fonction $f\colon \mathbb{R} \to F$ telle que $\lim_{t\to 0} f(t) = y \in F$. Nous posons

$$\begin{aligned} \varphi\colon E &\to F \\ h &\mapsto f(\|h\|). \end{aligned} \tag{11.353}$$

Alors φ admet une limite pour $h \to 0$ et elle est donnée par

$$\lim_{h\to 0} \varphi(h) = \lim_{t\to 0} f(t). \tag{11.354}$$

Démonstration. Soit $\epsilon > 0$. Il existe δ tel que si $t < \delta$ alors $\|f(t) - y\|_F < \epsilon$. Si $\|h\| < \delta$ nous avons

$$\|\varphi(h) - y\| = \|f(\|h\|) - y\| < \epsilon. \tag{11.355}$$

Donc c'est bon. □

Voici, dans le même ordre d'idée, un autre résultat qui permet de réduire le nombre de variables dans une limite lorsque la fonction ne dépend pas de certaines variables.

Lemme 11.153 ([1]).
Soit $f\colon \mathbb{R}^2 \to \mathbb{R}$ telle qu'il existe une fonction $g\colon \mathbb{R} \to \mathbb{R}$ vérifiant

(1) $f(x, y) = g(x)$,

(2) $\lim_{t\to a} g(t) = \ell$.

Alors

$$\lim_{\substack{(x,y)\to(a,b)\\ x\neq a}} f(x, y) = \ell. \tag{11.356}$$

Démonstration. Soit $\epsilon > 0$. Par hypothèse sur la limite de g en a, il existe $\delta > 0$ tel que $0 < |t - a| < \delta$ implique $|g(t) - \ell| < \epsilon$.

Attention : passage subtil[46]. Si $0 < \|(x, y) - (a, b)\| < \delta$, alors nous avons évidemment aussi $|x - a| < \delta$, mais pas spécialement $0 < |x - a| < \delta$ comme le requis pour utiliser la limite de g.

Dans le calcul de la limite restreinte à $x \neq a$, les points qui interviennent sont les valeurs de (x, y) dans $B\big((a, b), \delta\big)\backslash\{x = a\}$. Or pour celles-là nous avons bien $0 < |x-a| < \delta$. Le calcul suivant fonctionne donc :

$$|f(x, y) - \ell| = |g(x) - \ell| < \epsilon. \tag{11.357}$$

□

46. Je rejette déjà en bloc et d'un revers de main toute tentative de dire « la limite épointée, c'est mieux ». Voir aussi l'exemple 11.154.

Exemple 11.154.
Pourquoi prendre la limite $(x,y) \to (a,b)$ avec $x \neq a$ dans l'énoncé du lemme 11.153 ? Imaginons la fonction

$$g(x) = \begin{cases} 0 & \text{si } x \neq 0 \\ 1 & \text{si } x = 0. \end{cases} \tag{11.358}$$

Dans ce cas, le graphe de la fonction $f(x,y) = g(x)$ est tout plat sauf la ligne $x = 0$ qui est en hauteur. Nous avons donc $f(0,t) = 1$ pour tout t et donc nous n'avons pas $\lim_{(x,y)\to(0,0)} f(x,y) = 0$: tout voisinage de $(0,0)$ contient des points (x,y) tels que $f(x,y) = 1$ et des points (x,y) tels que $f(x,y) = 0$. △

Il existe de nombreuses façons de calculer des limites à plusieurs variables. Plus nous connaîtrons de mathématiques, plus nous aurons de techniques à notre disposition. Nous allons tout de suite voir quelques méthodes. Voir le thème 74 pour plus de techniques et d'exemples.

11.16.1 Caractérisation de la limite par les suites

Exemple 11.155.
Considérons la fonction

$$f(x,y) = \frac{xy}{x^2 + y^2}, \tag{11.359}$$

et remarquons que, quelle que soit la valeur de y, cette fonction est nulle lorsque $x = 0$. De la même manière, nous voyons que si $x = y$, alors la fonction vaut [47] $\frac{1}{2}$.

Il est impossible que la fonction ait une limite en $(0,0)$ parce qu'on ne peut pas trouver un ℓ dont on s'approche à la fois en suivant la ligne $x = 0$ et la ligne $x = y$.

Deux autres chemins avec encore deux autres valeurs sont dessinés sur la figure 11.8.

Cet exemple pourra être formalisé en utilisant le théorème 11.156. Voir l'exemple 11.157. △

Théorème 11.156 (Caractérisation de la limite par les suites)**.**
Une fonction $f\colon D \subset \mathbb{R}^m \to \mathbb{R}^n$ admet une limite ℓ en un point d'accumulation a de D si et seulement si pour toute suite (x_n) dans $D\backslash\{a\}$ convergente vers a, la suite $\big(f(x_n)\big)$ dans $\mathbb{R}^n$ converge vers ℓ.

Démonstration. Supposons d'abord que la fonction ait une limite ℓ lorsque $x \to a$, et considérons une suite (x_n) dans $D\backslash\{a\}$ convergente vers a. Nous devons montrer que la suite $y_n = f(x_n)$ converge vers ℓ, c'est-à-dire que si nous choisissons $\varepsilon > 0$ nous devons montrer qu'il existe un N tel que $n > N$ implique $\|y_n - \ell\| = \|f(x_n) - \ell\| < \varepsilon$.

Nous avons deux hypothèses. La première est la convergence de la fonction et la seconde est la convergence de la suite (x_n). L'hypothèse de convergence de la fonction nous dit que (le ε a déjà été choisi dans le paragraphe précédent)

$$\exists \delta \text{ tel que } 0 < \|x - a\| < \delta \Rightarrow \|f(x) - \ell\| < \varepsilon. \tag{11.360}$$

Une fois choisi ce δ qui « va avec » le ε qui a été choisi précédemment, la définition de la convergence de la suite nous enseigne que

$$\exists N \text{ tel que } n > N \Rightarrow \|x_n - a\| < \delta. \tag{11.361}$$

Récapitulons ce que nous avons fait. Nous avons choisi un ε, et puis nous avons construit un N. Lorsque $n > N$, nous avons $\|x_n - a\| < \delta$. Mais alors, par construction de ce δ, nous avons $\|f(x_n) - \ell\| < \varepsilon$. Au final, $n > N$ implique bien $\|y_n - \ell\| < \varepsilon$, ce qu'il nous fallait.

Nous supposons maintenant que la fonction f *ne* converge *pas* vers ℓ, et nous allons construire une suite d'éléments x_n qui converge vers a sans que $(y_n) = f(x_n)$ ne converge vers ℓ. La fonction

47. En fait ce que nous sommes en train de faire est de poser $\theta = \pi/2$ et $\theta = \pi/4$ dans (17.720).

f vérifie la condition (11.52). Nous prenons donc un ε tel que $\forall\delta$, il existe un x qui vérifie *en même temps* les deux conditions

$$\begin{cases} 0 < \|x-a\| < \delta & (11.362a) \\ \|f(x)-\ell\| > \varepsilon. & (11.362b) \end{cases}$$

Un tel x existe pour tout choix de δ. Choisissons un n arbitraire et $\delta = \frac{1}{n}$. Nous nommons x_n le x correspondant à ce choix de n. La suite (x_n) ainsi construite converge vers a parce que

$$\|x_n - a\| < \delta_n = \frac{1}{n}, \tag{11.363}$$

donc dès que n est grand, $\|x_n - a\|$ est petit. Mais la suite $y_n = f(x_n)$ ne converge pas vers ℓ parce que

$$\|f(x_n) - \ell\| > \varepsilon \tag{11.364}$$

pour tout n. La suite y_n ne s'approche donc jamais à moins d'une distance ε de ℓ. □

Exemple 11.157.
Reprenons l'exemple 11.155. Considérons les deux suites $x_n = (0, \frac{1}{n})$ et $y_n = (\frac{1}{n}, \frac{1}{n})$. Ce sont deux suites dans $\mathbb{R}^2$ qui tendent vers $(0,0)$. Si la fonction f convergeait vers ℓ, alors nous aurions au moins

$$\lim f(x_n) = \ell \tag{11.365a}$$
$$\lim f(y_n) = \ell, \tag{11.365b}$$

mais nous savons que pour tout n, $f(x_n) = f(0, \frac{1}{n}) = 0$ et $f(y_n) = f(\frac{1}{n}, \frac{1}{n}) = \frac{1}{2}$. Il n'y a donc aucun nombre ℓ qui vérifie les deux équations (11.365) parce que $\lim f(x_n) = 0$ et $\lim f(y_n) = \frac{1}{2}$. △

11.16.2 Règle de l'étau

Une première façon de calculer la limite d'une fonction est de la « coincer » entre deux fonctions dont nous connaissons la limite.

Théorème 11.158 (Règle de l'étau[247])**.**
Soit $\mathcal{O}$, un ouvert de $\mathbb{R}^m$ contenant le point a. Soient f, g et h, trois fonctions définies sur $\mathcal{O}$ (éventuellement pas en a lui-même). Supposons que pour tout $x \in \mathcal{O}$ (à part éventuellement a), nous ayons les inégalités

$$g(x) \leqslant f(x) \leqslant h(x). \tag{11.366}$$

Supposons de plus que

$$\lim_{x\to a} g(x) = \lim_{x\to a} h(x) = \ell. \tag{11.367}$$

Alors la limite $\lim_{x\to a} f(x)$ existe et vaut ℓ.

Nous insistons sur le fait que les deux fonctions entre lesquelles nous coinçons f doivent tendre vers *la même* valeur.

Cette méthode est très pratique lorsqu'on a des fonctions trigonométriques qui se factorisent parce qu'elles sont toujours majorables par 1 ; voir l'exemple 17.55.

Exemple 11.159.
Prouver la continuité en $(0,0)$ de la fonction

$$f(x,y) = \begin{cases} \frac{x|y|}{\sqrt{x^2+y^2}} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{sinon.} \end{cases} \tag{11.368}$$

Considérons une suite $(x_n, y_n) \in \mathbb{R}^2$ qui tend vers $(0,0)$. Étant donné que $\frac{|y|}{\sqrt{x^2+y^2}} < 1$ pour tout x et y, nous avons

$$0 \leqslant |f(x_n, y_n)| = \left| \frac{x_n |y_n|}{\sqrt{x_n^2 + y_n^2}} \right| \leqslant |x_n| \to 0. \tag{11.369}$$

Donc nous avons

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0 = f(0,0), \tag{11.370}$$

ce qui prouve que la fonction est continue en $(0,0)$ par la proposition 7.80. Nous avons utilisé la règle de l'étau (théorème 11.158). △

11.160.
Nous notons $f \sim g$ pour $x \to a$ lorsque $\lim_{x\to a} \frac{f(x)}{g(x)} = 1$.

Cela signifie que f et g tendent vers la même limite, à la même vitesse.

11.16.3 Méthode des chemins

Lorsque la limite n'existe pas, il y a une façon en général assez simple de le savoir, c'est la **méthode des chemins**.

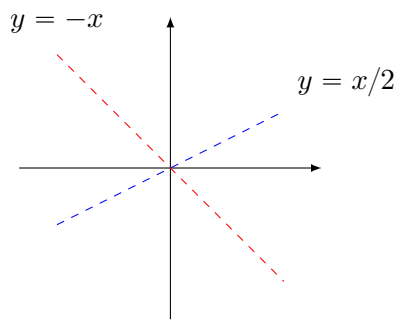


FIGURE 11.8 – Sur toute la droite $y = -x$, la fonction vaut $-1/2$, tandis que sur toute la droite $y = x/2$, elle vaut $\frac{2}{5}$. Il est donc impossible que la fonction ait une limite en $(0,0)$, parce que dans toute boule autour de zéro, il y aura toujours un point de chacune de ces deux droites.

C'est la proposition suivante qui va faire une grosse partie du travail.

Proposition 11.161 ([1]).
Soit $f\colon D \subset \mathbb{R}^m \to \mathbb{R}^n$ et a un point d'adhérence de D. Alors nous avons

$$\lim_{x\to a} f(x) = \ell \tag{11.371}$$

si et seulement si pour toute fonction $\gamma\colon \mathbb{R} \to \mathbb{R}^m$ telle que $\lim_{t\to 0} \gamma(t) = a$, nous avons

$$\lim_{t\to 0} (f \circ \gamma)(t) = \ell. \tag{11.372}$$

Démonstration. En deux parties.

Sens direct Soit une fonction $\gamma\colon \mathbb{R} \to \mathbb{R}^m$ telles que $\lim_{t\to 0} \gamma(t) = a$. Par le théorème 11.156, il suffit de montrer que $(f \circ \gamma)(t_n) \to \ell$ pour toute suite $t_n \to 0$ dans $\mathbb{R}$.

Nous savons que la suite $n \mapsto \gamma(t_n)$ est une suite qui converge vers a. Mais l'hypothèse $\lim_{x\to a} f(x) = \ell$ implique que pour toute suite $x_n \to a$ nous avons $f(x_n) \to \ell$. Cela est en particulier vrai pour la suite $n \mapsto \gamma(t_n)$. Donc :

$$\lim_{n\to\infty} f\big(\gamma(t_n)\big) = \ell, \tag{11.373}$$

ce qu'il fallait prouver.

Réciproque Pour les mêmes raisons de caractérisation séquentielle que précédemment, il faut prouver que $\lim_{n\to\infty} f(x_n) = \ell$ pour tout suite $x_n \to a$.

Un chemin Soit la fonction $\gamma\colon \mathbb{R} \to \mathbb{R}^m$ affine par morceaux et telle que

$$\gamma\left(\frac{1}{n}\right) = x_n. \tag{11.374}$$

Nous prolongeons γ par $\gamma(t) = a$ pour $t \leqslant 0$.

$\gamma(t) \to a$ Nous montrons que $\lim_{t\to 0} \gamma(t) = a$. Soient $\epsilon > 0$ et N tel que $x_n \in B(a, \epsilon)$ pour tout $n \geqslant N$. Si $t < \frac{1}{N}$ alors $t \in [\frac{1}{k+1}, \frac{1}{k}]$ pour un certain $k > N$. Donc

$$\gamma(t) \in [\gamma(\frac{1}{k+1}), \gamma(\frac{1}{k})] \tag{11.375}$$

et donc $\gamma(t) \in [x_{k+1}, x_k]$ parce que γ est formé de ces segments de droites. Mais comme $B(a, \epsilon)$ est convexe[48], nous avons

$$\gamma(t) \in [x_{k+1}, x_k] \subset B(a, \epsilon). \tag{11.376}$$

Nous avons donc bien $\lim_{t\to 0} \gamma(t) = a$.

Conclusion L'hypothèse nous donne alors $\lim_{t\to 0}(f \circ \gamma)(t) = \ell$. En particulier le critère de la caractérisation séquentielle de la limite dit que

$$\lim_{n\to\infty} f\big(\gamma(\frac{1}{n})\big) = \ell, \tag{11.377}$$

ce qui signifie $\lim_{n\to\infty} f(x_n) = \ell$.

□

Corolaire 11.162.
Soient $f\colon D \subset \mathbb{R}^m \to \mathbb{R}^n$ et a un point d'accumulation de D. Si nous avons deux fonctions $\gamma_1, \gamma_2\colon \mathbb{R} \to \mathbb{R}^m$ telles que

$$\lim_{t\to 0} \gamma_1(t) = \lim_{t\to 0} \gamma_2(t) = a \tag{11.378}$$

tandis que

$$\lim_{t\to 0}(f \circ \gamma_1)(t) \neq \lim_{t\to 0}(f \circ \gamma_2)(t), \tag{11.379}$$

ou bien que l'une des deux limites n'existe pas, alors la limite de $f(x)$ lorsque $x \to a$ n'existe pas.

Corolaire 11.163.
Soient $f\colon D \subset \mathbb{R}^m \to \mathbb{R}^n$ et a un point d'accumulation de D. Si il existe une fonction $\gamma\colon \mathbb{R} \to \mathbb{R}^m$ avec $\gamma(0) = a$ telle que la limite $\lim_{t\to 0}(f \circ \gamma)(t)$ n'existe pas, alors la limite $\lim_{x\to a} f(x)$ n'existe pas.

En ce qui concerne le calcul de limites, la méthode des chemins peut être utilisé de trois façons :

(1) Dès que l'on trouve une fonction $\gamma\colon \mathbb{R} \to \mathbb{R}^m$ telle que $\lim_{t\to 0}(f \circ \gamma)(t) = \ell$, alors nous savons que *si la limite* $\lim_{x\to a} f(x)$ *existe*, alors cette limite vaut ℓ.

(2) Dès que l'on a trouvé deux fonctions γ_i qui tendent vers a, mais dont les limites de $\lim_{t\to 0}(f \circ \gamma_i)(t)$ sont différentes, alors la limite $\lim_{x\to a} f(x)$ n'existe pas.

(3) Dès qu'on trouve une chemin le long duquel il n'y a pas de limite, alors la limite n'existe pas (corolaire 11.163).

La méthode des chemins ne permet donc pas de de calculer une limite quand elle existe. Elle permet uniquement de la « deviner », ou bien de prouver que la limite n'existe pas.

48. C'est la proposition 8.26.

Exemple 11.164.
Soit à calculer

$$\lim_{(x,y)\to(0,0)} \frac{x-y}{x+y}. \tag{11.380}$$

Si nous prenons le chemin $\gamma_1(t) = (t,t)$, nous avons bien $\lim_{t\to 0}\gamma_1(t) = (0,0)$, et nous avons

$$\lim_{t\to 0}(f\circ\gamma_1)(t) = \lim_{t\to 0}\frac{t-t}{t+t} = 0. \tag{11.381}$$

Donc si la limite (11.380) existait, elle vaudrait obligatoirement 0. Mais si nous considérons $\gamma_2(t) = (0,t)$, nous avons

$$(f\circ\gamma_2)(t) = \frac{-t}{t} = -1, \tag{11.382}$$

donc si la limite existe, elle doit obligatoirement valoir -1. Ne pouvant être égale à 0 et à -1 en même temps, la limite (11.380) n'existe pas. △

11.17 Dérivée directionnelle

Nous sommes capables de dériver une fonction de deux variables $f(x,y)$ par rapport à x et par rapport à y. C'est-à-dire que nous sommes capables de donner la variation de la fonction lorsqu'on bouge le long des axes horizontal et vertical. Il est évidemment souhaitable de parler de la variation de la fonction lorsqu'on se déplace le long d'autre droites.

Soit donc $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ un vecteur unitaire (c'est-à-dire $u_1^2 + u_2^2 = 1$), et considérons la fonction de une variable

$$\begin{aligned} \varphi\colon \mathbb{R} &\to \mathbb{R} \\ t &\mapsto f(a+tu_1, b+tu_2). \end{aligned} \tag{11.383}$$

La fonction φ n'est rien d'autre que la fonction f vue le long de la droite de direction donnée par le vecteur u. Nous pouvons aussi l'écrire $\varphi(t) = f(p+tu)$.

Soit $f\colon \mathbb{R}^2 \to \mathbb{R}$ une fonction de deux variables et soit $(a,b)\in\mathbb{R}^2$. La façon la plus naturelle de définir une dérivée à deux variables est de considérer les **dérivées partielles** définies par

$$\begin{aligned} \frac{\partial f}{\partial x}(a,b) &= \lim_{x\to a}\frac{f(x,b)-f(a,b)}{x-a} \\ \frac{\partial f}{\partial y}(a,b) &= \lim_{y\to b}\frac{f(a,y)-f(a,b)}{y-b}. \end{aligned} \tag{11.384}$$

Ces nombres représentent la façon dont le nombre $f(x,y)$ varie lorsque soit seul x varie soit seul y varie. Les dérivées partielles se calculent de la même façon que les dérivées normales. Pour calculer $\partial_x f$, on fait « comme si » y était une constante, et pour calculer $\partial_y f$, on fait comme si x était une constante.

11.17.1 Dérivée partielle et directionnelles

Soit une fonction $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$. Si $n \neq 1$, la notion de *dérivée* de la fonction f n'a plus de sens puisqu'on ne peut plus parler de pente de *la* tangente au graphe de f en un point. On introduit alors quelques notions qui feront, en dimension quelconque, le même travail que la dérivée en dimension un : les dérivées directionnelles et la différentielle. Nous allons voir qu'en dimension un, la différentielle coïncide avec la dérivée.

Définition 11.165.
Soit f une application de $U \subset \mathbb{R}^m$ dans $\mathbb{R}$, a un point dans U et v un vecteur de $\mathbb{R}^m$. On dit que f admet une ***dérivée suivant le vecteur*** *v* ***au point*** *a si la fonction $t \mapsto f(a+tv)$ admet une dérivée en $t=0$. La dérivée de f suivant le vecteur v au point a est alors cette dérivée, et f est dite dérivable suivant v en a,*

$$\partial_v f(a) = \lim_{\substack{t\to 0\\ t\neq 0}}\frac{f(a+tv)-f(a)}{t}.$$

Définition 11.166.
La fonction $f : U \subset \mathbb{R}^m \to \mathbb{R}^n$ de composantes $(f_1, \dots, f_n)$, est dite ***dérivable suivant*** *v* ***au point*** *a si toute ses composantes f_i, $i = 1, \dots, n$ sont dérivables suivant v au point a. Dans ce cas, nous écrivons*

$$\partial_v f(a) = (\partial_v f_1(a), \dots, \partial_v f_n(a)) . \tag{11.385}$$

Géométriquement, il s'agit du taux de variation instantané de f en a dans la direction du vecteur u, c'est-à-dire de la pente de la tangente dans la direction du vecteur u au graphe de f au point $(a, f(a))$.

Remarque 11.167.
De nombreuses sources parlent de de dérivée **dans la direction**u vecteur v en définissant (avec une certaine raison) une **direction** dans $\mathbb{R}^m$ comme étant un vecteur de norme 1.

Ces personnes ne définissent alors $\partial_u f$ que pour $\|u\| = 1$. Pourquoi ? Le but de la dérivée directionnelle dans la direction u est de savoir à quelle vitesse la fonction monte lorsque l'on se déplace en suivant la direction u. Cette information n'aura un caractère « objectif » que si l'on avance à une vitesse donnée. En effet, si on se déplace deux fois plus vite, la fonction montera deux fois plus vite. Par convention, on demande alors d'avancer à vitesse 1.

Ici, pour être plus souple en termes de notations et de manipulations, nous définissons $\partial_u f$ pour tout u (non nul). Nous devons cependant garder en tête que le nombre $(\partial_v f)(a)$ ne peut pas être interprété comme étant une « vitesse de croissance de f en a » de façon trop sérieuse.

Cas particulier où $n = 2$:

$a = (a_1, a_2)$, $u = (u_1, u_2)$ et

$$\frac{\partial f}{\partial u}(a_1, a_2) = \lim_{t \to 0} \frac{f(a_1 + tu_1, a_2 + tu_2) - f(a_1, a_2)}{t}$$

Un cas particulier des dérivées directionnelles est la dérivée partielle. Si nous considérons la base canonique e_i de $\mathbb{R}^n$, nous notons

$$\frac{\partial f}{\partial x_i} = \frac{\partial f}{\partial e_i}. \tag{11.386}$$

Dans le cas d'une fonction à deux variables, nous avons donc les deux dérivées partielles

$$\frac{\partial f}{\partial x}(a) \quad \text{et} \quad \frac{\partial f}{\partial y}(a) \tag{11.387}$$

qui correspondent aux dérivées directionnelles dans les directions des axes. Ces deux nombres représentent de combien la fonction f monte lorsqu'on part de a en se déplaçant dans le sens des axes X et Y.

11.17.1.1 Quelques propriétés et notations

(1) $\forall \alpha \in \mathbb{R}$, si $v = \alpha\, u$, alors $\frac{\partial f}{\partial v}(a) = \alpha \frac{\partial f}{\partial u}(a)$.

(2) Si on prend $u = e_j$ le $j$ème vecteur de la base canonique de $\mathbb{R}^n$, alors

$$\frac{\partial f}{\partial e_j}(a) = \frac{\partial f}{\partial x_j}(a)$$

c'est-à-dire que la dérivée de f au point a dans la direction e_j est la dérivée partielle de f par rapport à sa $j$ème variable.

(3) Une fonction peut être dérivable dans certaines directions mais pas dans d'autres (rappelez vous que si la limite à droite est différente de la limite à gauche, la limite n'existe pas).

(4) Même si une fonction est dérivable en un point dans toutes les directions, on n'est pas sûr qu'elle soit continue en ce point. La dérivabilité directionnelle n'est donc pas une notion suffisante pour assurer la continuité. C'est pourquoi on introduit le concept de *différentiabilité*.

Proposition 11.168.
Soit u un vecteur de norme 1 dans $\mathbb{R}^m$ et soit $v = \lambda u$, avec λ dans $\mathbb{R}$. La fonction f est dérivable suivant v au point a si et seulement si f est dérivable suivant u au point a, en outre

$$\partial_v f(a) = \lambda \partial_u f(a).$$

Démonstration.

$$\begin{aligned}\partial_v f(a) &= \lim_{\substack{t\to 0\\ t\neq 0}} \frac{f(a+tv)-f(a)}{t} = \lim_{\substack{t\to 0\\ t\neq 0}} \frac{f(a+t\lambda u)-f(a)}{t} = \\ &= \lambda \lim_{\substack{t\to 0\\ t\neq 0}} \frac{f(a+t\lambda u)-f(a)}{\lambda t} = \lambda \partial_u f(a).\end{aligned} \tag{11.388}$$

□

Définition 11.169.
*Soit f une application de $U \subset \mathbb{R}^m$ dans $\mathbb{R}$. On appelle **dérivées partielles de f au point** a les dérivées de f suivant les vecteurs de base $e_1, \dots, e_m$ au point a, si elles existent.*

Si $m = 2, 3$ on peut utiliser la notation f_x, ∂_x ou ∂_1 pour la dérivée partielle suivant e_1, f_y, ∂_y ou ∂_2 pour la dérivée partielle suivant e_2 et f_z, ∂_z ou ∂_3 pour la dérivée partielle suivant e_3. En général, nous écrivons ∂_i pour noter la la dérivée partielle suivant e_i.

Des exemples faisons intervenir les fonctions trigonométriques, exponentielles et logarithme sont les exemples 17.217, 14.65.

La fonction d'une seule variable qu'on obtient à partir de f en fixant les $p-1$ variables $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p$ et qui associe à x_i la valeur $f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_p)$, est appelée x_i-ème **section** de f en $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p$. L'$i$-ème dérivée partielle de f au point $a = (x_1, \dots, x_m)$ est la dérivée de l'i-ème section de f au point x_i. En pratique, pour calculer les dérivées partielles d'une fonction on fait une dérivation par rapport à la variable choisie en considérant les autres variables comme des constantes.

Exemple 11.170.
Considérons la fonction $f(x, y) = 2xy^2$. Lorsque nous calculons $\partial_x f(x, y)$, nous faisons comme si y était constant. Nous avons donc $\partial_x f(x, y) = 2y^2$. Par contre lors du calcul de $\partial_y f(x, y)$, nous prenons x comme une constante. La dérivée de y^2 par rapport à y est évidemment $2y$, et par conséquent, $\partial_y f(x, y) = 4xy$. △

Définition 11.171.
*Soit f une application de $U \subset \mathbb{R}^m$ dans $\mathbb{R}$ et u un vecteur de $\mathbb{R}^m$. La fonction f est **dérivable sur U suivant le vecteur** u, si f est dérivable suivant le vecteur u en tout point de U.*

Pour les fonctions d'une seule variable la dérivabilité en un point a implique la continuité en a. Cela n'est pas vrai pour les fonctions de plusieurs variables : il existe des fonctions f qui sont dérivables suivant tout vecteur au point a sans pour autant être continue en a.

Exemple 11.172.
Considérons la fonction $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4+y^2} & \text{si } (x, y) \neq (0, 0), \\ 0 & \text{sinon.} \end{cases} \tag{11.389}$$

Pour voir que f n'est pas continue en $(0, 0)$ il suffit de calculer la limite de f restreinte à la parabole $y = x^2$

$$\lim_{x\to 0} f(x, x^2) = \frac{1}{2} \neq 0.$$

Pourtant la fonction f est dérivable en $(0,0)$ dans toutes les directions. En effet, soit $v=(v_1,v_2)$. Si $v_2\neq 0$, alors

$$\partial_v f(a)=\lim_{\substack{t\to 0\\ t\neq 0}}\frac{t^3v_1^2v^2}{t^5v_1^4+t^3v_2^2}=\frac{v_1^2}{v_2},$$

tandis que si $v_2=0$, alors la valeur de $f(tv_1,0)$ est 0 pour tout t et v_1, donc la dérivée partielle de f par rapport à x en l'origine existe et est nulle. △

Exemple 11.173.
Pour une fonction réelle à variable réelle, la dérivabilité entraine la continuité. Il n'en va pas de même pour les fonctions à plusieurs variables, comme le montre l'exemple suivant :

$$f(x,y)=\begin{cases}0 & \text{si } x=0\\ \frac{y}{x}\sqrt{x^2+y^2} & \text{sinon.}\end{cases}\tag{11.390}$$

Nous avons tout de suite

$$\frac{\partial f}{\partial y}(0,0)=0.\tag{11.391}$$

De plus si $u_x\neq 0$ nous avons

$$\frac{\partial f}{\partial u}(0,0)=\frac{u_y}{u_x}\|u\|.\tag{11.392}$$

Donc toutes les dérivées directionnelles de f en $(0,0)$ existent alors que la fonction n'y est manifestement pas continue. En effet sous forme polaire,

$$f(r,\theta)=\frac{r\sin(\theta)}{\cos(\theta)},\tag{11.393}$$

et quelle que soit la valeur de r, en prenant θ suffisamment proche de $\pi/2$, la fraction peut être arbitrairement grande.

Nous verrons par la proposition 11.194 que la différentiabilité d'une fonction implique sa continuité. △

Théorème 11.174 (Accroissement finis pour les dérivées suivant un vecteur)**.**
Soit U un ouvert dans $\mathbb{R}^m$ et soit $f:U\to\mathbb{R}^n$ une fonction. Soient a et b deux points distincts dans U, tels que le segment[49] *$[a,b]$ soit contenu dans U. Soit u le vecteur*

$$u=\frac{b-a}{\|b-a\|_m}.$$

Si $\partial_u f(x)$ existe pour tout x dans $[a,b]$ on a

$$\|f(b)-f(a)\|_n\leqslant\sup_{x\in[a,b]}\|\partial_u f(x)\|_n\|b-a\|_m.$$

Démonstration. Nous considérons la fonction $g(t)=f\big((1-t)a-tb\big)$. Elle décrit la droite entre a et b parce que $g(0)=a$ et $g(1)=b$. En ce qui concerne la dérivée,

$$\begin{aligned}g'(t)&=\lim_{h\to 0}\frac{g(t+h)-g(t)}{h}\\&=\lim_{h\to 0}\frac{f\big((1-t-h)a-(t+h)b\big)}{h}\\&=\lim_{h\to 0}\frac{f\big(a+(t+h)(b-a)\big)-f\big(a+t(b-a)\big)}{h}\\&=\frac{\partial f}{\partial u}\big(a+t(b-a)\big)\|b-a\|.\end{aligned}\tag{11.394}$$

49. Définition 7.152.

Le dernier facteur $\|b-a\|$ apparaît pour la normalisation du vecteur u. En effet dans la limite, il apparaît $h(b-a)$, ce qui donnerait la dérivée le long de $b-a$, tandis que u vaut $(b-a)/\|b-a\|$.

Par le théorème des accroissements finis pour g, il existe $t_0 \in]0,1[$ tel que

$$g(1) = g(0) + g'(t_0)(1-0). \tag{11.395}$$

Donc

$$\|g(1)-g(0)\| \leqslant \sup_{t_0} \|g'(t_0)\| = \sum_{t_0 \in]0,1[} \left\| \frac{\partial f}{\partial u}(a + t_0(b-a)) \right\| \|b-a\|. \tag{11.396}$$

Mais lorsque t_0 parcours $]0,1[$, le point $a+t_0(b-a)$ parcours le segment $]a,b[$, d'où le résultat. □

Corolaire 11.175.
Dans les mêmes hypothèses, si $n=1$, alors il existe $\bar{x}$ dans $]a,b[$ tel que

$$f(b) - f(a) = \partial_u f(\bar{x}) \|b-a\|_m.$$

Définition 11.176.
Le nombre

$$\lim_{t \to 0} \frac{f\big(a+tu_1, b+tu_2\big) - f(a,b)}{t} \tag{11.397}$$

*est la **dérivée directionnelle** de f dans la direction de u au point (a,b). Il sera noté*

$$\frac{\partial f}{\partial u}(a,b), \tag{11.398}$$

ou plus simplement $\partial_u f(a,b)$.

Lorsque f est différentiable, la dérivée directionnelle est donnée par

$$\frac{\partial f}{\partial u}(p) = \nabla f(p) \cdot u. \tag{11.399}$$

11.17.2 Gradient : direction de plus grande pente

Étant donné que u est de norme 1, l'inégalité de Cauchy-Schwarz donne

$$\left| \nabla f(a,b) \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \right| \leqslant \|\nabla f(a,b)\|. \tag{11.400}$$

Donc

$$-\|\nabla f(p)\| \leqslant \nabla f(p) \cdot u \leqslant \|\nabla f(p)\|. \tag{11.401}$$

La norme de la dérivée directionnelle (qui est la valeur absolue du nombre au centre) est donc « coincée » entre $-\|\nabla f(p)\|$ et $\|\nabla f(p)\|$. Prenons par exemple

$$u = \frac{\nabla f(p)}{\|\nabla f(p)\|}. \tag{11.402}$$

Dans ce cas, nous avons exactement

$$\nabla f(p) \cdot u = \|\nabla f(p)\|, \tag{11.403}$$

qui est la valeur maximale que la dérivée directionnelle peut prendre.

La direction du gradient est donc la direction suivant laquelle la dérivée directionnelle est la plus grande. Pour la même raison, la dérivée directionnelle est la plus petite dans le sens opposé au gradient.

En termes bien clairs : lorsqu'on veut aller le plus vite possible au ski, on prend la direction du gradient de la piste de ski. C'est dans cette direction que ça descend le plus vite. Dans quelle direction vont les débutants ? Ils vont perpendiculairement à la pente (ce qui ennuie tout le monde,

mais c'est un autre problème). Les débutants vont donc dans la direction perpendiculaire au gradient. Prenons donc $u \perp \nabla f(p)$ et calculons la dérivée directionnelle de f dans la direction u en utilisant la formule 11.399 :

$$\frac{\partial f}{\partial u}(p) = \nabla f(p) \cdot u = 0 \tag{11.404}$$

parce que nous avons choisi $u \perp \nabla f(p)$. Nous voyons donc que les débutants en ski ont eu la bonne intuition que la direction dans laquelle la piste ne descend pas, c'est la direction perpendiculaire au gradient.

C'est aussi pour cela que l'on a tendance à faire du zig-zag à vélo lorsqu'on monte une pente très forte et qu'on est fatigué. C'est toujours pour cela que les routes de montagne font de longs lacets. La montée est moins rude en suivant une direction proche d'être perpendiculaire au gradient !

Théorème 11.177.
Le gradient des fonctions suit à peu près les mêmes règles que les dérivées. Soient f et g deux fonctions différentiables. Nous avons entre autres

(1) $\nabla(f+g) = \nabla f + \nabla g$;

(2) $\nabla(fg)(a,b) = g(a,b)\nabla f(a,b) + f(a,b)\nabla g(a,b)$;

(3) Dès que $g(a,b) \neq 0$, nous avons

$$\nabla \frac{f}{g} = \frac{g(a,b)\nabla f(a,b) - f(a,b)\nabla g(a,b)}{g(a,b)^2}. \tag{11.405}$$

11.17.3 Gradient : orthogonal au plan tangent

Vu que le gradient d'une fonction est la direction de plus grande pente et que le plan tangent est le plan de plus petite pente, quoi de plus naturel que de penser que le gradient est orthogonal au plan tangent ?

Lemme 11.178.
Soit $\phi\colon \mathbb{R}^n \to \mathbb{R}$ une fonction de classe C^1 et la partie

$$\Gamma = \{x \in \mathbb{R}^n \text{ tel que } \phi(x) = C\} \tag{11.406}$$

pour une certaine constante C.

Soit $x_0 \in \Gamma$. Le gradient de ϕ en x_0 est orthogonal au plan tangent à Γ en x_0.

Démonstration. Un vecteur tangent à Γ en x_0 est de la forme $\gamma'(0)$ où $\gamma\colon \mathbb{R} \to \Gamma$ vérifie $\gamma(0) = x_0$. Vu que ϕ est constante sur Γ nous avons

$$\frac{d}{ds}\Big[\phi\big(\gamma(s)\big)\Big]_{s=0} = 0, \tag{11.407}$$

ce qui donne

$$\sum_i \frac{\partial \phi}{\partial x_i}\big(\gamma(0)\big)\gamma_i'(0) = 0, \tag{11.408}$$

ce qui signifie exactement $\langle (\nabla\phi)(x_0), \gamma'(0)\rangle = 0$. Le vecteur $(\nabla\phi)(x_0)$ est donc perpendiculaire à tout vecteur tangent de Γ en x_0. □

11.18 Dérivée directionnelle de fonctions composées

Nous savons déjà comment dériver les fonctions composées de $\mathbb{R}$ dans $\mathbb{R}$. Si nous avons deux fonctions $f\colon \mathbb{R} \to \mathbb{R}$ et $u\colon \mathbb{R} \to \mathbb{R}$, nous formons la composée $\varphi = f \circ u\colon \mathbb{R} \to \mathbb{R}$ dont la dérivée vaut

$$\varphi'(a) = f'\big(u(a)\big)u'(a). \tag{11.409}$$

Considérons maintenant le cas un peu plus compliqué des fonctions $f\colon \mathbb{R} \to \mathbb{R}$ et $u\colon \mathbb{R}^2 \to \mathbb{R}$, et de la composée

$$\begin{aligned} \varphi\colon \mathbb{R}^2 &\to \mathbb{R} \\ \varphi(x,y) &= f\big(u(x,y)\big). \end{aligned} \tag{11.410}$$

Afin de calculer la dérivée partielle de φ par rapport à x, nous admettons que pour tout a, b et t, il existe $c \in [a, a+t]$ tel que

$$u(a+t,b) = u(a,b) + t\frac{\partial u}{\partial x}(c,b). \tag{11.411}$$

Cela est une généralisation immédiate du théorème 11.130. Nous devons calculer

$$\frac{\partial \varphi}{\partial x}(a,b) = \lim_{t\to 0} \frac{\varphi(a+t,b) - \varphi(a,b)}{t} = \lim_{t\to 0} \frac{f\big(u(a+t,b)\big) - g\big(u(a,b)\big)}{t}. \tag{11.412}$$

Étant donné l'hypothèse que nous avons faite sur u, nous avons

$$f\big(u(a+t,b)\big) = f\big(u(a,b) + t\frac{\partial u}{\partial x}(c,b)\big). \tag{11.413}$$

En utilisant le théorème des accroissements finis pour f, nous avons un point d entre $u(a,b)$ et $u(a,b) + t\frac{\partial u}{\partial x}(c,b)$ tel que

$$f\big(u(a,b) + t\frac{\partial u}{\partial x}(c,b)\big) = f\big(u(a,b)\big) + t\frac{\partial u}{\partial x}(c,b)f'(d). \tag{11.414}$$

Le numérateur de (11.412) devient donc

$$t\frac{\partial u}{\partial x}(c,b)f'(d). \tag{11.415}$$

Certes les points c et d sont inconnus, mais nous savons que c est entre a et $a+t$ ainsi que d se situe entre $u(a,b)$ et $u(a,b) + t\frac{\partial u}{\partial x}(c,b)$. Lorsque nous prenons la limite $t \to 0$, nous avons donc $\lim_{t\to 0} c = a$ et $\lim_{t\to 0} d = u(a,b)$. Nous avons alors

$$\lim_{t\to 0} \frac{t\frac{\partial u}{\partial x}(c,b)f'(d)}{t} = \frac{\partial u}{\partial x}(a,b)f'\big(u(a,b)\big). \tag{11.416}$$

La formule que nous avons obtenue (de façon pas très rigoureuse) est

$$\frac{\partial}{\partial x} f\big(u(x,y)\big) = \frac{\partial u}{\partial x}(x,y)f'\big(u(x,y)\big). \tag{11.417}$$

Prenons maintenant un cas un peu plus compliqué où nous voudrions savoir les dérivées partielles de la fonction φ donnée par

$$\varphi(x,y,z) = f\big(u(x,y), v(x,y,z)\big) \tag{11.418}$$

où $f\colon \mathbb{R}^2 \to \mathbb{R}$, $u\colon \mathbb{R}^2 \to \mathbb{R}$ et $v\colon \mathbb{R}^3 \to \mathbb{R}$.

Commençons par la dérivée partielle par rapport à z. Étant donné que φ ne dépend de z que via la seconde entrée de f, il est normal que seule la dérivée partielle de f par rapport à sa seconde entrée arrive dans la formule :

$$\frac{\partial \varphi}{\partial z}(x,y,z) = \frac{\partial f}{\partial v}\big(u(x,y), v(x,y,z)\big)\frac{\partial v}{\partial z}(x,y,z). \tag{11.419}$$

La dérivée partielle par rapport à y demande de tenir compte en même temps de la façon dont f varie avec sa première entrée et la façon dont elle varie avec sa seconde entrée ; cela nous fait deux termes :

$$\frac{\partial \varphi}{\partial y}(x,y,z) = \frac{\partial f}{\partial u}\big(u(x,y), v(x,y,z)\big)\frac{\partial u}{\partial y}(x,y) + \frac{\partial f}{\partial v}\big(u(x,y), v(x,y,z)\big)\frac{\partial v}{\partial y}(x,y,z). \tag{11.420}$$

Cette formule a une interprétation simple. Lançons un caillou du sommet d'une falaise. Son mouvement est une chute libre avec une vitesse initiale horizontale :

$$\begin{cases} x(t) = v_0 t & (11.421a) \\ y(t) = h_0 - \dfrac{gt^2}{2} & (11.421b) \end{cases}$$

où v_0 est la vitesse initiale horizontale et h_0 est la hauteur de la falaise. Si nous sommes intéressés à la distance entre le caillou et le bas de la falaise (point $(0,0)$), le théorème de Pythagore nous dit que

$$d(t) = \sqrt{x^2(t), y^2(t)}. \tag{11.422}$$

Pour trouver la variation de la distance par rapport au temps il faut savoir de combien la distance varie lorsque x varie et multiplier par la variation de x par rapport à t, et puis faire la même chose avec y.

Théorème 11.179.
Soit $g\colon \mathbb{R}^m \to \mathbb{R}^n$ une fonction différentiable[50] *en a, et $f\colon \mathbb{R}^n \to \mathbb{R}^p$ une fonction différentiable en $g(a)$. Si nous définissons $\varphi(x) = (f \circ g)(x)$, alors pour tout $i = 1, \dots, m$, nous avons*

$$\frac{\partial \varphi}{\partial x_i}(a) = \sum_{k=1}^{n} \frac{\partial f}{\partial y_k}\big(g(a)\big)\frac{\partial g}{\partial x_i} \tag{11.423}$$

où $\frac{\partial f}{\partial y_k}$ dénote la dérivée partielle de f par rapport à sa k-ième variable.

Donnons un exemple d'utilisation de cette formule. Si

$$\begin{aligned} g\colon \mathbb{R}^2 &\to \mathbb{R}^3 \\ f\colon \mathbb{R}^3 &\to \mathbb{R}, \end{aligned} \tag{11.424}$$

nous avons $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$. Les dérivées partielles de φ sont données par les formules

$$\frac{\partial \varphi}{\partial x}(x,y) = \frac{\partial f}{\partial x_1}\big(g(x,y)\big)\frac{\partial g_1}{\partial x}(x,y) + \frac{\partial f}{\partial x_2}\big(g(x,y)\big)\frac{\partial g_2}{\partial y}(x,y) + \frac{\partial f}{\partial x_3}\big(g(x,y)\big)\frac{\partial g_3}{\partial x}(x,y) \tag{11.425}$$

et

$$\frac{\partial \varphi}{\partial y}(x,y) = \frac{\partial f}{\partial x_1}\big(g(x,y)\big)\frac{\partial g_1}{\partial y}(x,y) + \frac{\partial f}{\partial x_2}\big(g(x,y)\big)\frac{\partial g_2}{\partial y}(x,y) + \frac{\partial f}{\partial x_3}\big(g(x,y)\big)\frac{\partial g_3}{\partial y}(x,y) \tag{11.426}$$

Notez que les dérivées de φ et des composantes de g sont calculées en (x,y), tandis que celles de f sont calculées en $g(x,y)$.

11.19 Formes différentielles

Nous parlerons de formes différentielles exactes et fermées dans la section 19.80.

11.19.1 Décomposition dans la base duale

Définition 11.180.
*Soit U, un ouvert dans $\mathbb{R}^n$. Une 1-**forme différentielle** ω sur U est une application*

$$\begin{aligned} \omega\colon U &\to (\mathbb{R}^n)^* \\ x &\mapsto \omega_x. \end{aligned} \tag{11.427}$$

50. Je ne suis pas certain que l'hypothèse de différentiabilité soit obligatoire.

Remarque 11.181.
L'ensemble des 1-formes différentielles forment un espace vectoriel avec les définitions

$$\begin{aligned}(\lambda\omega)_x(v) &= \lambda\omega_x(v)\\ (\omega+\mu)_x(v) &= \omega_x(v)+\mu_x(v).\end{aligned} \tag{11.428}$$

Nous connaissons la de $(\mathbb{R}^n)^*$ définie en 4.130. Nous allons noter ces formes par dx_i :

$$\begin{aligned}e_1^* &= dx_1\colon v\mapsto v_1\\ &\vdots\\ e_n^* &= dx_n\colon v\mapsto v_n\end{aligned} \tag{11.429}$$

Toute forme différentielle s'écrit

$$\omega_x = \sum_{i=0}^{n} a_i(x)dx_i \tag{11.430}$$

où $a_1,\dots,a_n$ sont les composantes de ω dans la base usuelle, et sont des fonctions à valeurs réelles.

Lemme 11.182.
Une 1*-forme différentielle est* ***continue*** *si les fonctions a_i sont continues. La forme sera C^k quand les a_i seront C^k.*

Pour un vecteur $v=(v_1,\dots,v_n)$ on a donc par définition de dx_i

$$\omega_x(v) = \sum_{i=0}^{n} a_i(x)v_i. \tag{11.431}$$

Ces fonctions a_i peuvent être trouvées en appliquant ω aux éléments de la base canonique de $\mathbb{R}^n$:

$$a_j(x) = \omega_x(e_j) \tag{11.432}$$

parce que $\omega_x(e_j) = \sum_i a_i(x)dx_i(e_i) = \sum_i a_i(x)\delta_{ij} = a_j(x)$.

11.19.2 L'isomorphisme musical

Si G est un champ de vecteur sur $\mathbb{R}^n$, et si $x\in\mathbb{R}^n$, nous pouvons définissons

$$\begin{aligned}G_x^\flat\colon \mathbb{R}^n &\to \mathbb{R}\\ v &\mapsto \langle G(x),v\rangle\end{aligned} \tag{11.433}$$

Pour chaque x, l'application $G_x^\flat$ est une forme sur $\mathbb{R}^n$, c'est-à-dire une application linéaire de $\mathbb{R}^n$ vers $\mathbb{R}$. Nous écrivons que

$$G_x^\flat \in \big(\mathbb{R}^n\big)^*. \tag{11.434}$$

Nous pouvons ainsi déterminer le développement de $G^\flat$ dans la base des dx_i en faisant le calcul

$$G_x^\flat(e_i) = \langle G(x),e_i\rangle = G_i(x), \tag{11.435}$$

donc les composantes de $G^\flat$ dans la base dx_i sont exactement les composantes de G dans la base e_i :

$$G_x^\flat = G_1(x)dx_1+\cdots+G_n(x)dx_n. \tag{11.436}$$

La construction inverse existe également. Si ω est une 1-forme différentielle, nous pouvons définir le champ de vecteur $\omega^\sharp$ par la formule (implicite)

$$\omega_x(v) = \langle \omega^\sharp(x),v\rangle \tag{11.437}$$

pour tout $v\in\mathbb{R}^n$. Par définition, $(\omega^\sharp)^\flat=\omega$.

Lemme 11.183.
En composantes nous avons :

$$\omega^\sharp(x) = \big(a_1(x),\dots,a_n(x)\big). \tag{11.438}$$

Si G est un champ de vecteurs, alors $(G^\flat)^\sharp = G$.

11.20 Différentielle

Nous avons déjà donné une définition abstraite de la différentielle dans la définition 10.127. Nous en voyons maintenant quelques motivations dans le cas de fonctions sur $\mathbb{R}^2$ ou $\mathbb{R}^n$.

11.20.1 Exemples introductifs

La notion de dérivée est associée à la recherche de la droite tangente à une courbe. Reprenons rapidement le cheminement. La dérivée de $f\colon \mathbb{R} \to \mathbb{R}$ au point a est un nombre $f'(a)$, qui définit donc une application linéaire dont le coefficients angulaire est $f'(a)$, et que nous notons df_a :

$$\begin{aligned} df_a\colon \mathbb{R} &\to \mathbb{R} \\ u &\mapsto f'(a)u. \end{aligned} \tag{11.439}$$

La droite donnée par l'équation

$$y(a+u) = f'(a)u \tag{11.440}$$

est parallèle à la tangente en a. Pour trouver la tangente, il suffit de la décaler de la hauteur qu'il faut. L'équation de la droite tangente au graphe de f au point $\big(a, f(a)\big)$ devient

$$y(x) = f(a) + f'(a)(x-a) = f(a) + df_a(x-a). \tag{11.441}$$

Nous nous proposons de généraliser cette formule au cas de la recherche du plan tangent à une surface.

Exemple 11.184.
Considérons $f(x,y) = x^2y + y^2e^x$. Les dérivées partielles sont

$$\begin{aligned} \frac{\partial f}{\partial x} &= 2xy + y^2e^x \\ \frac{\partial f}{\partial y} &= x^2 + 2ye^x. \end{aligned} \tag{11.442}$$

△

Cet exemple était l'exemple facile où tout se passe bien.

Exemple 11.185.
Les choses sont moins simples lorsqu'on considère la fonction suivante :

$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{si } (x,y) = (0,0). \end{cases} \tag{11.443}$$

On voit que pour tout x et tout y, nous avons $f(x,0) = f(0,y) = 0$. Donc cette fonction est nulle sur les axes horizontaux et verticaux. Nous avons en particulier

$$\begin{aligned} \frac{\partial f}{\partial x}(0,0) &= 0 \\ \frac{\partial f}{\partial y}(0,0) &= 0. \end{aligned} \tag{11.444}$$

Donc ces dérivées partielles existe.

Il n'est par contre pas question de dire que cette fonction « va bien » autour du point $(0,0)$. En effet si nous regardons sa valeur sur la droite diagonale $y = x$, nous avons

$$f(x,x) = \frac{x^2}{2x^2} = \frac{1}{2}. \tag{11.445}$$

Par conséquent si nous suivons la fonction le long de la droite $y = x$, la hauteur vaut $\frac{1}{2}$ en permanence, sauf juste en $(0,0)$ où la fonction fait un grand plongeon !

```
sage: var('x,y')
(x, y)
sage: f(x,y)=(x*y)/(x**2+y**2)
sage: plot3d(f,(x,-2,2),y(-2,2))
```

D'ailleurs elle fait un plongeon le long de toutes les droites (sauf verticale et horizontale). En effet si nous regardons la fonction le long de la droite $y = mx$, nous avons

$$f(x, mx) = \frac{mx^2}{x^2 + m^2x^2} = \frac{m}{1 + m^2}. \tag{11.446}$$

La fonction est donc *constante* sur chacune de ces droites. Il n'est donc pas question de dire que cette fonction est « dérivable » en $(0, 0)$, vu qu'elle fait des grands sauts dans presque toutes les directions. △

Nous devons donc trouver mieux que les dérivées partielles pour étudier le comportement des fonctions un peu problématiques.

11.20.2 Différentielle

Nous nous souvenons de l'équation (11.234) qui nous dit que pour une fonction d'une variable la dérivabilité signifiait qu'il existait un nombre ℓ et une fonction α tels que

$$f(x) = f(a) + \ell(x - a) + (x - a)\alpha(x - a) \tag{11.447}$$

et $\lim_{t\to 0} \alpha(t) = 0$.

En nous inspirant de cela, nous comprenons peut-être un peu le pourquoi de la définition 10.127.

11.186.
L'objet df_a est *en soi* une application $df_a \colon \mathbb{R}^m \to \mathbb{R}^n$. Nous notons $df_a(u)$ la valeur de df_a sur le vecteur $u \in \mathbb{R}^m$. En particulier, l'application df est une forme différentielle au sens de la définition 11.180.

11.187.
Les propositions 11.197 et 11.200 vont montrer qu'en étudiant bien les dérivées partielles, nous pouvons conclure à la différentiabilité d'une fonction. Attention cependant, nous verrons dans l'exemple 11.206 que l'existence des dérivées directionnelles partielles ne permettait pas de conclure à la différentiabilité.

11.20.3 Matrice de la différentielle

La différentielle est une application linéaire. Elle possède donc une matrice lorsque des bases sont fixées.

Proposition 11.188.
Soient une application différentiable $f \colon \mathbb{R}^m \to \mathbb{R}^n$ et $a \in \mathbb{R}^m$. Dans les bases canoniques de $\mathbb{R}^m$ et $\mathbb{R}^n$, la matrice de df_a est

$$(df_a)_{ij} = \frac{\partial f_i}{\partial x_j}(a). \tag{11.448}$$

Démonstration. Le lien entre matrice et application linéaire est fait dans la proposition 4.68. Dans le cas des bases canoniques de $\mathbb{R}^m$ et $\mathbb{R}^n$ nous savons qu'extraire une composante revient à prendre le produit scalaire. Nous avons donc

$$(df_a)_{ij} = \big(df_a(e_j)\big)_i = df_a(e_j) \cdot e_i. \tag{11.449}$$

La linéarité de la dérivation donne alors

$$(df_a)_{ij} = df_a(e_j) \cdot e_i = \frac{d}{dt}\Big[f(a + te_j)\Big]_{t=0} \cdot e_i = \frac{d}{dt}\Big[f_i(a + te_j)\Big]_{t=0} = \frac{\partial f_i}{\partial x_j}(a). \tag{11.450}$$

□

11.20.4 Quelques propriétés

Lemme 11.189.
La différentielle d'une application linéaire est l'application elle-même. Plus précisément : soit une application linéaire $f\colon E \to F$. Alors nous avons, pour tout $a \in E$ et $u \in E$:

$$df_a(u) = f(u). \tag{11.451}$$

Démonstration. Pour rappel, toujours bon à avoir en tête : $df\colon E \to \mathcal{L}(E, F)$. En posant $T(u) = f(u)$ nous avons :

$$\lim_{h\to 0} \frac{\|f(a+h) - f(a) - T(h)\|_F}{\|h\|_E} = 0 \tag{11.452}$$

parce que le numérateur est nul pour tout h. □

Lemme 11.190 ([1])**.**
Soit une fonction $g\colon \mathbb{R} \to \mathbb{R}$ de classe C^∞. Nous posons

$$\begin{aligned} f\colon \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto g(x). \end{aligned} \tag{11.453}$$

Alors f est de classe C^∞ sur $\mathbb{R}^2$.

Démonstration. Le problème lorsqu'il faut démontrer qu'une fonction est de classe C^∞, c'est que $d^k f$ sera une application de $\mathbb{R}^2$ vers un espace qui est un terrible emboîtement de $\mathcal{L}(\mathbb{R}^2, \ldots)$. Pour traiter cette difficulté, nous considérons les espaces suivants : $V_0 = \mathbb{R}$ et par récurrence $V_{k+1} = \mathcal{L}(\mathbb{R}^2, V_k)$.

Et nous considérons également les éléments

$$\begin{aligned} \alpha_1\colon \mathbb{R}^2 &\to \mathbb{R} \\ (u, v) &\mapsto u \end{aligned} \tag{11.454}$$

et plus généralement $\alpha_k \in V_k$ donné par

$$\begin{aligned} \alpha_k\colon \mathbb{R}^2 &\to V_{k-1} \\ (u, v) &\mapsto u\alpha_{k-1}. \end{aligned} \tag{11.455}$$

Notons que dans l'expression $u\alpha_{k-1}$, il s'agit d'un produit entre un scalaire $u \in \mathbb{R}$ et un vecteur $\alpha_{k+1} \in V_{k-1}$.

Nous prouvons maintenant par récurrence que $d^k f_{(a,b)} = g^{(k)}(a)\alpha_k$, en utilisant directement la définition.

<u>Initialisation</u> Pour $k = 1$, nous calculons

$$\frac{|f(a+h_1, b+h_2) - f(a,b) - g'(a)\alpha_1(h)|}{\|h\|} = \frac{|g(a+h_1) - g(a) - g'(a)h_1|}{\|h\|} \tag{11.456}$$

Notre but est de calculer la limite de cela lorsque $h \xrightarrow{\mathbb{R}^2} 0$ avec $h \neq 0$. L'hypothèse sur la dérivabilité de g nous indique que si $0 < |t| < \delta$, alors

$$\frac{|g(a+t) - g(a) - tg'(a)|}{|t|} < \epsilon. \tag{11.457}$$

Nous considérons donc la boule épointée de $\mathbb{R}^2$ de rayon δ : $B = B\big((0,0), \delta\big)\backslash\{(0,0)\}$, et nous considérons $h \in B$. Deux cas sont à distinguer : $h_1 = 0$ et $h_1 \neq 0$.

Si $h_1 = 0$, alors

$$\frac{|g(a+h_1) - g(a) - g'(a)h_1|}{\|h\|} = 0. \tag{11.458}$$

Sinon nous avons $0 < h_1 \leqslant \|h\| < \delta$ et donc

$$\frac{|g(a+h_1)-g(a)-g'(a)h_1|}{\|h\|} \leqslant \frac{|g(a+h_1)-g(a)-g'(a)h_1|}{|h_1|} < \epsilon \tag{11.459}$$

par la relation (11.457). Nous avons donc bien

$$\lim_{h\to 0} \frac{|f(a+h_1,b+h_2)-f(a,b)-g'(a)\alpha_1(h)|}{\|h\|} = 0. \tag{11.460}$$

Récurrence Nous supposons que $d^k f_{a,b} = g^{(k)}(a)\alpha_k$, et nous devons prouver que $d^k f$ est différentiable et que $d^{k+1} f_{(a,b)} = g^{(k+1)}(a)\alpha_{k+1}$. Pour cela nous introduisons tout dans la définition de la différentielle pour voir ce qui arrive.

Nous avons :

$$\begin{aligned}\frac{d^k f_{(a+h_1,b+h_2)} - d^k f_{(a,b)} - g^{(k+1)}(a)\alpha_{k+1}(h_1,h_2)}{\|h\|} \\ = \frac{g^{(k)}(a+h_1)\alpha_k - g^{(k)}(a)\alpha_k - g^{(k+1)}(a)h_1\alpha_k}{\|h\|}.\end{aligned} \tag{11.461}$$

Cela est, pour chaque $h \neq 0$, un élément V_k, mais le coefficient α_k se factorise de telle sorte que nous devons seulement calculer la limite (si elle existe)

$$\lim_{h\to 0} \frac{g^{(k)}(a+h_1) - g^{(k)}(a) - h_1 g^{(k+1)}(a)}{\|h\|}. \tag{11.462}$$

Le même jeu de séparation entre $h_1 = 0$ et $h_1 \neq 0$ que dans le cas $k = 1$ nous permet de déduire que cette limite existe et vaut zéro, grace à la définition de $g^{(k+1)}$.

Nous avons donc prouvé que f est différentiable autant que fois que souhaité. Elle est donc de classe C^∞ comme annoncé. □

11.20.5 Différentielle, dual et forme différentielle

11.20.5.1 Dans la base duale

Nous avons déjà parlé en (11.429) de la base $\{dx_i\}_{i=1,\dots,n}$ des formes différentielles sur $\mathbb{R}^n$.

Proposition 11.191.
La forme de base dx_i est la différentielle de la fonction de projection

$$\begin{aligned}\mathtt{proj}_i \colon \mathbb{R}^n &\to \mathbb{R} \\ v &\mapsto v_i.\end{aligned} \tag{11.463}$$

Autrement dit nous avons

$$d(\mathtt{proj}_i)_a = dx_i \tag{11.464}$$

pour tout i et pour tout a.

Démonstration. Le quotient

$$\frac{\mathtt{proj}_i(a+h) - \mathtt{proj}_i(a) - dx_i(h)}{\|h\|} \tag{11.465}$$

est toujours nul. La limite est a fortiori nulle. □

Nous avons donc $(d\,\mathtt{proj}_i)_a = dx_i$ pour tout a. Notons que les fonctions dx_i et $\mathtt{proj}_i$ sont les mêmes. Cela justifie la notation « dx_i » pour les formes différentielles de base, parce que ce sont les différentielles des fonctions « coordonnées » que nous pouvons noter x_i.

Étant donné une fonction f, il est légitime de nous demander comment (si elle existe) la différentielle se décompose en chaque point dans la base duale. C'est-à-dire fixer les fonctions a_i en termes des dérivées de f pour avoir

$$df_a = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) dx_i. \tag{11.466}$$

C'est ce que nous allons faire dans le corolaire 11.195. :5cnext

Exemple 11.192.
Si $F\colon \mathbb{R}^2 \to \mathbb{R}$ est une fonction C^2, sa différentielle est la forme

$$dF = \frac{\partial F}{\partial x} dx + \frac{\partial F}{\partial y} dy. \tag{11.467}$$

Si nous nommons f et g les fonctions $\partial_x F$ et $\partial_y F$, nous avons donc

$$Df = f dx + g dy, \tag{11.468}$$

qui vérifie

$$\partial_y f = \partial_x g, \tag{11.469}$$

parce que $\frac{\partial f}{\partial y} = \frac{\partial^2 F}{\partial x \partial y} = \frac{\partial^2 F}{\partial y \partial x} = \frac{\partial g}{\partial x}$. Ce que nous avons donc prouvé, c'est que △

Lemme 11.193.
Si $f dx + g dy$ est la différentielle d'une fonction de classe C^2 sur $\mathbb{R}^2$, alors $\partial_y f = \partial_x g$.

11.20.6 Ce n'est pas la différentielle extérieure

Il existe une notion de différentielle extérieure, mais ce n'est pas celle-là que nous utilisons la majorité du temps. En particulier si E et F sont des espaces vectoriels normés, lorsque $f\colon E \to F$ est une fonction, df est une application

$$df\colon E \to \mathcal{L}(E, F) \tag{11.470}$$

et la différentielle seconde est la différentielle de cette application-là. Chose faisable parce que $\mathcal{L}(E, F)$ est un espace vectoriel on ne peut plus respectable.

Soit $D \subset \mathbb{R}^n$. Par définition de la différentielle extérieure d'une 1-forme, nous avons une formule de Leibnitz

$$d(f\omega) = df \wedge \omega + f d\omega. \tag{11.471}$$

En particulier,

$$d(f dx) = df \wedge dx + f \underbrace{d(dx)}_{=0} = \frac{\partial f}{\partial x} dx \underbrace{dx \wedge dx}_{=0} + \frac{\partial f}{\partial y} dy \wedge dx. \tag{11.472}$$

Attention : la différentielle extérieure n'est pas la différentielle usuelle. Certes dans le cas d'une 0-forme (c'est-à-dire d'une fonction), les deux notions coïncident, mais ça ne va pas plus loin. La différentielle extérieure vérifie $d^2\omega = 0$ pour tout ω, y compris pour les fonctions : si $\omega = df$ alors $d\omega = 0$.

Nous mentionnerons la différentielle extérieure dans le cas de

(1) Théorème de Stockes 19.71.

11.20.7 Fonctions composées

Cette façon de voir la différentielle nous permet de jeter un nouveau regard sur la formule de différentiation des fonctions composées. Soient

$$\begin{aligned} f\colon \mathbb{R}^p &\to \mathbb{R}^n \\ g\colon \mathbb{R}^n &\to \mathbb{R}, \end{aligned} \tag{11.473}$$

et $h\colon \mathbb{R}^p \to \mathbb{R}$ définie par

$$h(u) = h\big(f(u)\big) = (g \circ f)(u). \tag{11.474}$$

Nous allons noter x les coordonnées de $\mathbb{R}^p$, a un point de $\mathbb{R}^p$ et u, un vecteur de $\mathbb{R}^p$ accroché au point a. Pour $\mathbb{R}^n$, les notations seront que les coordonnées sont y, b est un point de $\mathbb{R}^n$ et v est un vecteur « accroché » au point b.

Nous avons

$$dg_b(v) = \sum_{i=1}^{n} \frac{\partial g}{\partial y_i}(b) dy_i(v). \tag{11.475}$$

Ici $dy_i(v)$ signifie la $i$ème composante de v. C'est simplement v_i. Cette formule étant valable pour tout point $b \in \mathbb{R}^n$ et pour tout vecteur v, nous pouvons l'écrire en particulier pour

$$\begin{cases} b = f(a) & \text{(11.476a)} \\ v = df_a(u). & \text{(11.476b)} \end{cases}$$

Cela donne

$$dg_{f(a)}\big(df_a(u)\big) = \sum_{i=1}^{n} \frac{\partial g}{\partial y_i}\big(f(a)\big) dy_i\big(df_a(u)\big). \tag{11.477}$$

Mais

$$df_a(u) = \sum_{j=1}^{p} \frac{\partial f}{\partial x_j}(a) dx_j(u), \tag{11.478}$$

donc la $i$ème composante de ce vecteur est

$$\big(df_a(u)\big)_i = \sum_{j=1}^{p} \frac{\partial f_i}{\partial x_j}(a) dx_j(u). \tag{11.479}$$

En remplaçant $dy_i\big(df_a(u)\big)$ par cela dans l'expression (11.477), nous trouvons

$$dg_{f(a)}\big(df_a(u)\big) = \sum_{i=1}^{n} \frac{\partial g}{\partial y_i}\big(f(a)\big) \sum_{j=1}^{p} \frac{\partial f_i}{\partial x_j}(a) dx_j(u). \tag{11.480}$$

Nous pouvons vérifier que cela est la différentielle de $g \circ f$ au point a appliquée au vecteur u. En effet

$$d(g \circ f)_a(u) = \sum_{j=1}^{p} \frac{\partial (g \circ f)}{\partial x_j}(a) dx_j(u), \tag{11.481}$$

tandis que, par la dérivation de fonctions composées,

$$\frac{\partial (g \circ f)}{\partial x_j}(a) = \sum_{i=1}^{n} \frac{\partial g}{\partial y_i}\big(f(a)\big) \frac{\partial f_i}{\partial x_j}(a). \tag{11.482}$$

Au final, ce que nous avons prouvé est que

$$d(g \circ f)_a(u) = dg_{f(a)}\big(df_a(u)\big). \tag{11.483}$$

11.20.8 Continuité, dérivabilité et différentiabilité

Le théorème suivant reprend les principales propriétés d'une fonction différentiable. Il est à ne pas confondre avec le théorème 11.243 qui dira que si les dérivées partielles sont continues sur un voisinage de a, alors f est différentiable en a.

Proposition 11.194.
Si f est différentiable au point $a \in \mathbb{R}^n$ alors

(1) elle est continue en a,

(2) elle admet une dérivée dans toutes les directions de $\mathbb{R}^m$,

(3) toutes les dérivées directionnelles $\partial_u f(a)$ existent et nous avons l'égalité

$$\begin{aligned} df_a \colon \mathbb{R}^n &\to \mathbb{R}^m \\ u &\mapsto df_a(u) = \frac{\partial f}{\partial u}(a) = \sum_i \frac{\partial f}{\partial x_i}(a) u_i, \end{aligned} \tag{11.484}$$

si les u^i sont les composantes de u dans la base canonique de $\mathbb{R}^n$.

La dernière égalité sera de temps en temps utilisée sous la forme

$$df_a(u) = \frac{d}{dt}\Big[f(a+tu)\Big]_{t=0}. \tag{11.485}$$

Démonstration. La limite

$$\lim_{h \to 0_m} \frac{\|f(a+h) - f(a) - T(h)\|_n}{\|h\|_m} = 0,$$

implique que

$$\lim_{h \to 0_m} \|f(a+h) - f(a) - T(h)\|_n = 0.$$

Comme T est dans $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$, on a $\lim_{h \to 0} T(h) = 0$, d'où la continuité de f au point a.

Si u est un vecteur non nul, la différentiabilité de f au point a implique

$$\lim_{t \to 0} \frac{\|f(a+tu) - f(a) - T(tu)\|_n}{\|tu\|_m} = 0,$$

par la linéarité de T et par l'égalité $\|tu\|_m = |t|\|u\|_m$ on obtient

$$\lim_{t \to 0} \frac{f(a+tu) - f(a)}{|t|} = T(u).$$

Donc f est dérivable suivant le vecteur u et $\partial_u f(a) = T(u) = df_a(u)$. □

Corolaire 11.195.
Si f est différentiable, alors la forme différentielle df_a se décompose en

$$df_a f = \sum_i (\partial_i f)(a) dx_i. \tag{11.486}$$

Démonstration. Vue la définition des formes dx_i nous pouvons remplacer u_i par $dx_i(u)$ dans l'égalité (11.484) et écrire

$$df_a(u) = \sum_i (\partial_i f)(a) dx_i(u) \tag{11.487}$$

et donc écrire l'égalité demandée. □

Le lemme suivant regroupe quelques égalités avec lesquelles nous allons souvent travailler. Il explique comment sont liés les dérivées directionnelles, les dérivées partielles et la différentielle.

Lemme 11.196.
Si $f \colon \mathbb{R}^m \to \mathbb{R}^n$ est une fonction différentiable, alors

$$df_a(u) = \frac{\partial f}{\partial u}(a) = \frac{d}{dt}\Big[f(a+tu)\Big]_{t=0} = \sum_{i=1}^m u_i \frac{\partial f}{\partial x_i}(a) = \nabla f(a) \cdot u \tag{11.488}$$

pour tout vecteur $u \in \mathbb{R}^m$

Démonstration. La première égalité est la proposition 11.194, et la seconde est seulement la définition de la dérivée directionnelle avec des notations un peu plus snob. En particulier nous avons

$$df_a(e_i) = \frac{\partial f}{\partial x_i}(a). \tag{11.489}$$

Pour le reste c'est la linéarité de la différentielle qui joue : le vecteur u peut être écrit de façon unique comme combinaison linéaire des vecteurs de base

$$u = \sum_{i=1}^{m} u_i e_i, \qquad u_i \in \mathbb{R}, \forall i \in \{1, \dots, m\}.$$

Alors, la linéarité de df_a nous donne

$$df_a(u) = df_a\left(\sum_{i=1}^{m} u_i e_i\right) = \sum_{i=1}^{m} u_i \left(df_a e_i\right) = \sum_{i=1}^{m} u_i \frac{\partial f}{\partial x_i}(a). \tag{11.490}$$

Le lien avec le gradient est la définition du produit scalaire (9.269). □

La formule $df_a(u) = \frac{d}{dt}\Big[f(a+tu)\Big]_{t=0}$ est bien utile pour calculer des différentielles, mais elle ne permet pas de prouver que f est différentiable. Autrement dit, même si le calcul de la dérivée $\frac{d}{dt}\Big[f(a+tu)\Big]_{t=0}$ donne un résultat pour tout u, nous ne pouvons pas en déduire que f est différentiable au point a.

Proposition 11.197.
Soient f une fonction de x et y et un point $(a,b) \in \mathbb{R}^2$. Si les nombres $\partial_x f(a,b)$ et $\partial_y f(a,b)$ existent et s'il existe une fonction $\alpha\colon \mathbb{R} \to \mathbb{R}$ telle que

$$\begin{aligned} f(x,y) = f(a,b) &+ \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b) \\ &+ \|(x,y)-(a,b)\|\alpha\Big(\|(x,y)-(a,b)\|\Big) \end{aligned} \tag{11.491}$$

et

$$\lim_{t\to 0} \alpha(t) = 0, \tag{11.492}$$

alors f est différentiable en (a,b).

11.198.
Dans cet énoncé nous avons écrit $d\big((x,y),(a,b)\big)$ la distance entre (x,y) et (a,b), c'est-à-dire le nombre $\sqrt{(x-a)^2+(y-b)^2}$. Afin d'écrire l'équation (11.491) sous forme plus compacte, nous introduisons le vecteur

$$\nabla f(a,b) = \begin{pmatrix} \frac{\partial f}{\partial x}(a,b) \\ \frac{\partial f}{\partial y}(a,b). \end{pmatrix} \tag{11.493}$$

L'équation (11.491) devient alors

$$f(X) = f(P) + \nabla f(a,b) \cdot (X-P) + \|X-P\|\alpha\big(\|X-P\|\big). \tag{11.494}$$

Le vecteur $(\nabla f)(a,b)$ est appelé le **gradient** de f au point (a,b).

Remarque 11.199.
Nous avons introduit la notation ∇f pour le gradient d'une fonction f. Nous allons par la suite introduire $\nabla \cdot F$ pour la divergence du champ de vecteurs F et $\nabla \times F$ pour son rotationnel.

Toutes les formules pour ∇f, $\nabla \cdot F$ et $\nabla \times F$ peuvent facilement être mémorisées en pensant à ∇ comme étant le vecteur

$$\nabla = \begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \end{pmatrix}. \tag{11.495}$$

Nous allons ici cependant seulement penser à (11.495) comme un moyen mnémotechnique ; nous ne donnons pas de définitions à « ∇ » tout seul.

Proposition 11.200.
Soit f une fonction de deux variables admettant des dérivées partielles $\partial_x f(x,y)$ et $\partial_y f(x,y)$ qui sont elles-mêmes des fonctions continues de x et y. Alors la fonction f est différentiable partout.

Proposition 11.201.
Si f est différentiable en (a,b) alors pour tout vecteur u, la fonction

$$\begin{aligned}\varphi\colon \mathbb{R} &\to \mathbb{R}\\ t &\mapsto f(a+tu_1, b+tu_2)\end{aligned} \tag{11.496}$$

est dérivable en 0 et on a

$$\varphi'(0) = \nabla f(p) \cdot u \tag{11.497}$$

où nous avons noté $p=(a,b)$.

Démonstration. Récrivons la formule (11.494) sous la forme

$$f(x) = f(p) + \nabla f(p) \cdot (x-p) + \|x-p\|\alpha(\|x-p\|). \tag{11.498}$$

Cela étant vrai pour tout x, nous l'écrivons en particulier pour $x = p+tu$ où t est un réel et u est le vecteur unitaire choisi. Nous avons donc

$$f(p+tu) = f(p) + t\nabla f(p) \cdot u + \|tu\|\alpha(\|tu\|). \tag{11.499}$$

En utilisant le fait que u est unitaire, $\|tu\| = |t|\|u\| = |t|$. La dérivée de φ en 0 est alors donnée par

$$\lim_{t\to 0} \frac{f(p+tu)-f(p)}{t} = \lim_{t\to 0} \nabla f(p) \cdot u + \alpha(|t|). \tag{11.500}$$

Lorsque nous prenons la limite, le membre de gauche devient $\varphi'(0)$ tandis que dans le membre de droite, le second terme disparaît. Nous avons finalement

$$\varphi'(0) = \nabla f(p) \cdot u \tag{11.501}$$

□

11.20.9 Calcul de valeurs approchées

Si nous remplaçons les accroissements $x-a$ et $y-b$ par h et k, le critère de différentiabilité s'écrit

$$\begin{aligned} f(a+h,b+k) = f(a,b) + \frac{\partial f}{\partial x}(a,b)h + \frac{\partial f}{\partial y}(a,b)k \\ + \sqrt{h^2+k^2}\alpha\big(\sqrt{h^2+k^2}\big). \end{aligned} \tag{11.502}$$

Le dernier terme du membre de droite tend vers zéro à une vitesse double lorsque h et k tendent vers zéro : d'une part parce que $\sqrt{h^2+k^2}$ tend vers zéro et d'autre part parce que $\alpha\big(\sqrt{h^2+k^2}\big)$ tend vers zéro. Nous avons donc la « bonne » approximation

$$f(x,y) \simeq f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b). \tag{11.503}$$

lorsque (x,y) n'est pas trop loin de (a,b). Cette expression est évidemment une généralisation immédiate de l'équation (11.237). Elle exprime que l'on peut obtenir des informations sur la valeur d'une fonction en (x,y) si on peut calculer la fonction et ses dérivées en un point (a,b) non loin de (x,y).

Cette formule peut aussi être vue sous la forme suivante, plus pratique dans certains calculs :

$$f(a+\Delta x, b+\Delta y) \simeq f(a,b) + \Delta x\frac{\partial f}{\partial x}(a,b) + \Delta y\frac{\partial f}{\partial y}(a,b). \tag{11.504}$$

Exemple 11.202.
Prenons la fonction $f(x,y) = \cos(x)\sin(y)$ et calculons une approximation de

$$f\big(\frac{\pi}{3} + 0.01, \frac{\pi}{2} + 0.03\big). \tag{11.505}$$

D'abord les dérivées partielles sont

$$\begin{aligned} \frac{\partial f}{\partial x}(x,y) &= -\sin(x)\sin(y) \\ \frac{\partial f}{\partial y}(x,y) &= \cos(x)\cos(y). \end{aligned} \tag{11.506}$$

Nous allons utiliser l'approximation

$$f\big(\frac{\pi}{3} + 0.01, \frac{\pi}{2} + 0.03\big) \simeq f\big(\frac{\pi}{3}, \frac{\pi}{2}\big) + 0.01\frac{\partial f}{\partial x}\big(\frac{\pi}{3}, \frac{\pi}{2}\big) + 0.03\frac{\partial f}{\partial y}\big(\frac{\pi}{3}, \frac{\pi}{2}\big). \tag{11.507}$$

Nous avons

$$\begin{aligned} \frac{\partial f}{\partial x}\big(\frac{\pi}{3}, \frac{\pi}{2}\big) &= -\sin\frac{\pi}{3}\sin\frac{\pi}{2} = -\frac{\sqrt{3}}{2} \\ \frac{\partial f}{\partial y}\big(\frac{\pi}{3}, \frac{\pi}{2}\big) &= \cos\frac{\pi}{3}\cos\frac{\pi}{2} = 0. \end{aligned} \tag{11.508}$$

Par conséquent

$$f\big(\frac{\pi}{3} + 0.01, \frac{\pi}{2} + 0.03\big) \simeq \frac{1}{2} - 0.01\frac{\sqrt{3}}{2} = \frac{1}{2} - \frac{\sqrt{3}}{200}. \tag{11.509}$$

```
sage: var('x,y')
(x, y)
sage: f(x,y)=cos(x)*sin(y)
sage: a=f(pi/3+0.01,pi/2+0.03)
sage: numerical_approx(a)
0.491093815387986
sage: b=1/2-sqrt(3)/200
sage: numerical_approx(b)
0.491339745962156
sage: numerical_approx(a-b)
-0.000245930574169814
```

Cela fait une erreur de l'ordre du dix millième.

△

Remarque 11.203.
Les esprits les plus critiques diront que cette vérification pas Sage n'en est pas une parce que Sage a certainement utilisé un algorithme d'approximation qui se base sur la même idée que ce que nous venons de faire, et que par conséquent le fait qu'il obtienne le même résultat que nous est un peu tautologique.

Ils n'auront pas tort. Cependant, le code source de Sage est disponible publiquement [51] ; vous pouvez aller le lire et vérifier qu'il y a effectivement une *preuve* que le résultat fourni par Sage possède une bonne dizaine de décimales correctes.

Cette disponibilité publique du code source est une des nombreuses différences fondamentales entre Sage et votre calculatrice [52]. Dois-je vous rappeler qu'un des principes fondamentaux de l'éthique scientifique est que les résultats et les méthodes utilisés doivent être absolument ouverts à la vérification et à la critique de tous ?

$$df_p(u) = \nabla f(p) \cdot u. \tag{11.510}$$

51. Voir http://www.sagemath.org
52. et les autres logiciels de type fenêtre, pomme ou feuille d'érable.

11.20.10 Différentielle et tangente

La notion de dérivée partielle (ou de dérivée suivant un vecteur) pour une fonction de plusieurs variables n'est pas une généralisation de la notion de dérivée en une variable d'espace. En fait, du point de vue géométrique, la dérivée de la fonction $g : \mathbb{R} \to \mathbb{R}$ au point a est la pente de la ligne droite tangente au graphe de g au point $(a, g(a))$. Cette ligne, d'équation $r(x) = g'(a)x + g(a)$, est la meilleure approximation affine du graphe de g au point a, comme à la figure 11.9.

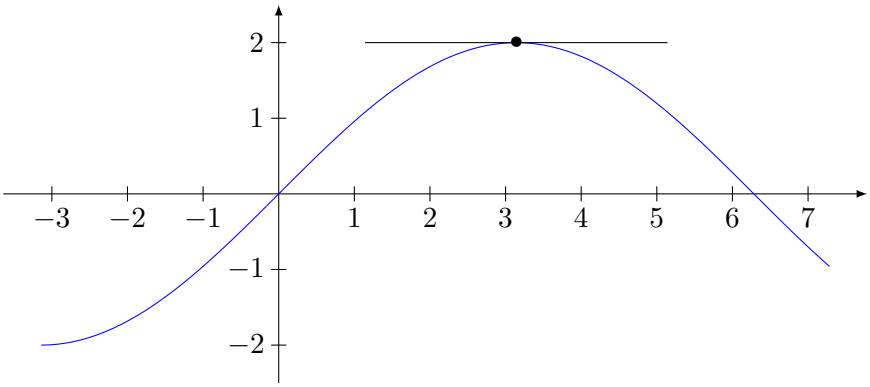


FIGURE 11.9 – Tangentes au graphe d'une fonction d'une variable

Le graphe d'une fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ est une surface de deux paramètres dans $\mathbb{R}^3$. Si l'approximation affine d'une telle surface au point $(x, y, f(x, y))$ existe, alors elle est un plan tangent. En dimension plus haute, le graphe de la fonction $f : \mathbb{R}^m \to \mathbb{R}$ est une surface de m paramètres dans $\mathbb{R}^{m+1}$ et son approximation affine (si elle existe) est un hyperplan de $\mathbb{R}^m$.

Nous allons voir que si f prend ses valeurs dans $\mathbb{R}^n$ l'approximation affine de f au point a est l'élément de $f(a) + \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ qui ressemble le plus à f au voisinage de a. Plus précisément, on utilise les définitions suivantes.

Définition 11.204.
Soient f et g deux applications d'un ouvert U de $\mathbb{R}^m$ dans $\mathbb{R}^n$. On dit que g est ***tangente*** *à f au point $a \in U$ si $f(a) = g(a)$ et*

$$\lim_{\substack{x \to a \\ x \neq a}} \frac{\|f(x) - g(x)\|_n}{\|x - a\|_m} = 0.$$

La relation de tangence est une relation d'équivalence. Nous sommes particulièrement intéressés par le cas où f admet une application affine tangente au point a.

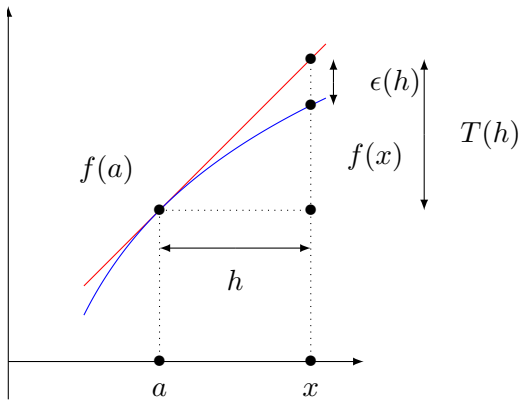


FIGURE 11.10 – Interprétation géométrique de la différentielle.

En ce qui concerne l'interprétation géométrique, si nous regardons la figure 11.10, et d'ailleurs aussi en voyant la définition 10.289, la fonction est différentiable et la différentielle est T s'il existe

une fonction α telle que

$$f(a+u) - f(a) - T(u) = \alpha(u) \tag{11.511}$$

où la fonction α satisfait

$$\lim_{u\to 0} \frac{\|\alpha(u)\|}{\|u\|} = 0 \tag{11.512}$$

C'est cela qui fait écrire $f(a+u) - f(a) - df_a(u) = o(\|u\|)$ à ceux qui n'ont pas peur de la notation o.

La différentielle df_a est donc la partie linéaire de l'application affine qui approxime au mieux la fonction f autour du point a. La notion de différentielle est la vraie généralisation du concept de dérivée pour fonctions de plusieurs variables, en outre elle nous permet d'expliciter la relation qui associe au vecteur u la dérivée $\partial_u f(a)$, pour f et a fixés.

Remarque 11.205.
Si on remplace les normes $\|\cdot\|_m$ et $\|\cdot\|_n$ par d'autres normes, l'existence et la valeur de la différentielle de f au point a ne sont pas remises en cause. En effet, soient $\|\cdot\|_M$ une norme sur $\mathbb{R}^m$ et $\|\cdot\|_N$ une norme sur $\mathbb{R}^n$. Par le théorème 7.224, ces normes sont équivalentes à $\|.\|_m$ et $\|.\|_n$ respectivement ; il existe donc des constantes k, K, l, $L > 0$ telles que pour tout vecteur u de $\mathbb{R}^m$ et tout vecteur v de $\mathbb{R}^n$

$$k\|u\|_M \leqslant \|u\|_m \leqslant K\|u\|_M,$$

$$l\|v\|_N \leqslant \|v\|_n \leqslant L\|v\|_N.$$

Les éléments de $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ sont les mêmes et on a

$$\begin{aligned} &\frac{l}{K}\frac{\|f(a+h) - f(a) - T(h)\|_N}{\|h\|_M} \leqslant \frac{\|f(a+h) - f(a) - T(h)\|_n}{\|h\|_m} \leqslant \\ &\leqslant \frac{L}{k}\frac{\|f(a+h) - f(a) - T(h)\|_N}{\|h\|_M}. \end{aligned} \tag{11.513}$$

Il est donc possible, pour démontrer la différentiabilité ou pour calculer la différentielle, d'utiliser le critère (10.289) avec une norme au choix. Parfois c'est utile.

11.20.11 Prouver qu'une fonction n'est pas différentiable

Chacun des points du théorème 11.194 est en soi un critère pour montrer qu'une fonction n'est pas différentiable en un point.

11.20.11.1 Continuité

Le premier critère à vérifier est donc la continuité. Si une fonction n'est pas continue en un point, alors elle n'y sera pas différentiable. Pour rappel, la continuité en a se teste en vérifiant si $\lim_{x\to a} f(x) = f(a)$.

11.20.11.2 Linéarité

Un second test est la linéarité de la dérivée directionnelle par rapport à la direction : l'application $u \mapsto \frac{\partial f}{\partial u}(a)$ doit être linéaire, sinon df_a n'existe pas.

Exemple 11.206.
Examinons la fonction

$$\begin{aligned} f\colon \mathbb{R}^2 &\to \mathbb{R} \\ (x,y) &\mapsto \begin{cases} \frac{xy^2}{x^2+y^4} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{sinon.} \end{cases} \end{aligned} \tag{11.514}$$

Prenons $u = (u_1, u_2)$ et calculons la dérivée de f dans la direction de u au point $(0,0)$:

$$\begin{aligned}\frac{\partial f}{\partial u}(0,0) &= \lim_{t\to 0}\frac{f(tu_1,tu_2)-f(0,0)}{t}\\ &= \lim_{t\to 0}\frac{1}{t}\left(\frac{tu_1t^2u_2}{t^2u_1^2+t^4u_2^4}\right)\\ &= \lim_{t\to 0}\left(\frac{u_1u_2^2}{u_1^2+t^2u_2^4}\right)\\ &= \begin{cases}\frac{u_2^2}{u_1} & \text{si } u_1\neq 0\\ 0 & \text{si } u_1 = 0.\end{cases}\end{aligned} \tag{11.515}$$

Cette application n'est pas linéaire par rapport à u. En effet, notons

$$\begin{aligned}A\colon \mathbb{R}^n &\to \mathbb{R}\\ u &\mapsto \frac{\partial f}{\partial u}(0,0),\end{aligned} \tag{11.516}$$

et vérifions que pour tout u et v dans $\mathbb{R}^n$ et $\lambda\in\mathbb{R}$, nous ayons $A(\lambda u) = \lambda A(u)$ et $A(u+v) = A(u)+A(v)$. Le premier fonctionne parce que

$$A(\lambda u) = A(\lambda u_1,\lambda u_2) = \frac{\lambda^2u_2^2}{\lambda u_1} = \lambda\frac{u_2^2}{u_1} = \lambda A(u). \tag{11.517}$$

Mais nous avons par exemple

$$A\big((0,1)+(2,3)\big) = A(2,4) = \frac{16}{2} = 8, \tag{11.518}$$

tandis que

$$A(0,1)+A(2,3) = 0+\frac{9}{2}\neq 8. \tag{11.519}$$

La fonction f n'est donc pas différentiable en $(0,0)$, parce que la candidate différentielle, $df_{(0,0)}(u) = \frac{\partial f}{\partial u}(0,0)$, n'est même pas linéaire.

△

Voici une autre façon de traiter la fonction de l'exemple 11.206.

Exemple 11.207.
La figure 11.11 représente le domaine d'une fonction $f\colon\mathbb{R}^2\to\mathbb{R}$, et sur chacune des parties, elle est définie différemment.

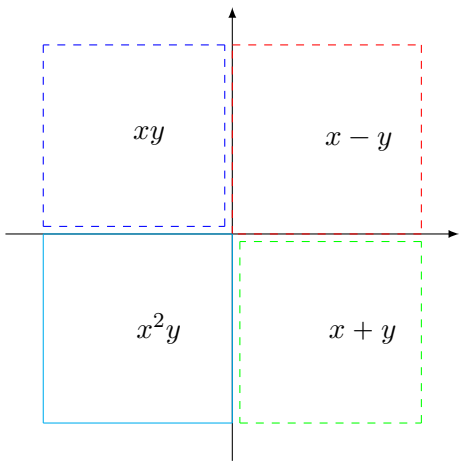


FIGURE 11.11 – La fonction de l'exemple 11.207.

L'expression de f est ici

$$f(x,y) = \begin{cases} xy & \text{si } x < 0 \text{ et } y > 0 \\ x - y & \text{si } x \geqslant 0 \text{ et } y \geqslant 0 \\ x^2 y & \text{si } x > 0 \text{ et } y < 0 \\ x + y & \text{sinon.} \end{cases} \tag{11.520}$$

On note que les deux axes forment une zone à problèmes. La zone hors des axes est un ouvert sur lequel f est différentiable car composée de polynômes. Analysons chacun des points de la forme (a,b) dans la zone à problèmes (c'est-à-dire si $ab = 0$).

Si $a = 0$ et $b > 0$ Un tel point $(0,b)$ est sur l'axe verticale, dans la moitié supérieure. Pour calculer la limite de f en ce point, on peut restreindre notre étude au demi-plan ouvert $y > 0$, ce qui revient à comparer la limite

$$\lim_{\substack{(x,y)\to(0,b)\\ y>0\\ x\geqslant 0}} f(x,y) = \lim_{\substack{(x,y)\to(0,b)\\ y>0\\ x\geqslant 0}} x - y = 0 - b = -b$$

avec la limite

$$\lim_{\substack{(x,y)\to(0,b)\\ y>0\\ x<0}} f(x,y) = \lim_{\substack{(x,y)\to(0,b)\\ y>0\\ x<0}} xy = 0b = 0$$

qui sont différentes puisque b est supposé non nul.

Conclusion : f n'est pas continue en un point du type $(0,b)$ avec $b > 0$.

Si $a = 0$ et $b < 0$ Un tel point $(0,b)$ est sur l'axe verticale, dans la moitié inférieure. Pour calculer la limite de f en ce point, on peut restreindre notre étude au demi-plan ouvert $y < 0$, ce qui revient à comparer la limite

$$\lim_{\substack{(x,y)\to(0,b)\\ y<0\\ x\geqslant 0}} f(x,y) = \lim_{\substack{(x,y)\to(0,b)\\ y<0\\ x\geqslant 0}} x^2 y = 0^2 b = 0$$

avec la limite

$$\lim_{\substack{(x,y)\to(0,b)\\ y<0\\ x<0}} f(x,y) = \lim_{\substack{(x,y)\to(0,b)\\ y<0\\ x<0}} x + y = 0 + b = b$$

qui sont différentes puisque b est supposé non nul.

Conclusion : f n'est pas continue en un point du type $(0,b)$ avec $b < 0$.

Si $a > 0$ et $b = 0$ Un tel point $(a,0)$ est sur l'axe horizontal, dans la moitié droite. Pour calculer la limite de f en ce point, on peut restreindre notre étude au demi-plan ouvert $x > 0$, ce qui revient à comparer la limite

$$\lim_{\substack{(x,y)\to(a,0)\\ x>0\\ y\geqslant 0}} f(x,y) = \lim_{\substack{(x,y)\to(a,0)\\ x>0\\ y\geqslant 0}} x - y = a - 0 = a$$

avec la limite

$$\lim_{\substack{(x,y)\to(a,0)\\ x>0\\ y<0}} f(x,y) = \lim_{\substack{(x,y)\to(a,0)\\ x>0\\ y<0}} x^2 y = a^2 0 = 0$$

qui sont différentes puisque a est supposé non nul.

Conclusion : f n'est pas continue en un point du type $(a,0)$ avec $a > 0$.

Si $a < 0$ et $b = 0$ Un tel point $(a, 0)$ est sur l'axe horizontal, dans la moitié gauche. Pour calculer la limite de f en ce point, on peut restreindre notre étude au demi-plan ouvert $x < 0$, ce qui revient à comparer la limite

$$\lim_{\substack{(x,y)\to(a,0)\\ x<0\\ y>0}} f(x,y) = \lim_{\substack{(x,y)\to(a,0)\\ x<0\\ y>0}} xy = a0 = 0$$

avec la limite

$$\lim_{\substack{(x,y)\to(a,0)\\ x<0\\ y\leqslant 0}} f(x,y) = \lim_{\substack{(x,y)\to(a,0)\\ x<0\\ y\leqslant 0}} x + y = a + 0 = a$$

qui sont différentes puisque a est supposé non nul.

Conclusion : f n'est pas continue en un point du type $(a, 0)$ avec $a < 0$.

Si $a = 0$ et $b = 0$ Le cas du point $(0, 0)$ est particulier, puisque il est adhérent aux quatre composantes du domaine où la fonction est définie différemment. Pour étudier la continuité, il faut donc étudier quatre limites. Ces limites ont déjà été étudiées ci-dessus et valent toutes 0, ce qui prouve la continuité de f en $(0, 0)$.

En ce qui concerne la différentiabilité, on sait qu'il est nécessaire que toutes les dérivées directionnelles existent. Calculons la dérivée dans la direction $(0, 1)$ (au point $(0, 0)$) :

$$\lim_{\substack{t\to 0\\ t\neq 0}} \frac{f((0,0)+t(0,1)) - f(0,0)}{t} = \lim_{\substack{t\to 0\\ t\neq 0}} \frac{f(0,t)}{t} = \dots$$

qu'on sépare en deux cas, car $f(0, t)$ possède une formule différente si $t < 0$ ou si $t \geqslant 0$:

$$\lim_{\substack{t\to 0\\ t\neq 0}} \frac{f(0,t)}{t} = \begin{cases} \lim_{\substack{t\to 0\\ t<0}} \frac{f(0,t)}{t} = \lim_{\substack{t\to 0\\ t<0}} \frac{0+t}{t} = 1 \\ \lim_{\substack{t\to 0\\ t\geqslant 0}} \frac{f(0,t)}{t} = \lim_{\substack{t\to 0\\ t\geqslant 0}} \frac{0-t}{t} = -1 \end{cases}$$

ce qui prouve que la limite n'existe pas, donc que la dérivée directionnelle n'existe pas, et finalement que la fonction n'est pas différentiable.

Conclusion : La fonction donnée est continue hors des axes et au point $(0, 0)$, mais discontinue partout ailleurs sur les axes. Elle est différentiable hors des axes, mais ne l'est pas sur les axes.

△

11.20.11.3 Cohérence des dérivées partielles et directionnelle

Dans la pratique, nous pouvons calculer $\partial_u f(a)$ pour une direction u générale, et puis en déduire $\partial_x f$ et $\partial_y f$ comme cas particuliers en posant $u = (1, 0)$ et $u = (0, 1)$. Une chose incroyable, mais pourtant possible est qu'il peut arriver que

$$\frac{\partial f}{\partial u}(a) \neq \sum_i \frac{\partial f}{\partial x_i}(a) u^i. \tag{11.521}$$

Ceci se produit lorsque f n'est pas différentiable en a. En voici un exemple.

11.20.11.4 Un candidat dans la définition (marche toujours)

Lorsqu'une fonction est donné, un candidat différentielle au point (a_1, a_2) est souvent assez simple à trouver en un point :

$$T(u_1, u_2) = \frac{\partial f}{\partial x}(a_1, a_2) u_1 + \frac{\partial f}{\partial y}(a_1, a_2) u_2. \tag{11.522}$$

L'application T est la candidate différentielle en ce sens que si la différentielle existe, alors elle est égale à T. Ensuite, il faut vérifier si

$$\lim_{(x,y)\to(a_1,a_2)} \frac{f(x,y) - f(a_1,a_2) - T\big((x,y)-(a_1,a_2)\big)}{\|(x,y)-(a_1,a_2)\|} = 0 \tag{11.523}$$

ou non. Si oui, alors la différentielle existe et $df_{(a,b)}(u) = T(u)$, sinon[53], la différentielle n'existe pas.

Attention : dans la ZAP, les dérivées partielles $\partial_x f$ et $\partial_y f$ ne peuvent en général pas être calculées en utilisant les règles de calcul (c'est bien pour ça que la ZAP est une zone à problèmes). Il faut d'office utiliser la définition

$$\frac{\partial f}{\partial x}(a_1,a_2) = \lim_{t\to 0} \frac{f(a_1+t,a_2) - f(a_1,a_2)}{t}, \tag{11.524}$$

et la définition correspondante pour $\partial_y f$.

Conclusion

Soient $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$, et $a \in int\, A$. Si f est différentiable en a,

$$(df_a(e_j))_i = d(f_i)_a(e_j) = \frac{\partial f_i}{\partial x_j}(a) = [Jac(f)_{|a}]_{ij}$$

et la matrice de l'application linéaire df_a est la matrice jacobienne $m \times n$ de f en a notée $Jac(f)_{|a}$.

Proposition 11.208 (Règles de calculs)**.**
Soient f et g des fonctions différentiables en $g(a)$ et a respectivement, alors la composée $f \circ g$ est différentiable en a et

$$d(f\circ g)_a = df_{g(a)} \circ dg_a$$

et de plus les jacobiennes correspondantes vérifient

$$J_{f\circ g}(a) = J_f\big(g(a)\big) J_g(a)$$

où le membre de droite est le produit (non-commutatif!) des deux matrices.

Corolaire 11.209 (Chain rule)**.**
Si $f : \mathbb{R}^p \to \mathbb{R}$ et $g : \mathbb{R} \to \mathbb{R}^p$, alors

$$(f\circ g)'(t) = \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(g(t)) g_i'(t).$$

Remarque 11.210.
Quelques remarques à propos de la règle de dérivation en chaine.

(1) Si $p = 1$, on retrouve la règle usuelle de dérivation de fonctions composées.

(2) Si g est à plusieurs variables, cette règle permet de déterminer les dérivées partielles de $f \circ g$, puisqu'une dérivée partielle peut être vue comme dérivée usuelle par rapport à une seule variable (voir 11.225).

(3) Si f est à valeurs vectorielles, cette formule permet de retrouver la jacobienne de $f\circ g$ puisqu'il suffit de traiter chaque composante de f séparément.

53. y compris si la limite (11.523) n'existe même pas.

11.20.12 Gradient

Définition 11.211.
Soit f une fonction différentiable de $\mathbb{R}^m$ dans $\mathbb{R}$. On appelle ***gradient*** *de f la fonction $\nabla f : \mathbb{R}^m \to \mathbb{R}^m$ de composantes*

$$\partial_1 f, \dots, \partial_m f.$$

Soit f une fonction de $\mathbb{R}^m$ dans $\mathbb{R}^n$, $f(a) = (f_1(a), \dots, f_n(a))^T$. On appelle ***matrice jacobienne*** *de f la fonction $J(f) : \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^n$ définie par*

$$a \mapsto \begin{pmatrix} \partial_1 f_1(a) & \dots & \partial_m f_1(a) \\ \vdots & \ddots & \vdots \\ \partial_1 f_n(a) & \dots & \partial_m f_n(a) \end{pmatrix} \tag{11.525}$$

11.20.13 Linéarité

La proposition suivante signifie que différentiation est une opération linéaire sur l'ensemble des fonctions différentiables.

Proposition 11.212.
Soient f et g deux fonctions de $U \subset \mathbb{R}^m$ dans $\mathbb{R}^n$ différentiables au point $a \in U$, et soit λ dans $\mathbb{R}$. Alors les fonctions $f + g$ et λf sont différentiables au point a et on a

$$\begin{aligned} d(f+g)(a) &= df(a) + dg(a), \\ d(\lambda f)(a) &= \lambda df(a), \end{aligned} \tag{11.526}$$

Démonstration.

$$\begin{aligned} &\lim_{h \to 0_m} \frac{\|(f(a+h) + g(a+h)) - (f(a) + g(a)) - df(a).h - dg(a).h\|_n}{\|h\|_m} \leqslant \\ &\lim_{h \to 0_m} \frac{\|f(a+h) - f(a) - df(a).h\|_n}{\|h\|_m} + \lim_{h \to 0_m} \frac{\|g(a+h) - g(a) - dg(a).h\|_n}{\|h\|_m} = 0. \end{aligned} \tag{11.527}$$

De même on démontre la propriété $d(\lambda f)(a) = \lambda df(a)$. □

11.21 Produit

Soient f et g deux fonctions de $\mathbb{R}^m$ dans $\mathbb{R}^n$. Nous notons $f \cdot g$ la fonction de $\mathbb{R}^n$ dans $\mathbb{R}$ donnée par le produit scalaire point par point, c'est-à-dire

$$(f \cdot g)(x) = f(x) \cdot g(x) \tag{11.528}$$

pour tout $x \in \mathbb{R}^m$. Le point dans le membre de droite est le produit scalaire dans $\mathbb{R}^n$. Le cas particulier $n = 1$ revient au produit usuel de fonctions :

$$(fg)(x) = f(x)g(x). \tag{11.529}$$

Lemme 11.213.
Si f et g sont des fonctions différentiables sur $\mathbb{R}^m$ à valeurs dans $\mathbb{R}$, alors la fonction produit fg est également différentiable et

$$(dfg)_a = g(a)fd_a + f(a)dg_a \tag{11.530}$$

au sens où pour chaque u dans $\mathbb{R}^m$,

$$(dfg)_a(u) = g(a)df_a(u) + f(a)dg_a(u). \tag{11.531}$$

Démonstration. Ce que nous devons faire pour vérifier la formule 11.530, c'est de vérifier le critère (10.289) en remplaçant f par fg et $T(h)$ par $g(a)df(a).h + f(a)dg(a).h$.

Ce que nous avons au numérateur est

$$\begin{aligned}\clubsuit &= (fg)(a+h) - (fg)(a) - g(a)df(a).h - f(a)dg(a).h \\ &= f(a+h)g(a+h) - f(a)g(a) - g(a)df(a).h - f(a)dg(a).h.\end{aligned} \tag{11.532}$$

Maintenant, nous allons faire apparaitre $\big(f(a+h)-f(a)-df(a)\big)g(a+h)$ en ajoutant et soustrayant ce qu'il faut pour conserver $\clubsuit$:

$$\begin{aligned}\clubsuit = &\big(f(a+h) - f(a) - df(a).h\big)g(a+h) \\ &+ f(a)g(a+h) + g(a+h)df(a).h \\ &- f(a)g(a) - g(a)df(a).h - f(a)dg(a).h.\end{aligned} \tag{11.533}$$

Nous mettons maintenant $f(a)$ et $fd(a).h$ en évidence là où c'est possible :

$$\begin{aligned}\clubsuit = &\big(f(a+h) - f(a) - df(a).h\big)g(a+h) \\ &+ f(a)\big(g(a+h) - g(a) - dg(a).h\big) \\ &+ \big(g(a+h) - g(a)\big)df(a).h.\end{aligned} \tag{11.534}$$

Nous devons maintenant considérer la limite

$$\lim_{h\to 0} \frac{\|\clubsuit\|}{\|h\|}. \tag{11.535}$$

Étant donné que f et g sont différentiables, les deux premiers termes sont nuls :

$$\begin{aligned}\lim_{h\to 0} \frac{\big(f(a+h) - f(a) - df(a).h\big)}{\|h\|} g(a+h) &= 0 \\ \lim_{h\to 0} f(a)\frac{\big(g(a+h) - g(a) - dg(a).h\big)}{\|h\|} &= 0.\end{aligned} \tag{11.536}$$

En ce qui concerne le troisième terme, en utilisant la norme d'une application linéaire, nous avons

$$\lim_{h\to 0} \frac{\|df(a).h\|}{\|h\|} \leqslant \sup_{h\in\mathbb{R}^m} \frac{\|df(a).h\|}{\|h\|} = \|df(a)\|, \tag{11.537}$$

et par conséquent

$$\begin{aligned}0 &\leqslant \lim_{h\to 0} \|g(a+h) - g(a)\| \frac{\|df(a).h\|\|h\|}{\|h\|} \\ &\leqslant \lim_{h\to 0} \|g(a+h) - g(a)\|\|df(a)\| = 0\end{aligned} \tag{11.538}$$

parce que g est continue (la limite du premier facteur est nulle tandis que la norme de $df(a)$ est un nombre constant). Nous avons donc bien prouvé que la formule (11.530) est la différentielle de fg au point a. □

Ce résultat se généralise pour des fonctions f et g de $\mathbb{R}^m$ dans $\mathbb{R}^n$ dans la proposition suivante qui généralise tout en même temps la proposition 11.119.

Proposition 11.214.
Soient f et g deux fonctions de $U \subset \mathbb{R}^m$ dans $\mathbb{R}^n$ différentiables au point $a \in U$. Alors la fonction $f \cdot g$ est différentiable au point a et on a

$$d(f \cdot g)(a) = g(a) \cdot df(a) + f(a) \cdot dg(a) \tag{11.539}$$

au sens où

$$d(f \cdot g)_a(u) = g(a) \cdot \big(df_a(u)\big) + f(a) \cdot \big(dg_a(u)\big) \tag{11.540}$$

pour tout $u \in \mathbb{R}^m$.

Démonstration. La preuve du cas $n = 1$ est déjà faite ; c'est la formule (11.530). Pour le cas général $n \geqslant 2$, nous passons au composantes en nous rappelant que

$$(f \cdot g)(a) = \sum_{i=1}^{n} f_i(a)g_i(a) = \sum_{i=1}^{n}(f_i g_i)(a). \tag{11.541}$$

En utilisant la linéarité de la différentiation, nous nous réduisons donc au cas des produits $f_i g_i$ qui sont des fonctions de $\mathbb{R}^m$ dans $\mathbb{R}$:

$$\begin{aligned} d(f \cdot g)(a) &= d\left(\sum_{i=1}^{n} f_i g_i\right)(a) \\ &= \sum_{i=1}^{n} \big(df_i(a)g_i(a) + f_i(a)dg_i(a)\big) \\ &= g(a) \cdot df(a) + f(a) \cdot dg(a). \end{aligned} \tag{11.542}$$

Ceci termine la preuve. □

11.21.1 Difficulté d'ordre supérieur

11.215.
Il serait tentant de faire une récurrence sur le lemme 11.213 pour dire que si f et g sont de classe C^p, alors le produit fg est également de classe C^p, parce que la formule de $d(fg)$ contient des produits de fonctions de classe C^p et C^{p-1}.

Le problème est que le lemme 11.213 est énoncé et prouvé pour des fonctions à valeurs dans $\mathbb{R}$, alors que déjà la formule

$$d(fg) = gdf + fdg \tag{11.543}$$

contient le produit de $g\colon E \to \mathbb{R}$ par $df\colon E \to \mathcal{L}(E, \mathbb{R})$. Lorsque nous montons dans les différentielles, la situation empire, et les produits dont sont composés les formules sont réellement à définir...

Oublions un instant les questions de régularité, et calculons sans ménagement, pour voir ce qu'il se passe. Nous considérons un espace vectoriel E ainsi que des des fonctions $f\colon E \to \mathbb{R}$ et $g\colon E \to V$ où V est un autre espace vectoriel.

Nous avons

$$d(fg)_a(u) = df_a(u)g(a) + f(a)dg_a(u). \tag{11.544}$$

Les deux termes sont des produits $\mathbb{R} \times V \to \mathbb{R}$. Montons un coup :

$$d(gdf)_a(u) = \frac{d}{dt}\Big[(gdf)(a+tu)\Big]_{t=0} = \frac{d}{dt}\Big[g(a+tu)df_{a+tu}\Big]_{t=0} = dg_a(u)df_a + g(a)(d^2 f)_a(u). \tag{11.545}$$

Un autre pour voir comment ça se passe plus haut :

$$d(dfdg)_a(u) = \frac{d}{dt}\Big[(dfdg)(a+tu)\Big]_{t=0} = \frac{d}{dt}\Big[df_{a+tu}dg_{a+tu}\Big]_{t=0} = (d^2 f)_a(y)dg_a + df_a(d^2 g)_a(u). \tag{11.546}$$

Là déjà vous noterez que nous sommes passés par le produit

$$df_{a+tu}df_{a+tu} \tag{11.547}$$

qui pour chaque t est un produit $\mathcal{L}(E, \mathbb{R}) \times \mathcal{L}(E, V)$ que nous n'avons pas réellement défini.

En continuant le calcul ainsi nous trouvons par exemple

$$\begin{aligned} (d^3 fg)_a(u) = {} & d^3 f_a(u)g(a) + d^2 f_a dg_a(u) \\ & + d^2 f_a(u)dg_a + df_a d^2 g_a(u) \\ & + d^2 f_a(u)dg_a + df_a(d^2 g)_a(u) \\ & + df_a(u)d^2 g_a + f(a)(d^3 g)_a(u). \end{aligned} \tag{11.548}$$

Vous noterez que cette formule contient trois termes que nous aurions eu envie de noter $d^2 fdg$. Or ces trois termes ne sont pas identiques : deux sont $d^2 f_a(u)dg_a$ et un est $(d^2 f)_a dg_a(u)$.

11.21.2 Solution : produit tensoriel

Afin de donner un sens à tous les produits, nous allons passer par les produits tensoriel. Nous avons déjà le théorème 10.143 qui fait pratiquement tout.

Proposition 11.216 ([1]).
Soient des fonctions $f\colon \mathbb{R}^n \to \mathbb{R}$ et $g\colon \mathbb{R}^n \to \mathbb{R}$ de classe C^p. Alors fg est de classe C^p.

Démonstration. Nous considérons l'application

$$\begin{aligned} \varphi\colon \mathbb{R}\otimes\mathbb{R} &\to \mathbb{R} \\ 1\otimes 1 &\mapsto 1 \end{aligned} \tag{11.549}$$

dont nous avons déjà parlé dans le lemme 10.125. En utilisant la notation $\tilde{\otimes}$ de la définition 10.139, nous avons

$$fg = \varphi\circ(f\tilde{\otimes}g). \tag{11.550}$$

La proposition 10.143 nous dit que $f\tilde{\otimes}g\colon \mathbb{R}^n \to \mathbb{R}\otimes\mathbb{R}$ est de classe C^p. Vu que φ est un isomorphisme d'espaces vectoriels, la proposition 10.136 nous dit que $\varphi\circ(f\tilde{\otimes}g)$ est encore de classe C^p.

Et voila. □

11.21.3 Formes bilinéaires

Nous avons aussi une formule importante pour la différentielle des formes bilinéaires.

Lemme 11.217.
Toute application bilinéaire

$$\begin{aligned} B\colon \mathbb{R}^m\times\mathbb{R}^n &\to \mathbb{R}^p \\ B(a_1,a_2) &= a_1\star a_2 \end{aligned} \tag{11.551}$$

est différentiable en tout point (a_1,a_2) de $\mathbb{R}^m\times\mathbb{R}^n$, et on a

$$dB(a_1,a_2).(h_1,h_2) = h_1\star a_2 + a_1\star h_2.$$

Démonstration.

$$\begin{aligned} &\frac{\|B(a_1+h_1,a_2+h_2)-B(a_1,a_2)-(h_1\star a_2+a_1\star h_2)\|_p}{\|(h_1,h_2)\|_{\mathbb{R}^m\times\mathbb{R}^n}} = \\ &= \frac{\|(a_1+h_1)\star(a_2+h_2)-a_1\star a_2-(h_1\star a_2+a_1\star h_2)\|_p}{\|(h_1,h_2)\|_{\mathbb{R}^m\times\mathbb{R}^n}} = \spadesuit \end{aligned} \tag{11.552}$$

on rajoute et on enlève la quantité $(a_1+h_1)\star a_2$ dans le numérateur, et on obtient

$$\begin{aligned} \spadesuit &= \frac{\|(a_1+h_1)\star h_2+h_1\star a_2-(h_1\star a_2+a_1\star h_2)\|_p}{\|(h_1,h_2)\|_{\mathbb{R}^m\times\mathbb{R}^n}} = \\ &= \frac{\|h_1\star h_2\|_p}{\|(h_1,h_2)\|_{\mathbb{R}^m\times\mathbb{R}^n}} \leqslant C\frac{\|h_1\|_m\|h_2\|_n}{\|(h_1,h_2)\|_{\mathbb{R}^m\times\mathbb{R}^n}} \leqslant \\ &\leqslant C\frac{\|(h_1,h_2)\|^2_{\mathbb{R}^m\times\mathbb{R}^n}}{\|(h_1,h_2)\|_{\mathbb{R}^m\times\mathbb{R}^n}} = C\|(h_1,h_2)\|_{\mathbb{R}^m\times\mathbb{R}^n}. \end{aligned} \tag{11.553}$$

Si on prend la limite de cette expression pour $(h_1,h_2)\to(0_m,0_n)$ on obtient 0, donc la preuve est complète. À noter, que dans l'avant-dernier passage on a utilisé la continuité des applications linéaires $\mathtt{proj}_m : \mathbb{R}^m\times\mathbb{R}^n \to \mathbb{R}^m$ et $\mathtt{proj}_n : \mathbb{R}^m\times\mathbb{R}^n \to \mathbb{R}^n$ qui à chaque point (a_1,a_2) de $\mathbb{R}^m\times\mathbb{R}^n$ associent a_1 et a_2 respectivement. □

Proposition 11.218.
Soit V et W deux espaces vectoriels et $\varphi\colon V\to W$ un isomorphisme. Soit $f\colon \mathbb{C}\to V$ une application telle que $\varphi\circ f\colon \mathbb{C}\to W$ soit différentiable.

Alors f est différentiable et $df = \varphi^{-1}\circ d(\varphi\circ f)$.

Démonstration. Si T est la différentielle de $\varphi \circ f$ au point z nous avons

$$\lim_{\substack{h\to 0\\ h\in\mathbb{C}}} \frac{(\varphi\circ f)(z+h)-(\varphi\circ f)(z)+T(h)}{h}=0. \tag{11.554}$$

En appliquant φ aux deux membres, et en permutant avec la limite (parce que φ est continue),

$$\varphi \lim_{h\to 0} \frac{f(z+h)-f(z)+\varphi^{-1}T(h)}{h}=0, \tag{11.555}$$

ce qui signifie que f est différentiable et que $df=\varphi^{-1}\circ T=\varphi^{-1}\circ d(\varphi\circ f)$. □

11.22 Différentielle de fonction composée

Une importante règles de différentiation est la règle de différentiation d'une fonction composée (*chain rule* dans les livres anglais et américains). Cette règle généralise la règle de dérivation pour fonctions de $\mathbb{R}$ dans $\mathbb{R}$.

Cette règle a déjà été donnée dans le théorème 10.134, mais si vous avez seulement envie d'entendre parler de $\mathbb{R}^n$, vous pouvez lire le lemme 11.219 suivit de la proposition 11.221.

Le lemme suivant est essentiellement une reformulation du lemme 10.130.

Lemme 11.219.
Soit U un ouvert de $\mathbb{R}^m$. La fonction $f:U\to\mathbb{R}^n$ est différentiable au point a dans U, si et seulement s'il existe une fonction $\sigma_f:U\times U\to\mathbb{R}^n$ telle que

$$\sigma_f(a,a)=\lim_{x\to a}\sigma_f(a,x)=0 \tag{11.556a}$$

$$f(x)=f(a)+T(x-a)+\sigma_f(a,x)\|x-a\|_m, \tag{11.556b}$$

pour une certaine application linéaire $T\in\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)$.

Démonstration. Si les conditions (11.556) sont satisfaites alors T est la différentielle de f en a. En effet, dans ce cas nous avons

$$f(a+h)=f(a)+T(h)+\sigma_f(a,a+h)\|h\|, \tag{11.557}$$

et la condition (10.289) devient

$$\lim_{h\to 0}\frac{\|\sigma_f(a,a+h)\|\|h\|}{\|h\|}=\lim_{h\to 0}\|\sigma_f(a,a+h)\|=0 \tag{11.558}$$

Si f est différentiable au point a il suffit de prendre $T=df(a)$ et

$$\sigma_f(a,x)=\frac{f(x)-f(a)-df(a).(x-a)}{\|x-a\|_m}.$$

□

Remarque 11.220.
La fonction $\sigma_f(a,x)\|x-a\|_m$ est ce qui avait été appelle $\epsilon(h)$ sur la figure 11.10.

Proposition 11.221.
Soient U un ouvert de $\mathbb{R}^m$ et V un ouvert de $\mathbb{R}^n$. Soient $f:U\to V$ et $g:V\to\mathbb{R}^p$ deux fonctions différentiables respectivement au point a dans U et $b=f(a)$ dans V. Alors la fonction composée $g\circ f:U\to\mathbb{R}^p$ est différentiable au point a et

$$d(g\circ f)_a=dg_{f(a)}\circ df_a. \tag{11.559}$$

Démonstration. En tenant compte du lemme 11.219 on peut écrire

$$f(a+h)-f(a)=df_a(h)+\sigma_f(a,a+h)\|h\|_m, \qquad \forall h\in U-a, \tag{11.560a}$$
$$g(b+k)-g(b)=dg_b(k)+\sigma_g(b,b+k)\|k\|_n, \qquad \forall k\in V-b. \tag{11.560b}$$

On sait que $f(a)=b$ et que $f(a+h)$ est un élément de V et $f(a+h)=f(a)+k$ pour $k=df(a).h+\sigma_f(a,a+h)\|h\|_m$. Par substitution dans la deuxième équation on obtient

$$\begin{aligned}
&g\big(f(a+h)\big)-g\big(f(a)\big)\\
&\qquad=dg_{f(a)}\Big(df_a(h)+\sigma_f(a,a+h)\|h\|_m\Big)\\
&\qquad\quad+\sigma_g\left(f(a),f(a+h)\right)\|df_a(h)+\sigma_f(a,a+h)\|h\|_m\|_n\\
&\qquad=g\circ f(a+h)-g\circ f(a)\\
&\qquad=dg_{f(a)}\circ df_a(h)\\
&\qquad\quad+\|h\|_m\Big[dg_{f(a)}\sigma_f(a,a+h)\\
&\qquad\quad+\sigma_g\left(f(a),f(a+h)\right)\Big\|df_a\frac{h}{\|h\|_m}+\sigma_f(a,a+h)\Big\|_n\Big],
\end{aligned} \tag{11.561}$$

donc

$$(g\circ f)(a+h)-(g\circ f)(a)=dg_f(a)\circ df_a(h)+S(a,a+h)\|h\|_m \tag{11.562}$$

où S représente le contenu du dernier grand crochet. Il ne reste plus qu'à prouver que $S(a,a+h)$ est $o(\|h\|_m)$. En tenant compte du fait que $\sigma_f(a,a+h)$ et $\sigma_g\left(f(a),f(a+h)\right)$ sont $o(\|h\|_m)$,

$$\begin{aligned}
\lim_{h\to 0_m}\frac{S(a,a+h)}{\|h\|_m}&=\lim_{h\to 0_m}\frac{dg_{f(a)}\sigma_f(a,a+h)}{\|h\|_m}+\\
+\lim_{h\to 0_m}&\frac{\sigma_g\left(f(a),f(a+h)\right)\Big\|df_a\frac{h}{\|h\|_m}+\sigma_f(a,a+h)\Big\|_n}{\|h\|_m}=0.
\end{aligned} \tag{11.563}$$

□

Remarque 11.222.
Note : la formule (11.559) est à comprendre de la façon suivante. Si $u\in\mathbb{R}^m$, alors

$$d(g\circ f)_a(u)=\underbrace{dg_{f(a)}}_{\in\mathcal{L}(\mathbb{R}^n,\mathbb{R}^p)}\Big(\underbrace{df_a(u)}_{\in\mathbb{R}^n}\Big)\in\mathbb{R}^p. \tag{11.564}$$

Le lemme suivant sert à prouver les théorèmes 14.6 et 19.72. Il est fondamentalement la raison de la formule définissant l'intégrale d'une forme sur un chemin (définition 19.47).

Lemme 11.223 ([1]).
Soient un espace vectoriel normé F de dimension finie, ainsi que E, un espace vectoriel normé. Nous considérons un chemin de classe C^1

$$\gamma\colon[a,b]\to E \tag{11.565}$$

et une application de classe C^1

$$f\colon E\to F. \tag{11.566}$$

Si $g=f\circ\gamma$, alors

$$g'(t)=(df)_{\gamma(t)}\big(\gamma'(t)\big) \tag{11.567}$$

pour tout $t\in[a,b]$.

Démonstration. Nous écrivons la dérivée de g de la façon suivante :

$$\begin{aligned} g'(t) &= \frac{d}{ds}\Big[g(t+s)\Big]_{s=0} && (11.568a)\\ &= \frac{d}{ds}\Big[(f\circ\gamma)(t+s)\Big]_{s=0} && (11.568b)\\ &= d(f\circ\gamma)_t(1) && (11.568c)\\ &= (df)_{\gamma(t)}\big(d\gamma_t(1)\big) && (11.568d)\\ & && (11.568e) \end{aligned}$$

Justifications :

— Pour (11.568c) : les formules (11.196). Notez que le 1 à qui s'applique la différentielle de $f\circ\gamma$ est le vecteur de $\mathbb{R}$ qui est multiplié par s dans l'expression $(f\circ\gamma)(t+s)$.

— Pour (11.568d) : la différentiation de fonction composées de la proposition 11.221.

Mais

$$d\gamma_t(1) = \frac{d}{ds}\Big[\gamma(t+s)\Big]_{s=0} = \gamma'(t). \tag{11.569}$$

En remettant au bout de (11.568), nous obtenons le résultat. □

11.23 Autres trucs sur la différentielle

11.23.1 Différentielle et dérivées partielles

Étant donné que pour tout vecteur u dans $\mathbb{R}^m$ on a $\partial_u f(a) = \nabla f(a)\cdot u$, le gradient de f nous donne la direction dans laquelle la croissance de f est maximale. Soit C une colline et soit f la fonction que a chaque point (x,y) de la Terre associe son altitude. Si nous voulons monter la colline le plus vite possible nous n'avons qu'a suivre la direction ∇f à chaque point. Elle est la projection sur le plan x-y de la direction de pente maximale. Au contraire, la direction $-\nabla f$ est la direction de croissance minimale.

La matrice jacobienne calculé au point a est la matrice associée canoniquement à l'application linéaire $df_a : \mathbb{R}^m \to \mathbb{R}^n$.

11.23.2 Plan tangent

On a dit au début de cette section que si f est une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ alors le graphe de f est une surface à deux paramètres et que l'application affine tangente au graphe de f au point $(a, f(a))$ est un plan. Maintenant on sait que ce plan est celui d'équation

$$T_a(x,y) = f(a_1,a_2) + \frac{\partial f}{\partial x}(a_1,a_2)(x-a_1) + \frac{\partial f}{\partial y}(a_1,a_2)(y-a_2). \tag{11.570}$$

Le plan tangent au graphe de f au point a est le graphe de cette fonction T_a.

Remarque 11.224.
Il existe cependant des fonctions différentiables dont les dérivées partielles ne sont pas continues. La construction d'un tel exemple est cependant délicate, et nous le ferons pas ici. Retenez cependant que si dans un exercice vous obtenez que les dérivées partielles ne sont pas continues, vous ne pouvez pas immédiatement en conclure que la fonction ne sera pas différentiable.

11.23.3 Calcul de différentielles

11.225.
En pratique, ayant une formule pour la fonction f, nous la dérivons par rapport à la variable x_i en utilisant les règles usuelle de dérivation en considérant que les autres (x_j avec $j \neq i$) sont des constantes.

Exemple 11.226.
Pour $f(x,y) = xy + x^2$, les dérivées partielles s'écrivent

$$\frac{\partial f}{\partial x} = y + 2x \quad \text{et} \quad \frac{\partial f}{\partial y} = x$$

△

Des *règles de calcul* sont d'application. En particulier, quand ces opérations existent, les sommes, différences, produits, quotients et compositions d'applications différentiables sont différentiables.

Toute application linéaire est différentiable, et sa différentielle en tout point est égale à l'application elle-même[54]. En particulier, les *projections canoniques*, c'est-à-dire les applications du type $(x,y,z) \mapsto y$, sont linéaires donc différentiables.

Exemple 11.227.
Les cas suivants sont faciles :

(1) En combinant les projections canoniques avec les règles de calculs, on obtient que toute fonction polynomiale à n variables est différentiable comme application de $\mathbb{R}^n$ dans $\mathbb{R}$.

(2) Toute fonction rationnelle, du type $f(x) \stackrel{\text{def}}{=} \frac{P(x)}{Q(x)}$ où P et Q sont des polynômes, est différentiable en tout point a tel que $Q(a) \neq 0$.

(3) Pour une fonction d'une variable $f : D \subset \mathbb{R} \to \mathbb{R}$, le caractère différentiable et le caractère dérivable coïncident. De plus, on a

$$df_a(u) = f'(a)u.$$

△

11.23.4 Notes idéologiques quant au concept de plan tangent

Notons G, le graphe d'une fonction f, c'est-à-dire

$$G = \{(x,y,z) \in \mathbb{R}^3 \text{ tel que } z = f(x,y)\}. \tag{11.571}$$

Première affirmation : si $\gamma\colon \mathbb{R} \to G$ est une courbe telle que $\gamma(0) = \big(a, f(a)\big)$, alors $\gamma'(0) \in \mathbb{R}^n$ est dans le plan tangent à G au point $\big(a, f(a)\big)$.

Plus fort : tous les éléments du plan tangent sont de cette forme.

Le plan tangent à G en un point $x \in G$ est donc constitué des vecteurs vitesse de tous les chemins qui passent par x.

Prenons maintenant S, une courbe de niveau de G, c'est-à-dire

$$S = \{(x,y) \in \mathbb{R}^2 \text{ tel que } f(x,y) = C\}. \tag{11.572}$$

Si nous prenons un chemin dans G qui est, de plus, contraint à S, c'est-à-dire tel que $\gamma(t) \in S$, alors $\gamma'(0)$ sera tangent à G (ça, on le savait déjà), mais en plus, $\gamma'(0)$ sera tangent à S, ce qui est logique.

La morale est que si vous prenez un chemin qui se ballade dans n'importe quoi, alors la dérivée du chemin sera un vecteur tangent à ce n'importe quoi.

En outre, si $\gamma(t) \in S$ et $\gamma(0) = a$, alors

$$\langle \nabla f(a), \gamma'(0) \rangle = 0, \tag{11.573}$$

c'est-à-dire que le vecteur tangent à la courbe de niveau est perpendiculaire au gradient. Cela est intuitivement logique parce que la tangente à la courbe de niveau correspond à la direction de *moins* grande pente.

54. Lemme 11.189.

11.23.5 Gradient et recherche du plan tangent

Nous avons maintenant en main les concepts utiles pour trouver l'équation du plan tangent à une surface.

De la même manière que la tangente à une courbe était la droite de coefficient directeur donné par la dérivée, maintenant, le plan tangent à une surface est le plan dont les vecteurs directeurs sont les dérivées partielles :

La généralisation de l'équation (11.441) est

$$T_a(x) = f(a) + \sum_i \frac{\partial f}{\partial x_i}(a)(x-a)^i \tag{11.574}$$

Nous introduisons aussi souvent l'opérateur différentiel abstrait **nabla**, noté ∇ et qui est donné par le vecteur

$$\nabla = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right). \tag{11.575}$$

Les égalités suivantes sont juste des notations, sommes toutes logiques, liées à ∇ :

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right), \tag{11.576}$$

et

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \frac{\partial f}{\partial x_2}(a), \dots, \frac{\partial f}{\partial x_n}(a)\right). \tag{11.577}$$

Ce dernier est un élément de $\mathbb{R}^n$: chaque entrée est un nombre réel.

Définition 11.228.
Le vecteur gradient de f au point a est le vecteur donné par la formule (11.577).

La notation ∇ permet d'écrire la différentielle sous forme un peu plus compacte. En effet, la formule (11.484) peut être notée

$$df_a(u) = \langle \nabla f(a), u \rangle. \tag{11.578}$$

En utilisant ce produit scalaire, l'équation (11.574) peut se récrire

$$T_a(x) = f(a) + \sum_i \frac{\partial f}{\partial x_i}(a)(x-a)^i = f(a) + \langle \nabla f(a), x-a \rangle. \tag{11.579}$$

Afin d'éviter les confusions, il est parfois souhaitable de bien mettre les parenthèses et noter $(\nabla f)(a)$ au lieu de $\nabla f(a)$.

Proposition 11.229.
$\nabla f(a) \perp S_a$

$$z = f(a) + \sum_i \frac{\partial f}{\partial f}(a)(x-a)^i. \tag{11.580}$$

Cas particulier où $n = 2$:

Le plan T_a avec $a = (a_1, a_2)$ a pour équation dans $\mathbb{R}^3$:

$$z = f(a_1, a_2) + \frac{\partial f}{\partial x}(a_1, a_2)\,(x - a_1) + \frac{\partial f}{\partial y}(a_1, a_2)\,(y - a_2). \tag{11.581}$$

Définition 11.230.
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable en un point a. Le plan tangent *au graphe de f en $(a, f(a))$ est l'ensemble des points*

$$\begin{aligned} T_a f &= \{(x,z) \in \mathbb{R}^n \times \mathbb{R} \text{ tel que } z = f(a) + df_a(x-a)\} \\ &= \{(x,z) \in \mathbb{R}^n \times \mathbb{R} \text{ tel que } z = f(a) + \langle \nabla f(a), x-a \rangle\} \end{aligned}$$

Nous avons vu que, de la même façon qu'en deux dimensions nous avions l'approximation (11.236) d'une fonction par sa tangente, en trois dimensions nous avons l'approximation suivante d'une fonction de deux variables :

$$f(x,y) \simeq f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b) \tag{11.582}$$

lorsque (x,y) n'est pas trop loin de (a,b). Cela signifie que le graphe de f ressemble au graphe de la fonction $T_{(a,b)}$ donnée par

$$T_{(a,b)}(x,y) = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(x-a). \tag{11.583}$$

En notations compactes :

$$T_p(x) = f(p) + \nabla f(p) \cdot (x-p). \tag{11.584}$$

Le graphe de la fonction T_p sera le **plan tangent** au graphe de f au point p. L'équation du plan tangent sera donc

$$z - f(p) = \nabla f(p) \cdot (x-p). \tag{11.585}$$

Remarque 11.231.
Lorsque nous utilisons la notation vectorielle, la lettre « x » désigne le vecteur (x,y). Il faut être attentif. Dans un cas x est un vecteur dans l'autre c'est une composante d'un vecteur.

11.23.6 Projection orthogonale

Le théorème suivant n'est pas indispensablissime parce qu'il est le même que le théorème de la projection sur les espaces de Hilbert [55]. Cependant la partie existence est plus simple en se limitant au cas de dimension finie.

Théorème 11.232 (Théorème de la projection)**.**
Soit E un espace vectoriel réel ou complexe de dimension finie, $x \in E$, et C un sous-ensemble fermé convexe de E.

(1) Les deux conditions suivantes sur $y \in E$ sont équivalentes :
  (1a) $\|x-y\| = \inf\{\|x-z\|$ tel que $z \in C\}$,
  (1b) pour tout $z \in C$, $\mathrm{Re}\langle x-y, z-y\rangle \leqslant 0$.

(2) Il existe un unique $y \in E$, noté $y = \mathtt{proj}_C(x)$ vérifiant ces conditions.

Démonstration. Nous commençons par prouver l'existence et l'unicité d'un élément dans C vérifiant la première condition. Ensuite nous verrons l'équivalence.

Existence Soit $z_0 \in C$ et $r = \|x - z_0\|$. La boule fermée $\overline{B(x,r)}$ est compacte [56] et intersecte C. Vu que C est fermé, l'ensemble $C' = C \cap \overline{B(x,r)}$ est compacte. Tous les points qui minimisent la distance entre x et C sont dans C' ; la fonction

$$\begin{aligned} C' &\to \mathbb{R} \\ z &\mapsto d(x,z) \end{aligned} \tag{11.586}$$

est continue sur un compact et donc a un minimum qu'elle atteint [57]. Un point P réalisant ce minimum prouve l'existence d'un point vérifiant la première condition.

Unicité Soient y_1 et y_2, deux éléments de C minimisant la distance avec x, et soit d ce minimum. Nous avons par l'identité du parallélogramme (7.171) que

$$\|y_1 - y_2\|^2 = -4\left\|\frac{y_1+y_2-x}{2}\right\|^2 + 2\|y_1-x\|^2 + 2\|y_2-x\|^2 \leqslant -4d + 2d + 2d = 0. \tag{11.587}$$

Par conséquent $y_1 = y_2$.

55. Théorème 24.5
56. C'est ceci qui ne marche plus en dimension infinie.
57. Théorème 7.154.

(1a)⇒ (1b) Soit $z \in C$ et $t \in]0,1[$; nous notons $P = \mathtt{proj}_C\, x$. Par convexité le point $z = ty + (1-t)P$ est dans C, et par conséquent,

$$\|x - P\|^2 \leqslant \|x - tz - (1-t)P\|^2 = \|(x-P) - t(z-P)\|^2. \tag{11.588}$$

Nous sommes dans un cas $\|a\|^2 \leqslant |a-b|^2$, qui implique $2\operatorname{Re}\langle a, b\rangle \leqslant \|b\|^2$. Dans notre cas,

$$2\operatorname{Re}\langle x - P, t(z-P)\rangle \leqslant t^2\|z-P\|^2. \tag{11.589}$$

En divisant par t et en faisant $t \to 0$ nous trouvons l'inégalité demandée :

$$2\operatorname{Re}\langle x - P, z - P\rangle \leqslant 0. \tag{11.590}$$

(1b)⇒ (1a) Soit un point $P \in C$ vérifiant

$$\operatorname{Re}\langle x - P, z - P\rangle \leqslant 0 \tag{11.591}$$

pour tout $z \in C$. Alors en notant $a = x - P$ et $b = P - z$,

$$\begin{aligned}\|x-z\|^2 = \|x - P + P - z\|^2 &= \|a+b\|^2\\ &= \|a\|^2 + \|b\|^2 + 2\operatorname{Re}\langle a, b\rangle\\ &= \|a\|^2 + \|b\|^2 - 2\operatorname{Re}\langle x-P, z-P\rangle\\ &\geqslant \|b\|^2,\end{aligned} \tag{11.592}$$

ce qu'il fallait.

□

11.24 Jacobienne

11.24.1 Rappels et définitions

Dans cette section nous considérons des fonctions $f : D \to \mathbb{R}^m$ où $D \subset \mathbb{R}^n$, et un point $a \in \operatorname{Int} D$ où f est différentiable.

Remarque 11.233.
La définition de continuité (resp. différentiabilité) pour une fonction à valeurs vectorielles est celle introduite précédemment, et on remarque que pour avoir la continuité (resp. différentiabilité) de f en un point, il faut et il suffit de chacune des composantes de $f = (f_1, \dots, f_m)$, vues séparément comme fonctions à n variables et à valeurs réelles, soit continue (resp. différentiable) en ce point.

Définition 11.234.
*La **jacobienne** de f en a est la matrice de l'application linéaire donnée par la différentielle. Elle a de nombreuses notations*

$$J_f(a) = \frac{\partial(f_1, \dots, f_m)}{\partial x_1, \dots, x_m} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a)\\ \vdots & & \vdots\\ \frac{\partial f_m}{\partial x_1}(a) & \dots & \frac{\partial f_m}{\partial x_n}(a)\end{pmatrix} \tag{11.593}$$

*Autrement dit, c'est la matrice composée de l'ensemble des dérivées partielles de f. Le **jacobien** de f au point a est le déterminant de cette matrice.*

*Si $m = 1$, cette matrice ne contient qu'une ligne ; c'est donc un vecteur appelé le **gradient** de f au point a et noté $\nabla f(a)$.*

Remarque 11.235. (1) Si la fonction est supposée différentiable, calculer la jacobienne revient à connaitre la différentielle. En effet, par linéarité de la différentielle et par définition des dérivées partielles, nous avons

$$df_a(u) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \dots & \frac{\partial f_1}{\partial x_n}(a) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \dots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

où $u = (u_1, \dots, u_n)$ et où le membre de droite est un produit matriciel

(2) Remarquons que la jacobienne peut exister en un point donné sans que la fonction soit différentiable en ce point !

11.236.
Le théorème de différentiation de fonctions composées 11.221 peut également se lire au niveau des matrices jacobiennes. La matrice jacobienne de $g \circ f$ au point a est le produit matriciel des matrices jacobiennes de f et de f. Plus précisément, nous avons

$$J_{g\circ f}(a) = J_g\big(f(a)\big) J_f(a). \tag{11.594}$$

Remarquez que nous considérons la matrice jacobienne de g au point $f(a)$.

Dans la cas particulier où $m = 1$ et f est une fonction d'un intervalle I dans $\mathbb{R}^n$, dérivable au point a, on a que la fonction composée $g \circ f$ est dérivable au point a si g est différentiable et alors

$$(g \circ f)'(a) = dg\left(f(a)\right).f'(a).$$

En fait, pour les fonctions d'une seule variable la dérivabilité coïncide avec la différentiabilité.

11.25 Fonctions de classe C^1

Soit f une fonction différentiable de U, ouvert de $\mathbb{R}^m$, dans $\mathbb{R}^n$. L'application différentielle de f est une application de $\mathbb{R}^m$ dans $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$

$$\begin{aligned} df \colon \mathbb{R}^m &\to \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n) \\ a &\mapsto df_a \end{aligned} \tag{11.595}$$

Nous savons que $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ est un espace vectoriel normé avec la définition 4.25. Si T est un élément dans $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$ alors la norme de T est définie par

$$\|T\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)} = \sup_{x\in\mathbb{R}^m} \frac{\|T(x)\|_n}{\|x\|_m} = \sup_{\substack{x\in\mathbb{R}^m \\ \|x\|_m \leqslant 1}} \|T(x)\|_n.$$

Lorsqu'il existe un $M > 0$ tel que $\|df(a)\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)} < M$ pour tout a dans U, nous disons que la différentielle de f est **bornée** sur U.

Définition 11.237.
*Une application $f \colon V \to E$ entre deux espaces vectoriels normés est dite être de **de classe** C^1 si son application différentielle $df \colon V \to \mathcal{L}(V, E)$ est continue*[58].

11.238.
Dans la suite, si $\{e_i\}$ est une base d'un espace vectoriel V, nous allons noter

$$\begin{aligned} \omega_i \colon V &\to \mathbb{R} \\ x &\mapsto x_i. \end{aligned} \tag{11.596}$$

58. Sur $\mathcal{L}(V, E)$ nous considérons la topologie de la norme opérateur 10.3.

Ce ω_i est ce qu'on appelle souvent e_i^*. Plus généralement, si I est le multiindice $(i_1, \dots, i_l)$ nous notons $\omega_I \in \mathcal{L}^l(V, \mathbb{R})$ par

$$\begin{aligned} \omega_I \colon V^l &\to \mathbb{R} \\ (x^{(1)}, \dots, x^{(l)}) &\mapsto x_{i_1}^{(1)} \dots x_{i_l}^{(l)}. \end{aligned} \tag{11.597}$$

Ce seront nos formes multilinéaires de base.

Afin de garder des notations très explicites, nous ne pouvons pas écrire des formules comme

$$df_a = \sum_i \frac{\partial f}{\partial x_i}(a)\omega_i$$

parce que si f prend ses valeurs dans $\mathcal{L}(V, \mathbb{R})$, lorsqu'on écrit $df_a(v)$, il n'y a aucune raison à priori de vouloir que v soit pris par ω_i au lieu de $\partial_i f(a)$.

Nous introduisons donc un produit fait exprès pour dire que « c'est celui de droite qui prend ».

Définition 11.239 ([1]).
Si W est un espace vectoriel, nous définissons le produit $\times_n$ par

$$\begin{aligned} \times_1 \colon W \times \mathcal{L}(V, \mathbb{R}) &\to \mathcal{L}(V, W) \\ (w \times_1 \alpha)(v) &= \alpha(v)w \end{aligned} \tag{11.598}$$

et par

$$\begin{aligned} \times_n \colon W \times \mathbb{R}_n &\to W_n \\ (w \times_n \alpha)(v) &= w \times_{n-1} \alpha(v) \end{aligned} \tag{11.599}$$

11.240.
Quelque précisions sur l'énoncé de la proposition 11.241. Lorsque nous parlons de $\partial_i f$, nous supposons donnée une base de V. Il n'y a aucune raison que la norme sur V soit adaptée à cette base. Nous allons donc utiliser une norme « euclidienne » adaptée à la base, et invoquer l'équivalence de toutes les normes pour dire que si une fonction est différentiable pour une norme elle l'est pour toutes les normes.

Proposition 11.241 ([147]).
Soient des espaces vectoriels normés de dimensions finies V et E ainsi qu'une application $f \colon V \to E$. Nous supposons que les dérivées partielles $\frac{\partial f}{\partial x_i}$ existent sur un voisinage de $a \in V$ et qu'elles sont continues en a.

Alors

(1) l'application f est différentiable en a,

(2) la différentielle est donnée par

$$df_a = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(a) \times_1 \omega_j. \tag{11.600}$$

Démonstration. L'espace V a une norme que nous notons $\|.\|_V$; nous n'allons presque pas l'utiliser. Nous nommons $\{e_i\}$ la base de V sous-entendue lorsque nous parlons des dérivées partielles, et nous considérons la norme $\|\|$ sur V donnée par $\|v\| = \sqrt{\sum_{i=1}^n x_i^2}$ pour $x = \sum_i x_i e_i$. Toutes les normes et boules dont nous allons parler dans la suite seront par rapport à cette norme.

Nous posons $\epsilon > 0$. La continuité des dérivées partielles en a nous permet de considérer $\delta > 0$ tel que

$$\|\frac{\partial f}{\partial x_j}(s) - \frac{\partial f}{\partial x_j}(a)\| < \epsilon \tag{11.601}$$

pour tout $s \in B(a, \delta)$, et pour tous les j en même temps [59].

59. Ce qui peut n'être pas possible en dimension infine. Je dis ça comme ça, juste pour faire remarquer que cette proposition n'est peut-être pas vraie en dimension infine. Voir aussi l'exemple de la fonction (24.131).

Soit $x \in B(a,\delta)$. Nous considérons les points

$$s_j = (x_1, \dots, x_j, a_{j+1}, \dots, a_n). \tag{11.602}$$

Précision : $s_n = x$.

Un peu de convexité Nous montrons que le segment $[s_j, s_{j-1}]$ est dans $B(a,\delta)$. Nous avons $s_j - a = (x_1 - a_1, \dots, x_j - a_j, 0, \dots, 0)$, et donc

$$\|s_j - a\| \leqslant \|x - a\| < \delta, \tag{11.603}$$

et donc $s_j \in B(a,\delta)$ pour tout j. Vu que la sphère $B(a,\delta)$ est convexe[60], tout le segment entre s_j et s_{j-1} est dedans.

Accroissements finis En nous souvenant que x et a ont été fixés, nous considérons les fonctions intermédiaires suivantes :

$$\begin{cases} g_1(t) = f(f, a_2, \dots, a_n) & \text{(11.604a)} \\ g_j(t) = f(x_1, \dots, x_{j-1}, t, a_{j+1}, \dots, a_n) & \text{(11.604b)} \\ g_n(t) = f(x_1, \dots, x_{n-1}, t). & \text{(11.604c)} \end{cases}$$

Elles vérifient en particulier $g_j(x_j) = f(s_j)$ et $g_j(a_j) = f(s_{j-1})$. Donc le théorème des accroissements finis 11.130 nous donne l'existence de τ_j entre x_j et a_j tel que

$$f(s_j) - f(s_{j-1}) = g_j(x_j) - g_j(a_j) = g_j'(\tau_j)(x_j - a_j). \tag{11.605}$$

Dérivées partielles Les fonctions g_j ont été construites de telle sorte à donner les dérivées partielles de f via une simple dérivation :

$$g_j'(t_0) = \frac{\partial f}{\partial x_j}(x_1, \dots, x_{j-1}, t_0, a_{j+1}, \dots, a_n). \tag{11.606}$$

Or lorsque[61] $t_0 \in [x_j, a_j]$, le point $(x_1, \dots, x_{j-1}, t_0, a_{j+1}, \dots, a_n)$ est dans $[s_j, s_{j-1}]$. Cela est le cas pour $t_0 = \tau_j$. Nous notons

$$\bar{s}_j = (x_1, \dots, x_{j-1}, \tau_j, a_{j+1}, \dots, a_n) \tag{11.607}$$

et nous avons

$$g_j'(\tau_j) = \frac{\partial f}{\partial x_j}(\bar{s}_j) \tag{11.608}$$

avec $\bar{s}_j \in [s_{j-1}, s_j]$. Tout cela pour récrire (11.605) sous la forme

$$f(s_j) - f(s_{j-1}) = \frac{\partial f}{\partial x_j}(\bar{s}_j)(x_j - a_j). \tag{11.609}$$

Une belle somme télescopique En nous souvenant que $s_0 = a$ et que $s_n = x$, nous avons cette somme télescopique

$$\begin{aligned} f(x) - f(a) &= \sum_{j=1}^{n} \big(f(s_j) - f(s_{j-1})\big) && \text{(11.610a)} \\ &= \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(\bar{s}_j)(x_j - a_j) && \text{(11.610b)} \\ &= \sum_{j=1}^{n} \big[(\partial_j f)(\bar{s}_j) - (\partial_j f)(a)\big](x_j - a_j) + \sum_{j=1}^{n} (\partial_i f)(a)(x_j - a_j). && \text{(11.610c)} \end{aligned}$$

Nous isolons les termes qui nous intéressent dans la définition de la différentielle :

$$f(x) - f(a) - \sum_{j=1}^{n} (\partial_j f)(a)(x_j - a_j) = \sum_{j=1}^{n} \big[(\partial_j f)(\bar{s}_j) - (\partial_j f)(a)\big](x_j - a_j). \tag{11.611}$$

60. Proposition 8.26.
61. Retournez éventuellement les intervalles sur $a_j < x_j$.

Et enfin : le quotient différentiel En utilisant la majoration (11.601), et en remarquant que $\|x-a\|$ majore $|x_j-a_j|$ pour chaque j, nous avons l'inégalité

$$\|f(x)-f(a)-\sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(a)(x_j-a_j)\| \leqslant n\epsilon\|x-a\|. \tag{11.612}$$

Autrement dit, en posant $T=\sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(a)\omega_j$ nous avons

$$\|f(a+h)-f(a)-T(h)\| \leqslant n\epsilon\|h\|, \tag{11.613}$$

et donc

$$\lim_{h\to 0}\frac{\|f(a+h)-f(a)-T(h)\|}{\|h\|}=0. \tag{11.614}$$

Et donc ? La limite (11.614) nous indique que f serait différentiable de différentielle T si $\|.\|$ était la norme sur V. C'est l'équivalence de toutes les normes [62] qui fait en sorte que la norme sur V n'est pas importante.

□

Proposition 11.242.
Soient des espaces vectoriels normés de dimensions finies V et E ainsi qu'une application $f\colon V\to E$. Nous supposons que les dérivées partielles $\frac{\partial f}{\partial x_i}$ existent et sont continues sur un ouvert $\mathcal{U}$ de V.

Alors f est de classe C^1 sur $\mathcal{U}$.

Démonstration. Nous notons $\mathcal{U}$ le voisinage sur lequel les dérivées partielles de f existent et sont continues. Il s'agit d'appliquer la proposition 11.241 en chaque point de $\mathcal{U}$. Nous avons alors

$$df_x=\sum_{j=1}^{n}\frac{\partial f}{\partial x_j}(x)\times_1\omega_j \tag{11.615}$$

pour tout $x\in\mathcal{U}$. Pour $a\in V$ et $h\in V$ supposé petit, nous avons

$$\|df_a-df_{a+h}\|_{\mathcal{L}(V,E)}=\sup_{\|v\|=1}\|df_a(v)-df_{a+h}(v)\|_E \tag{11.616a}$$

$$=\sup_{\|v\|=1}\|\sum_{j=1}^{n}\big(\frac{\partial f}{\partial x_j}(a)-\frac{\partial f}{\partial x_j}(a+h)\big)v_j\| \tag{11.616b}$$

$$\leqslant\sup_{\|v\|=1}\|\sum_{j=1}^{n}\big(\frac{\partial f}{\partial x_j}(a)-\frac{\partial f}{\partial x_j}(a+h)\big)\||v_i| \tag{11.616c}$$

$$\leqslant\|\sum_{j=1}^{n}\big(\frac{\partial f}{\partial x_j}(a)-\frac{\partial f}{\partial x_j}(a+h)\big)\| \tag{11.616d}$$

Justifications :

— Pour (11.616d), nous avons majoré $|v_i|\leqslant\|v\|=1$.

— Pour (11.616c), nous avons utilisé le lemme 10.10.

La continuité de $\partial_j f$ conclu que $\lim_{h\to 0}\|df_a-df_{a+h}\|=0$, ce qui signifie que df est continue et donc que f est de classe C^1. □

Théorème 11.243.
Soient des espaces vectoriels normés de dimensions finies V et E. Soit un ouvert $\mathcal{U}$ de V. Une application $f\colon V\to E$ est de classe C^1 sur $\mathcal{U}$ si et seulement si ses dérivées partielles existent et sont continues sur $\mathcal{U}$.

Dans ce cas, nous avons la formule

$$df_a=\sum_i\frac{\partial f}{\partial x_i}(a)\times_1\omega_i. \tag{11.617}$$

62. Sur un espace de dimension finie, 7.224.

Démonstration. Le premier sens, y compris la formule (11.617), a été fait dans la proposition 11.242. Nous supposons maintenant que f est de classe C^1 sur $\mathcal{U}$ et nous allons prouver que ses dérivées partielles existent et sont continues. Pour tout $t \neq 0$ nous avons l'égalité

$$\frac{f(a+te_i)-f(a)}{t} = \frac{f(a+te_i)-f(a)-df_a(fe_i)}{t} + df_a(e_i). \tag{11.618}$$

Par définition de la différentielle, prendre la limite $t \to 0$ à droite donne $df_a(e_i)$ parce que la fraction tend vers zéro. La limite définissant la dérivée partielle existe donc et vaut

$$\frac{\partial f}{\partial x_i}(a) = df_a(e_i). \tag{11.619}$$

Il nous reste à prouver que les dérivées partielles sont continues :

$$\begin{aligned}
\|\frac{\partial f}{\partial x_i}(a+h) - \frac{\partial f}{\partial x_i}(a)\| &= \|df_a(e_i) - df_{a+h}(e_i)\| && (11.620a)\\
&= \|(df_{a+h} - df_a)(e_i)\| && (11.620b)\\
&\leqslant \|df_{a+h} - df_a\|\|e_i\| && (11.620c)\\
&= \|df_{a+h} - df_a\|. && (11.620d)
\end{aligned}$$

Vu que f est de classe C^1, la limite $h \to 0$ de cela donne zéro. □

Voici une version du même théorème, démontré dans le cas seulement de dimension deux. Il rend peut-être aussi plus clairement pourquoi ça ne marche pas en dimension infinie.

Proposition 11.244.
Soit U un ouvert dans $\mathbb{R}^m$ et a un point dans U. Soit f une application de U dans $\mathbb{R}^n$. Si toutes les dérivées partielles de f existent sur U et sont continues au point a alors f est différentiable au point a.

Démonstration. On se limite au cas $m = 2$. Pour rendre les calculs plus simples on utilise ici la norme $\|\cdot\|_\infty$ dans l'espace $\mathbb{R}^2$, mais comme on a vu plus en haut, cela ne peut pas avoir des conséquences sur la différentiabilité de f. Si la différentielle de f au point a existe alors elle est définie par la formule

$$df_a(v) = \frac{\partial f}{\partial x}(a)v_1 + \frac{\partial f}{\partial y}(a)v_2$$

pour tout v dans $\mathbb{R}^m$.

On commence par prouver le résultat en supposant que les dérivées partielles de f au point a sont nulles. La différentiabilité de f signifie que pour toute constante $\varepsilon > 0$ il y a une constante $\delta > 0$ telle que si $\|v\|_\infty \leqslant \delta$ alors

$$\frac{\|f(a_1+v_1, a_2+v_2) - f(a_1,a_2)\|_n}{\|v\|_\infty} \leqslant \varepsilon.$$

On écrit alors

$$\begin{aligned}
&\|f(a_1+v_1, a_2+v_2) - f(a_1,a_2)\|_n = \\
&= \|f(a_1+v_1, a_2+v_2) - f(a_1+v_1, a_2) + f(a_1+v_1, a_2) - f(a_1,a_2)\|_n \leqslant \\
&\leqslant \|f(a_1+v_1, a_2+v_2) - f(a_1+v_1, a_2)\|_n + \|f(a_1+v_1, a_2) - f(a_1,a_2)\|_n.
\end{aligned} \tag{11.621}$$

Comme la dérivée partielle $\partial_x f$ est nulle au point a on sait que pour toute constante $\varepsilon > 0$ il y a une constante $\delta_1 > 0$ telle que si $|v_1| \leqslant \delta_1$ alors

$$\|f(a_1+v_1, a_2) - f(a_1,a_2)\|_n \leqslant \varepsilon|v_1|.$$

Pour l'autre terme on a, par la proposition 11.174,

$$\|f(a_1+v_1, a_2+v_2) - f(a_1+v_1, a_2)\|_n \leqslant \sup\{\|\partial_y f(x)\|_n \,|\, x \in S\}|v_2|. \tag{11.622}$$

où S est le segment d'extrémités (a_1+v_1, a_2) et (a_1+v_1, a_2+v_2). Comme la dérivée partielle $\partial_y f$ est continue et nulle au point a on sait que pour toute constante $\varepsilon > 0$ il existe une constante $\delta_2 > 0$ telle que si $\|(u_1,u_2)\|_\infty \leqslant \delta_2$ alors $\|\partial_y f(a_1+u_1, a_2+u_2)\|_n \leqslant \varepsilon$. Si on choisit $\delta = \min\{\delta_1, \delta_2\}$ le segment S est contenu dans la boule de rayon δ centrée au point a et on obtient

$$\|f(a_1+v_1, a_2+v_2) - f(a_1,a_2)\|_n \leqslant \varepsilon|v_1| + \varepsilon|v_2| \leqslant 2\varepsilon\|v\|_\infty.$$

Cela prouve que f est différentiable en (a_1,a_2) et que la différentielle est nulle :

$$df_{(a_1,a_2)} = 0. \tag{11.623}$$

Dans le cas général, où les dérivées partielles de f au point a ne sont pas spécialement nulles, on peut considérer la fonction[63]

$$g(x,y) = f(x,y) - \partial_1 f(a)x - \partial_2 f(a)y, \tag{11.624}$$

qui a dérivées partielles nulles au point a. La fonction g est donc différentiables. La fonction f est maintenant la somme de g et de la fonction linéaire et continue $(x,y) \mapsto \partial_1 f(a)x - \partial_2 f(a)y$. On verra dans la prochaine section que la somme de deux fonctions différentiables est une fonction différentiable. Par conséquent, la fonction f est différentiable. □

Remarque 11.245.
En dimension infinie, il n'est pas vrai que l'existence et la continuité de toutes les dérivées partielles en un point implique la différentiabilité en ce point. Pour donner un exemple, nous allons continuer l'exemple 10.28 avec la fonction 24.131 sur un espace de Hilbert.

En dimension infinie nous aurons le théorème 10.147 qui donnera quelque chose de moins fort.

11.26 Différentielle et dérivée complexe

11.246.
Nous commençons par donner quelques éléments à propos de dérivée et de différentielle pour des fonctions $\mathbb{C} \to \mathbb{C}$ parce que les séries entières vont souvent être des fonctions complexes. Le gros du chapitre sur les fonctions holomorphes est le chapitre 25.

Nous identifions $\mathbb{R}^2$ à $\mathbb{C}$ par l'application

$$\begin{aligned} \varphi\colon \mathbb{R}^2 &\to \mathbb{C} \\ (x,y) &\mapsto x+iy. \end{aligned} \tag{11.625}$$

Dans cette partie, nous désignons par Ω un ouvert de $\mathbb{C}$.

Définition 11.247.
*Une fonction $f\colon \Omega \to \mathbb{C}$ est $\mathbb{C}$-**dérivable** si la limite*

$$\lim_{\substack{h\to 0\\ h\in\mathbb{C}}} \frac{f(a+h)-f(a)}{h} \tag{11.626}$$

existe. Dans ce cas, cette limite est la dérivée de f et est notée f'.

Définition 11.248.
*Soit Ω un ouvert dans $\mathbb{C}$. Une fonction $f\colon \Omega \to \mathbb{C}$ est **holomorphe** si elle est $\mathbb{C}$-dérivable sur Ω.*

Définition 11.249.
*Si K est un compact de $\mathbb{C}$, nous disons qu'une fonction est **holomorphe** sur K si il existe un ouvert contenant K sur lequel f est holomorphe.*

Et si f n'est réellement définie que sur K, elle est holomorphe sur K si il existe une extension holomorphe de f vers un ouvert contenant K.

63. Vous verrez dans la discussion à propos de la fonction (24.131) pourquoi cette fonction ne fonctionne pas dans le cas de la dimension infinie.

Définition 11.250.
Une matrice de la forme

$$\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \tag{11.627}$$

*avec $\alpha, \beta \in \mathbb{R}$ est une **similitude**.*

Lemme 11.251.
En tant qu'application linéaire $\mathbb{C} \to \mathbb{C}$, l'opération de multiplication par $\alpha + \beta i$ est la matrice

$$\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}. \tag{11.628}$$

Démonstration. Cela est vite remarqué en calculant explicitement $(\alpha + \beta i)(u_1 + iu_2)$. □

Lemme 11.252.
Une application $A\colon \mathbb{C} \to \mathbb{C}$ est $\mathbb{C}$-linéaire si et seulement si elle est une similitude en tant qu'application $\mathbb{R}^2 \to \mathbb{R}^2$.

Dans ce cas, il existe $z_0 \in \mathbb{C}$ tel que $A(z) = z_0 z$ pour tout $z \in \mathbb{C}$.

Démonstration. Commençons par considérer l'application A sur $\mathbb{R}^2$. Elle est en particulier une application $\mathbb{R}$-linéaire et par conséquent il existe une matrice $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ telle que

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{11.629}$$

Nous voulons maintenant imposer la $\mathbb{C}$-linéarité, c'est-à-dire que nous voulons

$$A\big((a + bi)(x + iy)\big) = (a + bi)A(x + iy) \tag{11.630}$$

pour tout $a, b, x, y \in \mathbb{R}$. À gauche nous avons

$$A\big(ax - by + i(bx + ay)\big) \tag{11.631}$$

et à droite nous avons

$$(a + bi)\big(\alpha x + \beta y + i(\gamma x + \delta y)\big). \tag{11.632}$$

En égalant les deux expressions nous obtenons les équations

$$\begin{cases} \beta b = -b\gamma & \text{(11.633a)} \\ -\alpha b + \beta a = a\beta - b\delta & \text{(11.633b)} \\ \delta b = b\alpha & \text{(11.633c)} \\ -\gamma b + \delta a = b\beta + a\delta, & \text{(11.633d)} \end{cases}$$

dont nous tirons immédiatement que $\gamma = -b\beta$ et $\delta = \alpha$. La matrice de A est donc de la forme demandée.

Inversement nous devons prouver que la fonction

$$f(x + iy) = \alpha x + \beta y + i(-\beta x + \alpha y) \tag{11.634}$$

est $\mathbb{C}$-linéaire, c'est-à-dire qu'elle vérifie $f(z_0 z) = z_0 f(z)$ pour tout $z_0, z \in \mathbb{C}$. Cela est un simple calcul que nous confions à Sage : le code suivant affiche « 0 ».

```
#! /usr/bin/sage -python
# -*- coding: utf8 -*-

from sage.all import *
```

```

def f(z):
    var('alpha,beta')
    x=z.real_part()
    y=z.imag_part()
    return alpha*x+beta*y+I*( -beta*x+alpha*y  )


var('a,b,x,y')

A=a+b*I
Z=x+y*I

z1=f( A*Z  )
z2=A*f( Z )

rep=z1-z2
print(rep.full_simplify())
```

tex/frido/code_sage3.py

Pour conclure, notons que la fonction (11.634) est la fonction de multiplication par $\alpha - i\beta$. □

11.253.
Soient une fonction $f\colon \mathbb{C} \to \mathbb{C}$ et l'isomorphisme canonique $\varphi\colon \mathbb{C} \to \mathbb{R}^2$. La fonction f définit une la fonction

$$F = \varphi^{-1} \circ f \circ \varphi\colon \mathbb{R}^2 \to \mathbb{R}^2. \tag{11.635}$$

Cela est la fonction $\mathbb{R}^2 \to \mathbb{R}^2$ associée à f. Il serait tentant de croire que tout ce qui est vrai pour F est également vrai pour f. Eh bien non.

Par exemple, F peut être différentiable sans que f le soit. La proposition suivant donne une condition sur dF pour que f soit différentiable.

Proposition 11.254.
Une fonction $f\colon \Omega \to \mathbb{C}$ est $\mathbb{C}$-dérivable en $a \in \Omega$ si et seulement si elle est différentiable en a et si df_a est une similitude.

Plus précisément avec les notations de 11.253, la fonction f est $\mathbb{C}$-dérivable (donc holomorphe) au point $z_0 = x_0 + iy_0$ si et seulement si la fonction F est différentiable en (x_0, y_0) et si la matrice de dF est de la forme

$$dF = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}, \tag{11.636}$$

c'est-à-dire si $dF_{(x_0,y_0)}$ fournit une application $\mathbb{C}$-linéaire.

Dans ce cas, le lien entre $\mathbb{C}$-dérivée et différentielle est donné par

$$(df_{z_0})(z) = f'(z_0)z. \tag{11.637}$$

Démonstration. Nous décomposons f en parties réelles et imaginaires :

$$f(x + iy) = P(x, y) + iQ(x, y) \tag{11.638}$$

où P et Q sont des fonctions réelles. La jacobienne de F est la matrice

$$\begin{pmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} \\ \frac{\partial Q}{\partial x} & \frac{\partial Q}{\partial y} \end{pmatrix}, \tag{11.639}$$

et la condition dont nous parlons s'écrit comme le système

$$\begin{cases} \dfrac{\partial P}{\partial x} = \dfrac{\partial Q}{\partial y} & \text{(11.640a)} \\ \dfrac{\partial P}{\partial y} = -\dfrac{\partial Q}{\partial x}. & \text{(11.640b)} \end{cases}$$

Si F est différentiable en (x_0, y_0) alors nous avons

$$F\big((x_0, y_0) + (h, k)\big) = F(x_0, y_0) + dF_{(x_0,y_0)} \begin{pmatrix} h \\ k \end{pmatrix} + s(|h| + |k|) \tag{11.641}$$

où s est une fonction vérifiant $\lim_{t\to 0} \frac{s(t)}{t} = 0$. Soit

$$dF_{(x_0,y_0)} = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}. \tag{11.642}$$

Si nous posons $\sigma = \alpha - i\beta$ et $w = h + ik$, l'équation (11.641) s'écrit dans $\mathbb{C}$ sous la forme

$$f(z_0 + w) = f(z_0) + \sigma w + s(|w|), \tag{11.643}$$

ce qui implique que f est $\mathbb{C}$-dérivable en z_0.

Supposons maintenant que f soit $\mathbb{C}$-dérivable en z_0. Alors nous avons

$$f'(z_0) = \lim_{w\to 0} \frac{f(z_0 + w) - f(z_0)}{w} = \sigma \in \mathbb{C}, \tag{11.644}$$

ce qui se récrit sous la forme

$$\lim_{w\to 0} \frac{f(z_0 + w) - f(z_0) - \sigma w}{w} = 0. \tag{11.645}$$

Si nous posons $z_0 = x_0 + iy_0$, $w = h + ik$ et $\sigma = \alpha - i\beta$ nous avons

$$\lim_{(h,k)\to(0,0)} \left| \frac{F\big((x_0, y_0) + (h, k)\big) - F(x_0, y_0) - \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \begin{pmatrix} h \\ k \end{pmatrix}}{|w|} \right| = 0, \tag{11.646}$$

ce qui signifie que F est différentiable et que sa différentielle est la matrice

$$\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}. \tag{11.647}$$

La matrice (11.647) est, vue dans $\mathbb{R}^2$, la matrice de multiplication dans $\mathbb{C}$ par $\alpha - i\beta = f'(z_0)$. En d'autre termes, dans $\mathbb{C}$ nous avons

$$df_{z_0}(z) = f'(z_0)z, \tag{11.648}$$

et en particulier la différentielle est donnée par

$$df_{z_0} = f'(z_0)dz. \tag{11.649}$$

□

Exemple 11.255 (Une application C^∞ mais pas $\mathbb{C}$-dérivable)**.**
Nous considérons la fonction

$$\begin{aligned} f\colon \mathbb{C} &\to \mathbb{C} \\ x + iy &\mapsto x. \end{aligned} \tag{11.650}$$

Vu que c'est une application linéaire, elle est différentiable une infinité de fois et sa différentielle est elle-même. C'est donc une application C^∞.

Elle n'est cependant pas $\mathbb{C}$-dérivable. En effet le quotient différentiel est, pour $\epsilon \in \mathbb{C}$:

$$\frac{f(x + iy + \epsilon_x + i\epsilon_y) - f(x + iy)}{\epsilon} = \frac{\epsilon_x}{\epsilon}. \tag{11.651}$$

Cela n'a pas de limite lorsque $\epsilon \to 0$. Pour voir cela nous invoquons la méthode des chemins du corolaire 11.162 avec les chemins $\epsilon_1(t) = t$ et $\epsilon_2(t) = it$. Dans le premier cas, le quotient différentiel vaut 1 pour tout t, tandis que dans le second il vaut zéro pour tout t. △

11.26.1 Quelques règles de calcul

Lemme 11.256.
Si f et g sont deux fonctions holomorphes sur un ouvert $\Omega \subset \mathbb{C}$ et si g ne s'annule pas sur Ω, alors f/g est holomorphe sur Ω.

11.27 Théorèmes des accroissements finis

Nous avons déjà démontré (lemme 11.196) que si f est différentiable au point x alors $df_x(u) = \partial_u f(x)$. Une importante conséquence est le théorème des accroissements finis

Théorème 11.257 (Accroissements finis, inégalité de la moyenne)**.**
Soit U un ouvert dans $\mathbb{R}^m$ et soit $f : U \to \mathbb{R}^n$ une fonction différentiable. Soient a et b deux points dans U, $a \neq b$, tels que le segment $[a, b]$ soit contenu dans U. Alors

$$\|f(b) - f(a)\|_n \leqslant \sup_{x\in[a,b]} \|df(x)\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)} \|b - a\|_m. \tag{11.652}$$

Démonstration. On utilise le théorème 11.174 et le fait que

$$\|\partial_u f(x)\|_n \leqslant \|df(x)\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)} \|u\|_m,$$

pour tout u dans $\mathbb{R}^m$. □

La proposition suivante est une application fondamentale du théorème des accroissements finis 11.257.

Proposition 11.258.
Soit U un ouvert connexe par arcs de $\mathbb{R}^m$ et une fonction $f\colon U \to \mathbb{R}^n$. Les conditions suivantes sont équivalentes :

(1) f est constante ;

(2) f est différentiable et $df(a) = 0$ pour tout $a \in U$;

(3) les dérivées partielles $\partial_1 f, \dots, \partial_m f$ existent et sont nulles sur U.

Démonstration. Nous allons démonter les équivalences en plusieurs étapes. D'abord (1) ⇒ (2), puis (2) ⇒ (3), ensuite (3) ⇒ (2) et enfin (2) ⇒ (1).

Commençons par montrer que la condition (1) implique la condition (2). Si $f(x)$ est constante, alors la condition (10.289) est vite vérifiée en posant $T(h) = 0$.

Afin de voir que la condition (2) implique la condition (3), remarquons d'abord que la différentiabilité de f implique que les dérivées partielles existent (proposition 11.194) et que nous avons l'égalité $df(a).u = \sum_i u_i \partial_i f(a)$ pour tout $u \in \mathbb{R}^m$ (lemme 11.196). L'annulation de $\sum_i u_i \partial_i f(a)$ pour tout u implique l'annulation des $\partial_i f(a)$ pour tout i.

Prouvons maintenant que la propriété (3) implique la propriété (2). D'abord, par le théorème 11.243, l'existence et la continuité des dérivées partielles $\partial_i f(a)$ implique la différentiabilité de f. Ensuite, la formule $df(a).u = \sum_i u_i \partial_i f(a)$ implique que $df(a) = 0$.

Il reste à montrer que (2) implique la condition (1), c'est-à-dire que l'annulation de la différentielle implique la constance de la fonction. C'est ici que nous allons utiliser le théorème des accroissements finis. En effet, si a et b sont des points de U, le théorème 11.257 nous dit que

$$\|f(b) - f(a)\|_n \leqslant \sup_{x\in[a,b]} \|df(x)\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)} \|b - a\|_m. \tag{11.653}$$

Mais $\|df(x)\| = 0$ pour tout $x \in U$, donc ce supremum est nul et $f(b) = f(a)$, ce qui signifie la constance de la fonction. □

11.28 Fonctions Lipschitziennes

Définition 11.259.
Soient (E, d_E) et (F, d_F) deux espaces métriques[64]*, $f\colon E \to F$ une application et un réel k strictement positif. Nous disons que f est* ***Lipschitzienne*** *de constante k sur E si pour tout $x, y \in E$,*

$$d_F\big(f(x) - f(y)\big) \leqslant k d_E(x, y). \tag{11.654}$$

Soit f une fonction k-Lipschitzienne. Si $y \in \overline{B(x,\delta)}$ alors $\|x-y\| \leqslant \delta$ et donc $\big\|f(x)-f(y)\big\| \leqslant k\delta$. Cela signifie que la condition Lipschitz pour s'énoncer en termes de boules fermées par

$$f\big(\overline{B(x,\delta)}\big) \subset \overline{B\big(f(x), k\delta\big)} \tag{11.655}$$

tant que $\overline{B(x,\delta)}$ est contenue dans le domaine sur lequel f est Lipschitz.

Proposition 11.260.
Soit U un ouvert convexe de $\mathbb{R}^m$, et soit $f : U \to \mathbb{R}^n$ une fonction différentiable. La fonction f est Lipschitzienne sur U si et seulement si df est bornée sur U.

Démonstration. Le fait que l'application différentielle df soit bornée signifie qu'il existe un $M > 0$ dans $\mathbb{R}$ tel que $\|df_a\|_{\mathcal{L}(\mathbb{R}^m,\mathbb{R}^n)} \leqslant M$, pour tout a dans U. Si cela est le cas, alors le théorème 11.257 et la convexité[65] de U impliquent évidemment que f est de Lipschitz de constante plus petite ou égale à M.

Inversement, si f est Lipschitz de constante k, alors pour tout a dans U et u dans $\mathbb{R}^m$ on a

$$\left\| \frac{f(a+tu) - f(a)}{t} \right\|_n \leqslant k\|u\|_m,$$

En passant à la limite pour $t \to 0$ on a

$$\|\partial_u f(a)\|_n = \|df_a(u)\|_n \leqslant k\|u\|_m,$$

donc la norme de df_a est majorée par k pour tout a dans U. □

Notez cependant qu'une fonction peut être Lipschitzienne sans être différentiable.

Proposition 11.261.
Une fonction Lipschitzienne $f\colon \mathbb{R} \to \mathbb{R}$ est continue.

Démonstration. Nous utilisons la caractérisation de la continuité donnée par le théorème 7.74. Prouvons donc la continuité en $a \in \mathbb{R}$. Pour tout x nous avons

$$\big|f(x) - f(a)\big| \leqslant k|x - a|. \tag{11.656}$$

Si $\epsilon > 0$ est donné, il suffit de prendre $\delta < \frac{\epsilon}{k}$ pour avoir

$$\big|f(x) - f(a)\big| \leqslant k\frac{\epsilon}{k} = \epsilon. \tag{11.657}$$

Donc f est continue en a. □

Définition 11.262.
Une fonction

$$\begin{aligned} f\colon \mathbb{R}^n \times \mathbb{R}^m &\to \mathbb{R}^p \\ (t, y) &\mapsto f(t, y) \end{aligned} \tag{11.658}$$

64. Pour rappel, les espaces métriques sont définis par la définition 7.93 et le théorème 7.94 ; je précise que nous ne supposons pas que E soit vectoriel ; en particulier il peut être un ouvert de $\mathbb{R}^n$.

65. La convexité de U sert à assurer que la droite reliant a à b est contenue dans U ; c'est ce que nous utilisons dans la démonstration du théorème 11.257.

est ***localement Lipschitz*** *en y au point (t_0, y_0) s'il existe des voisinages V de t_0 et W de y_0 et un nombre $k > 0$ tels que pour tout $(t, y) \in V \times W$ on ait*

$$\big\|f(t_0, y_0) - f(t, y)\big\| \leqslant k\|y - y_0\|. \tag{11.659}$$

La fonction est localement Lipschitz sur un ouvert U de $\mathbb{R}^n \times \mathbb{R}^m$ si elle est localement Lipschitz en chaque point de U.

11.263.
Autrement dit, une fonction est localement Lipschitzienne en sa deuxième variable lorsque tout point admet un voisinage sur lequel elle est Lipschitzienne.

Proposition 11.264.
Une application Lipschitz[66] *est uniformément continue.*

Proposition 11.265.
Si f et g sont deux fonctions localement Lipschitz alors $f + g$ l'est.

Démonstration. Il s'agit d'un simple calcul avec une majoration standard :

$$\begin{aligned}
\|(f+g)(t_0, y_0) - (f+g)(t, y)\| &\leqslant \|f(t_0, y_0) - f(t, y)\| + \|g(t_0, y_0) - g(t, y)\| && (11.660a)\\
&\leqslant k_f\|y - y_0\| + k_g\|y - y_0\| && (11.660b)\\
&= (k_f + k_g)\|y - y_0\|. && (11.660c)
\end{aligned}$$

□

Lemme 11.266.
La fonction donné par

$$f(t, (x, y)) = xy \tag{11.661}$$

est localement Lipschitz en tout point.

Démonstration. Nous avons la majoration classique

$$|f\big(t, (x_0, y_0)\big) - f\big(t, (x, y)\big)| = |x_0y_0 - xy| \leqslant |x_0y_0 - x_0y| + |x_0y - xy| \leqslant |x_0||y_0 - y| + |y||x_0 - x|. \tag{11.662}$$

Vu que nous parlons de fonction *localement Lipschitzienne*, nous pouvons majorer $|y|$ et $|x_0|$ par un même nombre k dans un voisinage de (x_0, y_0). Cela donne

$$|f\big(t, (x_0, y_0)\big) - f\big(t, (x, y)\big)| \leqslant k\big(|y_0 - y| + |x_0 - x|\big) \leqslant \sqrt{2}k\left\|\begin{pmatrix} x_0 - x \\ y_0 - y \end{pmatrix}\right\|. \tag{11.663}$$

Nous avons utilisé l'équivalence de norme de la proposition 7.223(1). □

11.29 Différentielles d'ordre supérieur

11.29.1 Différentielle et dérivées partielles

Définition 11.267 ([248, 1])**.**
Soient deux espaces vectoriels normés V et W ainsi qu'une application $f\colon V \to W$. Nous disons que f est de classe C^1 si elle est différentiable et si sa différentielle $df\colon V \to \mathcal{L}(V, W)$ est une application continue.

Nous disons que f est de classe C^k si sa différentielle est de classe C^{k-1}.

Deux cas un peu particuliers :

— *l'application f est de classe C^0 si elle est continue,*

— *elle est de classe C^∞ si elle est de classe C^k pour tout k.*

66. Définition 11.259.

Soient deux espaces vectoriels normés V et W ainsi qu'une application $f\colon V\to W$. La différentielle est une application $df\colon V\to\mathcal{L}(V,W)$. Pour être clair, la différentielle seconde consiste à différentier df_x par rapport à x. C'est à dire que la différentielle seconde est une application $d(df)\colon V\to\mathcal{L}\big(V,\mathcal{L}(V,W)\big)$.

Et c'est là que commencent les problèmes. Les différentielles successives font intervenir des emboîtements de plus en plus profonds d'espaces comme $d^2f\colon V\to\mathcal{L}\Big(V,\mathcal{L}\big(V,\mathcal{L}(V,W)\big)\Big)$.

Nous introduisons maintenant quelque notations et lemmes pour traiter ces problèmes. Soient des espaces vectoriels normés V et W. Nous commençons par les espaces emboîtés par récurrence :

$$\begin{cases} E_0=E & \text{(11.664a)}\\ E_{k+1}=\mathcal{L}(V,E_k). & \text{(11.664b)}\end{cases}$$

Lemme 11.268 ([1, 249])**.**
À propos de dimensions,

(1) $\dim(E_n)=\dim(V)^n\dim(E)$.

(2) $\dim\big(\mathcal{L}^n(V,E)\big)=\dim(E)\dim(V)^n$.

Nous commençons par une sorte de projection ; pour $u\in V$ nous définissons

$$\begin{aligned}\mathtt{proj}_u\colon\mathcal{L}^n(V,E)&\to\mathcal{L}^{n-1}(V,E)\\ \mathtt{proj}_u(\omega)(v_1,\dots,v_{n-1})&=\omega(u,v_1,\dots,v_{n-1}).\end{aligned}\tag{11.665}$$

Lemme 11.269 ([1])**.**
Nous définissons les ϕ_k par récurrence. D'abord

$$\begin{aligned}\phi_1\colon\mathcal{L}^n(V,E)&\to\mathcal{L}\big(V,\mathcal{L}^{n-1}(V,E)\big)\\ \phi_1(\omega)u&=\mathtt{proj}_u(\omega),\end{aligned}\tag{11.666}$$

et ensuite

$$\begin{aligned}\phi_k\colon\mathcal{L}^n(V,E)&\to\mathcal{L}^{n-k}(V,E)_k\\ \phi_k(\omega)u&=\phi_{k-1}\big(\mathtt{proj}_u(\omega)\big).\end{aligned}\tag{11.667}$$

Les applications ϕ_k sont bijectives.

Démonstration. Nous prouvons par récurrence.

Injective, $k=1$ Soit $\omega\in\mathcal{L}^n(V,E)$ tel que $\phi_1(\omega)=0$. Pour tout $u\in V$ nous avons $\phi_1(\omega)u=0$, ce qui signifie que $\mathtt{proj}_u(\omega)=0$ ou encore que pour tout $v_1,\dots,v_{n-1}\in V$ nous avons $\omega(u,v_1,\dots,v_{n-1})=0$. Nous avons donc bien $\omega=0$.

Surjective, $k=1$ Soit $\alpha\in\mathcal{L}\big(V,\mathcal{L}^{n-1}(V,E)\big)$. Nous allons construire $\omega\in\mathcal{L}^n(V,E)$ tel que $\phi_1(\omega)=\alpha$. Nous posons

$$\omega(v_0,\dots,v_{n-1})=\alpha(v_0)(v_1,\dots,v_{n-1}).\tag{11.668}$$

Avec lui nous avons bien $\mathtt{proj}_{v_0}(\omega)=\alpha(v_0)$.

Injective, $k=k$ Nous supposons que $\phi_k(\omega)=0$, c'est à dire que pour tout $u\in V$ nous avons

$$\phi_{k-1}\big(\mathtt{proj}_u(\omega)\big)=0.\tag{11.669}$$

Cela implique $\mathtt{proj}_u(\omega)=0$ parce que ϕ_{k-1} est injective par hypothèse de récurrence. Nous déduisons que $\omega=0$, et que ϕ_k est injective.

Surjective, $k=k$ Nous allons montrer que $\phi_k\colon\mathcal{L}^n(V,E)\to\mathcal{L}^{n-k}(V,E)_k$ est une application linéaire injective entre deux espaces de même dimension.

Il s'agit d'utiliser le lemme 11.268. D'abord

$$\dim\big(\mathcal{L}^{n-k}(V,E)_k\big)=\dim(V)^k\dim\big(\mathcal{L}^{n-k}(V,E)\big)=\dim(V)^n\dim(E).\tag{11.670}$$

Ensuite $\dim\big(\mathcal{L}^n(V,E)\big)=\dim(V)^n\dim(E)$. Le compte est bon.

Le théorème du rang (formule (4.44)) avec $\dim\big(\ker(\phi_k)\big)=0$ nous dit que le rang de ϕ_k est maximal et donc que ϕ_k est surjective.

□

Lemme 11.270.
Soit $w \in W$. Nous avons

$$\big[w \times_k \phi_k(\omega_I)\big] \times_1 \omega_l = w \times_{k+1} \phi_{k+1}(\omega_{l,I}). \tag{11.671}$$

Démonstration. Il s'agit d'appliquer les deux membres à un élément $v \in V$. D'abord

$$\big[w \times_k \phi_k(\omega_I)\big] \times_1 \omega_l(v) = \omega_l(v) w \times_k \phi_k(\omega_I) \tag{11.672a}$$
$$= w \times_k \big(\phi_k(\omega_I) v_l\big) \tag{11.672b}$$

Pour (11.672b) nous avons utilisé le fait que $\omega_{l,I}(v) = v_l \in \mathbb{R}$ et que le produit $\times_k$ est linéaire. De l'autre côté, en utilisant les définitions de $\times_k$ et de ϕ_k, nous avons

$$\big[w \times_{k+1} \phi_{k+1}(\omega_{l,I})\big] v = w \times_k \big(\phi_{k+1}(\omega_{l,I}) v\big) \tag{11.673a}$$
$$= w \times_k \phi_k\big(\mathtt{proj}_v(\omega_{l,I})\big) \tag{11.673b}$$
$$= w \times_k \phi_k(v_l \omega_I) \tag{11.673c}$$

parce que $\mathtt{proj}_v(\omega_{l,I}) = v_l \omega_I$. □

Proposition 11.271 ([1]).
Soit une application $f\colon V \to E$ de classe C^n. Alors

(1) pour tout n-multiindice I, la dérivée partielle $\partial_I^n f$ existe,

(2) nous avons la formule

$$(d^n f)_a = \sum_{i_1,\dots,i_n} \frac{\partial^n f}{\partial x_{i_1} \dots \partial x_{i_n}}(a) \times_n \phi_n(\omega_{i_1,\dots,i_n}) \tag{11.674}$$

où la somme porte sur tous les multiindices de taille n, le produit $\times_n$ est la définition 11.239 et l'application ϕ_n est donnée dans l'énoncé du lemme 11.269.

Démonstration. Nous faisons tout cela par récurrence. Le cas $n = 1$ est déjà fait par le théorème 11.243 parce que $\phi_1(\omega) = \omega$. Nous supposons que le résultat est démontré pour une valeur k, et nous considérons f de classe C^n avec $n \geqslant k+1$. L'hypothèse de récurrence dit que

$$(d^k f)_a = \sum_I (\partial_I f)(a) \times_k \phi_k(\omega_I) \tag{11.675}$$

est continuement différentiable par rapport à a. Nous pouvons donc utiliser la formule (11.617) pour calculer la différentielle de (11.675) :

$$(d^{k+1} f)_a = \sum_l \frac{\partial (d^k f)}{\partial x_l}(a) \times_1 \omega_l \tag{11.676a}$$
$$= \sum_l \sum_I \big[(\partial_l \partial_I f)(a) \times_k \phi_k(\omega_I)\big] \times_1 \omega_l \tag{11.676b}$$
$$= \sum_{l,I} (\partial_{l,I} f)(a) \times_{k+1} \phi_{k+1}(\omega_{l,I}) \tag{11.676c}$$

où nous avons utilisé le lemme 11.270 pour obtenir (11.676c). □

Proposition 11.272.
Soit une application $f\colon V \to E$ telle que les dérivées partielles $\frac{\partial f}{\partial x_{i_1} \dots \partial x_{i_n}}$ existent et sont continues sur un ouvert $\mathcal{U}$. Alors f est de classe C^n sur $\mathcal{U}$.

Démonstration. Pour $n = 1$, cette proposition est déjà contenue dans le théorème 11.243. Nous pouvons donc directement passer au pas de récurrence. Nous définissons les objets

$$T_k(a) = \sum_I (\partial_I f)(a) \times_k \phi_k(\omega_I) \tag{11.677}$$

où la somme porte sur tous les k-multiindices. Ces applications T_k existent et sont continues par hypothèse.

Le théorème 11.243 nous indique que $T_1(a) = df_a$. Nous supposons maintenant que $(d^k f)_a = T_k(a)$ et nous allons prouver que T_{k+1} est la différentielle de $d^k f$. Pour cela nous devons prouver que

$$\lim_{h\to 0} \frac{T_k(a+h) - T_k(a) - T_{k+1}(a)h}{\|h\|} = 0. \tag{11.678}$$

En utilisant le lemme 11.270 nous pouvons dégrossir le terme $T_{k+1}(a)h$:

$$T_{k+1}(a)h = \sum_{l,I} \big[(\partial_{l,I} f)(a) \times_{k+1} \phi_{k+1}(\omega_{l,I})\big]h \tag{11.679a}$$

$$= \sum_{l,I} \big[(\partial_{l,I} f)(a) \times_k \phi_k(\omega_I)\big] \times_1 \omega_l(h) \tag{11.679b}$$

$$= \sum_{l,I} h_l (\partial_{l,I} f)(a) \times_k \phi_k(\Omega_I) \tag{11.679c}$$

$$= \sum_l h_l \big[\partial_l \sum_I (\partial_I f)(a)\big] \times_k \phi_k(\omega_I). \tag{11.679d}$$

Vu que $\partial_I f$ est une fonction dont les dérivées partielles existent et sont continues, le théorème 11.243 (toujours lui) nous dit qu'elle est différentiable et que

$$d(\partial_I f)_a(h) = \sum_l h_l \partial_l (\partial_I f)(a). \tag{11.680}$$

En mettant la somme sur I tout devant dans (11.679d), nous trouvons

$$T_{k+1}(a)h = \sum_I d(\partial_I f)_a(h) \times_k \phi_k(\omega_I). \tag{11.681}$$

C'est pas beau la vie ? Nous pouvons écrire le numérateur du quotient différentiel (11.678) :

$$T_k(a+h) - T_k(a) - T_{k+1}(a)h = \sum_I \big[(\partial_I f)(a+h) - (\partial_I f)(a)\big] \times_k \phi_k(\omega_I) \tag{11.682a}$$

$$- \sum_I d(\partial_I f)_a(h) \times_k \phi_k(\omega_I) \tag{11.682b}$$

$$= \sum_I \big[(\partial_I f)(a+h) - (\partial_I f)(a) - d(\partial_I f)(a)h\big] \times_l \phi_k(\omega_I). \tag{11.682c}$$

Nous utilisons le fait que l'application $w \mapsto w \times_k \phi_k(\omega_I)$ est linéaire et commute avec la limite, c'est à dire que

$$\lim_{h\to 0} \big(w(h) \times_k \phi_k(\omega_I)\big) = \big(\lim_{h\to 0} w(h)\big) \times_k \phi_k(\omega_I). \tag{11.683}$$

Et bien sûr, la somme sur les multiindices I commute également avec la limite. Donc

$$\lim_{h\to 0} \frac{T_k(a+h) - T_k(a) - T_{k+1}(a)h}{\|h\|} = \sum_I \left(\lim_{h\to 0} \frac{(\partial_I f)(a+h) - (\partial_I f)(a) - d(\partial_I f)_a h}{\|h\|}\right) \times_k \phi_k(\omega_I) \tag{11.684a}$$

$$= 0. \tag{11.684b}$$

Cela prouve que T_{k+1} est bien la différentielle de T_k. Par récurrence, la fonction f est bien n fois continuement différentiable. □

Maintenant que nous avons plein de lemmes et de résultats, il est facile de démontrer un très gros résultat en peu de lignes.

Théorème 11.273.
Soit une application $f\colon V \to E$ entre deux espaces vectoriels normés de dimension finies. Nous avons équivalence entre

(1) f est de classe C^n

(2) les dérivées partielles $\partial_i f$ sont de classe C^{n-1}

(3) les dérivées partielles $\partial_I^n f$ existent et sont continues pour tout multiiindice de longueur n.

Démonstration. En plusieurs implications.

(3) implique (1) C'est la proposition 11.272.

(1) implique (3) C'est la proposition 11.271.

(3) implique (2) En posant $g_i = \partial_i f$, l'hypothèse est que les $\partial_I^{n-1} g_i$ existent et sont continues pour tout multiindices I de longueur $n-1$. En appliquant « (3) implique (1) » à la fonction g_i, la fonction g_i est de classe C^{n-1}.

(2) implique (3) Les fonctions g_i déjà définies sont de classe C^{n-1}. Nous leur appliquons « (1) implique (3) », nous savons que les fonctions $\partial_I^{n-1} g_i$ sont continues, c'est à dire que les $\partial_I \partial_i f$. Vu que tout multiiindice de longueur n peut être écrit sous la forme iI où I est de longueur $n-1$, nous avons prouvé que les $\partial_J^n f$ existent et sont continues pour tout multiindice J de longueur n.

□

11.29.2 Espaces d'applications multilinéaires et identifications

Si V et E sont des espaces vectoriels de dimensions finies, la différentielle de $f\colon V \to E$ est une application $df\colon V \to \mathcal{L}(V,E)$. La différentielle seconde est une application $d(df)\colon V \to \mathcal{L}\big(V,\mathcal{L}(V,E)\big)$ et ainsi de suite.

Une grande difficulté de la manipulation des différentielles d'ordre supérieurs provient de cet emboîtement d'espaces d'applications linéaires. Nous nous attaquons à présent à la description de ces espaces emboîtés[67].

Pour la suite, nous considérerons des espaces vectoriels normés V et E de dimension finie. Nous notons $\mathcal{L}^n(V,E)$ l'espace des applications multilinéaires $V^n \to E$.

Nous introduisons le produit suivant[1] :

$$\begin{aligned} \cdot\ : W \times \mathcal{L}\big(V,\mathcal{L}(V,\mathbb{R})\big) &\to \mathcal{L}\big(V,\mathcal{L}(V,W)\big) \\ \big((w\cdot\psi)(u)\big)v &= \big(\psi(u)v\big)w. \end{aligned} \tag{11.685}$$

Dans la suite, pour économiser des parenthèses et des maux de tête, nous allons noter $\psi(u,v)$ le nombre $\psi(u)v$. Il n'est cependant pas question de dire que ψ est une application bilinéaire. En effet, identifier $\mathcal{L}\big(V,\mathcal{L}(V,W)\big)$ à l'espace des applications bilinéaires $V \times V \to W$ ne sert pas à grand chose pour l'instant parce qu'une telle identification a le prix de devoir prouver que toutes les propriétés des différentielles passent à travers l'identification, tâche qui est à priori (conservation de la difficulté) de la même nature que celle à laquelle nous nous attachons à présent.

Lemme 11.274 ([1]).
Soient deux espaces vectoriels normés V et E ainsi que $\psi \in \mathcal{L}\big(V,\mathcal{L}(V,\mathbb{R})\big)$. Pour tout $a \in E$ nous avons

$$\|a\cdot\psi\|_{\mathcal{L}\big(V,\mathcal{L}(V,E)\big)} = \|\psi\|_{\mathcal{L}\big(V,\mathcal{L}(V,\mathbb{R})\big)}\|a\|_E. \tag{11.686}$$

67. Toutes les constructions, tous les énoncés et les preuves qui suivent sont de l'invention personnelle de l'auteur de ces lignes. Je n'ai trouvé nulle part une source qui s'attaque réellement à le récurrence.

Démonstration. Il s'agit seulement d'un calcul :

$$\begin{aligned}
\|a \cdot \psi\| &= \sup_{\|v\|=1} \|(a \cdot \psi)(v)\|_{\mathcal{L}(V,E)} && (11.687a)\\
&= \sup_{\|v\|=1} \sup_{\|w\|=1} \|(a \cdot \psi)(v)w\|_E && (11.687b)\\
&= \sup_{\|v\|=1} \sup_{\|w\|=1} \|\big(\psi(v)w\big)a\|_E && (11.687c)\\
&= \sup_{\|v\|=1} \sup_{\|w\|=1} |\psi(v)w| \, \|a\|_E && (11.687d)\\
&= \sup_{\|v\|=1} \|\psi(v)\| \, \|a\| && (11.687e)\\
&= \|\psi\|\|a\|. && (11.687f)
\end{aligned}$$

Notez que dans (11.687c), $|\psi(v)w|$ est un simple réel ; c'est pourquoi nous le retrouvons hors de la norme $\|.\|_E$ dans (11.687d), muni d'une simple valeur absolue. □

Lemme 11.275 ([1]).
Soient $\psi \in \mathcal{L}\big(V, \mathcal{L}(V, \mathbb{R})\big)$ et une fonction continue $f\colon V \to E$. Alors la fonction

$$\begin{aligned}
g\colon V &\to \mathcal{L}\big(V, \mathcal{L}(V, E)\big)\\
x &\mapsto f(x) \cdot \psi
\end{aligned} \tag{11.688}$$

est continue.

Démonstration. Pour des raisons de notations, nous allons écrire g_x pour $g(x)$. Cela étant dit nous considérons $a \in E$, une suite $x_k \xrightarrow{E} a$ et nous calculons :

$$\begin{aligned}
\|g_a - g_{x_k}\| &= \sup_{\|u\|=1} \|g_a(u) - g_{x_k}(u)\|_{\mathcal{L}(V,E)} && (11.689a)\\
&= \sup_{\|u\|=1} \sup_{\|v\|=1} \|g_a(u)v - g_{x_k}(u)v\|_E && (11.689b)\\
&= \sup_{\|u\|=1} \sup_{\|v\|=1} \|f(a)\psi(u)v - f(x_k)\psi(u)v\| && (11.689c)\\
&= \sup_{\|u\|=1} \sup_{\|v\|=1} \|f(a) - f(x_k)\|\big|\psi(u)v\big| && (11.689d)\\
&= \big|f(a) - f(x_k)\big| \sup_{u,v} |\psi(u)v| && (11.689e)\\
&= \|f(a) - f(x_k)\| \, \|\psi\|. && (11.689f)
\end{aligned}$$

En prenant la limite $k \to \infty$, et en considérant que f est continue en a, nous obtenons

$$\lim_{k\to\infty} \|g_a - g_{x_k}\| = 0. \tag{11.690}$$

□

Lemme 11.276 ([1]).
Soient une application différentiable $f\colon V \to E$ ainsi que $\psi \in \mathcal{L}\big(V, \mathcal{L}(V, \mathbb{R})\big)$. Soit

$$\begin{aligned}
g\colon V &\to \mathcal{L}\big(V, \mathcal{L}(V, E)\big)\\
x &\mapsto f(x) \cdot \psi.
\end{aligned} \tag{11.691}$$

Alors g est différentiable et pour tout $a \in V$ nous avons

$$dg_a(h) = df_a(h) \cdot \psi. \tag{11.692}$$

Notons que nous n'avons pas $dg_a = df_a \cdot \psi$. En effet, $df_a \in \mathcal{L}(V, W)$, de telle sorte que $df_a \cdot \psi \in \mathcal{L}\big(V, \mathcal{L}\big(V, \mathcal{L}(V, W)\big)\big)$. Les espaces ne s'emboîtent pas dans le bon ordre.

Démonstration. Il s'agit de vérifier que $h \mapsto df_a(h) \cdot \psi$ vérifie la condition de la définition 10.127. En utilisant le fait que $(a+b) \cdot \psi = a \cdot \psi + b \cdot \psi$ ainsi que le lemme 11.274 nous écrivons

$$\frac{\|f(a+h) \cdot \psi - f(a) \cdot \psi - df_a(h) \cdot \psi\|}{\|h\|} = \|\frac{f(a+h) - f(a) - df_a(h)}{h} \cdot \psi\| \tag{11.693a}$$

$$= \|\frac{f(a+h) - f(a) - df_a(h)}{h}\|\|\psi\|. \tag{11.693b}$$

Vu que f est différentiable en a et que df_a est la différentielle, nous avons bien

$$\lim_{h \to 0} \|\frac{f(a+h) - f(a) - df_a(h)}{h}\|\|\psi\| = 0. \tag{11.694}$$

□

Démonstration. Nous faisons (1) par récurrence. D'abord $\dim(E_0) = \dim(E)$ et ensuite

$$\dim(E_{k+1}) = \dim \mathcal{L}(V, E_k) = \dim(V) \dim(E_k) = \dim(V)^{n+1} \dim(E). \tag{11.695}$$

Pour (2), si $\{e_i\}$ est une base de V, un élément $\omega \in \mathcal{L}^n(V, E)$ est déterminé par les valeurs de $\omega(e_{i_1}, \dots, e_{i_n})$ qui peuvent être n'importe quel vecteur de E. Donc la dimension est $\dim(V)^n \dim(E)$. □

Lemme 11.277 ([1]).
Soit $n \in \mathbb{N}$ nous définissons par récurrence

$$\begin{aligned} \psi_{n,0} \colon E_n &\to E_n \\ \alpha &\mapsto \alpha. \end{aligned} \tag{11.696}$$

et

$$\begin{aligned} \psi_{n,k} \colon E_n &\to \mathcal{L}^k(V, E_{n-k}) \\ \psi_{n,k}(\alpha)(v_1, \dots, v_k) &= \big(\psi_{n,k-1}(\alpha)(v_1, \dots, v_{k-1})\big)v_k. \end{aligned} \tag{11.697}$$

L'application

$$\psi_{n,n} \colon E_n \to \mathcal{L}^n(V, E) \tag{11.698}$$

est un isomorphisme isométrique.

Démonstration. Nous allons démontrer par récurrence sur k que tous les $\psi_{n,k}$ sont des isomorphismes isométriques. Pour $k = 1$ c'est évident parce que $\psi_{n,1}$ est l'identité.

Injective Soient $\alpha \in E_n$ tels que $\psi_{n,k+1}(\alpha) = O$. Cela signifie que pour tout $v_1, \dots, v_{k+1}$ nous avons

$$\psi_{n,k}(\alpha)(v_1, \dots, v_k)v_{k+1} = 0. \tag{11.699}$$

c'est à dire $\psi_{n,k}(\alpha)(v_1, \dots, v_k = 0)$. Vu que $\psi_{n,k}$ est injective (hypothèse de récurrence), nous avons $\alpha = 0$.

Surjective Soit $\omega \in \mathcal{L}^{k+1}(V, E_{n-k})$; nous cherchons $\alpha \in E_n$ tel que $\psi_{n,k+1}(\alpha) = \omega$. Cette condition s'exprime

$$\psi_{n,k+1}(\alpha)(v_1, \dots, v_{k+1}) = \psi_{n,k}(\alpha)(v_1, \dots, v_k)v_{k+1} \overset{!}{=} \omega(v_1, \dots, v_{k+1}). \tag{11.700}$$

Notez que

$$\psi_{n,k}(\alpha) \in \mathcal{L}^k(V, E_{n-k}) = \mathcal{L}^k\big(V, \mathcal{L}(V, E_{n-k-1})\big). \tag{11.701}$$

En considérant $\sigma \in \mathcal{L}^k\big(V, \mathcal{L}(V, E_{n-k-1})\big)$ donné par

$$\sigma(v_1, \dots, v_k)v_{k+1} = \omega(v_1, \dots, v_{k+1}), \tag{11.702}$$

il existe (hypothèse de récurrence sur k) un $\alpha \in E_n$ tel que $\psi_{n,k}(\alpha) = \sigma$.
Pour ce α, la condition (11.700) est satisfaite.

Isométrique Encore une fois par récurrence. Soit $\alpha \in E_n$. Nous avons

$$\begin{aligned}
\|\psi_{n,k}(\alpha)\|_{\mathcal{L}^k(V,E_{n-k})} &= \sup_{\|v_i\|=1} \|\psi_{n,k}(\alpha)(v_1,\dots,v_k)\|_{E_{n-k}} && (11.703a)\\
&= \sup_{\|v_i\|=1} \|\psi_{n,k-1}(\alpha)(v_1,\dots,v_{k-1})v_k\|_{E_{n-k}} && (11.703b)\\
&= \sup_{\substack{\|v_i\|=1\\ i=1,\dots,k-1}} \|\psi_{n,k-1}(v_1,\dots,v_{k-1})\|_{\mathcal{L}(V,E_{n-k-1})} && (11.703c)\\
&= \|\psi_{n,k-1}(\alpha)\| && (11.703d)\\
&= \|\alpha\|. && (11.703e)
\end{aligned}$$

La dernière égalité est l'hypothèse de récurrence. Notez la subtile utilisation du lemme 7.106 qui permet de donner un sens aux supremums, grace au fait que $\{v \in V \text{ tel que } \|v\| = 1\}$ est compact.

□

L'application $\psi_{n,n}$ est l'application inverse de l'isomorphisme ϕ_n donné dans le lemme 11.269.

Lemme 11.278.
Nous avons $\psi_{n,n} = \phi_n^{-1}$.

Démonstration. Pour nous échauffer nous posons $\omega \in \mathcal{L}^n(V,\mathbb{R})$, et nous calculons

$$\begin{aligned}
\psi_{n,2}\big(\phi_n(\omega)\big)(v_1,v_2) &= \Big(\psi_{n,1}\big(\phi_n(\omega)\big)v_1\Big)v_2 && (11.704a)\\
&= \big(\phi_n(\omega)v_1\big)v_2 && (11.704b)\\
&= \phi_{n-1}\big(\mathtt{proj}_{v_1}(\omega)\big) && (11.704c)\\
&= \phi_{n-2}\big(\mathtt{proj}_{v_2}\mathtt{proj}_{v_1}(\omega)\big). && (11.704d)
\end{aligned}$$

Cela étant dit, nous allons prouver ceci par récurrence :

$$\psi_{n,k}\big(\phi_n(\omega)\big)(v_1,\dots,v_k) = \phi_{n-k}\Big(\prod_{i=1}^{k}\mathtt{proj}_{v_i}(\omega)\Big). \tag{11.705}$$

Notez l'ordre du produit des projections. En ce qui concerne cet ordre, pour fixer les idées voici un exemple :

$$\mathtt{proj}_{v_2}\mathtt{proj}_{v_1}(\omega) = \big(\mathtt{proj}_{v_1}(\omega)\big)(v_2) = \omega(v_1,v_2). \tag{11.706}$$

Faisons maintenant la récurrence.

Pour $k = 2$ C'est le calcul (11.704).

Pour $k+1$ C'est encore un calcul, en faisant attention à l'ordre dans lequel viennent les projections :

$$\begin{aligned}
\psi_{n,k+1}\big(\phi_n(\omega)\big)(v_1,\dots,v_{k+1}) &= \Big(\psi_{n,k}\big(\phi_n(\omega)\big)(v_1,\dots,v_k)\Big)v_{k+1} && (11.707a)\\
&= \Big(\phi_{n-k}\big(\prod_{i=1}^{k}\mathtt{proj}_{v_i}(\omega)\big)\Big)v_{k+1} && (11.707b)\\
&= \phi_{n-k-1}\big(\mathtt{proj}_{v_{k+1}}\prod_{i=1}^{k}\mathtt{proj}_{v_i}(\omega)\big) && (11.707c)\\
&= \phi_{n-(k+1)}\big(\prod_{i=1}^{k+1}\mathtt{proj}_{v_i}(\omega)\big). && (11.707d)
\end{aligned}$$

Il ne reste qu'à écrire la formule démontrée avec $k = n$:

$$\psi_{n,n}\big(\phi_n(\omega)\big)(v_1,\dots,v_n) = \phi_{n-n}\big(\prod_{i=1}^{n}\mathtt{proj}_{v_i(\omega)}\big) = \omega(v_1,\dots,v_n). \tag{11.708}$$

Donc nous avons bien que $\psi_{n,n}\big(\phi_n(\omega)\big) = \omega$. □

Lemme 11.279.
Soient deux espaces vectoriels normés V et E. Nous considérons une application linéaire $A\colon V \to E$, une forme $\omega \in \mathcal{L}^n(V,\mathbb{R})$ ainsi que la forme

$$\begin{aligned}\alpha\colon V &\to E_n\\ u &\mapsto A(u)\psi^{-1}(\omega).\end{aligned} \tag{11.709}$$

Alors

$$\psi_{n+1,n+1}(\alpha)(v_1,\dots,v_{n+1}) = A(v_1)\omega(v_2,\dots,v_{n+1}). \tag{11.710}$$

Démonstration. Nous prouvons la formule suivante par récurrence :

$$\psi_{n+1,k}(\omega)(v_1,\dots,v_k) = A(v_1)\phi_{n-k+1}\big(\prod_{i=2}^{k}\mathtt{proj}_{v_i}(\omega)\big). \tag{11.711}$$

Nous commençons la récurrence avec $k = 2$, en tenant compte que $\psi^{-1} = \phi_n$ (lemme 11.278) :

$$\begin{aligned}\psi_{n+1,2}(\alpha)(v_1,v_2) &= \big(\psi_{n+1,1}(\alpha)v_1\big)v_2 &(11.712a)\\ &= \alpha(v_1)v_2 &(11.712b)\\ &= A(v_1)\big(\psi^{-1}(\omega)v_2\big) &(11.712c)\\ &= A(v_1)\big(\phi_n(\omega)v_2\big) &(11.712d)\\ &= A(v_1)\phi_{n-1}\big(\mathtt{proj}_{v_2}(\omega)\big). &(11.712e)\end{aligned}$$

C'est bon pour $k = 2$. Nous passons à la récurrence :

$$\begin{aligned}\psi_{n+1,k+1}(\omega)(v_1,\dots,v_{k+1}) &= \big(\phi_{n+1,k}(\omega)(v_1,\dots,v_k)\big)v_{k+1} &(11.713a)\\ &= A(v_1)\Big(\phi_{n-k+1}\big(\prod_{i=2}^{k}\mathtt{proj}_{v_i}\omega\big)\Big)v_{k+1} &(11.713b)\\ &= A(v_1)\phi_{n-k}\big(\mathtt{proj}_{v_{k+1}}\prod_{i=2}^{k}\mathtt{proj}_{v_i}\omega\big) &(11.713c)\\ &= A(v_1)\phi_{n-(k+1)}\big(\prod_{i=2}^{k+1}\mathtt{proj}_{v_i}\omega\big). &(11.713d)\end{aligned}$$

La récurrence est terminée. Nous écrivons la formule pour $k = n+1$ en tenant compte de $\phi_0 = \mathrm{Id}$ pour terminer la preuve de ce lemme :

$$\psi_{n+1,n+1}(\omega)(v_1,\dots,v_{n+1}) = A(v_1)\phi_0\big(\prod_{i=2}^{n+1}\mathtt{proj}_{v_i}\omega\big) = A(v_1)\omega(v_2,\dots,v_{n+1}). \tag{11.714}$$

□

11.29.3 Identification des espaces d'applications multilinéaires (vieux)

Problèmes et choses à faire

Les choses dites en 11.29.3 sont obsolètes par le lemme 11.268. Il faut soit les retirer soit les déplacer.

La différentielle de la différentielle de f est notée

$$d(df)(a) = d^2f(a),$$

et est une application de U dans $\mathcal{L}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n))$. Comme on a vu dans la proposition 10.47, l'espace $\mathcal{L}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n))$ est isométriquement isomorphe à l'espace $\mathcal{L}(\mathbb{R}^m \times \mathbb{R}^m, \mathbb{R}^n)$. On verra comment cette propriété est utilisé dans l'exemple 11.281.

Soient V et W deux espaces vectoriel normés de dimension finie et $\mathcal{O}$ un ouvert autour de $x \in V$. D'une part l'espace des applications linéaires $\mathcal{L}(V, W)$ est lui-même un espace vectoriel normé de dimension finie, et on peut identifier $\mathcal{L}\big(V, \mathcal{L}^{(k)}(V, W)\big)$ avec $\mathcal{L}^{(k+1)}(V, W)$, ce qui nous permet de dire que la k^{e} différentielle est une application

$$d^k f \colon \mathcal{O} \to \mathcal{L}^{(k)}(V, W). \tag{11.715}$$

Plus précisément, l'identification se fait de la façon suivante : si $\omega \in \mathcal{L}\big(V, \mathcal{L}^{(k)}(V, W)\big)$, alors ω vu dans $\mathcal{L}^{(k+1)}(V, W)$ est définie par

$$\omega(u_1, \dots, u_{k+1}) = \omega(u_1)(u_2, \dots, u_{k+1}). \tag{11.716}$$

Cela étant posé nous pouvons donner des définitions.

11.29.4 Fonctions différentiables plusieurs fois

Définition 11.280.
*Un C^k-**difféomorphisme** est une application inversible de classe C^k dont l'inverse est également de classe C^k.*

Exemple 11.281.
Soit $B : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n$ une application bilinéaire. On définit $f : \mathbb{R}^m \to \mathbb{R}^n$ par $f(x) = B(x, x)$. Le lemme 11.217 nous dit que B est différentiable. Cela implique la différentiabilité de f. Pour trouver la différentielle de la fonction f, nous écrivons $f = B \circ s$ où $s \colon \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m$ est l'application $s(x) = (x, x)$. En utilisant la règle de différentiation de fonctions composées,

$$df(a) = dB\big(s(a)\big) \circ ds(a). \tag{11.717}$$

Mais $ds(a).u = (u, u)$ parce que $s(a + h) - s(a) - (h, h) = 0$. Par conséquent,

$$df(a).u = dB\big(s(a)\big)(u, u) = B(u, a) + B(a, u) \tag{11.718}$$

où nous avons utilisé la formule du lemme 11.217. La formule (11.718) peut être écrite sous la forme compacte

$$df(a) = B(\ \cdot\ , a) + B(a,\ \cdot\ ) \tag{11.719}$$

La fonction $df(a)$ ainsi écrite est linéaire par rapport à a, donc différentiable. En outre elle coïncide avec sa différentielle, comme on a vu dans le lemme 11.189, au sens que la différentielle de df au point a sera l'application que à chaque x dans $\mathbb{R}^m$ associe l'application linéaire $B(x,\ \cdot\ ) + B(\ \cdot\ , x)$. On voit bien que $d^2 f$ au point a est une application de $\mathbb{R}^m$ vers l'espace des applications linéaires $\mathcal{L}(\mathbb{R}^m, \mathbb{R}^n)$. On peut utiliser d'autre part l'isomorphisme des espaces $\mathcal{L}(\mathbb{R}^m, \mathcal{L}(\mathbb{R}^m, \mathbb{R}^n))$ et $\mathcal{L}(\mathbb{R}^m \times \mathbb{R}^m, \mathbb{R}^n)$ et dire que, une fois que a est fixé, l'application $d^2 f(a)$ est une application bilinéaire sur $\mathbb{R}^m \times \mathbb{R}^m$. On écrit alors $d^2 f(a)(x, y) = B(x, y) + B(y, x)$. △

11.29.5 Différentielle seconde, fonction de classe C^2

La différentielle seconde dans l'exemple 11.281 est symétrique, c'est-à-dire que $d^2 f(a)(x_1, x_2) = d^2 f(a)(x_2, x_1)$. En fait toute différentielle seconde est symétrique.

Théorème 11.282 (Schwarz)**.**
Soit U un ouvert de $\mathbb{R}^m$ et $f : U \subset \mathbb{R}^m \to \mathbb{R}^n$ une fonction de classe $\mathcal{C}^2$. Alors, pour tout couple i, j d'indices dans $\{1, \dots, m\}$ et pour tout point a dans U, on a

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a).$$

Démonstration. Pour simplifier nous nous limitons ici au cas $m = 2$. Soit (h, g) un vecteur fixé dans $\mathbb{R}^2$. Pour tout $v = (x, y)$ dans $\mathbb{R}^2$ on note

$$\begin{aligned}\Delta_h f(v) &= f(v + he_1) - f(v) = f(x + h, y) - f(x, y),\\ \Delta_g f(v) &= f(v + ge_2) - f(v) = f(x, y + g) - f(x, y),\end{aligned} \tag{11.720}$$

Nous avons

$$\begin{aligned}\Delta_g\Delta_h f(v) &= \left(f(x + h, y + g) - f(x, y + g)\right) - \left(f(x + h, y) - f(x, y)\right),\\ \Delta_h\Delta_g f(v) &= \left(f(x + h, y + g) - f(x + h, y)\right) - \left(f(x, y + g) - f(x, y)\right),\end{aligned} \tag{11.721}$$

donc,

$$\frac{1}{g}\Delta_g\left(\frac{1}{h}\Delta_h f(v)\right) = \frac{1}{h}\Delta_h\left(\frac{1}{g}\Delta_g f(v)\right). \tag{11.722}$$

On utilise alors le théorème des accroissements finis 11.130

$$\frac{1}{h}\Delta_h f(v) = \frac{1}{h}\big(f(x + h, y) - f(x, y)\big) = \frac{1}{h}\partial_1 f(x + t_1 h, y)h = \partial_1 f(x + t_1 h, y), \tag{11.723}$$

pour un certain t_1 dans $]0, 1[$. De même on obtient

$$\frac{1}{g}\Delta_g f(v) = \partial_2 f(x, y + t_2 g),$$

pour un certain t_2 dans $]0, 1[$. Alors

$$\frac{1}{g}\Delta_g\big(\partial_1 f(x + t_1 h, y)\big) = \frac{1}{h}\Delta_h\big(\partial_2 f(x, y + t_2 g)\big). \tag{11.724}$$

En appliquant encore une fois le théorème des accroissements finis on a

$$\partial_2\partial_1 f(x + t_1 h, y + s_1 g) = \partial_1\partial_2 f(x + s_2 h, y + t_2 g). \tag{11.725}$$

Il suffit maintenant de passer à la limite pour $(h, g) \to (0, 0)$ et de se souvenir du fait que f est $\mathcal{C}^2$ seulement si ses dérivées partielles secondes sont continues [68] pour avoir $\partial_2\partial_1 f(v) = \partial_1\partial_2 f(v)$. □

Si f est deux fois différentiable $d^2 f(a)$ est l'application bilinéaire associée avec la matrice symétrique

$$H_f(a) = \begin{pmatrix} \partial_1^2 f(a) & \dots & \partial_1\partial_m f(a)\\ \vdots & \ddots & \vdots\\ \partial_1\partial_m f(a) & \dots & \partial_1^2 f(a), \end{pmatrix} \tag{11.726}$$

Cette matrice est dite la matrice **hessienne** de f.

Exemple 11.283.
Montrons qu'il n'existe pas de fonctions f de classe $\mathcal{C}^2$ telles que

$$\begin{cases} \partial_x f(x, y) = 5\sin x & \text{(11.727a)}\\ \partial_y(x, y) = 6x + y. & \text{(11.727b)}\end{cases}$$

Ceci est vite fait en appliquant le théorème de Schwarz, 11.282 ; ce que nous trouvons est

$$\partial_y(\partial_x f) = 0$$

alors que

$$\partial_x(\partial_y f) = 6. \tag{11.728}$$

Vu que $\partial_x(\partial_y f) \neq \partial_y(\partial_x f)$, le théorème 11.282 dit que f ne peut pas être de classe C^2. △

68. Théorème 11.273.

Soit une fonction de classe C^2 $f\colon V \to \mathbb{R}$ où V est un espace vectoriel de dimension $n < \infty$. Nous avons

$$f\colon V \to \mathbb{R} \tag{11.729a}$$
$$df\colon V \to \mathcal{L}(V,\mathbb{R}) \tag{11.729b}$$
$$d^2f\colon V \to \mathcal{L}\Big(V,\mathcal{L}(V,\mathbb{R})\Big), \tag{11.729c}$$

avec, en suivant les différentes formules du lemme 11.196,

$$df_a(u) = \frac{d}{dt}\Big[f(v+tu)\Big]_{t=0} \tag{11.730}$$

et

$$(d^2f)_a(u) = \frac{d}{dt}\Big[df_{v+tu}\Big]_{t=0} \tag{11.731}$$

pour tout $a,u \in V$. Notons que dans le deuxième cas, il s'agit d'une limite dans $\mathcal{L}(V,\mathbb{R})$. Si $\dim(V) = n$, alors $\dim \mathcal{L}(V,\mathbb{R}) = n$ et avec un choix de base, nous pouvons trouver une matrice $n \times n$ pour $(d^2f)_a$.

Soit une base $\{e_i\}$ de V et la base duale $\{e_i^*\}$ de $\mathcal{L}(V,\mathbb{R})$. Nous allons chercher la matrice de $(d^2f)_a$ pour ces bases. L'élément de matrice

$$\big[(d^2f)_a\big]_{ij} \tag{11.732}$$

est la composante e_j^* de $(d^2f)_a$ appliqué à e_i. Trouver cette composante e_j^* revient à appliquer l'élément $(d^2f)_a e_i$ de $\mathcal{L}(V,\mathbb{R})$ à e_j. Le calcul est donc :

$$\big[(d^2f)_a\big]_{ij} = \big((d^2f)_a e_i\big)(e_j) \tag{11.733a}$$
$$= \frac{d}{dt}\Big[df_{a+te_i}(e_j)\Big]_{t=0} \tag{11.733b}$$
$$= \frac{d}{dt}\Big[\frac{d}{ds}\Big[f(a+te_i+se_j)\Big]_{s=0}\Big]_{t=0} \tag{11.733c}$$
$$= \frac{\partial^2 f}{\partial x_i \partial x_j}(a). \tag{11.733d}$$

Attention : le passage à (11.733b) n'est pas une trivialité. Le fait est que si $t \mapsto A(t)$ est une application continue $\mathbb{R} \to \mathcal{L}(V,\mathbb{R})$ alors

$$\lim_{t\to 0}\big(A(t)v\big) = \big(\lim_{t\to 0} A(t)\big)v. \tag{11.734}$$

Donc la matrice de d^2f est la matrice des dérivées secondes. Il s'agit d'une matrice symétrique par le théorème de Schwarz 11.282.

11.284.
Si $a \in v$, nous pouvons aussi voir $(d^2f)_a$ comme une forme bilinéaire sur V grâce à la proposition 10.47. Si $u,v \in V$ nous notons

$$(d^2f)_a(u,v) = (d^2f)_a(u)v. \tag{11.735}$$

À droite, il s'agit de la définition réelle de d^2f sans abus de notations, et à gauche, il s'agit d'une notation. Cette application bilinéaire $(d^2f)_a \in \mathcal{L}^{(2)}(V,\mathbb{R})$ a pour matrice symétrique la matrice des dérivées secondes calculées en a.

Exemple 11.285.
Voyons comment la différentielle seconde fonctionne entre deux espaces vectoriels. Soient deux espaces vectoriels de dimension finie V et W. Pour que les choses soient claires, nous avons :

$$f\colon V \to W \tag{11.736a}$$
$$df\colon V \to \mathcal{L}(V,W) \tag{11.736b}$$
$$d^2f\colon V \to \mathcal{L}\Big(V,\mathcal{L}(V,W)\Big). \tag{11.736c}$$

Si $a \in V$, alors $(d^2 f)_a$ est une application $V \to \mathcal{L}(V, W)$. Il faut donc l'appliquer à $u \in V$ et ensuite à $v \in V$ pour obtenir un élément de W :

$$(d^2 f)_a(u)v = \frac{d}{dt}\Big[df_{a+tu}\Big]_{t=0} v \tag{11.737a}$$
$$= \frac{d}{dt}\Big[df_{a+tu}(v)\Big]_{t=0} \tag{11.737b}$$
$$= \frac{d}{dt}\Big[\frac{d}{ds}\Big[f(a+tu+sv)\Big]_{s=0}\Big]_{t=0} \tag{11.737c}$$
$$= \frac{\partial^2 f}{\partial u \partial v}(a). \tag{11.737d}$$

Par conséquent nous voyons

$$\begin{aligned} d^2 f \colon V &\to \mathcal{L}^{(2)}(V, W) \\ d^2 f_a(u, v) &= \frac{\partial^2 f}{\partial u \partial v}(a). \end{aligned} \tag{11.738}$$

Dans le cas d'une fonction $f \colon \mathbb{R} \to \mathbb{R}$, nous avons une seule direction et par linéarité de (11.738) par rapport à u et v, nous avons

$$d^2 f_a(u, v) = f''(a)uv \tag{11.739}$$

où les produits sont des produits usuels dans $\mathbb{R}$ et f'' est la dérivée seconde usuelle. △

Tout ceci est un peu résumé dans la proposition suivante.

Proposition 11.286.
Soit une fonction $f \colon \mathbb{R}^n \to \mathbb{R}$ de classe C^2. Alors en désignant par $H_a f$ sa matrice hessienne au point a nous avons

$$(d^2 f)_a(u, v) = \frac{\partial^2 f}{\partial u \partial v}(a) = \langle (H_a f)u, v \rangle \tag{11.740}$$

pour tout $u, v \in \mathbb{R}^n$.

Démonstration. La première égalité est l'équation (11.739) déjà faite. Pour la seconde, il faut se rappeler du lien entre dérivée partielle et dérivée directionnelle, donné en le lemme 11.196. En particulier ici nous avons

$$\frac{\partial^2 f}{\partial u \partial v} = \sum_{kl} \frac{\partial^2 f}{\partial x_k \partial x_l}(a) u_k v_l = \langle (H_a f)u, v \rangle. \tag{11.741}$$

□

En particulier, la matrice hessienne $H_a f$ est symétrique et donc diagonalisable (théorème spectral 9.150). Si e_i est un vecteur propre unitaire pour la valeur propre λ_i nous avons

$$(d^2 f)_a(e_i, e_i) = \langle (H_a f)e_i, e_i \rangle = \lambda_i \langle e_i, e_i \rangle = \lambda. \tag{11.742}$$

Enfin pour celles qui aiment les notations matricielles de tout poil, il y a cette façon-ci d'écrire :

$$(d^2 f)_a(\alpha, \beta) = \begin{pmatrix} \alpha & \beta \end{pmatrix} \begin{pmatrix} \partial_x^2 f(a) & \partial_{xy}^2 f(a) \\ \partial_{xy}^2 f(a) & \partial_y^2 f(a) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}. \tag{11.743}$$

11.29.6 Ordre supérieur

Intuitivement, une fonction est de classe C^p si elle est p fois continûment différentiable. Nous posons la définition suivante.

Définition 11.287.
*Une fonction $f \colon E \to F$ est **de classe** C^0 si elle est continue.*

*Nous disons que la fonction $f \colon E \to F$ est **de classe** C^p si elle est différentiable et si sa différentielle $df \colon E \to \mathcal{L}(E, F)$ est de classe C^{p-1}.*

Ce qui est terrible avec les différentielles d'ordre supérieurs, c'est l'emboîtement des espaces. Pour une fonction $f\colon E \to F$, nous allons souvent poser

$$V_0 = F \tag{11.744a}$$

$$V_{k+1} = \mathcal{L}(E, V_k), \tag{11.744b}$$

de telle sorte à avoir

$$df\colon E \to \mathcal{L}(E, F) = V_1 \tag{11.745}$$

et

$$d^2 f\colon E \to \mathcal{L}(E, V_1) = V_2, \tag{11.746}$$

ce qui donne en général

$$d^k f\colon E \to \mathcal{L}(E, v_{k-1}) = V_k. \tag{11.747}$$

La proposition suivante lie une bonne fois pour toute dérivée et différentielle dans le cadre de fonctions $\mathbb{R} \to \mathbb{R}$.

Proposition 11.288 ([1]).
Une fonction $f\colon \mathbb{R} \to \mathbb{R}$ est de classe C^p si et seulement si elle est p fois continûment dérivable.

Démonstration. Nous commençons par poser un certain nombre de notations. Comme souvent nous posons $V_0 = \mathbb{R}$ et

$$V_{k+1} = \mathcal{L}(\mathbb{R}, V_k). \tag{11.748}$$

De plus nous considérons $M_1 \in \mathcal{L}(\mathbb{R}, \mathbb{R})$ donnée par $M_1(t) = t$. Et par récurrence

$$M_{k+1}(t) = tM_k. \tag{11.749}$$

Nous avons $M_1 \in V_1$ et $M_k\colon \mathbb{R} \to V_{k-1}$, c'est-à-dire $M_k \in V_k$.

Les formules que nous allons prouver sont que d'une part,

$$df_a = f'(a)M_1. \tag{11.750}$$

et que d'autre part, plus généralement,

$$(d^k f)_a = f^{(k)}(a)M_k. \tag{11.751}$$

En plusieurs parties et par récurrence.

Si f est continûment dérivable, alors f est C^1 Nous montrons que l'application T donnée par $T(h) = hf'(a)$ vérifie la condition pour être la différentielle de f en a :

$$\frac{f(a+h) - f(a) - f'(a)h}{\|h\|} = \frac{f(a+h) - f(a)}{\|h\|} - 1'_h f(a). \tag{11.752}$$

où nous avons noté 1_h le vecteur unité dans la direction de h, c'est-à-dire $1_h = h/\|h\|$. Vu que $h \in \mathbb{R}$, c'est simplement

$$1_h = \begin{cases} 1 & \text{si } h > 0 \\ -1 & \text{si } h < 0 \end{cases} \tag{11.753}$$

et nous ne définissons pas 1_h si $h = 0$.

C'est le moment de prendre la limite de (11.752) pour $h \to 0^+$ et $h \to 0^-$ séparément. Lorsque $h \to 0^+$, nous avons $\|h\| = h$ et $1_h = h$. Vu que f est supposée dérivable, la limite du quotient existe et vaut $f'(a)$. Donc le tout a une limite nulle :

$$\lim_{h\to 0^+} \frac{f(a+h) - f(a) - f'(a)h}{\|h\|} = \lim_{h\to 0^+} \frac{f(a+h) - f(a)}{h} -' f(a) = 0. \tag{11.754}$$

En ce qui concerne la limite $h \to 0^-$, nous avons $\|h\| = -h$ et $1_h = -1$, et à nouveau une limite nulle. La proposition 11.37 nous permet alors de conclure que la limite existe et est nulle. Les limites à gauche et à droite étant nulles, la limite existe et est nulle par la proposition 11.37.

Si $f^{(p)}$ est continue alors d^pf aussi Nous passons à la récurrence de notre preuve. Sous l'hypothèse que $f^{(p)}$ existe et est continue, nous allons montrer que d^pf existe, est continue et vaut

$$(d^pf)_a = f^{(p)}(a)M_p. \tag{11.755}$$

Soit $k < p$ tel que ce soit bon (pour $k = 1$ c'est déjà fait). Nous avons $(d^kf)_a = f^{(k)}(a)M_k$, et pour prouver que $(d^{k+1}f)_a = f^{(k+1)}(a)M_{k+1}$ nous l'y mettons dans la définition de la différentielle. Nous avons :

$$\frac{(d^kf)_{a+h} - (d^kf)_a - f^{(k+1)}(a)M_{k+1}(h)}{\|h\|} = \frac{f^{(k)}(a+h)M_k - f^{(k)}(a)M_k - hf^{(k+1)}(a)M_k}{\|h\|}. \tag{11.756}$$

La limite $h \to 0$ est une limite dans V_k, et elle se traite comme précédemment. Elle vaut zéro parce que $f^{(k+1)}$ est la dérivée de $f^{(k)}$. Cela justifie les faits que d^kf est différentiable en a et que la différentielle est donné par la formule voulue.

Par hypothèse, $k + 1 \leqslant p$, donc $f^{(k+1)}$ est continue. Par conséquent l'application $a \mapsto f^{(k+1)}(a)M_{k+1}$ est continue.

Si f est de classe C^1 alors f' existe et est continue Dire que f est de classe C^1 revient à dire que la différentielle $df\colon \mathbb{R} \to \mathcal{L}(\mathbb{R}, \mathbb{R})$ existe et est continue. Soyons conscient que $df_a(\epsilon) = \epsilon df_a(1)$ et calculons

$$\frac{f(a+\epsilon) - f(a) - df_a(\epsilon)}{\epsilon} = \frac{f(a+\epsilon) - f(a)}{\epsilon} - df_a(1). \tag{11.757}$$

La définition de la différentielle est que la limite de cela pour $\epsilon \to 0$ soit nulle. Cela implique que la limite suivante existe et vaut

$$\lim_{\epsilon \to 0} \frac{f(a+\epsilon) - f(a)}{\epsilon} = df_a(1). \tag{11.758}$$

Nous avons prouvé que $f'(a) = df_a(1)$.

La fonction $a \mapsto df_a$ est continue. Pouvons-nous en déduire que f' est continue ? Nous avons

$$f' = ev_1 \circ df \tag{11.759}$$

où ev_1 est l'application d'évaluation dont le lemme 10.11 a déjà donné la continuité. Donc f' est continue comme composée d'applications continues.

f est C^p. Récurrence Nous supposons que f est de classe C^p, et nous allons montrer par récurrence que $f^{(k)}$ existe et est continue pour tout $k \leqslant p$. Posons exactement l'énoncé de notre récurrence.

Pour $k = 1$ c'est fait. Nous supposons que la formule soit correcte pour un certain $k \leqslant p$ et nous y allons pour $k + 1$. Nous avons

$$\frac{(d^kf)(a+h) - (d^kf)(a) - f^{(k+1)}(a)M_{k+1}(h)}{\|h\|} = \frac{\big[f^{(k)}(a+h) - f^k(a) - hf^{(k+1)}(a)\big]M_k}{\|h\|} \tag{11.760a}$$

$$= \Big[\frac{f^{(k)}(a+h) - f^{(k)}(a)}{\|h\|} - 1_hf^{(k+1)}(a)\Big]M_k. \tag{11.760b}$$

où nous avons aussi tenu compte que $M_{k+1}(h) = hM_k$.

C'est le moment de calculer séparément les limites $h \to 0^+$ et $h \to 0^-$. Cela fonctionne comme toutes les autres fois.

□

Soit une fonction $f\colon \mathbb{R}^n \to \mathbb{R}$ différentiable l fois. L'application

$$d^l f\colon \mathbb{R}^n \to \mathcal{L}\Big(\mathbb{R}^n, \mathcal{L}\big(\mathbb{R}^n, \mathcal{L}(\dots)\big)\Big) \tag{11.761}$$

au point x appliquée à $v^{(1)}$ appliquée au point $v^{(2)}$, ..., appliquée à $v^{(l)}$ est notée

$$(d^l f)_x(v^{(1)}, \dots, v^{(l)}) \in \mathbb{R}. \tag{11.762}$$

Proposition 11.289.
Soit une fonction $f\colon \mathbb{R}^n \to \mathbb{R}$ différentiable l fois. Avec la notation (11.762) nous avons

$$(d^l f)_x(v^{(1)}, \dots v^{(l)}) = \sum_{k_1,\dots,k_l} v^{(1)}_{k_1} \dots v^{(l)}_{k_l} \frac{\partial^l f}{\partial x_{k_1} \dots \partial x_{k_l}}(x). \tag{11.763}$$

Démonstration. La preuve se fait par récurrence sur l, en sachant que la formule est déjà vraie pour $l = 1$ et $l = 2$. Si la formule est valable pour l, nous avons

$$(d^{l+1} f)_x(v^{(1)}, \dots, v^{(l+1)}) = \frac{d}{dt}\Big[(d^l)_{x+tv^{(l+1)}}(v^{(1)}, \dots, v^{(l)})\Big]_{t=0} \tag{11.764a}$$

$$= \sum_{k_1 \dots k_l} v^{(1)}_{k_1} \dots v^{(l)}_{k_l} \frac{d}{dt}\Big[\frac{\partial^l f}{\partial x_1 \dots \partial x_l}(x + tv^{l+1})\Big]_{t=0} \tag{11.764b}$$

$$= \sum_{k_1 \dots k_l} v^{(1)}_{k_1} \dots v^{(l)}_{k_l} \sum_i \frac{\partial}{\partial x_i} \frac{\partial^l f}{\partial x_{k_1} \dots \partial x_k}(x). \tag{11.764c}$$

Cela donne le résultat attendu. □

11.290.
La formule de la proposition 11.289 nous permet d'écrire de jolies formules comme

$$(d^3 f)_x(h, h, h) = \sum_{ijk} h_i h_j h_k (\partial^3_{ijk} f)(x). \tag{11.765}$$

Proposition 11.291 ([1]).
Soient des espaces vectoriels E, V et W de dimension fine, et une fonction $f\colon E \to V$ de classe C^p. Si $\varphi\colon V \to W$ est linéaire, alors

$$\varphi \circ f\colon E \to W \tag{11.766}$$

est de classe C^p.

Démonstration. En utilisant le théorème de différentiation de fonctions composées 10.134,

$$f(\varphi \circ f)_a(u) = d\varphi_{f(a)} df_a(u), \tag{11.767}$$

et donc, parce que φ est linéaire,

$$d(\varphi \circ f)_a = \varphi \circ df_a. \tag{11.768}$$

Nous pouvons exprimer cela de façon un peu différente en posant $\varphi_1\colon \mathcal{L}(E, V) \to \mathcal{L}(E, W)$,

$$\varphi_1(\alpha)(a) = (\varphi \circ \alpha)(a). \tag{11.769}$$

Cela nous permet d'écrire $\varphi \circ df_a = (\varphi_1 \circ df)(a)$ et donc

$$d(\varphi \circ f) = \varphi_1 \circ df \tag{11.770}$$

où φ_1 est encore une application linéaire. Une récurrence semble possible. Nous posons $V_0 = V$ et $W_0 = W$ puis

$$V_{k+1} = \mathcal{L}(E, V_k) \tag{11.771a}$$

$$W_{k+1} = \mathcal{L}(E, W_k) \tag{11.771b}$$

et

$$\begin{aligned}\varphi_k\colon \mathcal{L}(E,V_{k-1}) &\to \mathcal{L}(E,W_{k-1})\\ g &\mapsto \varphi_{k-1}\circ g.\end{aligned}\tag{11.772}$$

Avec tout cela, nous prétendons que $d^k(\varphi\circ f)=\varphi_k\circ d^k f$ avec φ_k linéaire.

φ_k est linéaire Soient $\alpha_1,\alpha_2\in\mathcal{L}(E,V_{k-1})$, ainsi que $\lambda,\mu\in\mathbb{K}$. Nous avons, en utilisant la linéarité de φ_{k-1} :

$$\varphi_k(\lambda\alpha_1+\mu\alpha_2)(a)=\varphi_{k-1}\big((\lambda\alpha_1+\mu\alpha_2)(a)\big)\tag{11.773a}$$
$$=\varphi_{k-1}\big(\lambda\alpha_1(a)\big)+\mu\varphi_{k-1}\big(\alpha_2(a)\big)\tag{11.773b}$$
$$=\lambda\varphi_k(\alpha_1)a+\mu\varphi_k(\alpha_2)(a).\tag{11.773c}$$

Donc φ_k est linéaire pour tout k.

La relation La relation

$$d^k(\varphi\circ f)=\varphi_k\circ d^k f\tag{11.774}$$

se démontre par récurrence, chaque pas étant justifié de la même manière que (11.770).

□

11.30 Suites et séries : généralités

11.30.1 Convergence uniforme

11.30.1.1 Critère de Cauchy uniforme

Définition 11.292 ([147]).
Soient un espace (dont la nature n'est pas très importante) Ω, une partie A de Ω et un espace normé V. Lorsque g est une fonction $g\colon\Omega\to V$, nous notons

$$\|g\|_A=\sup_{x\in A}\|g(x)\|\tag{11.775}$$

C'est la norme supremum limitée à la partie A.

Nous disons qu'une suite de fonctions (f_n) définies sur un ensemble A ***converge uniformément sur*** *A vers la fonction f si*

$$\lim_{n\to\infty}\|f_n-f\|_A=0.\tag{11.776}$$

Proposition 11.293 (Critère de Cauchy uniforme[250]).
Soit X un espace topologique et (Y,d) un espace topologique complet. La suite de fonctions $f_n\colon X\to Y$ converge uniformément sur A si et seulement si pour tout $\epsilon>0$ il existe $N\in\mathbb{N}$ tel que si $k,l>N$ alors

$$d\big(f_k(x),f_l(x)\big)\leqslant\epsilon\tag{11.777}$$

pour tout $x\in X$.

Grosso modo, cela dit que si qu'une suite de Cauchy pour la norme uniforme est une suite uniformément convergente. Le fait que la suite converge fait partie du résultat et n'est pas une hypothèse. Ce critère sera utilisé pour montrer que $\big(C(K),\|.\|_\infty\big)$ est complet, proposition 11.296.

Démonstration. Si $f_n\xrightarrow{unif}f$ alors le critère est satisfait ; c'est dans l'autre sens que la preuve est intéressante.

Soit donc une suite de fonctions satisfaisant au critère et montrons qu'elle converge uniformément. Pour tout $x\in X$ la suite $n\mapsto f_n(x)$ est de Cauchy dans l'espace complet Y ; nous avons donc convergence ponctuelle $f_n\to f$. Nous devons prouver que cette convergence est uniforme. Soit $\epsilon>0$ et $N\in\mathbb{N}$ tel que si $k,l>N$ alors

$$d\big(f_k(x),f_l(x)\big)\leqslant\epsilon\tag{11.778}$$

pour tout $x \in X$. Si nous nous fixons un tel k et un $x \in A$ nous considérons l'inégalité

$$d\big(f_k(x), f_l(x)\big) \leqslant \epsilon \tag{11.779}$$

qui est vraie pour tout l. En passant à la limite $l \to \infty$ (limite qui commute avec la fonction distance par définition de la topologie) nous avons

$$d\big(f_k(x), f(x)\big) \leqslant \epsilon. \tag{11.780}$$

Cette inégalité étant valable pour tout $x \in X$, cela signifie que $f_n \xrightarrow{unif} f$. □

Théorème 11.294 (Limite uniforme de fonctions continues)**.**
Soit A, un ensemble mesuré et $f_n \colon A \to \mathbb{R}^n$, une suite de fonctions continues convergeant uniformément vers f. Si les fonctions f_n sont toutes continues en $x_0 \in A$, alors f est continue en x_0.

Démonstration. Soit $\epsilon > 0$. Si $x \in A$ nous avons, pour tout n, la majoration

$$\begin{aligned}\|f(x) - f(x_0)\| &\leqslant \|f(x) - f_n(x)\| + \|f_n(x) - f_n(x_0)\| + \|f_n(x_0) - f(x_0)\| && (11.781a)\\ &\leqslant \|f_n(x) - f_n(x_0)\| + 2\|f_n - f\|_\infty. && (11.781b)\end{aligned}$$

Grâce à l'uniforme convergence, nous considérons $N \in \mathbb{N}$ tel que $\|f_n - f\| \leqslant \epsilon$ pour tout $n \geqslant N$. Pour de tels n, nous avons

$$\|f(x) - f(x_0)\| \leqslant 2\epsilon + \|f_n(x) - f_n(x_0)\|. \tag{11.782}$$

La continuité de f_n nous fournit un $\delta > 0$ tel que $\|f_n(x_0) - f_n(x)\| < \epsilon$ dès que $\|x - x_0\| < \delta$. Pour ce δ, nous avons alors $\|f(x) - f(x_0)\| < \epsilon$.

Donc lorsque $\|x - x_0\| < \delta$ et $n \geqslant N$ nous avons

$$\|f(x) - f(x_0)\| \leqslant 3\epsilon, \tag{11.783}$$

où vous remarquerez qu'il n'y a plus de dépendance en n. Cela prouve la continuité de f en x_0. □

11.30.1.2 Complétude avec la norme uniforme

Proposition 11.295 (Limite uniforme de fonctions continues)**.**
Soit X un espace topologique et (Y, d) un espace métrique. Si une suite de fonctions $f_n \colon X \to Y$ continues converge uniformément, alors la limite est séquentiellement continue[69]*.*

Démonstration. Soit $a \in X$ et prouvons que f est séquentiellement continue en a. Pour cela nous considérons une suite $x_n \to a$ dans X. Nous savons que $f(x_n) \xrightarrow{Y} f(x)$. Pour tout $k \in \mathbb{N}$, tout $n \in \mathbb{N}$ et tout $x \in X$ nous avons la majoration

$$\begin{aligned}\big\|f(x_n) - f(x)\big\| &\leqslant \big\|f(x_n) - f_k(x_n)\big\| + \big\|f_k(x_n) - f_k(x)\big\| + \big\|f_k(x) - f(x)\big\| && (11.784a)\\ &\leqslant 2\|f - f_k\|_\infty + \big\|f_k(x_n) - f_k(x)\big\|. && (11.784b)\end{aligned}$$

Soit $\epsilon > 0$. Si nous choisissons k suffisamment grand la premier terme est plus petit que ϵ. Et par continuité de f_k, en prenant n assez grand, le dernier terme est également plus petit que ϵ. □

Proposition 11.296.
Soit X un espace topologique métrique (Y, d) un espace espace métrique complet. Alors les espaces

(1) $\big(C_b^0(X, Y), \|.\|_\infty\big)$ des fonctions continues et bornées $X \to Y$,
(2) $\big(C_0^0(X, Y), \|.\|_\infty\big)$ des fonctions continues et s'annulant à l'infini
(3) $\big(C_0^k(X, Y), \|.\|_\infty\big)$ des fonctions de classe C^k et s'annulant à l'infini

69. Si X est métrique, alors c'est la continuité usuelle par la proposition 7.80.

sont complets.

Démonstration. Soit (f_n) une suite de Cauchy dans $C(X,Y)$, c'est-à-dire que pour tout $\epsilon > 0$ il existe $N \in \mathbb{N}$ tel que si $k, l > N$ nous avons $\|f_k - f_l\|_\infty \leqslant \epsilon$. Cette suite vérifie le critère de Cauchy uniforme 11.293 et donc converge uniformément vers une fonction $f\colon X \to Y$. La continuité (ou l'aspect C^k) de la fonction f découle de la convergence uniforme et de la proposition 11.295 (c'est pour avoir l'équivalence entre la continuité séquentielle et la continuité normale que nous avons pris l'hypothèse d'espace métrique).

Si les fonctions f_k sont bornées ou s'annulent à l'infini, la convergence uniforme implique que la limite le sera également. □

Notons que si X est compact, les fonctions continues sont bornées par le théorème 7.92 et nous pouvons simplement dire que $C^0(X,Y)$ est complet, sans préciser que nous parlons des fonctions bornées.

Lemme 11.297.
Soient un espace topologique compact A et un espace complet B. L'ensemble des fonctions continues de A vers B muni de la norme uniforme est complet.

Dit de façon courte : $\big(C(A,B), \|.\|_\infty\big)$ est complet.

Démonstration. Soit (f_k) une suite de Cauchy de fonctions dans $C(A,B)$. Pour chaque $x \in A$ nous avons

$$\|f_k(x) - f_l(x)\|_B \leqslant \|f_k - f_l\|_\infty, \tag{11.785}$$

de telle sorte que la suite $(f_k(x))$ est de Cauchy dans B et converge donc vers un élément de B. La suite de Cauchy (f_k) converge donc ponctuellement vers une fonction $f\colon A \to B$. Nous devons encore voir que cette fonction est continue ; ce sera l'uniformité de la norme qui donnera la continuité. En effet soit $x_n \to x$ une suite dans A qui converge vers $x \in A$. Pour chaque $k \in \mathbb{N}$ nous avons

$$\|f(x_n) - f(x)\| \leqslant \|f(x_n) - f_k(x_n)\| + \|f_k(x_n) - f_k(x)\| + \|f_k(x) - f(x)\|. \tag{11.786}$$

En prenant k et n assez grands, cette expression peut être rendue aussi petite que l'on veut ; le premier et le troisième terme par convergence ponctuelle $f_k \to f$, le second terme par continuité de f_k. La suite $f(x_n)$ est donc convergente vers $f(x)$ et la fonction f est continue. □

Problèmes et choses à faire

Il serait sans doute bon de revoir cette preuve à la lumière du critère de Cauchy uniforme 11.293.

11.298 ([251]).
Le théorème de Stone-Weierstrass indique que les polynômes sont denses pour la topologie uniforme dans les fonctions continues. Donc il existe des limites uniformes de fonctions C^∞ qui ne sont même pas dérivables. Les espaces de type C^p munis de $\|.\|_\infty$ ne sont donc pas complets sans quelques hypothèses. Voir la proposition 11.296 et le thème 25.

Théorème 11.299 (Théorème de Dini[252]).
Soient un espace métrique complet D et une suite de fonctions $f_n \in C(D,\mathbb{R})$ telle que

(1) $f_n \to g$ ponctuellement,

(2) $g \in C(D,\mathbb{R})$,

(3) la suite (f_n) est croissante, c'est-à-dire que pour tout $x \in D$ et pour tout $n \geqslant 0$ nous avons $f_{n+1}(x) \geqslant f_n(x)$.

Alors la convergence est uniforme.

Démonstration. Soit $x \in D$ et $\epsilon > 0$. Il existe $N(x) \in \mathbb{N}$ tel que

$$g(x) - \epsilon \leqslant f_{N(x)} \leqslant g(x). \tag{11.787}$$

De plus g et $f_{N(x)}$ sont des fonctions continues, donc il existe $\eta(x)$ tel que si $y \in B\big(x, \eta(x)\big)$ alors

$$g(y) \in B\big(g(x), \epsilon\big) \tag{11.788a}$$
$$f_{N(x)}(y) \in B\big(f_{N(x)}(x), \epsilon\big). \tag{11.788b}$$

Si $n \geqslant N(x)$ et si $y \in B(x, \eta(x))$ alors nous avons les majorations

$$g(y) \geqslant f_n(y) \geqslant f_{N(x)}(y) \geqslant f_{N(x)}(x) - \epsilon \geqslant g(x) - 2\epsilon \geqslant g(y) - 3\epsilon. \tag{11.789}$$

Justifications :

(1) Les deux premières inégalités sont la croissance de la suite.
(2) La suivante est (11.788b).
(3) Ensuite il y a le choix de $N(x)$.
(4) Et enfin il y a (11.788a).

Nous retenons que si $x \in D$ et si $n \geqslant N(x)$ alors

$$g(y) \geqslant f_n(y) \geqslant g(y) - 3\epsilon \tag{11.790}$$

pour tout $y \in B(x, \eta(x))$.

Nous utilisons maintenant la compacité de D. Pour chaque $x \in D$ nous pouvons considérer la boule ouverte $B\big(x, \eta(x)\big)$; ces boules recouvrent D. Nous en extrayons un sous-recouvrement fini, c'est-à-dire un ensemble fini d'éléments $x_1, \dots, x_K$ tels que

$$D = \bigcup_{k=1}^{K} B\big(x_k, \eta(x_k)\big). \tag{11.791}$$

Si à ce moment vous ne comprenez pas pourquoi c'est une égalité au lieu d'une inclusion, il faut lire l'exemple 7.19. Considérons

$$n \geqslant N = \max\{N(x_1), \dots, N(x_K)\}. \tag{11.792}$$

Pour tout $y \in D$ il existe $k \in \{1, \dots, K\}$ tel que $y \in B\big(x_k, \eta(x_k)\big)$, et vu que $n \geqslant N(x_k)$ nous reprenons la majoration (11.790) :

$$g(y) \geqslant f_n(y) \geqslant g(y) - 3\epsilon. \tag{11.793}$$

Pour le n choisi nous avons ces inégalités pour tout $y \in D$, c'est-à-dire que nous avons $\|f_n - g\| \leqslant 3\epsilon$ et donc la convergence uniforme. □

Proposition 11.300 ([1]).
Soient une suite de fonctions continues $u_i \colon \mathbb{R} \to \mathbb{R}$ et une fonction continue u telle que $u_i \to u$ simplement. Alors la convergence est uniforme sur tout compact.

Démonstration. Soit un compact K ; nous notons $\|.\|$ la norme uniforme sur K. Supposons que la limite ne soit pas uniforme, c'est-à-dire qu'il existe un $\epsilon > 0$ tel que

$$\|u_i - u\| > 2\epsilon \tag{11.794}$$

pour tout i. Cela permet de considérer pour tout i un élément $x_i \in K$ tel que [70]

$$\|u_i(x_i) - u(x_i)\| > \epsilon. \tag{11.795}$$

Pour cela, il faut noter que K est compact et que la fonction $x \mapsto \|u_i(x) - u(x)\|$ est continue sur K. Elle est donc bornée et atteint son maximum (c'est le théorème de Weierstrass 7.105).

70. Notez l'inégalité stricte, obetenue en considérant 2ϵ plus haut.

La suite $i \mapsto x_i$ est une suite dans un compact, et quitte à prendre une sous-suite, nous supposons qu'elle converge vers $a \in K$ (ça, c'est Bolzano-Weierstrass 7.102).

La convergence ponctuelle $u_i \to u$, prise en a, dit qu'il existe un N tel que $|u_i(a) - u(a)| < \epsilon$ pour tout $i \geqslant N$. Pour un tel i, nous avons aussi

$$|u_i(x) - u(x)| < \epsilon \tag{11.796}$$

sur un voisinage de a, parce que $u_i - u$ est continue. Mais tout voisinage de a contient un élément x_j pour lequel

$$|u_i(x_j) - u(x_j)| > \epsilon. \tag{11.797}$$

Contradiction. □

11.30.2 Série de fonctions

Les séries de fonctions sont des cas particuliers de suites.

Définition 11.301.
Si (f_n) est une suite de fonctions, nous définissons la somme des f_n de la façon suivante :

$$\sum_{n=1}^{\infty} f_n = \lim_{N\to\infty} \sum_{n=1}^{N} f_n. \tag{11.798}$$

Le membre de droite est une définition de la notation introduite dans le membre de gauche.

Avant de vous lancer, relisez une bonne fois les définitions de convergence absolue (définition 10.51) et de convergence uniforme (équation 10.98).

Lemme 11.302.
Soient des fonctions $u_n \colon \Omega \to \mathbb{C}$. Si il existe une suite réelle positive $(a_n)_{n\in\mathbb{N}}$ telle que

(1) pour tout $z \in \Omega$ et pour tout $n \in \mathbb{N}$ nous avons $|u_n(z)| \leqslant a_n$ (c'est-à-dire $a_n \geqslant \|u_n\|_\infty$),

(2) la somme $\sum_n a_n$ converge,

alors la série de fonctions $\sum_{n=0}^{\infty} u_n$ converge normalement[71].

Démonstration. Découle du lemme de comparaison 10.61. □

Théorème 11.303.
Soit $\sum_{k=1}^{\infty} g_k(x)$, une série de fonctions complexes où $g_k(x) = \varphi_k(x)\psi_k(x)$. Supposons que

(1) $\varphi_k \colon A \to \mathbb{C}$ et $|\sum_{k=1}^{K} \varphi_k(x)| \leqslant M$ où M est indépendant de x et K,

(2) $\psi_k \colon A \to \mathbb{R}$ avec $\psi_k(x) \geqslant 0$ et pour tout x dans A, $\psi_{k+1}(x) \leqslant \psi_k(x)$, et enfin supposons que $\psi_k(x)$ converge uniformément vers 0.

Alors $\sum_{k=1}^{\infty} g_k$ est uniformément convergente.

Théorème 11.304.
Si la série de puissances (réelle) converge en $x = x_0 + R$, alors elle converge uniformément sur $[x_0 - R + \epsilon, x_0 + R]$ ($\epsilon > 0$) vers une fonction continue.

Proposition 11.305.
Soit (u_n) une suite de fonctions continues $u_n \colon \Omega \subset \mathbb{C} \to \mathbb{C}$. Si la série $\sum_n u_n$ converge normalement alors la somme est continue.

71. Définition 10.52.

Démonstration. Nous posons $u(z) = \lim_{N\to\infty} \sum_{n=0}^{N} u_n(z)$, et nous vérifions que la fonction ainsi définie sur Ω est continue. Soit $z \in \Omega$. Prouvons la continuité de u au point z. Pour tout z' dans un voisinage de z nous avons

$$\left|u(z) - u(z')\right| = \left|\sum_{n=0}^{N} u_n(z) - \sum_{n=0}^{N} u_n(z') + \sum_{n=N+1}^{\infty} u_n(z) - \sum_{n=N+1}^{\infty} u_n(z')\right| \tag{11.799a}$$

$$\leqslant \left|\sum_{n=0}^{N} u_n(z) - \sum_{n=0}^{N} u_n(z')\right| + \sum_{n=N+1}^{\infty} |u_n(z)| + \sum_{n=N+1}^{\infty} |u_n(z')|. \tag{11.799b}$$

Étant donné que les sommes partielles sont continues, en prenant N suffisamment grand, le premier terme peut être rendu arbitrairement petit. Si N est suffisamment grand, le second terme est également petit. Par contre, cet argument ne tient pas pour le troisième terme parce que nous souhaitons une majoration pour tout z' dans une boule autour de z. Nous devons donc écrire

$$\sum_{n=N}^{\infty} |u_n(z)| \leqslant \sum_{n=N+1}^{\infty} \|u_n\|_\infty. \tag{11.800}$$

Ce dernier est arbitrairement petit lorsque N est grand. Notons que nous avons utilisé l'hypothèse de convergence normale. □

La même propriété, avec la même démonstration, tient dans le cas d'espaces vectoriels normés.

Proposition 11.306.
Soient E et F, deux espaces vectoriels normés, Ω une partie ouverte de E et une suite de fonctions $u_n\colon \Omega \to F$ convergeant normalement sur Ω, c'est-à-dire que $\sum_n \|u_n\|_\infty$ converge, la norme $\|.\|_\infty$ devant être comprise comme la norme supremum sur Ω. Alors la fonction $u = \sum_n u_n$ est continue sur Ω.

Démonstration. Soit $x, x' \in \Omega$ en supposant que $\|x - x'\|$ est petit. Soit encore $\epsilon > 0$. Nous allons montrer la continuité en x. Pour cela nous savons que pour tout N l'inégalité suivante est correcte :

$$\|u(x) - u(x')\| \leqslant \left\|\sum_{n=0}^{N} u_n(x) - \sum_{n=0}^{N} u_n(x')\right\| + \sum_{n=N+1}^{\infty} \|u_n(x)\| + \sum_{n=N+1}^{\infty} \|u_n(x')\|. \tag{11.801}$$

Les deux derniers termes sont majorés par $\sum_{n=N+1}^{\infty} \|u_n\|_\infty$ qui, par hypothèse, peut être rendu aussi petit que souhaité en choisissant N assez grand. Nous choisissons donc un N tel que ces deux termes soient plus petits que ϵ. Ce N étant fixé, la fonction $\sum_{n=0}^{N} u_n$ est continue et nous pouvons choisir x' assez proche de x pour que le premier terme soit majoré par ϵ. □

Théorème 11.307.
Si les g_k sont continues et si $\sum g_k$ converge uniformément, alors $\sum g_k$ est continue.

Le corolaire suivant permet de considérer des séries de fonctions indexées par exemple par $\mathbb{Z}$ plutôt que par $\mathbb{N}$.

Corolaire 11.308.
Une famille dénombrable de fonctions continues convergeant normalement converge vers une fonction continue.

Démonstration. Soit I dénombrable. Considérons une famille de fonctions continues $(f_n)_{n\in I}$ telles que la famille $(\|f_i\|_\infty)_{i\in I}$ soit sommable. Le proposition 10.99 nous permet d'utiliser une bijection entre I et $\mathbb{N}$. Le théorème 11.305 s'applique alors. □

Théorème 11.309 (Critère de Weierstrass)**.**
Soit une suite de fonctions $f_k\colon A \to \mathbb{C}$ telles que $|f_k(x)| \leqslant M_k \in \mathbb{R}$, $\forall x \in A$. Si $\sum_{k=1}^{\infty} M_k$ converge, alors $\sum_{k=1}^{\infty} f_k$ converge absolument et uniformément.

Démonstration. La convergence normale est facile : l'hypothèse dit que $\|f_k\|_\infty \leqslant M_k$, et donc que

$$\sum_{k=1}^{\infty} \|f_k\|_\infty \leqslant \sum_k M_k < \infty. \tag{11.802}$$

La convergence uniforme est à peine plus subtile. Nous nommons F la fonction somme. Pour tout x et pour tout N, nous avons

$$\left\| \sum_{n=1}^{N} f_n(x) - F(x) \right\| = \| \sum_{n=N}^{\infty} f_n(x) \| \tag{11.803a}$$

$$\leqslant \sum_{n=N}^{\infty} \|f_k(x)\| \tag{11.803b}$$

$$\leqslant \sum_{n=N}^{\infty} \|f_n\|_\infty. \tag{11.803c}$$

La convergence normale étant assurée, la série $\sum_{n_1}^{\infty} \|f_n\|_\infty$ est finie, ce qui implique que la queue de somme $\sum_{n=N}^{\infty} \|f_n\|_\infty$ tend vers zéro lorsque $N \to \infty$. Pour tout ϵ, il existe donc un N (non dépendant de x) tel que

$$\| \sum_{n=1}^{N} f_n(x) - F(x) \| \leqslant \epsilon. \tag{11.804}$$

En prenant le supremum sur $x \in A$ nous trouvons la convergence uniforme. □

Remarque 11.310.
Il n'y a pas de critère correspondant pour les suites. Il n'est pas vrai que si $\lim_{n\to\infty} \|f_n\|$ existe, alors $\lim_{n\to\infty} f_n$ existe, comme le montre l'exemple

$$f_n(x) = \begin{cases} 1 & \text{si } x \in [0,1] \text{ et } n \text{ est pair} \\ 1 & \text{si } x \in [1,2] \text{ et } n \text{ est impair} \\ 0 & \text{sinon.} \end{cases} \tag{11.805}$$

11.31 Permuter limite et dérivée

Pour permuter différentielle et limite, ce sera le théorème 14.6.

Théorème 11.311 ([147, 228])**.**
Soient une suite de fonctions $f_i \colon \mathbb{R} \to \mathbb{R}$, une fonction $f \colon \mathbb{R} \to \mathbb{R}$ et une fonction $g \colon \mathbb{R} \to \mathbb{R}$ telles que

(1) f_i est de classe C^1 pour tout i,

(2) $f_i \to f$ simplement,

(3) $f_i' \to g$ uniformément sur tout compact.

Alors

(1) f est de classe C^1,

(2) $f' = g$,

(3) $f_i \to f$ uniformément sur tout compact.

Démonstration. On commence par démonter la convergence uniforme sur les compacts des f_i. Soit $K \subseteq \mathbb{R}$ un intervalle compact contenant x. Il suffit de montrer que la suite $(f_i)_i$ restreinte à K est une suite de Cauchy pour la norme uniforme. Soit $\epsilon > 0$. On note ω_i le module de continuité de f_i'. Par le lemme 10.167, il existe $N \geqslant 0$ tel que pour tout $\delta > 0$ nous ayons

$$\omega_i(\delta) \leqslant \omega_g(\delta) + \epsilon. \tag{11.806}$$

Soit $y \in K$, $n \in \mathbb{N}$ et $\alpha_n = \frac{x-y}{n+1}$. Pour tout $i \geqslant 0$,

$$f_i(y) = f_i(x) + \sum_{k=0}^{n} \left(f_i(x + (k+1)\alpha_n) - f_i(x + k\alpha_n)\right), \tag{11.807}$$

c'est une somme télescopique. Par le théorème des accroissements finis, il existe pour tout $0 \leqslant k \leqslant n$ un réel $u_{n,i,k} \in [k, (k+1)\alpha_n]$ tel que

$$f_i(x + (k+1)\alpha_n) - f_i(x + k\alpha_n) = \alpha_n f_i'(x + u_{n,i,k}), \tag{11.808}$$

de sorte que

$$f_i(y) = f_i(x) + \alpha_n \sum_{k=0}^{n} f_i'(x + u_{n,i,k}). \tag{11.809}$$

Et pour tout $i, j \geqslant 0$, on obtient

$$|f_i(y) - f_j(y)| \leqslant |f_i(x) - f_j(x)| + \alpha_n \sum_{k=0}^{n} \left|f_i'(x + u_{n,i,k}) - f_j'(x + u_{n,j,k})\right|. \tag{11.810}$$

Or

$$\left|f_i'(x + u_{n,i,k}) - f_j'(x + u_{n,j,k})\right| \leqslant \left|f_i'(x + u_{n,i,k}) - f_i'(x + u_{n,j,k})\right| + \left|f_i'(x + u_{n,j,k}) - f_j'(x + u_{n,j,k})\right| \tag{11.811}$$

$$\leqslant \omega_i(\alpha_n) + \|f_i' - f_j'\|, \tag{11.812}$$

et il suit que pour tout $i, j \geqslant N$,

$$|f_i(y) - f_j(y)| \leqslant |f_i(x) - f_j(x)| + |x - y| \left(\omega_i(\alpha_n) + \|f_i' + f_j'\|_K\right) \tag{11.813}$$

$$\leqslant |f_i(x) - f_j(x)| + |x - y| \left(\omega_g(\alpha_n) + \epsilon + \|f_i' + f_j'\|_K\right), \tag{11.814}$$

où la dernière inégalité vient de (11.806). Comme g est continue (limite uniforme de fonctions continues) sur un compact, elle est uniformément continue et $\lim_{\delta \to 0} \omega_g(\delta) = 0$. Donc l'inégalité (11.814), avec $n \to \infty$ devient

$$|f_i(y) - f_j(y)| \leqslant |f_i(x) - f_j(x)| + |x - y| \left(\epsilon + \|f_i' + f_j'\|_K\right). \tag{11.815}$$

Comme K est borné, $|x - y|$ est borné par une quantité M ne dépendant que de K. En prenant dans (11.815) le supremum par rapport à y, on obtient

$$\|f_i - f_j\|_K \leqslant |f_i(x) - f_j(x)| + |x - y| \left(\epsilon + \|f_i' + f_j'\|_K\right). \tag{11.816}$$

Étant donné que $(f_i(x))_i$ est une suite de Cauchy, que $(f_i')_i$ est une suite de Cauchy pour la norme uniforme sur K, il suit que $(f_i|_K)_i$ est une suite de Cauchy. Donc elle converge uniformément vers une limite $f|_K$.

Montrons maintenant que la limite f de la suite $(f_i)_i$ est dérivable et que $f' = g$. Soit $y \in \mathbb{R}$ et soit K un voisinage compact de y. On reprend les notations précédentes. Pour tout $\delta \in \mathbb{R}$ suffisamment petit $y + \delta \in K$, et pour tout $i \geqslant 0$ on a alors

$$\left|\frac{1}{\delta}\big(f(y+\delta) - f(y)\big) - g(y)\right| \tag{11.817a}$$

$$\leqslant \frac{1}{\delta}\left|f(y) - f_i(y)\right| + \frac{1}{\delta}\left|f(y+\delta) - f_i(y+\delta)\right| + \left|\frac{1}{\delta}\left(f_i(y+\delta) - f_i(y)\right) - g(y)\right| \tag{11.817b}$$

$$\leqslant \frac{2}{\delta}\|f_i - f\|_K + \left|\frac{1}{\delta}\left(f_i(y+\delta) - f_i(y)\right) - g(y)\right|. \tag{11.817c}$$

Par le théorème des accroissements finis, il existe $u \in [y - |\delta|, y + |\delta|]$ tel que

$$\frac{1}{\delta}\left(f_i(y+\delta) - f_i(y)\right) = f_i'(u). \tag{11.818}$$

Il suit,

$$\left|\frac{1}{\delta}\left(f(y+\delta) - f(y)\right) - g(y)\right| \leqslant \frac{2}{\delta}\|f_i - f\|_K + \left|f_i'(u) - f_i'(y)\right| + \|f_i' - g\| \tag{11.819}$$

$$\leqslant \frac{2}{\delta}\|f_i - f\|_K + \omega_i(\delta) + \|f_i' - g\|. \tag{11.820}$$

En prenant $i \to \infty$, on obtient

$$\left|\frac{1}{\delta}\left(f(y+\delta) - f(y)\right) - g(y)\right| \leqslant \omega_g(\delta). \tag{11.821}$$

Or $\lim_{\delta\to 0}\omega_g(\delta) = 0$, donc f est dérivable en y et $f'(y) = g(y)$. □

Théorème 11.312.
Soit $U \subset \mathbb{R}^n$ ouvert, $f_k \colon U \to \mathbb{R}$ et f_k de classe C^1. Supposons que f_k converge simplement vers f et que $\partial_i f_k$ converge uniformément sur tout compact vers une fonction g_i pour $i = 1, \dots, n$. Alors f est de classe C^1 et $\partial_i f = g_i$. De plus, f_k converge vers f uniformément.

11.32 Densité des polynômes

11.32.1 Théorème de Stone-Weierstrass

Voir le thème 13.

Note : le lemme 11.313 est utilisé dans la démonstration du théorème 11.317 ; c'est pour cela que nous l'avons isolé.

Lemme 11.313.
Il existe une suite de polynômes sur $[0, 1]$ convergeant uniformément vers la fonction racine carrée.

Démonstration. Nous donnons cette suite par récurrence :

$$P_0(t) = 0 \tag{11.822a}$$

$$P_{n+1}(t) = P_n(t) + \frac{1}{2}\big(t - P_n(t)^2\big). \tag{11.822b}$$

Nous commençons par montrer que pour tout $t \in [0, 1]$, $P_n(t) \in [0, \sqrt{t}]$. Pour P_0, c'est évident. Ensuite nous avons

$$P_{n+1}(t) - \sqrt{t} = P_n(t) - \sqrt{t} + \frac{1}{2}(t - P_n(t)^2) \tag{11.823a}$$

$$= \big(P_n(t) - \sqrt{t}\big)\left(1 - \frac{1}{2}\frac{t - P_n(t)^2}{P_n(t) - \sqrt{t}}\right) \tag{11.823b}$$

$$= \big(P_n(t) - \sqrt{t}\big)\left(1 - \frac{\sqrt{t} + P_n(t)}{2}\right) \tag{11.823c}$$

$$\leqslant 0 \tag{11.823d}$$

parce que $\sqrt{t} \leqslant 1$ et $P_n(t) \leqslant 1$ par hypothèse de récurrence.

Nous savons au passage que $P_n(t)$ est une suite réelle croissante parce que $t - P_n(t)^2 \geqslant t - (\sqrt{t})^2 = 0$. La suite $P_n(t)$ est donc croissante et majorée par $\sqrt{t}$; elle converge donc. Les candidats limites sont déterminés par l'équation

$$\ell = \ell + \frac{1}{2}(t - \ell^2), \tag{11.824}$$

dont les solutions sont $\ell = \pm\sqrt{t}$. La suite étant positive, nous avons une convergence ponctuelle de P_n vers la racine carrée. Cette suite étant une suite croissante de fonctions continues sur un compact, convergeant ponctuellement vers une fonction continue, la convergence est uniforme par le théorème de Dini 11.299. □

Lemme 11.314.
Soit K, un compact de $\mathbb{R}$ et f_n une suite de fonctions sur K convergeant uniformément vers f. Soit $g\colon X \to K$ une fonction depuis un espace topologique K. Alors $f_n \circ g$ converge uniformément vers $f \circ g$.

Démonstration. En effet, pour tout $x \in X$ nous avons

$$\|(f_n \circ g) - (f \circ g)\|_\infty = \sup_{x\in X} \|f_n\big(g(x)\big) - f\big(g(x)\big)\| \leqslant \|f_n - f\|_\infty. \tag{11.825}$$

Par conséquent, si ϵ 0 est donné, il suffit de choisir n de telle sorte à avoir $\|f_n - f\|_\infty < \epsilon$ et nous avons $\|(f_n \circ g) - (f \circ g)\|_\infty \leqslant \epsilon$. □

Définition 11.315.
Nous disons qu'une algèbre A de fonctions sur un espace X ***sépare les points*** *de X si pour tout $x_1 \neq x_2$ il existe $g \in A$ telle que $g(x_1) \neq g(x_2)$.*

Nous pouvons maintenant énoncer et démontrer une forme nettement plus générale du théorème de Stone-Weierstrass. Le théorème 11.317 le donne pour $C(X, \mathbb{C})$ et le théorème 11.316 le donne pour $C(X, \mathbb{R})$.

Théorème 11.316 (Stone-Weierstrass[253])**.**
Soient X, un espace compact et Hausdorff. Soit A, une sous-algèbre de $C(X, \mathbb{R})$ contenant une fonction constante non nulle. Alors A est dense dans $\Big(C(X, \mathbb{R}), \|.\|_\infty\Big)$ si et seulement si A sépare les points de X.

Démonstration. Nous allons écrire la démonstration en plusieurs étapes (dont la première est le lemme 11.313). Nous commençons par la première partie, sur les réels.

Première étape Pour tout $x \neq y \in X$ et pour tout $\alpha, \beta \in \mathbb{R}$, il existe une fonction $f \in A$ telle que $f(x) = \alpha$ et $f(y) = \beta$.

En effet, vu que A sépare les points nous pouvons considérer une fonction $g \in A$ telle que $g(x) \neq g(y)$ et ensuite poser

$$f(z) = \alpha + \frac{\alpha - \beta}{g(y) - g(x)}\big(g(z) - g(x)\big). \tag{11.826}$$

Les constantes faisant partie de A, cette fonction f est encore dans A.

Seconde étape Pour tout n-uples de fonctions $f_1, \dots, f_n$ dans $\bar{A}$, les fonctions $\min(f_1, \dots, f_n)$ et $\max(f_1, \dots, f_n)$ sont dans $\bar{A}$.

Nous le démontrons pour $n = 2$; le reste allant évidemment par récurrence. Soient $f, g \in \bar{A}$. Étant donné que

$$\max(f, g) = \frac{f + g}{2} + \frac{|f - g|}{2} \tag{11.827a}$$

$$\min(f, g) = \frac{f + g}{2} - \frac{|f - g|}{2}, \tag{11.827b}$$

if suffit de montrer que si $f \in \bar{A}$ alors $|f| \in \bar{A}$. Si f est nulle, c'est évident ; supposons que $f \neq 0$ et posons $M = \|f\|_\infty \neq 0$. Pour tout $x \in X$ nous avons

$$\frac{f(x)^2}{M^2} \in [0, 1]. \tag{11.828}$$

Nous considérons alors la suite

$$h_n = P_n \circ \frac{f^2}{M^2} \tag{11.829}$$

où P_n est une suite de polynômes convergent uniformément vers la racine carrée (voir lemme 11.313). Le lemme 11.314 nous assure que h_n converge uniformément vers $\frac{|f|}{M}$ dans $C(X, \mathbb{R})$. Étant donné que $\bar{A}$ est également une algèbre, h_n est dans $\bar{A}$ pour tout n et la limite s'y trouve également (pour rappel, la fermeture $\bar{A}$ est celle de la topologie de la convergence uniforme).

Troisième étape Soit $\epsilon > 0$, $f \in C(X, \mathbb{R})$ et $x \in X$. Il existe une fonction $g_x \in \bar{A}$ telle que

$$\begin{cases} g_x(x) = f(x) & \text{(11.830a)} \\ g_x(y) \leqslant f(y) + \epsilon & \text{(11.830b)} \end{cases}$$

pour tout $y \in X$.

Soit $z \in X \backslash \{x\}$ et une fonction h_z telle que $h_z(x) = f(x)$ et $h_z(z) = f(z)$. Une telle fonction existe par une des étapes précédentes. Étant donné que f et h_z sont continues, il existe un voisinage ouvert V_z de z sur lequel

$$h_z(y) \leqslant f(y) + \epsilon \tag{11.831}$$

pour tout $y \in V_z$. Nous pouvons sélectionner un nombre fini de points $z_1, \dots, z_n$ tels que les ouverts $V_{z_1}, \dots, V_{z_n}$ recouvrent X (parce que X est compact, de tout recouvrement par des ouverts, nous extrayons un sous recouvrement fini.). Nous posons

$$g_x = \min(h_{z_1}, \dots, h_{z_n}) \in \bar{A}. \tag{11.832}$$

Si $y \in X$, nous sélectionnons le i tel que $h_{z_i}(y) \leqslant f(y) + \epsilon$ et nous avons

$$g_x(y) \leqslant h_{z_i}(y) \leqslant f(y) + \epsilon. \tag{11.833}$$

Étape finale Soit $\epsilon > 0$ et $f \in C(X, \mathbb{R})$. Pour chaque $x \in X$ nous considérons une fonction $g_x \in \bar{A}$ telle que

$$\begin{cases} g_x(x) = f(x) & \text{(11.834a)} \\ g_x(y) \leqslant f(y) + \epsilon & \text{(11.834b)} \end{cases}$$

pour tout $y \in X$. Les fonctions f et g_x sont continues, donc il existe un voisinage ouvert W_x de x sur lequel

$$g_x(y) \geqslant f(y) - \epsilon. \tag{11.835}$$

De ces W_x nous extrayons un sous recouvrement fini de X : $W_{x_1}, \dots, W_{x_m}$ et nous posons

$$\varphi = \max(g_{x_1}, \dots, g_{x_n}) \in \bar{A}. \tag{11.836}$$

Si $y \in X$, il existe un i tel que

$$\varphi(y) \geqslant g_{x_i}(y) \geqslant f(y) - \epsilon. \tag{11.837}$$

La première inégalité est le fait que φ est le maximum des g_{x_k}, et la seconde est le choix de i. Donc pour tout $y \in X$ nous avons

$$f(y) - \epsilon \leqslant \varphi(y) \leqslant f(y) + \epsilon. \tag{11.838}$$

La première inégalité est ce que l'on vient de faire. La seconde est le fait que pour tout i nous ayons $g_{x_i}(y) \leqslant f(y) + \epsilon$; le fait que φ soit le maximum sur les i ne change pas l'inégalité.

Le fait que les inégalités (11.838) soient vraies pour tout $y \in X$ signifie que $\|\varphi - f\|_\infty \leqslant \epsilon$, et donc que $f \in \mathrm{Adh}\big(\mathrm{Adh}(A)\big) = \mathrm{Adh}(A)$.

Tout cela prouve que $C(X,\mathbb{R}) \subset \mathrm{Adh}(A)$. L'inclusion inverse est le fait que $C(X,\mathbb{R})$ est fermé pour la norme $\|.\|_\infty$, étant donné qu'une limite uniforme de fonctions continues est continue.

Nous pouvons maintenant nous tourner vers l'énoncé concernant $C(X,\mathbb{C})$. □

Théorème 11.317 (Stone-Weierstrass[1]).
Soit X, un espace compact et Hausdorff. Soit une sous-algèbre A stable par conjugaison[72] *A de $C(X,\mathbb{C})$ contenant une fonction constante non nulle. Alors A est dense dans $\Big(C(X,\mathbb{C}),\|.\|_\infty\Big)$ si et seulement si A sépare les points de X.*

Entendons-nous bien : ici A et $C(X,\mathbb{C})$ sont des algèbres à coefficients dans $\mathbb{C}$.

Démonstration. La preuve de cette version dans $C(X,\mathbb{C})$ va bien entendu fortement reposer sur le cas dans $C(X,\mathbb{R})$ que nous venons de prouver. Soit donc A, une sous-algèbre vérifiant les hypothèses.

$\mathrm{Re}(A) \subset A$ Nous prouvons que si $f \in A$, alors $\mathrm{Re}(f) \in A$. En effet, vu que A est stable par conjugaison, si $f \in A$, alors $\bar{f} \in A$ et $f + \bar{f} = 2\,\mathrm{Re}(f) \in A$.

Nous posons

$$A_1 = \{\mathrm{Re}(g) \text{ tel que } g \in A\}. \tag{11.839}$$

A_1 est une sous-algèbre de A Le fait que les élément de A_1 soient dans A est déjà fait. Pour le produit, si $g_1, h_1 \in A_1$, alors il existe $h, h \in A$ tels que $g_1 = \mathrm{Re}(g)$ et $h_1 = \mathrm{Re}(h)$. Nous avons

$$(g_1 + ig_2)(h_1 + ih_2) = g_1h_2 - g_2h_2 + i(g_1h_2 + g_2h_1) \in A. \tag{11.840}$$

La partie réelle de cela est dans A_1, donc

$$g_1h_2 - g_2h_1 \in A_1. \tag{11.841}$$

Mais comme $g_1 + ig_2 \in A$, nous avons aussi $g_1 - ig_2 \in A$ et donc

$$(g_1 - ig_2)(h_1 + ih_2) = g_1h_1 + g_2h_2 + i(g_1h_2 - g_2h_1) \in A. \tag{11.842}$$

La partie réelle de cela est dans A_1. Donc

$$g_1h_2 + g_2h_1 \in A_1. \tag{11.843}$$

En comparant avec (11.841), nous avons $g_1h_1 \in A_1$.

A_1 sépare les points de X Soient $x, y \in X$ ainsi que $f \in A$ séparant les points x et y, c'est-à-dire

$$f(x) \neq f(y). \tag{11.844}$$

Supposons $f_1(x) = f_2(y)$. Vu que f sépare, si ce ne sont pas les parties réelles, ce sont les parties imaginaires. C'est-à-dire que $f_2(x) \neq f_2(y)$. Mais d'autre part, $if = f_2 + if_1 \in A$, donc en réalité $f_2 \in A_1$ également.

Le partie A_1 dans $C(X,\mathbb{R})$ vérifie les hypothèses de Stone-Weierstrass réel 11.316, donc A_1 est dense dans $C(X,\mathbb{R})$. Le même raisonnement montre que A_2 est également dense dans $C(X,\mathbb{R})$[73]

Soit maintenant le vif de la preuve : $f \in C(X,\mathbb{C})$ avec $f = u + iv$, les fonctions u et v étant dans $C(X,\mathbb{R})$. Nous avons des suites $u_k \xrightarrow{unif} u$ et $v_k \xrightarrow{unif} v$ pour des suites (u_k) et (v_k) dans $C(X,\mathbb{R})$.

Par le même genre de raisonnements que nous avons déjà fait, nous nous convainquons que $u_k + iv_k \in A$ pour chaque k. Nous avons

$$\|u_k + iv_k - u - iv\|_\infty \leqslant \|u_k - u\|_\infty + \|v_k - v\|_\infty \tag{11.845}$$

En prenant k assez grand, les deux termes peuvent être rendus plus petit que ϵ. □

72. Pour tout $g \in A$, nous avons $\bar{g} \in A$.

73. Il me semble même que $A_1 = A_2$ et qu'il y a un raccourci possible dans cette preuve en exploitant ce fait. Écrivez-moi pour dire ce que vous en pensez.

Corolaire 11.318 ([1]).
Soit B, la boule fermée de centre 0 et de rayon 1 dans $\mathbb{R}^n$. La partie $C^\infty(B,\mathbb{R}^n)$ est dense dans $\big(C(B,B),\|.\|_\infty\big)$.

Démonstration. Soit $f\in C(B,B)$ et $\epsilon>0$. La fonction donnant la composante i est une fonction $f_i\in C(B,\mathbb{R})$ et il existe donc, par le théorème de Stone-Weierstrass 11.317, une fonction $g_i\in C^\infty(B,\mathbb{R})$ telle que $\|g_i-f_i\|_\infty\leqslant\epsilon$.

La fonction g dont les composantes sont les g_i ainsi construits vérifie $\|g-f\|_\infty\leqslant n\epsilon$. □

Attention toutefois que rien n'assure que les fonctions construites par le corolaire 11.318 prennent leurs valeurs dans B.

Le théorème suivant est un des énoncés les plus classiques de Stone-Weierstrass. Il découle évidemment du théorème général 11.317 (encore qu'il faut alors bien comprendre qu'il faut traiter la fonction $x\mapsto\sqrt{x}$ séparément). Il en existe cependant une preuve indépendante.

Théorème 11.319.
Soit f, une fonction continue de l'intervalle compact $[a,b]$ à valeurs dans $\mathbb{R}$. Alors pour tout $\epsilon>0$, il existe un polynôme P tel que $\|P-f\|_\infty<\epsilon$.

Autrement dit, les polynômes sont denses dans $C[a,b]$ pour la norme uniforme.

11.33 Primitive de fonction continue

Proposition 11.320 ([254]).
Soit un intervalle compact K de $\mathbb{R}$ et une suite (f_n) de fonctions continues sur K telles que $f_n\xrightarrow{unif}f$. Si chacune des fonctions f_n a une primitive sur K alors f également.

Démonstration. Soit $x_0\in K$ et les primitives F_n choisies[74] pour avoir $F_n'f_n$ et $F_n(x_0)=0$. Nous allons voir que (F_n) est une suite de Cauchy dans $\big(K,\|.\|_\infty\big)$. Soient $n,m\in\mathbb{N}$ et $x\in K$. Nous avons

$$\begin{aligned}\|F_n-F_m\|_\infty&\leqslant\|F_n(x)-F_m(x)\|&(11.846a)\\&=\|(F_n-F_m)(x)\|&(11.846b)\\&\leqslant\|F_n'-F_m'\|_{[x,x_0]}\|x-x_0\|&(11.846c)\end{aligned}$$

où nous avons utilisé le théorème des accroissements finis 10.145. Vu que $x\in K$ et que K est borné, $\|x-x_0\|$ est majoré par $\operatorname{diam}(K)$ et

$$\|F_n-F_m\|_K\leqslant\|f_n-f_m\|_K\operatorname{diam}(K).\tag{11.847a}$$

Vu que (f_n) est de Cauchy, si n et m sont assez grands, cela tend vers zéro. La suite (F_n) converge donc vers une certaine fonction F.

Le théorème 11.312 nous permet de permuter la limite et la dérivée pour conclure que $F'=f$ et donc que f a une primitive sur K. □

Proposition 11.321 ([254]).
Soit un intervalle ouvert I de $\mathbb{R}$ et une fonction $f\colon I\to\mathbb{R}$ qui admet une primitive sur tout compact de I. Alors f a une primitive sur I.

Démonstration. Nous considérons une suite exhaustive[75] de compacts K_n pour I et $x_0\in K_0$. Nous considérons aussi F_n la primitive de f sur K_n telle que $F_n(x_0)=0$ (possible parce que $x_0\in K_n$ pour tout n). Les fonctions F_n sont des restrictions les unes des autres, et nous pouvons définir

$$\begin{aligned}F\colon I&\to\mathbb{R}\\x&\mapsto F_n(x)\text{ si }x\in K_n.\end{aligned}\tag{11.848}$$

74. Les fonctions F_n étant dérivables sont continues.
75. Voir le lemme 7.312.

Nous avons évidemment $F(x_0) = 0$ et nous allons prouver que F est une primitive de f sur I. Soit $x \in I$ vu que I est ouvert, nous pouvons choisir n_0 tel que $x \in \mathrm{Int}(K_{n_0})$. Les fonctions F et F_{n_0} sont égales sur K_n et donc sur un ouvert autour de x. Par conséquent F est dérivable en x et $F'(x) = F'_{n_0}(x) = f(x)$. □

Théorème 11.322.
Soit I un intervalle ouvert de $\mathbb{R}$. Une fonction continue sur I admet une primitive[76] *sur I.*

Démonstration. Sur chaque compact de I, la fonction f est limite uniforme de polynômes[77] (théorème de Stone-Weierstrass 11.319). Donc f est primitivable sur tout compact de I (proposition 11.320) et donc sur I par la proposition 11.321. □

11.34 La fonction puissance

Si x et y sont des réels, définir x^y n'est pas une mince affaire. Pour l'instant nous savons déjà définir x^n lorsque $x \in \mathbb{R}$ et $n \in \mathbb{N}$. Voir la définition 1.95 et le thème 33.

Pour la suite nous notons

$$\begin{cases} f_\alpha(x) = x^\alpha & \text{(11.849a)} \\ g_a(x) = a^x & \text{(11.849b)} \end{cases}$$

pour autant que ces fonctions sont définies[78].

11.34.1 Sur les naturels

Définition 11.323.
La fonction puissance définie sur $\mathbb{N}$ s'étend à $\mathbb{Z}$ de la façon suivante :

$$x^{-n} = \frac{1}{x^n} \tag{11.850}$$

pour $n \geqslant 0$. Cela donne donne donc x^n pour $x \in \mathbb{R}$ et $n \in \mathbb{Z}$ a l'exception de $x = 0$ lorsque $n < 0$.

Nous étudions quelques propriétés de cette fonction pour $n > 0$ fixé.

Proposition 11.324.
Soit $n \in \mathbb{N}\backslash\{0\}$; nous posons $f_n(x) = x^n$.

Si n est pair,

$$f_n \colon [0, \infty[\to [0, \infty[\tag{11.851}$$

est bijective.

Si n est impair,

$$f_n \colon \mathbb{R} \to \mathbb{R} \tag{11.852}$$

est bijective.

Toutes les fonctions f_n sont continues sur $\mathbb{R}$.

Démonstration. En plusieurs morceaux, pas spécialement dans l'ordre auquel on s'attend.

Continuité Soit $x \in \mathbb{R}$. En vertu de 11.63 nous allons prouver que $\lim_{\epsilon \to 0} f_n(x+\epsilon) = f_n(x)$. Pour cela nous utilisons la formule du binôme 3.47 avec $x, h > 0$:

$$f_n(x+h) = (x+h)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} h^k. \tag{11.853}$$

76. Définition 11.136.
77. Si tu veux te passer de Stone-Weierstrass, tu peux prouver que toute fonction continue sur un compact est limite uniforme de fonctions affines par morceaux, par exemple. Voir [254].
78. L'objet des pages suivantes est de déterminer pour quelles valeurs de a, α et x nous pouvons trouver des définitions raisonnables pour ces fonctions.

Nous fixons $x_0 \in \mathbb{R}$. Calcul :

$$|f_n(x_0+h)-f_n(x)| = |\sum_{k=1}^{n}\binom{n}{k}x_0^{n-k}h^k| \tag{11.854a}$$

$$\leqslant \sum_{k=1}^{n}\binom{n}{k}|x_0|^{n-k}|h|^k \tag{11.854b}$$

$$= h\sum_{k=1}^{n}\binom{n}{k}|x_0|^{n-k}|h|^{k-1} \tag{11.854c}$$

$$\leqslant h\sum_{k=1}^{n}\binom{n}{k}|x_0|^{n-k}. \tag{11.854d}$$

Justifications :

— Le terme $k=0$ est égal à $x^n = f_n(x)$ parce que $\binom{n}{0}=1$.

— Dans la somme nous avons majoré $|h|$ par 1, opération justifiée par le fait que nous ayons dans l'idée de faire $h \to 0$.

Nous avons donc

$$\lim_{h\to 0}|f_n(x_0+h)-f_n(x)| \leqslant \lim_{h\to 0} h\sum_{k=1}^{n}\binom{n}{k}|x_0|^{n-k} = 0. \tag{11.855}$$

D'où la continuité de f_n en tout point $x_0 \in \mathbb{R}$.

Pour n pair ou impair, bijection sur les positifs Ceci sera déjà le résultat complet pour les n pairs, et a moitié du résultat pour les n impairs.

Stricte croissance Soit $n \neq 0$ dans $\mathbb{N}$. Nous commençons par prouver que f_n est strictement croissante sur $[0,\infty[$. Nous repartons de la formule du binôme, mais cette fois, nous séparons les termes $k=0$ et $k=n$ des autres (si $n=1$, il y a un peu de réécriture) en tenant compte de $\binom{n}{0}=\binom{n}{n}=1$:

$$f_n(x+h) = x^n + h^n + \sum_{k=1}^{n-1}\binom{n}{k}x^{n-k}h^k > x^n = f_n(x). \tag{11.856}$$

Vous noterez que l'inégalité est stricte même si $n=1$.

Vu que nous avons stricte monotonie, le théorème 11.61(2) nous dit que

$$f_n\colon [0,\infty[\to f_n\big([0,\infty[\big) \tag{11.857}$$

est une bijection.

Bijection Nous prouvons que $f_n\big([0,\infty[\big) = [0,\infty[$. Si $x>0$ alors $f_n(x)>0$, cela prouve une inclusion.

Pour l'autre inclusion nous savons que $f_n(x) > x$ dès que $x>1$. Donc $\lim_{x\to\infty} f_n(x) = \infty$. Si $y \in [0,\infty[$, alors il existe x_0 tel que $f_n(x_0) > y$. Étant donné que $f_n(0)=0$ et que nous avons déjà prouvé que f_n était continue (proposition 11.324), le théorème des valeurs intermédiaires 11.50 nous indique l'existence de $x_1 \in [0,x_0[$ tel que $f_n(x_1)=y$.

Nous avons prouvé que pour tout n, la fonction

$$f_n\colon [0,\infty[\to [0,\infty[\tag{11.858}$$

est une bijection.

Pour n impair Nous montrons à présent que si n est impair, alors

$$f_n\colon]-\infty,0] \to]-\infty,0] \tag{11.859}$$

est une bijection.

Tout se base sur le fait que si $x>0$ alors $f_n(-x) = -f_n(x)$. Le fait que (11.858) soit injective et surjective montre alors tout de suite le fait que (11.859) soit également injective et surjective.

□

Vous noterez que la continuité de f_n démontrée dans la proposition 11.324 est indépendant de la proposition 11.72 qui sera invoquée plus tard pour définir a^x lorsque $a > 0$ dans $\mathbb{R}$.

11.34.2 Sur les rationnels, racines

L'existence, pour tout réel $a \geqslant 0$, d'une réel r tel que $r^2 = a$ est déjà faite en la proposition 1.166.

Définition 11.325 (Exposant rationnels)**.**
La proposition 11.324 nous dit entre autres que pour tout $n \in \mathbb{N}$, la fonction

$$\begin{aligned} f_n \colon [0,\infty[&\to [0,\infty[\\ x &\mapsto x^n \end{aligned} \tag{11.860}$$

est bijective. Nous définissions alors, pour $a \in [0,\infty[$,

$$a^{1/n} = f_n^{-1}(a). \tag{11.861}$$

Autrement dit, le nombre $a^{1/n}$ est l'unique solution positive de

$$x^n = a. \tag{11.862}$$

11.326.
Nous ne définissons pas $a^{1/n}$ pour $a < 0$, du moins pas encore. Vu que f_3 est bijective sur $\mathbb{R}$, il serait tentant de définir $(-1)^{1/3} = f_3^{-1}(-1) = -1$.

Cela causera un certain nombre de problèmes plus tard vu que nous aurons envie de deux choses en même temps :

— d'une part $\ln(-1) = i\pi$,
— d'autre part, $a^x = e^{x\ln(a)}$.

De cette façon, nous devrions avoir

$$(-1)^{1/3} = e^{i\pi/3}, \tag{11.863}$$

qui est un nombre complexe non réel. Voici un exemple de ce que ça donne avec Sage :

```
sage: a=(-1)**(1/3)
sage: a.real_part()
1/2
sage: a.imag_part()
1/2*sqrt(3)
```

tex/sage/sageSnip019.sage

Définition 11.327 (Racince)**.**
Pour $n \in \mathbb{N}$ nous définissons $\sqrt[n]{x} = f_n^{-1}(x)$. Lorsque n est pair, la fonction $x \mapsto \sqrt[n]{x}$ n'est définie que sur $\mathbb{R}^+$, et lorsque n est impair, elle est définie sur tout $\mathbb{R}$.

11.328.
Notons que les fonctions $x \mapsto \sqrt[3]{x}$ et $x \mapsto x^{1/3}$ ne sont pas les mêmes : la première est définie sur tout $\mathbb{R}$ et donne des valeurs réelles tandis que la seconde n'est (pour l'instant) définie que sur les positifs, et donnera (quand on l'aura définie par l'exponentielle) des nombres complexes sur les négatifs.

En suivant cette convention, c'est-à-dire en réservant la notation $\sqrt{}$ pour l'inverse de f_2, nous ne devrions pas écrire des choses comme « $\sqrt{-1} = i$ », mais plutôt « $(-1)^{1/2} = i$ ». En effet, $\sqrt{-1}$ n'est pas défini et ne sera jamais défini alors que $(-1)^{1/2}$ n'est pas encore défini, mais sera défini par

$$(-1)^{1/2} = e^{\frac{1}{2}\ln(-1)} = e^{i\pi/2} = i. \tag{11.864}$$

En résumé, nous avons les fonctions suivantes :

(1) $\sqrt[n]{}: \mathbb{R} \to \mathbb{R}$ si n est impair,

(2) $\sqrt[n]{}: [0, \infty[\to [0, \infty[$ si n est pair,

(3) $x^{1/n}: [0, \infty[\to [0, \infty[$ pour tout $n \in \mathbb{N}$.

Cependant nous n'hésiterons pas à utiliser la notation $\sqrt{x}$ pour $x^{1/2}$ même lorsque x est négatif, parce c'est une notation très pratique. Il faut garder en tête que cette façon de faire est incohérente parce qu'elle inciterait à penser que $\sqrt[3]{-1} = e^{i\pi/3}$ au lieu de $\sqrt[3]{-1} = -1$.

Pour toute la suite de cette section, nous allons considérer a^x uniquement pour $a > 0$.

Définition 11.329.
Pour $m, n \in \mathbb{N}$ nous définissons

$$a^{m/n} = (a^m)^{1/n}, \tag{11.865}$$

ce qui définit la fonction puissance sur $\mathbb{Q}^+$. Enfin nous posons

$$a^{-q} = \frac{1}{a^q} \tag{11.866}$$

lorsque $q \in \mathbb{Q}^+$.

Et avec tout ça, lorsque $a > 0$ nous avons défini a^q pour tout $q \in \mathbb{Q}$.

Nous allons souvent noter la définition (11.865) sous la forme

$$f_{m/n}(x)^n = x^m. \tag{11.867}$$

Lemme 11.330 ([1]).
Pour $a > 0$ et $p, q \in \mathbb{Z}$ nous avons :

$$a^{p/q} = (a^p)^{1/q} = (a^{1/q})^p. \tag{11.868}$$

Démonstration. Nous divisons la preuve en fonction de la positivité du numérateur et du dénominateur.

Numérateur et dénominateurs positifs Nous commençons avec $p, q \in \mathbb{N}$. La première égalité est la définition 11.325. Pour la seconde, la définition de $(a^p)^{1/q}$ est d'être le $x > 0$ tel que

$$x^q = a^p. \tag{11.869}$$

La définition de $a^{1/q}$ est d'être le $y > 0$ tel que

$$y^q = a. \tag{11.870}$$

Ce y vérifie donc aussi $y^{pq} = a^p$ et donc $(y^p)^q = a^p$. Autrement dit, $y^p = x$, c'est-à-dire exactement

$$(a^{1/q})^p = (a^p)^{1/q}. \tag{11.871}$$

Le lemme est prouvé dans le cas où $p, q \in \mathbb{N}$.

Numérateur et dénominateur négatifs Si p et q sont tous les deux négatifs, nous remarquons que $p/q = (-p)/(-q)$ et nous sommes dans le même cas qu'avant.

Numérateur négatif, dénominateur positif Pour simplifier les notations nous supposons toujours $p, q \in \mathbb{N}$ mais nous considérons $a^{(-p)/q}$. Nous avons d'une part :

$$a^{(-p)/q} = a^{-(p/q)} = \frac{1}{a^{p/q}} = \frac{1}{(a^{1/q})^p} = (a^{1/q})^{-p}. \tag{11.872}$$

Dans ce calcul, nous avons utilisé au dénominateur le résultat dans le cas positif.

Et d'autre part nous avons :

$$(a^{-p})^{1/q} = \left(\left(\frac{1}{a}\right)^p\right)^{1/q} = \left(\left(\frac{1}{a}\right)^{1/q}\right)^p = \left(\frac{1}{a^{1/q}}\right)^p = \frac{1}{(a^{1/q})^p} = (a^{1/q})^{-p} \tag{11.873}$$

où nous avons utilisé le résultat avec $1/a$ en guise de a.

Numérateur positif, dénominateur négatif Nous traitons maintenant $a^{p/(-q)}$. Nous avons d'une part

$$a^{p/(-q)} = a^{-(p/q)} = \frac{1}{a^{p/q}} = \frac{1}{(a^p)^{1/q}} = (a^p)^{-(1/q)} = (a^p)^{1/(-q)}. \tag{11.874}$$

Et d'autre part :

$$a^{p/(-q)} = \frac{1}{a^{p/q}} = \frac{1}{(a^{1/q})^p} = \left(\frac{1}{a^{1/q}}\right)^p = \left(a^{-(1/q)}\right)^p = (a^{1/(-q)})^p. \tag{11.875}$$

□

Le lemme suivant montre que la définition sur $\mathbb{Q}^-$ est cohérente avec celle sur $\mathbb{Q}^+$, au sens où finalement nous retrouvons que $a^{m/n}$ vérifie $x^n = a^m$ quel que soient les signes de m et n.

Lemme 11.331 ([1]).
Le nombre $y = a^{-m/n}$ vérifie l'équation $y^{-n} = a^m$

Démonstration. Nous posons $x = a^{m/n}$, c'est-à-dire $x^n = a^m$. Nous avons, par définition $y = a^{-m/n} = \frac{1}{x}$. Alors

$$y^{-n} = \frac{1}{\left(\frac{1}{x}\right)^n} = x^n = a^m, \tag{11.876}$$

donc c'est bon. □

Lemme 11.332 ([1]).
Pour $a > 0$ et $q, q' \in \mathbb{Q}$ nous avons

$$a^q a^{q'} = a^{q+q'}. \tag{11.877}$$

Démonstration. Nous mettons q et q' au même dénominateur. Soient $q = s/c$ et $q' = r/c$ avec $s, r \in \mathbb{Z}$ et $c \in \mathbb{N}$. En utilisant les égalités du lemme 11.330 nous trouvons

$$a^{s/c}a^{r/c} = (a^{1/c})^s(a^{1/c})^r = (a^{1/c})^{s+r} = a^{(s+r)/c} = a^{q+q'}. \tag{11.878}$$

□

Lemme 11.333 ([1]).
La fonction puissance prend les valeurs suivantes.

(1) Si $a = 1$ alors $a^q = 1$ pour tout $q \in \mathbb{Q}$.

(2) Si $a > 1$ alors
- *$a^q > 1$ si $q > 0$*
- *$a^q < 1$ si $q < 0$*
- *$a^0 = 1$.*

(3) Si $a < 1$ alors
- *$a^q < 1$ si $q > 0$*
- *$a^q > 1$ si $q < 0$*
- *$a^0 = 1$.*

Démonstration. Si $a = 1$ alors $a^k = 1$ pour tout $k \in \mathbb{N}$. Ensuite, pour $m, n \in \mathbb{N}$, $a^{n/m}$ est solution de $x^m = a^n = 1$, donc $x = 1$. En ce qui concerne les puissances négatives, $1/1 = 1$.

Si $a > 1$ alors $a^k > 1$ pour tout $k \in \mathbb{N}$. De plus pour $q > 0$ nous avons $q = m/n$ avec $m, n \in \mathbb{N}$. Alors $a^{m/n}$ est solution de $x^m = a^n > 1$. Or pour $x \leqslant 1$ nous avons $x^m \leqslant 1$, donc la solution à $x^m = a^n$ vérifie forcément $x > 1$.

Toujours avec $a > 1$, si $q < 0$ nous posons $q = -q'$ avec $q' > 0$. Alors

$$a^q = q^{-q'} = \frac{1}{a^{q'}}. \tag{11.879}$$

Mais $a^{q'} > 1$, donc l'inverse est inférieur à 1.

En ce qui concerne les cas $a < 1$, ils sont obtenus en posant $b = 1/a$ et en calculant

$$a^q = \left(\frac{1}{b}\right)^q = \frac{1}{b^q} = b^{-q}. \tag{11.880}$$

□

Proposition 11.334 ([1]).
Si $a > 1$ et si $M > 0$, il existe $n \in \mathbb{N}$ tel que $a^n > M$.

Démonstration. Soit $a = 1 + h$. Alors en utilisant la formule du binôme,

$$a^n = (1+h)^n = \sum_{k=0}^{n} \binom{n}{k} h^{n-k}. \tag{11.881}$$

Tous les termes de la somme sont strictement positifs. Prenons le terme $k = n - 1$. Il vaut

$$\binom{n}{n-1} h = nh. \tag{11.882}$$

Donc $a^n \geqslant nh$, donc oui, cela peut être rendu arbitrairement grand avec n sans toucher à a. □

Proposition 11.335 ([1]).
Pour $a > 0$ nous considérons la fonction

$$\begin{aligned} g_a \colon \mathbb{Q} &\to \mathbb{R} \\ q &\mapsto a^q. \end{aligned} \tag{11.883}$$

(1) Si $a \in]0, 1[$ alors g_a est décroissante et

$$\lim_{q\to\infty} g_a(q) = 0, \qquad \lim_{q\to-\infty} g_a(q) = \infty. \tag{11.884a}$$

(2) Si $a > 1$ alors g_a est croissante et

$$\lim_{q\to\infty} g_a(q) = \infty, \qquad \lim_{q\to-\infty} g_a(q) = 0. \tag{11.885a}$$

Démonstration. Nous prouvons le cas $a > 1$. L'autre cas s'en déduit en posant $b = 1/a$. Pour la croissance, soient $q \in \mathbb{Q}$ et $r > 0$ dans $\mathbb{Q}$. En utilisant le lemme 11.332, nous avons

$$a^{q+r} = a^q a^r > a^q \tag{11.886}$$

parce que $a^r > 1$ par le lemme 11.333.

En ce qui concerne la limite $q \to \infty$, la fonction g_a est croissante et non bornée par la proposition 11.334. Donc sa limite est ∞.

Pour la limite $q \to -\infty$, nous avons

$$\lim_{q\to-\infty} a^q = \lim_{q\to\infty} a^{-q} = \lim_{q\to\infty} \frac{1}{a^q} = 0. \tag{11.887}$$

□

Proposition 11.336 ([1]).
Soit $a > 0$. Nous avons

$$\lim_{q\to 0} a^q = 1. \tag{11.888}$$

Notons que cette limite est une limite dans $\mathbb{Q}$ parce que nous n'avons même pas encore défini a^x lorsque x est irrationnel.

Démonstration. Nous notons, comme à l'accoutumée, $g_a(x) = a^x$. Soit une suite $x_k \to 0$ (avec $x_k \neq 0$ pour tout k). En définissant y_k par $x_k = 1/y_k$ nous savons que a^{1/y_k} est la solution de $x^{y_k} = a$.

Nous posons $t_k = a^{x_k}$ et notre but est de prouver que $t_k \to 1$. Pour tout k nous avons la relation

$$t_k^{y_k} = a. \tag{11.889}$$

Soit $s > 1$. Il existe un $M > 0$ tel que $y_k > M$ implique $s^{y_k} > a$ (proposition 11.334). Donc dès que $y_k > M$ nous avons $t_k < s$.

De la même manière, si $r < 1$, il existe un $R > 0$ tel que $y_k > R$ implique $r^{y_k} < a$. Donc dès que $y_k > R$ nous avons $t_k > r$.

Soit donc un voisinage $]r, s[$ de 1 (avec $r < 1$ et $s > 1$). Nous avons les nombres M et R correspondant et nous posons $L = \max\{M, R\}$. Soit K tel que $k > K$ implique $y_k > L$. Alors pour $k > K$ nous avons aussi $t_k < s$ et $t_k > r$, c'est-à-dire $t_k \in]r, s[$.

Cela prouve que $t_k \to 1$.

Donc pour toute suite $x_k \to 0$ nous avons $g_a(x_k) \to 1$. Par le critère séquentiel de la limite (proposition 7.263) nous avons $\lim_{x\to 0} g_a(x) = 1$. □

Lemme 11.337.
Soit $a > 0$. La fonction

$$\begin{aligned} g_a \colon \mathbb{Q} &\to \mathbb{R} \\ x &\mapsto a^x \end{aligned} \tag{11.890}$$

est continue.

Démonstration. Soient $x \in \mathbb{Q}$ et une suite $x_k \to 0$ (toujours dans $\mathbb{Q}$) et utilisons le lemme 11.332 :

$$a^{x+x_k} = a^x a^{x_k}. \tag{11.891}$$

Cela est, dans $\mathbb{R}$, le produit entre une constante (a^x) et une suite. La limite est donc le produit de cette constante et la limite de la suite (si elle existe). Par la proposition 11.336 nous avons la limite $a^{x_k} \to 1$, et donc

$$\lim_{k\to\infty} a^{x+x_k} = a^x, \tag{11.892}$$

ce qui prouve la continuité (caractérisation séquentielle, proposition 7.80) de g_a. □

Proposition 11.338.
Soit $a > 0$ dans $\mathbb{R}$. La fonction

$$\begin{aligned} g_a \colon \mathbb{Q} &\to \mathbb{R} \\ q &\mapsto a^q \end{aligned} \tag{11.893}$$

est Cauchy-continue.

Démonstration. En quelque étapes.

<u>**Pour** $a > 1$</u> Avant de nous lancer dans la preuve directe, nous prouvons une petite formule. Soit $\epsilon > 0$. Vu que, par la proposition 11.336, $\lim_{q\to 0} g_a(q) = 1$, il existe $\delta > 0$ tel que $0 < q < \delta$ implique $|1 - g_a(q)| < \epsilon$.

Soient maintenant $p, q \in \mathbb{Q}$ tels que $|p - q| < \delta$. En utilisant de plus la définition (11.866) et la formule du lemme 11.332,

$$|g_a(q) - g_a(p)| = |g_a(q)| \left| 1 - \frac{g_a(p)}{g_a(q)} \right| = |g_a(p)||1 - g_a(p - q)| \leqslant |g_a(q)|\epsilon. \tag{11.894a}$$

Nous y allons pour la preuve directe. Soit une suite de Cauchy (q_n) dans $\mathbb{Q}$. Nous devons prouver que la suite $n \mapsto g_a(q_n)$ est de Cauchy dans $\mathbb{R}$. Soit $\epsilon > 0$.

La suite (q_n) étant de Cauchy dans $\mathbb{Q}$, elle l'est également dans $\mathbb{R}$, elle est bornée parce que convergente vu que $\mathbb{R}$ est complet[79]. Vu que g_a est croissante[80] et que (q_n) est bornée, il existe M tel que $|g_a(q_n)| \leqslant M$ pour tout n.

Nous considérons δ tel que $0 < q < \delta$ implique $|1 - g_a(q)| \leqslant \epsilon$, ainsi que N tel que $i, j > N$ implique $|q_i - q_j| \leqslant \delta$ (là nous utilisons le fait que (q_n) est de Cauchy). Pour de tels N, i, j nous avons

$$|g_a(q_i) - g_a(q_j)| \leqslant M\epsilon. \tag{11.895}$$

Donc la suite $g_a(g_n)$ est de Cauchy.

Pour $a = 1$ La fonction g_a est constante.

Pour $0 \leqslant a < 1$ J'imagine que ça se fait comme $a > 1$, mais en renversant quelque inégalités[81].

□

11.339.
L'ingrédient magique qui fait fonctionner la proposition 11.338 est le fait que $g_a(x+y) = g_a(x)g_a(y)$ couplé au fait que $\lim_{q\to 0} g_a(q) = 1$. C'est cela qui débloque la situation pour étendre la fonction puissance de $\mathbb{Q}$ vers $\mathbb{R}$ en utilisant le lemme 11.72.

Le chemin suivit par [255] pour étendre la fonction puissance de $\mathbb{Q}$ vers $\mathbb{R}$ est un peu différent : il définit $a^x = \sup\{a^q \text{ tel que } q < x, q \in \mathbb{Q}\}$. La preuve que cette définition donne $x \mapsto a^x$ continue sur $\mathbb{R}$ repose, elle aussi, essentiellement sur le fait que $\lim_{q\to 0} a^q = 1$.

Il y a donc une certaine justice.

Proposition-définition 11.340 (Fonction puissance[1])**.**
Si $a > 0$ et $x \in \mathbb{R}$, la fonction

$$\begin{aligned} g_a \colon \mathbb{Q} &\to \mathbb{R} \\ x &\mapsto a^x. \end{aligned} \tag{11.896}$$

est Cauchy-continue par la proposition 11.338. Si $x \in \mathbb{R}$ nous définissons

$$a^x = \tilde{g}_a(x) \tag{11.897}$$

où $\tilde{g}_a$ est l'extension de g_a donnée par le lemme 11.72.

Nous allons la noter g_a également, et écrire a^x la valeur de g_a même lorsque x n'est pas un rationnel.

Proposition 11.341 ([1])**.**
Quelques propriétés de la fonction puissance.

(1) Pour $a > 0$, la fonction $g_a \colon x \mapsto a^x$ est continue sur $\mathbb{R}$.

(2) Pour $a > 1$, la fonction $g_a \colon x \mapsto a^x$ est croissante.

(3) Pour $a > 0$ et $x, y \in \mathbb{R}$ nous avons

$$a^x a^y = a^{x+y}. \tag{11.898}$$

En particulier,

$$a^{-x} = \frac{1}{a^x}. \tag{11.899}$$

Démonstration. La continuité de $x \mapsto a^x$ est par construction. Le point (1) est fait.

Pour le point (2), lorsque $a > 1$, la fonction $f_a \colon \mathbb{Q} \to \mathbb{R}$ est croissante (proposition 11.335). Donc par la proposition 11.74, la fonction $x \mapsto a^x$ est croissante sur $\mathbb{R}$.

79. Théorème 7.294.
80. Proposition 11.335(2).
81. Je n'ai pas essayé. Faites-le et écrivez-moin pour me dire si ça marche.

Et enfin pour le point (3), il faut faire un peu plus attention. Soient des suites $x_i \to x$ et $y_i \to y$ dans $\mathbb{Q}$. Calculons :

$$\begin{aligned} a^x a^y &= (\lim_i a^{x_i}) a^y && (11.900a)\\ &= \lim_i \big(a^{x_i} a^y\big) && (11.900b)\\ &= \lim_i \big(\lim_k a^{x_i} a^{y_k}\big) && (11.900c)\\ &= \lim_i \big(\lim_k a^{x_i+y_k}\big) && (11.900d)\\ &= \lim_i a^{x_i+y} && (11.900e)\\ &= a^{x+y}. && (11.900f) \end{aligned}$$

Justifications :

— Pour 11.900a. Définition de a^x lorsque $x \in \mathbb{R}$.

— Pour 11.900b. Nous entrons le nombre a^y dans la limite. Entrer un facteur dans une limite convergente dans $\mathbb{R}$ est un acte anodin.

— Pour 11.900c. Définition de a^y, et renter le nombre réel a^{x_i} dans la limite sur k.

— Pour 11.900d. Utilisation du lemme 11.332, valable pour $x_i, y_k \in \mathbb{Q}$.

— Pour 11.900e. Pour i fixé, la suite $k \mapsto x_i + x_k$ est une suite de rationnels qui converge vers le réel $x_i + y$. Par définition 11.340 de la fonction puissance nous avons alors $\lim_k a^{x_i+y_k} = a^{x_i+y}$.

— Pour 11.900f. La suite de réels $i \mapsto x_i + y$ converge dans $\mathbb{R}$ vers le réel $x + y$. Par la continuité de $t \mapsto a^t$ (ça fait partie du lemme 11.72 définissant la fonction puissance sur $\mathbb{R}$) nous avons $\lim_i a^{x_i+y} = a^{x+y}$.

Vous remarquerez que les limites sur k et sur i ne s'enlèvent pas tout à fait avec la même justification. Nous aurions pu invoquer la continuité sur $\mathbb{R}$ de $t \mapsto a^t$ pour les deux limites. Mais cette continuité, dans le cas d'une suite purement constituée de rationnels, est la définition de la prolongation vers $\mathbb{R}$. □

Lemme 11.342.
Soient $a, b > 0$. Si $1 < x < y$ alors

$$a - b < ay - bx. \tag{11.901}$$

Démonstration. Nous posons $y = x + s$ avec $s > 0$. Alors

$$ay - bx = a(x + s) - bx = (a - b)x + as > (a - b)x > a - b \tag{11.902}$$

parce que $as > 0$ et $x > 1$. □

Proposition 11.343 ([1]).
Pour $q > 0$ dans $\mathbb{Q}$, la fonction

$$\begin{aligned} f_q \colon \mathbb{Q}^+ &\to \mathbb{R}\\ x &\mapsto x^q \end{aligned} \tag{11.903}$$

est strictement croissante.

Démonstration. Division selon la généralité de q.

Si q est entier positif Soit $q = n \in \mathbb{N}$. Si $s > 0$ alors l'inégalité $(x + s)^n > x^n$ découle du binôme de Newton de la proposition 3.47.

Si q est rationnel Soient un rationnel $q = m/n$ et un nombre strictement positif s. Nous avons, par la définition 11.325 sous la forme (11.867) :

$$f_{m/n}(x + s)^n = (x + s)^m > x^m = f_{m/n}(x)^n. \tag{11.904}$$

Nous avons utilisé la stricte croissance de $x \mapsto x^m$. Cela donne

$$f_{m/n}(x + s)^n > f_{m/n}(x)^n. \tag{11.905}$$

En utilisant encore la stricte croissance de $x \mapsto x^n$, nous avons le résultat.

□

Corolaire 11.344.
Soient $1 < b < a$ dans $\mathbb{R}$ et des rationnels strictement positifs $p < q$. Alors

$$a^p - b^p < a^q - b^q \tag{11.906}$$

Démonstration. Nous notons $q = p + r$ avec $r > 0$ dans $\mathbb{Q}$. Par la proposition 11.343,

$$a^r > b^r. \tag{11.907}$$

Cela nous permet d'utilise le lemme 11.342 pour écrire

$$a^p - b^p < a^p a^r - b^p b^r = a^q - b^q. \tag{11.908}$$

□

Proposition 11.345 ([1]).
Soient $a, b > 0$ et $\alpha \in \mathbb{R}$. Nous avons :

$$a^\alpha b^\alpha = (ab)^\alpha. \tag{11.909}$$

Démonstration. Nous supposons que c'est bon pour $\alpha \in \mathbb{N}$ et $\alpha \in \mathbb{Z}$. Pour les autres, nous donnons plus de détails.

<u>$\mathbb{Q}^+$</u> Soit $q = m/n$ avec $m, n \in \mathbb{N}$. Si $a^{m/n} = x$ et $b^{m/n} = y$, alors

$$x^n = a^m \tag{11.910a}$$
$$y^n = b^m \tag{11.910b}$$

par (11.867). Nous multiplions (11.910a) par y^n à gauche et par b^m à droite : $x^n y^n = a^m b^m$. En tenant compte du résultat pour $\mathbb{N}$, nous avons

$$(xy)^n = (ab)^m, \tag{11.911}$$

ce qui signifie que le nombre xy est $(ab)^{m/n}$.

<u>Pour $\mathbb{Q}^-$</u> Soit $q \in \mathbb{Q}^+$, nous avons le calcul

$$a^{-q} b^{-q} = \frac{1}{a^q b^q} = \frac{1}{(ab)^q} = (ab)^{-q}. \tag{11.912}$$

<u>Pour $\mathbb{R}$</u> Soit une suite de rationnels $\alpha_i \to \alpha$. Nous avons

$$a^\alpha b^\alpha = \big(\lim_i a^{\alpha_i}\big)\big(\lim_j b^{\alpha_j}\big) = \lim_i \big(a^{\alpha_i} b^{\alpha_i}\big) = \lim_i (ab)^{\alpha_i} = (ab)^\alpha. \tag{11.913}$$

Justifications :
- la proposition 7.132 pour le produit des limites,
- le résultat dans $\mathbb{Q}$ que nous venons de prouver,
- la définition de $(ab)^\alpha$ comme limite de $(ab)^{\alpha_i}$.

□

Pour rappel, la proposition suivantes, dans le cas de $\alpha \in \mathbb{Q}^+$ est la proposition 11.343.

Proposition 11.346 ([1]).
Pour $\alpha > 0$, la fonction

$$\begin{aligned} f_\alpha \colon]0, \infty[&\to \mathbb{R} \\ x &\mapsto x^\alpha \end{aligned} \tag{11.914}$$

est strictement croissante.

Aussi, la fonction

$$\begin{aligned} f_\alpha \colon]-\infty, 0[&\to \mathbb{R} \\ x &\mapsto x^\alpha \end{aligned} \tag{11.915}$$

est strictement décroissante.

Démonstration. Nous rappellons que le cas $\alpha \in \mathbb{Q}^+$ est déjà traité par la proposition 11.343. Soient $x \in]0,\infty[$ et $s > 0$. Nous allons montrer que $f_\alpha(x+s) - f_\alpha(x) > 0$. Pour cela nous décomposons en plusieurs cas.

<u>$x > 1$</u> Par la proposition 1.152, nous considérons une suite strictement croissante de rationnels strictement positifs $\alpha_i \to \alpha$. Pour tout i nous avons $\alpha_i > \alpha_0$.

En utilisant la stricte croissance de f_{α_0} et le lemme 11.333(2), nous avons les inégalités $1 < x^{\alpha_0} < (x+s)^{\alpha_0}$, et en particulier

$$0 < (x+s)^{\alpha_0} - x^{\alpha_0}. \tag{11.916}$$

De plus nous avons $1 < x < x+s$ et $\alpha_0 < \alpha_i$ pour tout i. Donc le corolaire 11.344 s'applique et nous avons, pour tout i :

$$0 < (x+s)^{\alpha_0} - x^{\alpha_0} < (x+s)^{\alpha_i} - x^{\alpha_i}. \tag{11.917}$$

C'est le moment de passer à la limite $i \to \infty$. La seconde inégalité devient non stricte, mais la première reste :

$$0 < (x+s)^{\alpha_0} - x^{\alpha_0} \leqslant (x+s)^{\alpha} - x^{\alpha}. \tag{11.918}$$

Nous avons donc bien la stricte croissance de f_α sur $]1,\infty[$.

<u>$x \leqslant 1$</u> Nous choisissons encore $\alpha_i \to \alpha$ strictement croissante dans $\mathbb{Q}$. Pour chaque i, nous avons encore

$$(x+s)^{\alpha_i} - x^{\alpha_i} > 0. \tag{11.919}$$

Le passage à la limite change l'inégalité stricte en inégalité large, et ne permet donc pas de conclure immédiatement. Nous devons donc ruser. Soit $k \in \mathbb{N}$ tel que $k(x+s) > 1$ et $kx > 1$ (existence parce que $\mathbb{R}$ est archimédien, proposition 1.143). Nous avons :

$$\big(k(x+s)\big)^{\alpha} - (kx)^{\alpha} > 0 \tag{11.920}$$

par la partie « $x > 1$ » que nous venons de prouver. Grâce à la proposition 11.345 nous pouvons factoriser k^α :

$$0 < \big(k(x+s)\big)^{\alpha} - (kx)^{\alpha} = k^{\alpha}\big((x+s)^{\alpha} - x^{\alpha}\big). \tag{11.921}$$

Vu que $k^\alpha > 0$, cela implique $(x+s)^\alpha - x^\alpha > 0$, ce qu'il fallait.

Nous avons fini de prouver que la fonction f_α était strictement croissante sur $]0,\infty[$. En ce qui concerne la fonction f_α sur $]-\infty,0[$, nous avons, pour $x > 0$ que

$$f_\alpha(-x) = \frac{1}{f_\alpha(x)}, \tag{11.922}$$

et donc stricte décroissance. □

Nous prouvons à présent que f_α est localement injective ; nous en avons besoin pour prouver la continuité. Or cette continuité est nécessaire à prouver que f_α est localement bijective. Donc nous ne pouvons pas énoncer la bijectivité ici. Ce sera la proposition 11.350.

Proposition 11.347.
Soient $\alpha \in \mathbb{R}$ et $x \in \mathbb{R}\backslash\{0\}$. Il existe un voisinage V de x sur lequel

$$f_\alpha \colon V \to f_\alpha(V) \tag{11.923}$$

est injective.

Démonstration. Soit $x > 0$; nous considérons un voisinage V de x inclu à $]0,\infty[$. Soit $y \in V$; pour fixer les idées nous supposons $y < x$. Par la stricte croissance de f_α sur $]0,\infty[$ (proposition 11.346), nous avons $f_\alpha(y) < f_\alpha(x)$ et en particulier $f_\alpha(x) \neq f_\alpha(y)$.

Le cas $x < 0$ se traite de façon analogue, avec la stricte décroissance de f_α sur $]-\infty,0[$. □

Notons que les voisinages sur lesquels f_α est injective sont assez grands. Ils peuvent être toute une demi-droite, si l'on veut.

Lemme 11.348.
Soient $\alpha > 0$, une suite de rationnels strictement décroissante $\alpha_i \to \alpha$ ainsi que les fonctions

$$\begin{aligned} f_{\alpha_i} \colon]1,\infty[&\to \mathbb{R} \\ x &\mapsto x^{\alpha_i}. \end{aligned} \tag{11.924}$$

La famille $\{f_{\alpha_i}\}_{i\in\mathbb{N}}$ est équicontinue[82].

Démonstration. Soient $x > 1$, et $\alpha > 0$. Nous allons montrer que $\{f_{\alpha_i}\}$ est équicontinue en x. Soit s tel que $1 < x < x + s$; le corolaire 11.344 nous enseigne que

$$(x+s)^p - x^p < (x+s)^q - x^q \tag{11.925}$$

dès que $p < q$. En particulier, f_p étant croissante par la proposition 11.346,

$$0 < (x+s)^{\alpha_i} - x^{\alpha_i} < (x+s)^{\alpha_0} - x^{\alpha_0}. \tag{11.926}$$

Soit $\epsilon > 0$ et δ tel que $s < \delta$ implique $|(x+s)^{\alpha_0} - x^{\alpha_0}| < \epsilon$. Alors nous avons aussi, pour de tels σ et s :

$$|(x+s)^{\alpha_i} - x^{\alpha_i}| < |(x+s)^{\alpha_0} - x^{\alpha_0}| < \epsilon. \tag{11.927}$$

En procédant de même pour $s < 0$, nous trouvons bien que

$$|y^{\alpha_i} - x^{\alpha_i}| \leqslant \epsilon \tag{11.928}$$

pour tout $y \in B(x,\delta)$.

Cela signifie que $\{f_i\}$ est équicontinue. □

Proposition 11.349 ([1]).
Soit $\alpha > 0$ dans $\mathbb{R}$. La fonction

$$\begin{aligned} f_\alpha \colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto x^\alpha \end{aligned} \tag{11.929}$$

est continue (sauf pour $x = 0$ si $\alpha < 0$).

Démonstration. Nous allons subdiviser quelque cas.

Pour $\alpha \in \mathbb{N}$ Nous supposons que ce cas va bien.

Pour $\alpha \in \mathbb{Q}^+$ Soit $q = m/n$ avec $m, n \in \mathbb{N}$. Soit aussi $\epsilon > 0$. Nous avons :

$$f_{m/n}(x)^n = x^m \tag{11.930a}$$

$$f_{m/n}(x+\epsilon)^n = (x+\epsilon)^m. \tag{11.930b}$$

L'équation (11.930b) s'écrit aussi bien sous la forme

$$f_n\big(f_{m/n}(x+\epsilon)\big) = (x+\epsilon)^m. \tag{11.931}$$

En prenant la limite,

$$\lim_{\epsilon\to 0}\big[f_n\big(f_{m/n}(x+\epsilon)\big)\big] = x^m = f_{m/n}(x)^n. \tag{11.932}$$

Vu que f_n est continue, nous pouvons la permuter avec la limite dans le membre de gauche tout en écrivant $f_{m/n}(x)^n = f_n\big(f_{m/n}(x)\big)$ dans le membre de droite :

$$f_n\big[\lim_{\epsilon\to 0} f_{m/n}(x+\epsilon)\big] = f_n\big(f_{m/n}(x)\big). \tag{11.933}$$

La fonction f_n étant injective dans un voisinage autour de x (proposition 11.347),

$$\lim_{\epsilon\to 0} f_{m/n}(x+\epsilon) = f_{m/n}(x), \tag{11.934}$$

ce qui est la continuité de $f_{m/n}$ en x.

82. Définition 7.321.

Pour $\alpha \in \mathbb{R}^+$ Nous prouvons séparément le cas $x < 1$ et le cas $x \geqslant 1$. Commençons par $x \in]1, \infty[$.

Soit une suite $\alpha_i \to \alpha$ strictement décroissante dans $\mathbb{Q}^+$. Le lemme 11.348 nous dit que l'ensemble de fonctions $\{f_{\alpha_i} \colon]1, \infty[\to \mathbb{R}\}_{i \in \mathbb{N}}$ est équicontinu. La convergence simple $f_{\alpha_i} \to f_\alpha$ étant par définition, la proposition 7.322 nous dit que la fonction $f_\alpha \colon]1, \infty[\to \mathbb{R}$ est continue.

Soit maintenant $x \in]0, 1]$. Il existe $k \in \mathbb{N}$ tel que $kx > 1$, $k(x/2) > 1$ et $k^\alpha > 1$ (si vous pensez bien, seule la première condition est utile).

Nous considérons ϵ tel que $x + \epsilon > x/2$; de toutes façons nous comptions faire $\epsilon \to 0$. Nous avons :

$$\big|(x+\epsilon)^\alpha - x^\alpha\big| \leqslant k^\alpha \big|(x+\epsilon)^\alpha - x^\alpha\big| = \big|[k(x+\epsilon)]^\alpha - (kx)^\alpha\big|. \tag{11.935}$$

Nous prenons le δ qui correspond à ϵ en kx dans la continuité de f_α déjà démontée pour $kx > 1$. Alors si $\epsilon < \delta$ nous avons

$$\big|(x+\epsilon)^\alpha - x^\alpha\big| \leqslant \epsilon. \tag{11.936}$$

Pour $\alpha \in \mathbb{R}^-$ Si $\alpha > 0$, la fonction $f_{-\alpha}$ est donnée par

$$f_{-\alpha}(x) = \frac{1}{f_\alpha(x)} \tag{11.937}$$

et est donc continue (sauf en $x = 0$ où elle n'existe pas).

□

Proposition 11.350.
Soit $\alpha > 0$. La fonction

$$\begin{aligned} f_\alpha \colon [0, \infty[&\to [0, \infty[\\ x &\mapsto x^\alpha \end{aligned} \tag{11.938}$$

est bijective.

Proposition 11.351 ([1]).
Soient $a > 0$ ainsi que $x, y \in \mathbb{R}$. Alors

$$(a^x)^y = (a^y)^x = a^{xy}. \tag{11.939}$$

Démonstration. Nous découpons en fonction de la nature de x et y.

x rationnel, y naturel Si $q \in \mathbb{Q}$ et $n \in \mathbb{N}$ alors la formule

$$(a^q)^n = a^{nq} \tag{11.940}$$

découle seulement d'une récurrence sur la formule 11.898.

$x, y \in \mathbb{Q}$ Soient $y = m/n$ avec $n \in \mathbb{Z}$, $m \in \mathbb{N}$ et $q \in \mathbb{Q}$. Nous avons, en utilisant la partie déjà démontrée et le lemme 11.330,

$$(a^q)^p = (a)^{m/n} = \big((a^q)^m\big)^{1/n} = (a^{mq})^{1/n} = a^{mq/n} = a^{pq}. \tag{11.941}$$

x, y irrationnels Soient des suites des rationnels $x_i \to x$ et $y_i \to y$. En utilisant les définitions,

$$(a^x)^y = \lim_i (a^x)^{y_i} = \lim_i \big(\lim_j a^{x_j}\big)^{y_i}. \tag{11.942}$$

Fixons un i pour commencer. Nous avons, par la continuité de f_{y_i} (proposition 11.349)

$$\big(\lim_j a^{x_j}\big)^{y_i} = f_{y_i}\big(\lim_j a^{x_j}\big) = \lim_j \big(f_{y_i}(a^{x_j})\big) = \lim_j a^{x_j y_i}. \tag{11.943}$$

Nous avons utilisé le résultat déjà démontré dans le cas des rationnels. La suite $j \mapsto x_j y_i$ est une suite dans $\mathbb{Q}$ qui converge vers le réel xy_i, donc la limite sur j redonne la fonction puissance :

$$\big(\lim_j a^{x_j}\big)^{y_i} = \lim_j a^{x_j y_i} = a^{xy_i}. \tag{11.944}$$

Le résultat découle maintenant de la prise de limite dans (11.942) qui revient à prendre la limite $i \to \infty$ de l'expression dans (11.944) :

$$(a^x)^y = \lim_i \big(\lim_j a^{x_j}\big)^{y_i} = \lim_i a^{xy_i} = a^{xy}. \tag{11.945}$$

□

Le lemme suivant montre en gros que x^y croît plus rapidement en y qu'en x.

Lemme 11.352.
Pour tout $\alpha > 0$ et $a < 1$ nous avons la limite

$$\lim_{n\to\infty} n^\alpha a^n = 0 \tag{11.946}$$

Démonstration. Soit $k \in \mathbb{N}$ plus grand que α. Soit la suite numérique $s_n = n^k a^n$. Tous ses termes sont positifs et

$$\frac{s_n}{s_{n+1}} = \left(\frac{n}{n+1}\right)^k \frac{1}{a}. \tag{11.947}$$

Étant donné que $n/n+1 \to 1$ et que $a < 1$, il existe un certain rang à partir duquel la suite (s_n) est décroissante. Deux conclusions :

— Elle est majorée par une constante M.
— Elle est convergente par le lemme 7.137.

Soit l tel que $ka^l < 1$ et $n > l$ alors

$$s_{n+l} = (n+l)^k a^{n+l} \leqslant kn^k a^n a^l = ka^l s_n \leqslant ka^l M. \tag{11.948}$$

La majoration est due au fait que dans $(n+l)^k$ nous avons k termes tous plus petits que n^k. De la même façon,

$$s_{2n+2l} \leqslant ka^{2l} s_{2n} \leqslant ka^{2l} M. \tag{11.949}$$

En posant $\varphi(i) = in + il$ nous avons

$$s_{\varphi(i)} \leqslant ka^i M, \tag{11.950}$$

qui est une sous-suite convergente vers 0. Or si une suite est convergente (ce qui est le cas de (s_n)), toutes les sous-suites convergent vers la même limite. Nous en concluons que $s_n \to 0$. □

11.353.
Une conséquence est que si vous voulez choisir un mot de passe fort, la longueur du mot est plus importante que la taille de l'alphabet choisit : il est plus efficace de choisir une combinaison longue qu'une combinaisons mélangeant des lettres, chiffres et symboles spéciaux.

Exemple : si vous choisissez un mot de passe contenant majuscules, minuscules, chiffres et symboles spéciaux complètement mélangés (ne mentez pas, vous ne le faites pas), mais que vous ne le choisissez que de taille 6, vous avez 72^6 possibilités (en supposant un jeu de 10 symboles spéciaux).

Eh bien, en seulement 8 lettres minuscules, vous avez plus de possibilités : $26^8 > 72^6$.

De nombreux sites font l'erreur de considérer que

— « ggzxzheaiynshunxuydajkwyohgqxz » est un mot de passe faible,
— « azerty.2019A » est un mot de passe fort.

Il n'en est rien. Le premier est considérablement meilleur que le second, même si le second, très superficiellement, mélange les lettres majuscules, minuscules, chiffres et symboles spéciaux.

Voila voila. La prochaine fois qu'un site vous refusera un mot de passe de 30 lettres minuscules mélangées, vous saurez pourquoi il n'y a rien qui marche en informatique, et en particulier pourquoi la sécurité générale de nos systèmes d'informations est désastreuse.

11.34.3 Dérivation de la fonction puissance (première)

Nous n'allons pas complètement résoudre la question de la dérivation de la fonction $x \mapsto a^x$; il faudrait des logarithmes, et nous ne les avons pas encore défini. Le logarithme sera introduit comme fonction inverse de l'exponentielle en 14.58.

Proposition 11.354 ([255]).
Soit la fonction puissance

$$\begin{aligned} g_a \colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto a^x. \end{aligned} \tag{11.951}$$

(1) La fonction g_a est dérivable.

(2) La dérivée vérifie l'équation

$$g_a'(x) = g_a'(0) g_a(x). \tag{11.952}$$

Démonstration. La fonction g_a est continue par 11.341(1). La proposition 11.322 nous dit donc que la fonction g_a admet une primitive sur $\mathbb{R}$. Nous notons F une telle primitive.

Soit $x \in \mathbb{R}$. En posant $F_x(t) = F(x+t)$, nous avons une primitive de $t \mapsto a^x a^t$. En effet

$$F_x'(t) = F'(x+t) \frac{d}{dt}\Big[x+t\Big]_{t=0} = a^{a+t} = a^x a^t. \tag{11.953}$$

Par ailleurs la fonction $t \mapsto a^x F(t)$ est également une primitive de $t \mapsto a^x a^t$. Donc il existe un nombre $C(x)$ tel que

$$F_x(t) = F(x+t) = a^x F(t) + C(x). \tag{11.954}$$

Le nombre $F(1) - F(0)$ est un nombre sans histoires. Nous avons :

$$\begin{aligned} g_a(x)\big(F(1) - F(0)\big) &= g_a(x)F(1) - g_a(x)F(0) && \text{(11.955a)} \\ &= F_x(1) - C(x) - F_x(0) + C(x) && \text{(11.955b)} \\ &= F_x(1) - F_x(0) && \text{(11.955c)} \\ &= F(1+x) - F(x). && \text{(11.955d)} \end{aligned}$$

La fonction F étant dérivable, nous en déduisons que g_a est dérivable.

Vu que nous n'avons aucune idée de la forme de F, nous ne pouvons pas tirer beaucoup d'informations d'une dérivation des membres de gauche et de droite de (11.955).

En ce qui concerne la formule, nous écrivons la fameuse équation fonctionnelle [83]

$$g_a(x+y) = g_a(x) g_a(y) \tag{11.956}$$

Nous fixons x et dérivons par rapport à y en $y = y_0$:

$$g_a'(x+y_0) = g_a(x) g_a'(y_0). \tag{11.957}$$

En posant $y_0 = 0$ nous trouvons le résultat demandé. □

11.355.
La démonstration donnée dans [255] s'assure d'abord de l'existence d'une intégrale (lemme 13.226), pose ensuite $A = \int_0^1 g_a(t)dt$ et fait le calcul suivant :

$$A g_a(x) = \int_0^1 g_a(x) g_a(t) dt = \int_0^1 g_a(x+t) dt = \int_x^{x+1} g_a(t) dt. \tag{11.958}$$

Vu que le membre de droite est une fonction dérivable de x, nous concluons que g_a est dérivable. Cela demande donc toute la théorie de l'intégration pour prouver la *dérivabilité* d'une fonction.

La démonstration donnée ici est à peine mieux. Elle utilise l'existence d'une primitive et donc tout le théorème de Stone-Weierstrass 11.319.

83. Pour rappel, proposition 11.341(3).

Dans les deux cas, je trouve que la situation n'est pas fameuse. Si vous êtes capable de montrer l'existence de la limite

$$\lim_{\epsilon\to 0}\frac{a^\epsilon-1}{\epsilon} \tag{11.959}$$

sans recourir à autre chose que des astuces sur les limites, je suis preneur. Ou, au contraire, si vous avez un argument pour dire que c'est impossible, dites-le moi également. Écrivez-moi.

Nous posons une définition

Définition 11.356.
*Soit $a>0$. Nous nommons l'**équation fonctionnelle** l'équation*

$$\begin{cases} f(x+y)=f(x)f(y) & (11.960a)\\ f(1)=a & (11.960b)\end{cases}$$

pour la fonction inconnue $f\colon\mathbb{R}\to\mathbb{R}$.

Définition 11.357.
*Soit $a>0$. Nous nommons l'**équation exponentielle** l'équation*

$$\begin{cases} y'=y & (11.961a)\\ y(1)=a & (11.961b)\end{cases}$$

pour la fonction inconnue $y\colon\mathbb{R}\to\mathbb{R}$.

Proposition 11.358 (Unicité de l'exponentielle)**.**
Si elle existe, la solution au problème

$$\begin{cases} y'=y & (11.962a)\\ y(0)=1 & (11.962b)\end{cases}$$

est unique.

Démonstration. Soient y et g deux solutions et considérions la fonction $h(x)=g(x)y(-x)$. Un calcul immédiat donne

$$h'(x)=0 \tag{11.963}$$

et donc h est constante. Vu que $h(0)=1$ nous avons $g(x)y(-x)=1$ pour tout x, c'est-à-dire

$$g(x)=\frac{1}{y(-x)}=y(x). \tag{11.964}$$

□

11.359.
Nous savons qu'il existe une unique solution de l'équation exponentielle avec $a=1$. Avec la relation

$$g_a'(x)=g_a(x)g_a'(0), \tag{11.965}$$

de la proposition 11.354, nous n'en sommes pas loin. Il faut encore savoir si il existe un $a>0$ tel que $g_a'(0)=1$. Notre culture générale nous dit qu'un tel réel existe et est la fameuse constante e.

Nous nous attelons maintenant à la tâche de montrer l'existence de la chose.

11.34.4 Équation fonctionnelle

Il n'est un secret pour personne (proposition 11.341(3)) que la fonction

$$\begin{aligned} g_a \colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto a^x \end{aligned} \tag{11.966}$$

vérifie l'équation fonctionnelle (11.960). Nous pouvons nous demander à quel point cette propriété caractérise la fonction puissance.

Proposition 11.360 ([255]).
Encore plusieurs résultats sur la fonction g_a avec $a > 0$.

(1) La fonction g_a vérifie l'équation fonctionnelle.

(2) La dérivée vérifie $g_a'(0) \neq 0$.

(3) Pour tout a, en posant $\alpha = 1/g_a'(0)$ nous avons

$$g_{a^\alpha}'(0) = 1. \tag{11.967}$$

(4) Il existe un unique $e > 0$ tel que

$$g_e' = g_e. \tag{11.968}$$

(5) Pour la valeur de e donnée en (4), la fonction g_e vérifie l'équation exponentielle (11.961)

$$\begin{cases} g_e' = g_e & \text{(11.969a)} \\ g_e(1) = e. & \text{(11.969b)} \end{cases}$$

Démonstration. Un point à la fois.

Pour (1) Le fait que g_a vérifie l'équation fonctionnelle est la proposition 11.341(3).

Pour (2) La formule (11.952) de la proposition 11.354 nous assure que

$$g_a'(x) = g_a(x)g_a'(0). \tag{11.970}$$

Donc $g_a'(0) = 0$ impliquerait que $g_a = 0$, ce qui n'est pas le cas.

Pour (3) Par ailleurs la proposition 11.351 nous permet d'écrire

$$g_a(\alpha x) = g_{a^\alpha}(x). \tag{11.971}$$

En dérivant des deux côtés,

$$\alpha g_a'(\alpha x) = g_{a^\alpha}'(x). \tag{11.972}$$

En posant donc $\alpha = g_a'(0)$ et en évaluant (11.972) en $x = 0$ nous trouvons le résultat.

Pour (4), existence Pour les valeurs de α données par le point (3), nous avons $g_{a^\alpha}'(0) = 1$, et l'équation (11.970) nous donne alors

$$g_{a^\alpha}(x) = g_{a^\alpha}(x). \tag{11.973}$$

Comme de plus $g_{a^\alpha}(0) = 1$, cette fonction vérifie bien l'équation exponentielle.

Pour (4), unicité Si a et b font en sorte que $g_a' = g_a$ et $g_b' = g_b$, alors nous avons aussi $g_a'(0) = g_b'(0) = 1$ à cause de (11.970). Donc g_a et g_b vérifient l'équation de la proposition 11.358 dont la solution est unique. Donc $g_a = g_b$.

Pour tout x nous avons $g_a(x) = g_b(x)$. En particulier pour $x = 1$ nous avons $a = b$.

□

Proposition 11.361 ([255]).
Soit $a > 0$. Nous considérons l'équation fonctionnelle 11.356 et l'équation exponentielle 11.357 pour une fonction $f \colon \mathbb{R} \to \mathbb{R}$.

(1) Si f vérifie l'équation fonctionnelle, alors

$$f(q) = a^q \tag{11.974}$$

pour tout $q \in \mathbb{Q}$.

(2) Si f vérifie l'équation fonctionnelle et est monotone, alors $f = g_a$.

(3) Si f vérifie l'équation fonctionnelle et est continue, alors $f = g_a$.

Démonstration. En beaucoup de parties. Nous commençons par prouver (1). Nous supposons que $f\colon \mathbb{R} \to \mathbb{R}$ est une fonction vérifiant l'équation fonctionnelle[84].

$f(x) \geqslant 0$ Quel que soit $x \in \mathbb{R}$ nous avons

$$f(x) = f(\frac{x}{2} + \frac{x}{2}) = f(\frac{x}{2})^2 \geqslant 0. \tag{11.975}$$

Vous noterez que cet argument ne fonctionne pas si f est à valeurs dans $\mathbb{C}$ au lieu de $\mathbb{R}$.

Pour $n \in \mathbb{N}$ Soit $n \in \mathbb{N}$. Je vous laisse rédiger la récurrence correctement, mais l'idée est que $f(1) = a$ et ensuite que

$$f(n+1) = f(n)f(1) = f(1)^n f(1) = f(1)^{n+1}. \tag{11.976}$$

Pour $m \in \mathbb{Z}$ Nous avons d'une part que $f(-m+m) = f(0) = 1$, mais d'autre part que $f(-m+m) = f(-m)f(m)$. Donc $1 = f(-m)f(m)$; et nous concluons que

$$f(-m) = \frac{1}{f(m)}. \tag{11.977}$$

Pour $q = 1/n$ Nous savons que $f(1) = a$, mais $1 = \frac{1}{n} + \dots \frac{1}{n}$ (avec n termes), donc

$$a = f(\frac{1}{n} + \dots \frac{1}{n}) = f(\frac{1}{n})^n. \tag{11.978}$$

Cela implique que $f(\frac{1}{n})^n = a$. La proposition 11.324 indique qu'il existe un unique $x > 0$ tel que $x^n = a$. Vu que nous savons déjà que f est partout positive[85], cette contrainte fixe $f(1/n)$ et la définition 11.325 nous permet d'écrire

$$f(\frac{1}{n}) = a^{1/n}. \tag{11.979}$$

Pour $q \in \mathbb{Q}$ Nous posons $q = m/n$. Le nombre q peut être écrit sous la forme $\frac{m}{n} = \frac{1}{n} + \dots + \frac{1}{n}$ avec m termes. Donc

$$f(\frac{m}{n}) = f(\frac{1}{n} + \dots \frac{1}{n}) = f(\frac{1}{n})^m = (a^{1/n})^m = a^{m/n} \tag{11.980}$$

où nous avons utilisé le lemme 11.330.

La preuve de (1) est terminée.

Démonstration de (2) Nous faisons maintenant la preuve de (2). Nous supposons que f vérifie l'équation fonctionnelle et qu'elle est monotone. Pour fixer les idées, nous supposons qu'elle est monotone croissante[86].

84. Bien que ce ne soit pas strictement nécessaire ici, nous rappelons qu'une telle fonction existe par la proposition 11.360.

85. C'est ici que l'hypothèse de fonction à valeurs dans $\mathbb{R}$ est cruciale. Pour $f\colon \mathbb{R} \to \mathbb{C}$ ceci ne fonctionne pas, et de loin.

86. Si f est monotone décroissante, soit vous adaptez la preuve, soit vous essayez de voir si on ne peut pas recycler le cas croissant en l'appliquant à $-f$.

Nous considérons les parties [87]

$$A = \{q \in \mathbb{Q} \text{ tel que } q < x\} \tag{11.981a}$$
$$B = \{q \in \mathbb{Q} \text{ tel que } q > x\}. \tag{11.981b}$$

Vu que f est croissante, nous avons $f(x) \geqslant f(q)$ pour tout $q \in A$ et $f(x) \leqslant f(q)$ pour tout $q \in B$. En passant au supremum et à l'infimum,

$$\sup_{q\in A} f(q) \leqslant f(x) \leqslant \inf_{q\in B} f(q). \tag{11.982}$$

Mais il existe dans A une suite strictement croissante convergente q_i vers x (parce que $x = \sup(A)$), donc

$$a^x = \lim_i a^{q_i} \tag{11.983}$$

par la définition 11.340. Et de même, il existe une suite r_i décroissante dans B telle que $x = \lim r_i$. Cette suite donne aussi

$$a^x = \lim_i a^{r_i}. \tag{11.984}$$

Nous avons donc l'encadrement

$$a^x \leqslant f(x) \leqslant a^x, \tag{11.985}$$

qui implique que $f(x) = a^x$.

Démonstration de (3) Nous ne supposons plus que f est monotone. Au lieu de cela nous supposons qu'elle est continue. Nous avons déjà vu en (1) que $f = g_a$ sur $\mathbb{Q}$. Mais par hypothèse f est continue et par la proposition 11.341, g_a est continue. La proposition 11.73 conclu que $f = g_a$ sur $\mathbb{R}$.

□

Proposition 11.362 ([255]).
Si y vérifie l'équation exponentielle, alors elle est continue, monotone et vérifie l'équation fonctionnelle.

11.34.5 Dérivation de la fonction puissance (seconde)

La proposition suivante donne la dérivée de $x \mapsto x^q$ pour tout $q \in \mathbb{Q}$. La formule donnée est encore valable pour $x \mapsto x^\alpha$ pour tout $\alpha \in \mathbb{R}$, mais elle demandera plus de théorie pour être démontrée, voir la proposition 13.245.

Proposition 11.363 ([1]).
Pour tout $\alpha \in \mathbb{Q}$, si $f_\alpha(x) = x^\alpha$ alors

$$f'_\alpha(x) = \alpha x^{\alpha-1}. \tag{11.986}$$

En particulier, f_α est de classe C^∞ sur $\mathbb{R}\backslash\{0\}$.

Démonstration. Petit à petit.

Naturel Nous prouvons que $(x^n)' = nx^{n-1}$ par récurrence en utilisant la règle de Leibnitz de la propositon 11.115(3).

D'abord pour $n = 1$ nous avons $f_1(x) = x$ et donc

$$f'_1(x) = \lim_{\epsilon\to 0} \frac{(x+\epsilon) - x}{\epsilon} = 1. \tag{11.987}$$

Supposons que $f'_k(x) = kx^{k-1}$ pour un certain $k \in \mathbb{N}$. Nous prouvons que $f'_{k+1}(x) = (k+1)x^k$. Nous avons

$$x^{k+1} = xx^k. \tag{11.988}$$

87. Il du meilleur gout de citer le lemme 1.144 pour dire qu'ils sont non vides.

En utilisant la règle de Leibnitz et l'hypothèse de récurrence,

$$\left(x^{k+1}\right)' = (x)'x^k + x\left(x^k\right)' = x^k + x\left(kx^{k-1}\right) = x^k + kx^k = (k+1)x^k, \tag{11.989}$$

ce qu'il fallait démontrer.

Rationnel positif Soit donc $\alpha = p/q$ avec $p, q \in \mathbb{N}$. Le lemme 11.330 nous permet d'écrire $f_{p/q}(x) = x^{p/q} = (x^p)^{1/q}$. Cela donne

$$f_{p/q}(x)^q = x^p. \tag{11.990}$$

Nous dérivons cette relation par rapport à x en utilisant à la fois la règle pour les entiers et la règle des fonctions composées [88] :

$$qf'_{p/q}(x)^{q-1}f'_{p/q}(x) = px^{p-1}. \tag{11.991}$$

En isolant $f'_{p/q}(x)$ dans cette expression et en utilisant le fait que $\frac{x^a}{x^b} = x^{a-b}$, nous trouvons le résultat.

Rationnels négatifs Soit $\alpha = -p/q$ avec $p, q \in \mathbb{N}$. Nous avons $x^{-p/q} = \frac{1}{f_{p/q}(x)}$. En utilisant la proposition 11.115(5) et le point déjà prouvé sur les rationnels positifs,

$$f'_{p/q} = -\frac{f'_{-p/q}}{f^2_{p/q}} = -\frac{(-p/q)x^{-p/q-1}}{x^{-2p/q}} = (p/q)x^{p/q-1}. \tag{11.992}$$

Notez l'utilisation de la proposition 11.351 au dénominateur.

Irrationnel Ah ah ! On vous a bien eu. Les irrationnels, c'est pour la proposition 13.245.

En ce qui concerne le fait que la fonction f_α est de classe C^∞ sur $\mathbb{R}\backslash\{0\}$, c'est simplement une récurrence. Attention : si le rationnel α est négatif, $f_\alpha(0)$ n'est pas défini. Mais, lorsque α est positif non entier, à partir d'un certain ordre, les dérivées font intervenir x^β avec $\beta < 0$. D'où la restriction à $\mathbb{R}\backslash\{0\}$ du domaine sur lequel f_α est de classe C^∞.

Si α est positif entier, alors f_α est de classe C^∞ sur tout $\mathbb{R}$ parce que toutes les dérivées sont nulles à partir d'un certain ordre. □

11.34.6 Hölder

Notre étude de la fonction puissance permet de démontrer quelques inégalités de Hölder. Voir le thème 7.

Théorème 11.364 ([256]).
Si $1 \leqslant p < \infty$, alors pour tout $x \in \mathbb{R}^n$ nous avons

$$\|x\|_\infty \leqslant \|x\|_p \leqslant n^{1/p}\|x\|_\infty. \tag{11.993}$$

Démonstration. Vu que la fonction $t \mapsto |t|^p$ est croissante pour les t positifs [89], pour chaque i nous avons

$$|x_i| \leqslant \left(\sum_k |x_k|^p\right)^{1/p} = \|x\|_p. \tag{11.994}$$

Cela montre que

$$\|x\|_\infty = \max\{|x_i|\} \leqslant \|x\|_p. \tag{11.995}$$

D'autre part, pour chaque i nous avons $|x_i| \leqslant \|x\|_\infty$, donc

$$\|x\|_p \leqslant \left(n\|x\|_\infty\right)^{1/p} = n^{1/p}\|x\|_\infty. \tag{11.996}$$

□

88. Proposition 11.115(4).
89. Parce que $p \geqslant 1$ et la proposition 11.346.

Le corolaire suivant donne une façon de majorer une norme ℓ^p par une norme ℓ^q moyennant un facteur. Notons cependant que l'inégalité de Hölder de la proposition 16.106 est plus précise. Ce corolaire est suffisant pour prouver l'équivalence des normes ℓ^p.

Corolaire 11.365.
Soient $p \geqslant 1$ et $q \leqslant \infty$. Pour tout $x \in \mathbb{R}^n$ nous avons

$$\|x\|_p \leqslant n^{1/p}\|x\|_q. \tag{11.997}$$

Démonstration. Il suffit d'utiliser les deux inégalités du théorème 11.364. D'abord la seconde avec p, et ensuite la première avec q. □

11.34.7 Vers les complexes

Nous avons déjà vu la proposition 11.360 qui dit essentiellement que si une fonction continue $f\colon \mathbb{R} \to \mathbb{R}$ vérifie $f(x+y) = f(x)f(y)$, alors $f(x) = a^x$. Comme indiqué durant la preuve, cette proposition (et en particulier sa preuve) ne fonctionne pas pour les fonctions à valeurs complexes. L'endroit où cela coinçait est que la contrainte

$$f(\frac{1}{n})^n = a \tag{11.998}$$

n'implique pas grand chose lorsque f est à valeurs complexes.

Nous allons maintenant attaquer ce problème.

Lemme 11.366.
Soit $\alpha \in \mathbb{C}$. Si elle existe, la solution au problème

$$\begin{cases} y' = \alpha y & \text{(11.999a)} \\ y(0) = 1 & \text{(11.999b)} \end{cases}$$

pour $y\colon \mathbb{R} \to S^1$ est unique.

Démonstration. Soient deux solutions y_1 et y_2. Nous posons $h(x) = y_1(x)y_2(-x)$. Une dérivation donne

$$h'(x) = y_1'(x)y_2(-x) - y_1(x)y_2'(-x). \tag{11.1000}$$

En y substituant $y_1'(x) = \alpha y_1(x)$ et $y_2'(-x) = \alpha y_2(x)$ nous trouvons $h'(x) = 0$. Donc h est constante et nous avons

$$y_1(x)y_2(-x) = 1 \tag{11.1001}$$

pour tout x. Notons que cette identité est encore valable avec $y_1 = y_2$. Nous avons en particulier les égalités $y_1(x)y_1(-x) = 1$ et $y_2(x)y_2(-x) = 1$, et nous notons au passage que $y_1(x)$ et $y_2(x)$ ne s'annulent pas.

En substituant dans (11.1001) la valeur $y_2(-x) = \frac{1}{y_2(x)}$ nous trouvons

$$\frac{y_1(x)}{y_2(x)} = 1, \tag{11.1002}$$

ce qui signifie $y_1(x) = y_2(x)$. □

Dans la proposition suivante, S^1 désigne l'ensemble des nombres complexes de norme 1, dont un paramétrage est donnée dans la proposition 17.48 :

$$S^1 = \{e^{ix} \text{ tel que } x \in \mathbb{R}\} = \{e^{ix} \text{ tel que } x \in [0, 2\pi[\}. \tag{11.1003}$$

Proposition 11.367 ([1]).
Soit une fonction continue $f\colon \mathbb{R} \to S^1$ vérifiant

$$f(x+y) = f(x)f(y). \tag{11.1004}$$

Alors

(1) f est dérivable,

(2) f satisfait au système

$$\begin{cases} f'(x) = f'(0)f(x) & (11.1005a) \\ f(0) = 1, & (11.1005b) \end{cases}$$

(3) il existe $m \in \mathbb{R}$ tel que $f(x) = e^{imx}$.

Démonstration. Pour chaque $m \in \mathbb{R}$, la fonction

$$g_m(x) = e^{imx} \tag{11.1006}$$

vérifie évidemment toutes les conditions. Le but de cette démonstration est de montrer que les conditions imposées à f la déterminent de façon univoque (à part ce m).

La condition (11.1004) nous dit que $f(0) = 1$. Soit une primitive F de f. Il existe $s > 0$ tel que $F(s) > F(0)$ parce que $F' = f$ et $f(0) = 1$.

Soit $a \in \mathbb{R}$. La fonction G_a donnée par $G_a(x) = F(x + a)$ est une primitive de $x \mapsto f(x)f(a)$. Donc $G_a(x) = f(a)F(x)$. Cela dit nous avons

$$f(x)\big(F(s) - F(0)\big) = f(x)F(s) - f(x)F(0) = G_1(x) - G_0(x). \tag{11.1007}$$

Le membre de droite est évidemment dérivable, et $F(s) - F(0) \neq 0$. Donc f est dérivable.

Nous dérivons maintenant la relation $f(x+y) = f(x)f(y)$ par rapport à y en $y = 0$. Cela donne

$$f'(x) = f'(0)f(x). \tag{11.1008}$$

Donc il existe $\alpha \in \mathbb{C}$ tel que $f'(x) = \alpha f(x)$.

Jusqu'ici nous avons prouvé qu'il existe $\alpha \in \mathbb{C}$ tel que

$$\begin{cases} f'(x) = \alpha f(x) & (11.1009a) \\ f(0) = 1. & (11.1009b) \end{cases}$$

Or le lemme 11.366 donne l'unicité de la solution à ce système, et il ne faut pas chercher loin : la solution est

$$f(x) = e^{\alpha x}. \tag{11.1010}$$

Pour avoir $f(x) \in S^1$, nous devons de plus imposer que α soit imaginaire pur. Donc, en posant $\alpha = im$, nous avons $m \in \mathbb{R}$ tel que $f(x) = e^{imx}$. □

11.35 Polynômes de Taylor

Définition 11.368.
*Soit un ouvert $\Omega \subset \mathbb{C}$. Une fonction $f\colon \Omega \to \mathbb{C}$ est $\mathbb{C}$-**analytique** sur Ω si pour tout $z_0 \in \Omega$, il existe une suite complexe (c_n) et $r > 0$ tels que*

$$f(z) = \sum_{n=0}^{\infty} c_n (z - z_0)^n \tag{11.1011}$$

pour tout $z \in B(z_0, r)$.

Définition 11.369.
*Soit une fonction $f\colon \mathbb{R} \to \mathbb{R}$. Si il existe, nous définissons le n^e **polynôme de Taylor** de f au point $a \in \mathbb{R}$ par*

$$P_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x - a)^k. \tag{11.1012}$$

*Et la **série de Taylor** de f est la limite :*

$$T(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x - a)^k \tag{11.1013}$$

dans la mesure où la somme converge.

Tant que f est n fois dérivable, le polynôme P_n existe et vérifie $P_n(a) = f(a)$. Nous ne pouvons rien en dire de plus pour l'instant. En particulier, si f est de classe C^∞ il ne faudrait pas croire que

$$\lim_{n\to\infty} P_n(x) = f(x) \tag{11.1014}$$

pour tout x dans un voisinage de a. Autrement dit, même si toutes les dérivées de f existent, la série entière T n'est pas garantie de

— un rayon de convergence[90] plus grand que zéro,

— et même avec un grand rayon de convergence, que la limite soit les valeurs de f.

11.370.
Il n'est pas très compliqué de construire une fonction f telles que $f(0) = 0$ et telle que $f^{(k)}(0) = 0$ pour tout k, sans pour autant que f soit nulle partout (voir les fonctions plateaux 14.13.1). Les polynômes de Taylor d'une telle fonction sont tous identiquement nuls.

Ceci pour dire qu'en posant

$$T(x) = \sum_{n=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k, \tag{11.1015}$$

nous n'avons aucune garantie de $T = f$, même pas sur le rayon de convergence de la série entière définissant P. Et nous n'avons pas de garanties d'avoir un rayon de convergence plus grand que 0.

Notons toutefois que les polynômes étant denses pour la norme supremum parmi les fonctions continues[91], pour tout compact, il existe une suite de polynômes qui converge uniformément uniformément f. Mais ces polynômes ne sont pas spécialement ceux de Taylor.

11.371.
Ce que nous venons de dire en 11.370 n'est pas vrai pour les fonctions analytiques[92]. Une fonction analytique $f\colon \Omega \to \mathbb{R}$ s'écrit, autour de 0, sous la forme

$$f(x) = \sum_{n=0}^{\infty} a_n x^n. \tag{11.1016}$$

Demander $f^{(k)}(0) = 0$ pour tout k implique $a_n = 0$ pour tout n, et donc $f = 0$ sur un voisinage de 0.

La condition d'analycité est donc très rigide.

Le théorème de Taylor que nous démontrons à présent n'est pas un résultat que va dans le sens de $\lim_{n\to\infty} P_n(x) = f(x)$. C'est un résultat qui dit juste que $\lim_{x\to a} P_n(x) = f(a)$, et que la limite va d'autant plus vite que n est grand.

Le théorème de Taylor généralise le développement limité au premier ordre de la proposition 11.113.

11.372.
Lorsque le contexte n'est pas ambigu, nous notons simplement P_n le polynôme d'ordre n de f au point a. De même nous notons le reste

$$R_n(x) = f(x) - P_n(x). \tag{11.1017}$$

Proposition 11.373 ([257]).
Soit une fonction f qui est n fois dérivable sur l'intervalle ouvert $I \subset \mathbb{R}$ contenant a. Alors

$$\lim_{x\to a} \frac{R_n(x)}{(x-a)^n} = \lim_{x\to a} \frac{f(x) - P_n(x)}{(x-a)^n} = 0 \tag{11.1018}$$

où P_n est le n^e polynôme de Taylor de f autour de $x = a$.

90. Définition 14.8.
91. Théorème 11.319.
92. Définition 11.368 qu'il faut écrire pour $\mathbb{R}$ au lieu de $\mathbb{C}$

Démonstration. Pour tout $k = 0, \dots, n$ nous avons $f^{(k)}(a) = P_n^{(k)}(a)$ et donc

$$R_n^{(k)}(a) = 0 \tag{11.1019}$$

pour $k = 0, \dots, n$. En posant d'autre par $s(x) = (x-a)^n$ nous avons $s^{(k)}(a) = 0$ pour tout $k = 0, \dots, n-1$. Par conséquent la règle de l'Hospital de la proposition 11.132 s'applique au quotient $R_n(x)/s(x)^n$. En l'utilisant n fois,

$$\lim_{x\to a} \frac{R_n(x)}{s(x)} = \lim_{x\to a} \frac{R^{(k)}(x)}{k!(x-a)^0} = \frac{0}{k!} = 0. \tag{11.1020}$$

□

Nous démontrons à présent que le polynôme de Taylor est le seul à avoir la propriété de la proposition 11.373.

Proposition 11.374 ([257]).
Soit f, une fonction n fois dérivable sur l'intervalle $I \subset \mathbb{R}$ contenant 0. Soit un polynôme Q de degré n (ou moins) tel que

$$\lim_{x\to a} \frac{f(x) - Q(x)}{(x-a)^n} = 0. \tag{11.1021}$$

Alors Q est le polynôme de Taylor de degré n pour f en a ci-après simplement noté P_n.

Démonstration. D'après la proposition 11.373, la fonction $f - P_n$ vérifie la même limite que $f - Q$. DOnc $P_n - Q$ vérifie également la limite

$$\lim_{x\to 0} \frac{(P_n - Q)(x)}{x^n} = 0. \tag{11.1022}$$

Nous notons $P_n(x) = \sum_{k=0}^n a_k x^k$ et $Q(x) = \sum_{k=0}^n b_k x^k$. La relation (11.1022) donne en particulier

$$\lim_{x\to 0} (P_n - Q)(x) = 0 \tag{11.1023}$$

qui donne $a_0 - b_0 = 0$. Nous continuons par récurrence en supposant que $a_i = b_i$ pour $i = 0, \dots, k$. Alors

$$0 = \lim_{x\to 0} \frac{(P_n - Q)(x)}{x^{k+1}} = \lim_{x\to 0} \sum_{l=k+1}^n (a_l - b_l) x^{l-(k+1)}. \tag{11.1024}$$

Le seul terme non nul à droite est celui vérifiant $l - (k+1) = 0$. Et ce terme donne l'équation

$$a_{k+1} - b_{k+1} = 0, \tag{11.1025}$$

c'est-à-dire $a_{k+1} = b_{k+1}$. La récurrence continue ainsi jusqu'à $k = n$, et nous pouvons conclure que $Q = P_n$. □

L'intérêt de cette proposition est que si l'on trouve, par n'importe quel moyen, un polynôme Q vérifiant la condition (11.1021), alors nous savons que c'est le polynôme de Taylor.

Théorème 11.375 (Théorème de Taylor[147, 258]).
Soit $I \subset$ un intervalle non vide et non réduit à un point de $\mathbb{R}$ ainsi que $a \in I$. Soit une fonction $f\colon I \to \mathbb{R}$ telle que $f^{(n)}(a)$ existe. Alors il existe une fonction α définie sur I et à valeurs dans $\mathbb{R}$ vérifiant les deux conditions suivantes :

$$f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!}(x-a)^k + \alpha(x)(x-a)^n, \tag{11.1026a}$$

$$\lim_{t\to a} \alpha(t) = 0 \tag{11.1026b}$$

pour tout $x \in I$. Ici $f^{(k)}$ dénote la k-ième dérivée de f (en particulier, $f^{(0)} = f$, $f^{(1)} = f'$).

Démonstration. Si $R_n(x) = f(x) - P_n(x)$, il suffit de poser

$$\alpha(x) = R_n(x)(x-a)^n \tag{11.1027}$$

et d'utiliser la proposition 11.373. □

Remarque 11.376.
Quelque remarques.

(1) La formule (11.1026b) est une égalité, et non une approximation. Ce qui serait une approximation serait de récrire la formule dans le terme contenant α.

(2) Nous avons l'égalité (11.1026b) uniquement sur I. Pour les x hors de I, le polynôme existe évidemment, mais nous n'avons pas spécialement de fonction α, et d'ailleurs la fonction f n'est pas spécialement définie.

11.377.
Les conditions (11.1026) sont souvent aussi énoncées sous la forme qu'il existe une fonction α telle que

$$\begin{cases} \lim_{t\to 0} \dfrac{\alpha(t)}{t^n} = 0 & \text{(11.1028a)} \\ f(a+h) = f(a) + hf'(a) + \dfrac{h^2}{2}f''(a) + \cdots + \dfrac{h^n}{n!}f^{(n)}(a) + \alpha(h). & \text{(11.1028b)} \end{cases}$$

Le théorème suivant donne une expression pas tout à fait explicite, mais pas mal quand même pour le reste de Taylor.

Théorème 11.378.
Soient un intervalle ouvert $I \subset \mathbb{R}$ ainsi que $a \in I$. Soit encore une fonction de classe C^{k+1} sur I. Pour tout $x \in I$, il existe un $c \in]a, x[$ tel que l'égalité

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \tag{11.1029}$$

soit vérifiée.

Démonstration. Pour les besoins de la preuve, nous allons démontrer la formule (11.1029) pour un $b \in I$ au lieu de x. C'est juste que nous allons écrire b au lieu de x parce que nous aurons besoin de la notation x dans le courant de la preuve.

Nous posons

$$R(x) = f(x) - P_n(x) = f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k. \tag{11.1030}$$

Cela vérifie $R(a) = f(a) - f(a) = 0$ et même

$$R^{(j)}(a) = 0 \tag{11.1031}$$

pour tout $j = 1, \ldots, n$. Nous posons encore

$$F(x) = R(x) - \frac{R(b)}{(b-a)^{n+1}}(x-a)^{n+1}. \tag{11.1032}$$

Nous avons $F^{(j)}(a) = 0$ pour tout $j = 0, \ldots, n$ ainsi que

$$F(b) = R(b) - \frac{R(b)}{(b-a)^{n+1}}(b-a)^{n+1} = 0. \tag{11.1033}$$

et aussi

$$F(a) = R(a) - 0 = 0. \tag{11.1034}$$

Bref, la fonction F vérifie les conditions de la généralisation 11.129 du lemme de Rolle. Il existe donc $c \in]a,b[$ tel que $F^{(n+1)}(c) = 0$. Mais vu que $P_n^{(n+1)}(x) = 0$, nous avons $R^{(n+1)}(x) = f^{(n+1)}(x)$, de telle sorte que

$$f^{(n+1)}(c) = \frac{R(b)(n+1)!}{(b-a)^{n+1}}. \tag{11.1035}$$

En injectant cela dans la définition de F

$$F(x) = R(x) - \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}. \tag{11.1036}$$

En évaluant en $x = b$, et en nous souvenant que $F(b) = 0$, nous trouvons

$$0 = R(b) - \frac{f^{(n+1)}(c)}{(n+1)!}(b-a)^{n+1} \tag{11.1037}$$

qui est ce que nous voulions prouver. □

Voici un énoncé pour les fonctions à plusieurs variables.

Théorème 11.379 ([259]).
Si $f\colon E \to \mathbb{R}$ est une application n fois différentiable en $a \in E$ alors il existe une fonction $\epsilon\colon \mathbb{R} \to \mathbb{R}$ telle que

$$\begin{cases} f(a+h) = f(a) + df_a(h) + \frac{1}{2}(d^2f)_a(h,h) + \ldots + & \text{(11.1038a)} \\ \qquad + \ldots + \frac{1}{n!}(d^nf)_a(h,\ldots,h) + \|h\|^n\epsilon(\|h\|) & \text{(11.1038b)} \\ \lim_{t\to 0}\epsilon(t) = 0. & \text{(11.1038c)} \end{cases}$$

11.35.1 Fonctions « petit o »

Nous voulons formaliser l'idée d'une fonction qui tend vers zéro « plus vite » qu'une autre. Nous disons que $f \in o\big(\varphi(x)\big)$ si

$$\lim_{x\to 0}\frac{f(x)}{\varphi(x)} = 0. \tag{11.1039}$$

En particulier, nous disons que $f \in o(x)$ lorsque $\lim_{x\to 0} f(x)/x = 0$.

En termes de notations, nous définissons l'ensemble $o(x)$ l'ensemble des fonctions f telles que

$$\lim_{x\to 0}\frac{f(x)}{x} = 0. \tag{11.1040}$$

Plus généralement si g est une fonction telle que $\lim_{x\to 0} g(x) = 0$, nous disons $f \in o(g)$ si

$$\lim_{x\to 0}\frac{f(x)}{g(x)} = 0. \tag{11.1041}$$

De façon intuitive, l'ensemble $o(g)$ est l'ensemble des fonctions qui tendent vers zéro « plus vite » que g.

Nous pouvons donner un énoncé alternatif au théorème 11.375 en définissant $h(x) = \epsilon(x+a)x^n$. Cette fonction est définie exprès pour avoir

$$h(x-a) = \epsilon(x)(x-a)^n, \tag{11.1042}$$

et donc

$$\lim_{x\to 0}\frac{h(x)}{x^n} = \lim_{x\to 0}\epsilon(x-a) = \lim_{x\to a}\epsilon(x) = 0. \tag{11.1043}$$

Donc $h \in o(x^n)$.

Le théorème dit donc qu'il existe une fonction $\alpha \in o(x^n)$ telle que

$$f(x) = T_{f,n}^a(x) + \alpha(x-a). \tag{11.1044}$$

pour tout $x \in I$.

Remarque 11.380.
À titre personnel, l'auteur de ces lignes déconseille d'utiliser cette notation qui est un peu casse-figure pour qui ne la maîtrise pas bien.

Exemple 11.381.
Le développement en série du cosinus sera traité dans la proposition 17.57. △

Proposition 11.382 (Ordre deux sur $\mathbb{R}^n$[1])**.**
Soit un ouvert Ω de $\mathbb{R}^n$ et $a \in \Omega$ ainsi qu'une fonction $f\colon \Omega \to \mathbb{R}$ de classe C^2. Alors il existe une fonction $\alpha\colon \mathbb{R}^n \to \mathbb{R}$ telle que

$$\begin{cases} f(a+h) = f(a) + df_a(h) + \frac{1}{2}(d^2f)_a(h,h) + \|h\|^2\alpha(h) & (11.1045a) \\ \lim_{h\to 0} \alpha(h) = 0. & (11.1045b) \end{cases}$$

Ici, la notation $(d^2f)_a(h,h)$ réfère à ce qui est expliqué en 11.284.

Démonstration. Dans la suite nous considérons t et h tels que toutes les expressions suivantes aient un sens, c'est-à-dire que tous les trucs comme $a + th$ restent dans Ω. Pour $h \in \mathbb{R}^n$ nous nommons e_h le vecteur unitaire dans la direction de h, c'est-à-dire $e_h = h/\|h\|$ et nous posons

$$k_h(t) = f(a + te_h). \tag{11.1046}$$

et nous lui appliquons Taylor 11.375 à l'ordre deux : il existe une fonction β_h telle que

$$k_h(x) = k_h(0) + xk_h'(0) + \frac{x^2}{2}k_h''(0) + x^2\beta_h(x). \tag{11.1047}$$

avec $\lim_{x\to 0} \beta_h(x) = 0$.

En ce qui concerne les dérivées de k_h nous avons

$$k_h'(0) = df_a(e_h) \tag{11.1048}$$

et

$$k_h''(0) = (d^2f)_a(e_h, e_h). \tag{11.1049}$$

Il est maintenant temps d'écrire $f(a+h) = k(\|h\|)$ et de substituer les dérivées de k par les différentielles de f dans (11.1047) :

$$f(a+h) = k(\|h\|) = f(a) + df_a(h) + \frac{1}{2}(d^2f)_a(h,h) + \|h^2\|\beta_h(\|h\|). \tag{11.1050}$$

Il reste à voir que la fonction $\alpha\colon h \mapsto \beta_h(\|h\|)$ tend vers zéro pour $h \to 0$. En prenant la limite $h \to 0$ dans (11.1050), il est manifeste que la limite du membre de gauche existe et vaut $f(a)$. Donc la limite du membre de droite doit exister et valoir également $f(a)$. Nous en déduisons que la limite de

$$df_a(h) + \frac{1}{2}(d^2f)_a(h,h) + \|h\|^2\beta_h(\|h\|) \tag{11.1051}$$

existe et vaut zéro. La limite des deux premiers termes existe et vaut zéro, donc la limite du troisième existe et vaut zéro :

$$\lim_{h\to 0} \|h\|^2\beta_h(\|h\|) = 0. \tag{11.1052}$$

□

Proposition 11.383.
Soit un ouvert Ω de $\mathbb{R}^n$ et une fonction $f\colon \Omega \to \mathbb{R}$ de classe C^1 et deux fois différentiable sur $]x, x+h[$. Alors il existe $\theta \in]0,1[$ tel que

$$f(x+h) = f(x) + df_x(h) + \frac{1}{2}(d^2f)_{x+\theta h}(h,h). \tag{11.1053}$$

11.35.2 Autres formulations

Exemple 11.384.
Une des façons les plus courantes d'utiliser les formules (11.1026) est de développer $f(a+t)$ pour des petits t en posant $x = a+t$ dans la formule :

$$f(a+t) = f(a) + f'(a)t + f''(a)\frac{t^2}{2} + \epsilon(a+t)t^2 \tag{11.1054}$$

avec $\lim_{t\to 0}\epsilon(a+t) = 0$. Ici, la fonction T dont on parle dans le théorème est $T^a_{f,2}(a+t) = f(a) + f'(a)t + f''(a)\frac{t^2}{2}$.

Lorsque x et y sont deux nombres « proches [93] », nous pouvons développer $f(y)$ autour de $f(x)$:

$$f(y) = f(x) + f'(x)(y-x) + f''(x)\frac{(y-x)^2}{2} + \epsilon(y-x)(y-x)^2, \tag{11.1055}$$

et donc écrire

$$f(x) - f(y) = -f'(x)(y-x) - f''(x)\frac{(y-x)^2}{2} - \epsilon(y-x)(y-x)^2. \tag{11.1056}$$

De cette manière nous obtenons une formule qui ne contient plus que y dans la différence $y-x$. △

11.35.3 Formule et reste

Proposition 11.385.
Soient $f\colon I \subset \mathbb{R} \to \mathbb{R}$ et $a \in \mathrm{Int}(I)$. Soit un entier $k \geqslant 1$. Si f est k fois dérivable en a, alors il existe un et un seul polynôme P de degré $\leqslant k$ tel que

$$f(x) - P(x-a) \in o\big(|x-a|^k\big) \tag{11.1057}$$

lorsque $x \to a$, $x \neq a$. Ce polynôme est donné par

$$P(h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \cdots + \frac{f^{(k)}(a)}{k!}h^k. \tag{11.1058}$$

Notons encore deux façons alternatives d'écrire le résultat. Si $f \in C^k$ il existe une fonction α telle que $\lim_{t\to 0}\alpha(t) = 0$ et

$$f(x) = \sum_{n=0}^{k}\frac{f^{(n)}(a)}{n!}(x-a)^n + (x-a)^n\alpha(x-a). \tag{11.1059}$$

Si $f \in C^{k+1}$ alors

$$f(x) = \sum_{n=0}^{k}\frac{f^{(n)}(a)}{n!}(x-a)^n + (x-a)^{n+1}\xi(x-a) \tag{11.1060}$$

où ξ est une fonction telle que $\xi(t)$ tend vers une constante lorsque $t \to 0$.

La proposition suivant donne une intéressante façon de trouver le reste d'un développement de Taylor.

Proposition 11.386.
Soient I, un intervalle dans $\mathbb{R}$ et $f\colon I \to \mathbb{R}$ une fonction de classe C^k sur I telle que $f^{(k+1)}$ existe sur I. Soient $a \in \mathrm{Int}(I)$ et $x \in I$. Alors il existe $c \in \,]x, a[$ tel que

$$f(x) = \sum_{k=0}^{n}\frac{f^{(k)}(a)}{k!}(x-a)^k + \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}. \tag{11.1061}$$

93. par exemple dans une limite $(x,y) \to (h,h)$.

11.35.4 Reste intégral

Comme son nom l'indique, le « reste intégral » demande de savoir les intégrales. La formule du reste intégral sera donc pour après la définition des intégrales, proposition 19.147.

11.36 Développement limité autour de zéro

Dans cette sections nous supposons toujours que les fonctions sont définies sur un intervalle ouvert de $\mathbb{R}$, I, contenant 0.

11.36.1 Généralités

Définition 11.387.
Soit $f\colon I \to 0$ une fonction définie sur un ouvert I autour de zéro. Nous disons que f admet un ***développement limité*** *autour de 0 à l'ordre n s'il existe une fonction $\alpha\colon I \to \mathbb{R}$ telle que*

$$\begin{cases} f(x) = P_n(x) + x^n\alpha(x) & \text{(11.1062a)} \\ \lim\limits_{x\to 0} \alpha(x) = 0 & \text{(11.1062b)} \end{cases}$$

où $P(x) = a_0 + a_1x + \cdots + a_nx^n$ est une polynôme de degré n. Le polynôme P_n est appelé la ***partie régulière*** *du développement.*

La fonction α est appelé le **reste** du développement et sera parfois noté α_f. Lorsque P est la partie régulière d'un développement limité de f nous notons parfois $f \sim P$.

Proposition 11.388 (Troncature)**.**
Si f admet un développement limité d'ordre n alors il admet également un développement limité d'ordre n' pour tout $n' < n$. Ce dernier s'obtient en tronquant le polynôme d'ordre n à l'ordre n'.

Proposition 11.389 (Unicité)**.**
Si f admet une développement limité alors ce dernier est unique : il existe un unique polynôme P_n d'ordre n et une unique fonction α vérifiant simultanément les deux conditions

$$\begin{cases} f(x) = P_n(x) + x^n\alpha(x), & \text{(11.1063a)} \\ \lim\limits_{x\to 0} \alpha(x) = 0. & \text{(11.1063b)} \end{cases}$$

Exemple 11.390.
En ce qui concerne les séries géométriques de raison x nous savons les formules

$$1 + x + x^2 + \cdots + x^n = \frac{1 - x^{n+1}}{1 - x} \tag{11.1064}$$

et

$$1 + x + x^2 + x^3 + \cdots = \frac{1}{1 - x} \tag{11.1065}$$

pour tout $x \in]-\infty, 1[$. Comparant les deux, il est naturel d'essayer de prendre $1 + x + x^2 + \cdots + x^n$ comme développement limité de la fonction $f(x) = \frac{1}{1-x}$. Pour voir si cela fonctionne, il faut vérifier si « le reste » est bien de la forme $x^n\alpha(x)$ avec $\lim_{x\to 0} \alpha(x) = 0$.

Le reste en question est donné par

$$\frac{1}{1 - x} - 1 - x - x^2 - \ldots - x^n = \frac{1}{1 - x} - \frac{1 - x^{n+1}}{1 - x} = \frac{x^{n+1}}{1 - x} = x^n\frac{x}{1 - x}. \tag{11.1066}$$

En posant $\alpha(x) = \frac{x}{1-x}$ nous avons donc bien

$$f(x) = \frac{1}{1 - x} = 1 + x + x^2 + \cdots + x^n + x^n\frac{x}{1 - x} \tag{11.1067}$$

et $\lim_{x\to 0} \frac{x}{1-x} = 0$. Cela est le développement limité de f à l'ordre n autour de 0. △

La formule des accroissements finis est un cas particulier de développement fini. Supposons que f soit dérivable en 0. En effet nous pouvons facilement trouver la fonction α qui convient. Sachant que $f(0) + xf'(0)$ donne l'approximation affine de f autour de 0, nous cherchons α en écrivant

$$f(x) = f(0) + xf'(0) + x\alpha(x). \tag{11.1068}$$

Cela nous pousse à définir

$$\alpha(x) = \frac{f(x) - f(0)}{x} - f'(0). \tag{11.1069}$$

Notons que cette fonction n'est pas définie en $x = 0$, mais cela n'a pas d'importance : seule la limite $\lim_{x\to 0} \alpha(x)$ nous intéresse. Par définition de la dérivée,

$$\lim_{x\to 0} \alpha(x) = \lim_{x\to 0} \frac{f(x) - f(0)}{x} - f'(0) = 0. \tag{11.1070}$$

En conclusion si f est dérivable, son développement limité à l'ordre 1 est donné par

$$f(x) = f(0) + xf'(0) + x\alpha(x) \tag{11.1071}$$

où $\alpha(x)$ est donnée par la formule (11.1069).

11.36.2 Formule de Taylor-Young

Plus généralement nous avons la proposition suivante qui donne le développement limité de toute fonction dérivable n fois.

Proposition 11.391 (Formule de Taylor-Young)**.**
Soit f une fonction n fois dérivable sur un intervalle I contenant 0. Alors il existe une fonction $\alpha\colon I \to \mathbb{R}$ telle que

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + x^n\alpha(x) \tag{11.1072}$$

et

$$\lim_{x\to 0} \alpha(x) = 0. \tag{11.1073}$$

Cette proposition nous permet de calculer facilement des développements limités tant que nous sommes capables de calculer les dérivées successives de la fonction à développer. Dans l'exemple 11.390 nous avons dû utiliser des astuces et des formules pour déterminer le développement limité de $\frac{1}{1-x}$. Au contraire la formule (11.1072) nous permet de trouver le polynôme en appliquant mécaniquement une formule simple.

Exemple 11.392.
Utilisation de la formule (11.1072) pour déterminer le développement limité de la fonction

$$f(x) = \frac{1}{1-x}. \tag{11.1074}$$

Il faut calculer les dérivées successives de f :

$$f(x) = \frac{1}{1-x} \tag{11.1075a}$$

$$f'(x) = \frac{1}{(1-x)^2} \tag{11.1075b}$$

$$f''(x) = \frac{2}{(1-x)^3} \tag{11.1075c}$$

Avec ces résultats, nous devinons que

$$f^{(n)}(x) = \frac{n!}{(1-x)^{n+1}}. \tag{11.1076}$$

Pour en être sûr nous le prouvons par récurrence. La dérivée de $\frac{n!}{(1-x)^{n+1}}$ est donnée par

$$\frac{n!(n+1)(1-x)^n}{(1-x)^{2n+2}} = \frac{(n+1)!}{(1-x)^{n+2}}. \tag{11.1077}$$

Évaluées en $x = 0$, les dérivées successives de f sont $f(0) = 0$, $f'(0) = 1$, $f''(0) = 2$,...,$f^{(n)}(0) = n!$. Utilisant la formule (11.1072) nous avons

$$f(x) = 1 + x + x^2 + \cdots + x^n + x^n\alpha(x), \tag{11.1078}$$

conformément à ce que nous avions déjà trouvé. △

Exemple 11.393.
Soient $r \in \mathbb{Q}$ et la fonction donnée par

$$f(x) = (1 + x)^r. \tag{11.1079}$$

Nous notons I le domaine de cette fonction : c'est $\mathbb{R}$ si $r > 0$ ou $[-1, \infty]$ si $r < 0$. Si par contre $r = 0$, la fonction est constante et le domaine est $I = \mathbb{R}$.

En ce qui concerne les dérivées[94] : $f'(x) = r(1 + x)^{r-1}$ et plus généralement

$$f^{(k)}(x) = r(r - 1)\dots(r - k + 1)(1 + x)^{r-k} \tag{11.1080}$$

si $k > 0$. Pour $k = 0$ nous avons $f^{(k)}(0) = 1$. Le développement de Taylor-Young est alors

$$(1 + x)^r = 1 + \sum_{k=1}^{n} \frac{r(r - 1)\dots(r - k + 1)}{k!}x^k + x^n\alpha(x). \tag{11.1081}$$

Notons que que si r est un entier, pour $k = r$, le produit au numérateur s'annule et le développement s'arrête.

Dans le développement de $(1 + x)^r$, nous reconnaissons la formule de $\binom{k}{r}$, sauf que nous ne pouvons pas l'écrire avec cette notation lorsque r n'est pas entier. △

Cet exemple fonctionnera encore avec $r \in \mathbb{R}$ au lieu de $r \in \mathbb{Q}$, mais il faudra la proposition 14.66 pour la dérivée

Remarque 11.394.
Pour alléger la notation et ne pas écrire $\dots + x^n\alpha(x)$ nous pouvons aussi écrire

$$f(x) \sim 1 + x + x^2 + \cdots + x^n, \tag{11.1082}$$

mais il est interdit d'écrire

$$f(x) = 1 + x + x^2 + \cdots + x^n \tag{11.1083}$$

en mettant un signe d'égalité entre une fonction et son développement limité[95].

Notons cependant que la proposition 11.391 ne donne pas de moyen simple de trouver la fonction α. Si la fonction f est très régulière dans l'intervalle I on a le résultat suivant.

Proposition 11.395 (Reste dans la forme de Lagrange)**.**
Si la fonction f est dérivable $n + 1$ fois dans I alors il existe $\bar{x}$ dans l'intervalle $[0, x]$ tel que

$$f(x) = P_n(x) + \frac{1}{(n + 1)!}f^{n+1}(\bar{x})x^{n+1}. \tag{11.1084}$$

94. Nous utilisons la proposition 11.363.

95. Il faut cependant être très prudents avec la notation abrégée. Elle pourrait nous faire oublier des informations importantes, voir les développements des fonctions trigonométriques pour un exemple.

11.36.3 Règles de calcul

Les règles suivantes permettent de calculer les développements limités des fonctions qu'on peut écrire comme combinaison de fonctions dont nous savons déjà le développement.

Il est toujours possible de calculer le développement limité d'une fonction par la formule de Taylor-Young (proposition 11.391). Les règles suivantes peuvent nous economiser de l'effort et du temps.

11.36.3.1 Linéarité des développements limités

L'opération qui consiste à prendre le développement limité d'une fonction est une opération linéaire : connaissant les développements limités de f et de g, il suffit de les sommer pour obtenir celui de $f+g$. De même, si λ est une constante, le développement limité de λf est le développement limité de f fois λ.

Proposition 11.396.
Soient λ et μ dans $\mathbb{R}$. Si f et g sont deux fonctions acceptant des développements limités d'ordre n

$$f(x) = P(x) + x^n \alpha_f(x) \tag{11.1085a}$$
$$g(x) = Q(x) + x^n \beta(x) \tag{11.1085b}$$

avec $\lim_{x\to 0} \alpha(x) = \lim_{x\to 0} \beta(x) = 0$, alors la fonction $\lambda f + \mu g$ admet le développement limité

$$(f+g)(x) = (\lambda P + \mu Q)(x) + (\lambda\alpha + \mu\beta)(x). \tag{11.1086}$$

Remarque 11.397.
La forme explicite du reste ne nous interesse pas. Dans la pratique on écrira toujours $(f+g)(x) = (P+Q)(x) + \alpha(x)$, où on appelle α une fonction apportune telle que $\lim_{x\to 0} \alpha(x) = 0$.

Démonstration. Vu les définitions (11.1085) des polynômes P, Q et des restes α et β, l'égalité (11.1086) est une conséquence de la linéarité de la dérivation et de la proposition 11.391

De plus $P+Q$ est un polynôme de degré n dès que P et Q sont des polynômes de degré n, et

$$\lim_{x\to 0} (\lambda\alpha + \mu\beta)(x) = \lim_{x\to 0} \lambda\alpha(x) + \lim_{x\to 0} \mu\beta(x) = 0. \tag{11.1087}$$

Par conséquent $\lambda\alpha + \mu\beta$ est la fonction de reste de $\lambda f + \mu g$. □

Exemple 11.398.
Calculer le développement de la fonction

$$f(x) = 3\sqrt[3]{1+x} + e^{-2x}. \tag{11.1088}$$

Le développement de $\sqrt[3]{1+x}$ est donné par la formule de l'exemple 11.393 avec $\alpha = \frac{1}{3}$. Nous avons donc dans un premier temps

$$\sqrt[3]{1+x} = 1 + \frac{1}{3}x + \frac{\frac{1}{3}\left(\frac{1}{3}-1\right)}{2}x^2 + \frac{\frac{1}{3}\left(\frac{1}{3}-1\right)\left(\frac{1}{3}-2\right)}{6}x^3 + x^3\alpha(x) \tag{11.1089a}$$
$$= 1 + \frac{1}{3}x - \frac{1}{9}x^2 + \frac{5}{81}x^3 + x^3\alpha(x). \tag{11.1089b}$$

Nous avons alors

$$3\sqrt[3]{1+x} + e^{-2x} = 3\Big[1 + \frac{1}{3}x - \frac{1}{9}x^2 + \frac{5}{81}x^3 + x^3\alpha(x)\Big] + 1 - 2x + 2x^2 - \frac{4}{3}x^3 + x^3\beta(x) \tag{11.1090a}$$
$$= 4 - x + \frac{5}{3}x^2 - \frac{31}{27}x^3 + x^3\big(\alpha(x) + \beta(x)\big). \tag{11.1090b}$$

△

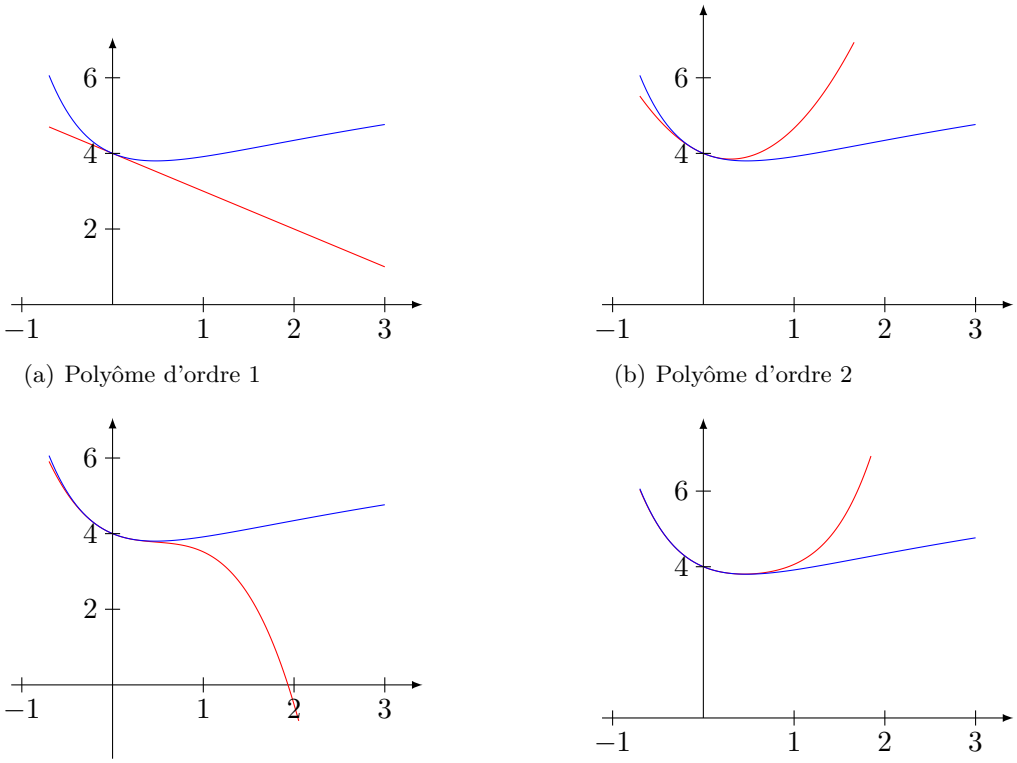
La condition $\lim_{x\to 0}\alpha(x) = 0$ signifie que l'approximation qui consiste à remplacer $f(x)$ par le polynôme n'est pas une trop mauvaise approximation lorsque x est petit. Cela ne signifie rien de plus. En particulier si x est grand, l'approximation polynomiale peut-être (et est souvent) très mauvaise.

À ce propos, notez qu'un polynôme tend toujours vers $\pm\infty$ lorsque x est grand. Une approximation polynomiale d'une fonction bornée est donc toujours (très) mauvaise pour les grandes valeurs de x.

À titre d'exemple nous avons tracé sur la figure 11.12 la fonction

$$f(x) = 3\sqrt[3]{x+1} + e^{-2x} \tag{11.1091}$$

et ses développements limités d'ordre 1 à 3. Il est particulièrement visible que l'approximation est assez bonne pour la partie gauche du graphe sur laquelle la fonction est bien croissante, alors qu'elle est franchement mauvaise sur la droite où le graphe ressemble plutôt à une constante [96].



(a) Polyôme d'ordre 1

(b) Polyôme d'ordre 2

(c) Polyôme d'ordre 3

(d) Polyôme d'ordre 4

FIGURE 11.12 – Les développements limités d'ordre de plus en plus grand de la fonction de l'exemple 11.398. La fonction est en bleu et les « approximations » sont en rouge.

11.36.3.2 Développement limité d'un quotient

Proposition 11.399.
Si P_f est le polynôme du développement limité de f à l'ordre n et P_g celui de g, alors nous obtenons le développement limité de f/g à l'ordre n en effectuant la division selon les puissances croissantes de P_f par P_g.

Attention : il s'agit bien de faire une division selon les puissances croissantes, et non une divisions euclidienne. La division euclidienne de A par B consiste à écrire $A = BQ + R$ avec le reste R de degré le plus *petit* possible. Ici nous voulons avoir un reste de degré le plus *grand* possible.

96. Pouvez-vous cependant dire que vaut $\lim_{x\to\infty} f(x)$?

11.36.3.3 Développement limité d'une fonction composée

Proposition 11.400.
Soient f et g des fonctions admettant des développements limités d'ordre n au voisinage de 0. Nous supposons que $\lim_{x\to 0} g(x) = 0$. Alors la composée $f\big(g(x)\big)$ admet un développement limité d'ordre n au voisinage de 0 qui s'obtient en substituant le développement de g à chaque «x» du développement de f, et en supprimant tous les termes de degré plus élevé que n.

11.37 Développement ailleurs qu'à l'origine

Il est intéressant de développer une fonction au voisinage de zéro lorsque nous nous intéressons à son comportement pour les x pas très grands. Il est toutefois souvent souhaitable de savoir le comportement d'une fonction au voisinage d'autres valeurs que zéro.

Pour développer la fonction f autour de x_0, nous considérons la fonction $h \mapsto f(x_0 + h)$ que nous développons autour de zéro (pour h). L'objectif est de trouver une polynôme P et une fonction α tels que

$$\begin{cases} f(x) = P(x) + (x - x_0)^n \alpha(x) & \text{(11.1092a)} \\ \lim\limits_{x\to x_0} \alpha(x) = 0. & \text{(11.1092b)} \end{cases}$$

En pratique, le développement limité à l'ordre n d'une fonction autour d'un point x_0 quelconque à l'intérieur de son domaine prend la forme suivante, qui généralise la formule de Taylor-Young vue dans la proposition 11.391

Proposition 11.401 (Formule de Taylor-Young, cas général)**.**
Soit f une fonction n fois dérivable sur un intervalle I contenant x_0. Alors il existe une fonction $\alpha\colon I \to \mathbb{R}$ telle que

$$\begin{aligned} f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \\ + \frac{f^{(3)}(x_0)}{3!}(x - x_0)^3 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + (x - x_0)^n \alpha(x - x_0) \end{aligned} \tag{11.1093}$$

et

$$\lim_{t\to 0} \alpha(t) = 0. \tag{11.1094}$$

11.38 Développement au voisinage de l'infini

Il est souvent utile de connaitre le comportement d'une fonction pour les grandes valeurs de x et de déterminer ses asymptotes éventuelles. La technique que nous allons utiliser consiste à poser $x = \frac{1}{h}$ et de développer la fonction "auxiliaire" $g(h) = f(1/h)$ autour de $h = 0$. La limite avec $h \to 0^+$ donnera le comportement pour $x \to \infty$ et la limite $h \to 0^-$ donnera le comportement pour $x \to -\infty$.

Dans le cas d'une développement autour de $\pm\infty$ nous ne parlons plus de développement *limité* mais de **développement asymptotique**.

11.38.1 La fonction puissance : remarques pour la suite

Il y a encore de nombreuses choses à dire sur la fonction puissance. Pour savoir lesquelles, voir le thème 33.

11.39 Fonctions réelles de deux variables réelles

Une **fonction réelle de 2 variables réelles** est une fonction $f : A \subset \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto z = f(x, y)$.

Le **graphe de** f, noté $\operatorname{Gr} f$, est un sous-ensemble de $\mathbb{R}^3$:

$$\operatorname{Gr} f = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y) \in A \text{ et } z = f(x, y)\}$$

Les **courbes de niveau** de la fonction f sont obtenues en posant $f(x, y) = \lambda$.

11.39.1 Limites de fonctions à deux variables

Ici nous n'allons pas entrer dans tous les détails, mais simplement mentionner les quelques techniques les plus courantes.

Théorème 11.402.
Soient deux fonctions $f\colon \mathbb{R}^n \to \mathbb{R}^p$ et $g\colon \mathbb{R}^p \to \mathbb{R}^q$. Si a est un point adhérent au domaine de $g \circ f$ et si

$$\begin{aligned} \lim_{x\to a} f(x) &= b \\ \lim_{y\to b} g(y) &= c, \end{aligned} \tag{11.1095}$$

alors

$$\lim_{x\to a} (g \circ f)(x) = c. \tag{11.1096}$$

Les techniques usuelles sont

(1) La règle de l'étau. Cette technique demande un peu plus d'imagination parce qu'il faut penser à un « truc » différent pour chaque exercice. En revanche, la justification est facile : il y a un théorème qui dit que ça marche.

(2) Lorsqu'on applique la règle de l'étau, penser à

$$|x| = \sqrt{x^2} \leqslant \sqrt{x^2 + y^2}. \tag{11.1097}$$

Cela permet de majorer le numérateur. Attention : ce genre de majoration fonctionne seulement au numérateur : agrandir le dénominateur ferait diminuer la fraction.

(3) Il n'est pas vrai que

$$|x| = \sqrt{x^2} \leqslant \sqrt{x^4} \leqslant \sqrt{x^4 + 2y^4}. \tag{11.1098}$$

En effet, si x est petit, alors $x^2 > x^4$, et non le contraire.

Une technique très efficace pour les limites $(x, y) \to (0, 0)$ est le passage aux coordonnées polaires. Il s'agit de poser

$$\begin{cases} x = r\cos(\theta) & \text{(11.1099a)} \\ y = r\sin(\theta) & \text{(11.1099b)} \end{cases}$$

et puis de faire la limite $r \to 0$.

Si la limite obtenue **ne dépend pas de** θ, alors c'est la limite cherchée. Voici quelque exemples.

Exemple 11.403.
Calculer les limites suivantes :

(1) $\lim_{(x,y)\to(0,0)} \frac{x-y}{x+y}$

(2) $\lim_{(x,y)\to(0,0)} \frac{(xy)^2}{(x+y)^2+(x-y)^2}$

(3) $\lim_{(x,y)\to(0,0)} \frac{xy^3}{x^2+y^2}$

(4) $\lim_{(x,y)\to(0,0)} \frac{x\sin(y)}{\sqrt{x^2+y^2}}$

Tentez de les faire par vous-même avant de regarder la solution qui suit.

(1) Ici la méthode des chemins pour est particulièrement éclairante. Regardons d'abord la fonction sur la droite $x = y$. Nous avons

$$f(x,y) = \frac{x-x}{2x} = 0. \tag{11.1100}$$

Donc la fonction est nulle sur toute la ligne.
Si nous regardons maintenant la ligne verticale $x = 0$, nous avons

$$f(0,y) = \frac{-y}{y} = -1, \tag{11.1101}$$

donc la fonction vaut -1 sur toute la ligne verticale.

(2)

(3) Regardons la technique des coordonnées polaires. Nous remplaçons x par $r\cos(\theta)$ et y par $r\sin(\theta)$:

$$f(r,\theta) = \frac{r^4\cos(\theta)\sin^3(\theta)}{r^2} = r^2\cos(\theta)\sin^3(\theta). \tag{11.1102}$$

Cette fonction tend vers zéro quand $r \to 0$. Nous avons donc

$$\lim_{(x,y)\to(0,0)} f(x,y) = 0. \tag{11.1103}$$

Pour cet exercice nous pouvons aussi utiliser la règle de l'étau en écrivant d'abord

$$0 \leqslant |f(x,y)| \leqslant \frac{|x||y^3|}{|x^2+y^2|}. \tag{11.1104}$$

Mais on a $|x| \leqslant \sqrt{x^2+y^2}$, $|y| \leqslant \sqrt{x^2+y^2}$ et $|x^2+y^2| = \left(\sqrt{x^2+y^2}\right)^2$, donc

$$0 \leqslant |f(x,y)| \leqslant \frac{\sqrt{x^2+y^2}\left(\sqrt{x^2+y^2}\right)^3}{\left(\sqrt{x^2+y^2}\right)^2} = \left(\sqrt{x^2+y^2}\right)^2 \to 0. \tag{11.1105}$$

(4) En passant aux polaires, nous avons

$$f(r,\theta) = \frac{r\cos\theta\sin\big(r\sin\theta\big)}{r} = \cos(\theta)\sin\big(r\sin\theta\big). \tag{11.1106}$$

La limite de cette dernière fonction lorsque $r \to 0$ vaut zéro.
Une autre façon de procéder consiste à multiplier et diviser par y de telle façon à faire apparaitre $\sin(y)/y$ dont nous connaissons la limite :

$$f(x,y) = \frac{\sin(y)}{y} \cdot \frac{xy}{\sqrt{x^2+y^2}}. \tag{11.1107}$$

La limite du premier facteur est 1, tandis que le second peut être traité de façon classique en prenant la valeur absolue et en majorant $|x|$ par $\sqrt{x^2+y^2}$.

△

11.39.2 Dérivées partielles

La **dérivée partielle** par rapport à x au point (x,y) est notée

$$\frac{\partial f}{\partial x}(x,y) \tag{11.1108}$$

et se calcule en dérivant f par rapport à x en considérant que y est constante.

De la même manière, la dérivée partielle par rapport à y au point (x,y) est notée

$$\frac{\partial f}{\partial y}(x,y) \tag{11.1109}$$

et se calcule en dérivant f par rapport à y en considérant que x est constante.

Pour les dérivées partielles secondes,

— $f''_{xx}(x,y) = (f'_x)'_x = \frac{\partial^2 f}{\partial x^2}(x,y) = \frac{\partial}{\partial x}(\frac{\partial f}{\partial x})$.
— $f''_{yy}(x,y) = (f'_y)'_y = \frac{\partial^2 f}{\partial y^2}(x,y) = \frac{\partial}{\partial y}(\frac{\partial f}{\partial y})$.
— $f''_{xy}(x,y) = (f'_x)'_y = (f'_y)'_x = f''_{yx}(x,y)$ ou $\frac{\partial^2 f}{\partial x \partial y}(x,y) = \frac{\partial}{\partial x}(\frac{\partial f}{\partial y}) = \frac{\partial}{\partial y}(\frac{\partial f}{\partial x}) = \frac{\partial^2 f}{\partial y \partial x}(x,y)$.

11.39.3 Différentielle et accroissement

La **différentielle totale** de f au point (a,b) est donnée, quand elle existe (!), par la formule

$$df(a,b) = \frac{\partial f}{\partial x}(a,b)dx + \frac{\partial f}{\partial y}(a,b)dy. \tag{11.1110}$$

De la même façon que la formule des accroissements finis disait que $f(x+a) \simeq f(x) + af'(x)$, en deux dimensions nous avons que l'**accroissement** approximatif de f au point (a,b) pour des accroissements Δx et Δy est

$$f(x+\Delta x, y+\Delta y) = f(x,y) + \Delta x \frac{\partial f}{\partial x}(x,y) + \Delta y \frac{\partial f}{\partial y}(x,y). \tag{11.1111}$$

Le **plan tangent** au graphe de f au point $\big(a,b,f(a,b)\big)$ est

$$T_{(a,b)}(x,y) = f(a,b) + \frac{\partial f}{\partial x}(a,b)(x-a) + \frac{\partial f}{\partial y}(a,b)(y-b) \tag{11.1112}$$

essayez d'écrire l'équation de la droite tangente au graphe de $f(x)$ au point $x = a$ en termes de la dérivée de f, et comparez votre résultat à cette formule.

Un des principaux théorèmes pour tester la différentiabilité d'une fonction est le suivant.

Théorème 11.404.
Soit une fonction $f\colon \mathbb{R}^m \to \mathbb{R}^p$. Si les dérivées partielles existent dans un voisinage de a et donc continues en a, alors f est différentiable en a.

Le plus souvent, nous prouvons qu'une fonction est différentiable en calculant les dérivées partielles et en montrant qu'elles sont continues.

Dérivation implicite : Soit $F(x, f(x)) = 0$ la représentation implicite d'une fonction $y = f(x)$ alors

$$y' = f'(x) = -\frac{F'_x}{F'_y}.$$

11.40 Les fonctions à valeurs vectorielles

Jusqu'à présent nous avons vu des fonctions de plusieurs variables qui prenaient leurs valeurs dans $\mathbb{R}$. Nous allons maintenant voir ce qu'il se passe lorsque les fonctions prennent leurs valeurs dans $\mathbb{R}^3$.

Une fonction d'une variable est dite **à valeurs vectorielles** lorsque

$$\begin{aligned} f\colon I \subset \mathbb{R} &\to \mathbb{R}^3 \\ f(x) &= \begin{pmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{pmatrix}. \end{aligned} \tag{11.1113}$$

Les fonctions $f_i\colon \mathbb{R} \to \mathbb{R}$ sont les **composantes** de f. Ce que nous avons raconté à propos des dérivées passe facilement :

$$\frac{f(a+\epsilon) - f(a)}{\epsilon} = \begin{pmatrix} \frac{f_1(a+\epsilon)-f_1(a)}{\epsilon} \\ \frac{f_2(a+\epsilon)-f_2(a)}{\epsilon} \\ \frac{f_3(a+\epsilon)-f_3(a)}{\epsilon} \end{pmatrix}. \tag{11.1114}$$

En particulier dès que les fonctions f_i sont dérivables, nous avons

$$f'(a) = \begin{pmatrix} f_1'(a) \\ f_2'(a) \\ f_3'(a) \end{pmatrix} \tag{11.1115}$$

comme dérivée de la fonction. Cette dérivée est un vecteur.

Exemple 11.405.
Si

$$f\colon x \in \mathbb{R} \mapsto \begin{pmatrix} x^2 e^x \\ \cos(x^2) \\ x^3 + x \end{pmatrix}, \tag{11.1116}$$

alors

$$f'(x) = \begin{pmatrix} 2xe^x + x^2 e^x \\ -2x\sin(x^2) \\ 3x^2 + 1 \end{pmatrix}. \tag{11.1117}$$

△

11.41 Fonctions vectorielles de plusieurs variables

Ce sont les fonctions de la forme

$$\begin{aligned} f\colon \mathbb{R}^3 &\to \mathbb{R}^3 \\ \begin{pmatrix} x \\ y \\ z \end{pmatrix} &\mapsto \begin{pmatrix} f_1(x,y,z) \\ f_2(x,y,z) \\ f_3(x,y,z) \end{pmatrix}. \end{aligned} \tag{11.1118}$$

En ce qui concerne les dérivées, tout se passe comme avant. Si les dérivées partielles des composantes f_i existent au point $a \in \mathbb{R}^3$, alors

$$\frac{\partial f}{\partial x}(a) = \begin{pmatrix} \partial_x f_1(a) \\ \partial_x f_2(a) \\ \partial_x f_3(a) \end{pmatrix}, \quad \frac{\partial f}{\partial y}(a) = \begin{pmatrix} \partial_y f_1(a) \\ \partial_y f_2(a) \\ \partial_y f_3(a) \end{pmatrix}, \quad \frac{\partial f}{\partial z}(a) = \begin{pmatrix} \partial_z f_1(a) \\ \partial_z f_2(a) \\ \partial_z f_3(a) \end{pmatrix}. \tag{11.1119}$$

11.42 Limites à plusieurs variables

Proposition 11.406.
Soit $f\colon D \subset \mathbb{R}^m \to \mathbb{R}^n$. Nous avons

$$\lim_{x \to a} f(x) = \ell \tag{11.1120}$$

si et seulement si

$$\lim_{x \to a} f_i(x) = \ell_i \tag{11.1121}$$

pour tout $i \in \{1, \dots, n\}$ où $f_i(x)$ dénote la i-ème composante de $f(x)$ et ℓ_i la i-ème composante de $\ell \in \mathbb{R}^n$.

Cette proposition revient à dire que la convergence d'une fonction est équivalente à la convergence de chacune de ses composantes.

Démonstration. L'élément clef de la preuve est le fait que pour tout vecteur $u \in \mathbb{R}^p$, nous ayons l'inégalité

$$|u_i| \leqslant \sqrt{\sum_{k=1}^{p} |u_k|^2} = \|u\|. \tag{11.1122}$$

La norme (dans $\mathbb{R}^p$) d'un vecteur est plus grande ou égale à la valeur absolue de chacune de ses composantes.

Supposons que nous ayons une fonction dont chacune des composantes a une limite en a : $\lim_{x\to a} f_i(x) = \ell_i$. Montrons que dans ce cas la fonction f tend vers ℓ. Si nous considérons $\varepsilon > 0$, par définition de la limite de chacune des fonctions f_i, il existent des δ_i tels que

$$\|x - a\|_{\mathbb{R}^m} < \delta_i \Rightarrow |f_i(x) - \ell_i| < \varepsilon. \tag{11.1123}$$

Notez que la norme à gauche est une norme dans $\mathbb{R}^m$ et que celle à droite est une simple valeur absolue dans $\mathbb{R}$. Considérons $\delta = \min\{\delta_i\}_{i=1,\dots n}$. Si $\|x - a\| < \delta$, alors

$$\|f(x) - \ell\| = \sqrt{\sum_{i=1}^{n} |f_i(x) - \ell_i|^2} < \sqrt{\sum_{i=1}^{n} \varepsilon^2} = \sqrt{n\varepsilon^2} = \sqrt{n}\varepsilon. \tag{11.1124}$$

Nous voyons qu'en choisissant les δ_i tels que $|f_i(x) - \ell_i| < \varepsilon$, nous trouvons $\|f(x) - \ell\| < \sqrt{n}\varepsilon$. Afin d'obtenir $\|f(x) - \ell\| < \varepsilon$, nous choisissons donc les δ_i de telle manière a avoir $|f_i(x) - \ell_i| < \varepsilon/\sqrt{n}$.

Nous avons donc prouvé que la limite composante par composante impliquait la limite de la fonction. Nous devons encore prouver le sens inverse.

Supposons donc que $\lim_{x\to a} f(x) = \ell$, et prouvons que nous ayons $\lim_{x\to a} f_i(x) = \ell_i$ pour chaque i. Soit $\varepsilon > 0$ et $\delta > 0$ tel que $\|x - a\| < \delta$ implique $\|f(x) - \ell\| < \varepsilon$. Avec ces choix, nous avons

$$|f_i(x) - \ell_i| \leqslant \|f(x) - \ell\| < \varepsilon \tag{11.1125}$$

où nous avons utilisé la majoration (11.1122) avec $f(x) - \ell$ en guise de u. □

De même, pour la continuité nous avons la proposition suivante :

Proposition 11.407.
Soit une fonction $f\colon D \subset \mathbb{R}^m \to \mathbb{R}^n$ et $a \in D$. La fonction f est continue en a si et seulement si chacune de ses composantes l'est, c'est-à-dire si et seulement si chacune des fonctions $f_i\colon D \to \mathbb{R}$ est continue en a.

Essayez de prouver cette proposition directement par la définition de la continuité, en suivant pas à pas la démonstration de la proposition 11.406.

Proposition 11.408.
Soit $f\colon \mathbb{R}^m \to \mathbb{R}$ et a, un point du domaine de f telle que $f(a) > 0$. Alors il existe un rayon r tel que $f(x) > 0$ pour tout x dans $B(a, r)$.

Cette proposition signifie que si la fonction est strictement positive en un point, alors elle restera strictement positive en tous les points « pas trop loin ».

Démonstration. Prenons $\varepsilon = f(a)/2$ dans la définition de la continuité. Il existe donc un rayon δ tel que pour tout x dans $B(a, \delta)$,

$$|f(x) - f(a)| \leqslant \frac{f(a)}{2}, \tag{11.1126}$$

en d'autres termes, $f(x) \in B\big(f(a), \frac{f(a)}{2}\big)$. évidemment aucun nombre négatif ne fait partie de cette dernière boule lorsque $f(a)$ est strictement positif. □

Corolaire 11.409.
Si $f\colon \mathbb{R}^m \to \mathbb{R}$ est une fonction continue, alors l'ensemble

$$A = \{x \in \mathbb{R}^m \text{ tels que } f(x) \neq 0\} \tag{11.1127}$$

est ouvert.

Démonstration. Soit $x \in A$. Si $x > 0$ (le cas $x < 0$ est laissé en exercice), alors il existe une boule autour de x sur laquelle f reste strictement positive (proposition 11.408). Cette boule est donc contenue dans A. Étant donné qu'autour de chaque point de A nous pouvons trouver une boule contenue dans A, ce dernier est ouvert. □

Exemple 11.410.
Soit $GL_n(\mathbb{R})$ l'ensemble des matrices $n \times n$ inversibles. Nous allons montrer que $GL_n(\mathbb{R})$ est un ouvert de $\mathbb{R}^{n^2}$. L'identification entre les vecteurs et les matrices consiste simplement à « déplier » la matrice pour en faire un vecteur. Par exemple, en dimension deux,

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \mapsto \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \in \mathbb{R}^4. \tag{11.1128}$$

En dimension 3,

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \mapsto \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{pmatrix} \in \mathbb{R}^9. \tag{11.1129}$$

Une matrice est inversible si et seulement si son déterminant est non nul. Or le déterminant est un polynôme en les composantes de la matrice. En dimension deux, nous avons

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc, \tag{11.1130}$$

mais en écriture « dépliée », nous pouvons aussi bien écrire

$$\det \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = ad - bc. \tag{11.1131}$$

En dimension 3, le déterminant est donc un polynôme des 9 variables qui apparaissent dans le vecteur « déplié ». En général, dans $\mathbb{R}^{n^2}$, nous considérons donc le polynôme $\det\colon \mathbb{R}^{n^2} \to \mathbb{R}$ qui à un vecteur $X \in \mathbb{R}^{n^2}$ fait correspondre le déterminant de la matrice obtenue en « repliant » le vecteur X.

Donc dans $\mathbb{R}^{n^2}$, l'ensemble des matrices inversibles est donné par l'ensemble des vecteurs sur lesquels le polynôme det ne s'annule pas, c'est-à-dire

$$\{X \in \mathbb{R}^{n^2} \text{ tels que } \det(X) \neq 0\}. \tag{11.1132}$$

Mais le déterminant est un polynôme, et donc une fonction continue. Cet ensemble est par conséquence ouvert par le corolaire 11.409. △

La proposition suivante montre que la limite peut « passer à travers » les fonctions continues.

Proposition 11.411 (limite de fonction composée)**.**
Soit $f\colon \mathbb{R}^n \to \mathbb{R}^q$ et $g\colon \mathbb{R}^m \to \mathbb{R}^n$ telles que

$$\lim_{x \to a} g(x) = p \tag{11.1133a}$$

$$\lim_{y \to p} f(y) = q \tag{11.1133b}$$

Alors nous avons $\lim_{x \to a} (f \circ g)(x) = q$.

Démonstration. Comme presque toute preuve à propos de limite ou de continuité, nous commençons par choisir $\varepsilon > 0$. Nous devons montrer qu'il existe un δ tel que $\|x - a\| \leqslant \delta$ implique $\|f\big(g(x)\big) - q\| \leqslant \varepsilon$.

La limite (11.1133b) impose l'existence d'un $\tilde{\delta}$ tel que $\|y - p\| \leqslant \tilde{\delta}$ implique $\|f(y) - q\| \leqslant \varepsilon$, tandis que la limite (11.1133a) donne un δ tel que $\|x - a\| \leqslant \delta$ implique $\|g(x) - p\| \leqslant \tilde{\delta}$ (nous avons pris $\tilde{\delta}$ en guise de ε dans la définition de la limite pour g).

Avec ces choix, si $\|x - a\| \leqslant \delta$, alors $\|g(x) - p\| \leqslant \tilde{\delta}$, et par conséquent,

$$\|f\big(g(x)\big) - q\| \leqslant \varepsilon, \tag{11.1134}$$

ce que nous voulions. □

De façon pragmatique, la proposition 11.411 nous fournit une formule pour les limites de fonctions composée :

$$\lim_{x\to a}(f \circ g)(x) = \lim_{y\to \lim_{x\to a} g(x)} f(y) \tag{11.1135}$$

lorsque f est continue.

Remarque 11.412.
La formule (11.1135) ne peut pas être utilisée à l'envers. Il existe des cas où $\lim_{x\to a}(g \circ f)(x) = q$, et $\lim_{x\to a} f(x) = p$ sans pour autant avoir $\lim_{y\to q} g(y) = q$. Par exemple

$$g(x) = \begin{cases} 2 & \text{si } x \geqslant 0 \\ 0 & \text{si } x < 0 \end{cases} \tag{11.1136a}$$

$$f(x) = |x|. \tag{11.1136b}$$

Nous avons $(g \circ f)(x) = 2$ pour tout x, ainsi que $\lim_{x\to 0} f(x) = 0$, mais la limite $\lim_{y\to 0} g(y)$ n'existe pas.

En général, lorsqu'un ensemble est donné par des inégalités, prendre la fermeture consiste à transformer les inégalités strictes en inégalités non strictes ; prendre l'intérieur consiste à rendre stricte toutes les inégalités ; la frontière consiste *en gros* à transformer toutes les inégalités en égalités (nous allons voir que pour la frontière, c'est un peu plus de travail). Comprenez bien que cela n'est vrai que « en général ». Il faut toujours bien regarder sur chaque exemple s'il n'y a pas l'un ou l'autre point problématique.

La proposition 11.408 sera une des clefs pour dire que si une inégalité stricte est satisfaite en un point, alors elle sera satisfaite en tout point dans un voisinage. Voir aussi l'exemple 11.413.

Exemple 11.413.
Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue. Montrer que le sous-ensemble A de $\mathbb{R}^2$ défini par $A = \big\{(x, y) \in \mathbb{R}^2;\ y > f(x)\big\}$ est ouvert.

Prouver que l'ensemble $B = \{(x, y) \in \mathbb{R}^2 \text{ tels que } y \geqslant f(x)\}$ est fermé.

Nous considérons la fonction $g\colon \mathbb{R}^2 \to \mathbb{R}$ donnée par $g(x, y) = f(x) - y$. Cela est une fonction continue parce que c'est une différence de fonctions continues. Par définition,

$$A = \{(x, y) \in \mathbb{R}^2 \text{ tel que } g(x, y) > 0\}. \tag{11.1137}$$

Par la proposition 11.408, autour de chaque point (x, y) tel que $g(x, y) > 0$ (c'est-à-dire autour de chaque point de A), il existe une boule sur laquelle g reste strictement positive. L'ensemble A est donc ouvert.

Pour prouver que l'ensemble B est fermé, prouver que le complémentaire est ouvert, c'est-à-dire que les points tels que $y - f(x) < 0$ forment un ouvert. Cela revient au même que ce que nous avons fait pour A.

△

11.43 Champs de vecteurs

Un champ de vecteur est une fonction $f\colon \mathbb{R}^3 \to \mathbb{R}^3$. Géométriquement, il s'agit simplement de mettre un vecteur en chaque point de l'espace. Cela arrive très souvent en physique.

Exemple 11.414.
Si un fluide (eau, gaz) coule dans un tube, en tout point le point a une vitesse, qui sera un vecteur généralement dirigé le long du tube. △

Exemple 11.415.
La force d'attraction de la Terre sur une masse m située au point $r = (x, y, z)$ est donnée par

$$F(r) = -G\frac{Mmr}{\|r\|^3}. \tag{11.1138}$$

Dans cette expression, tant r que $F(r)$ sont des vecteurs. Nous l'avons représenté sur la figure 11.13.

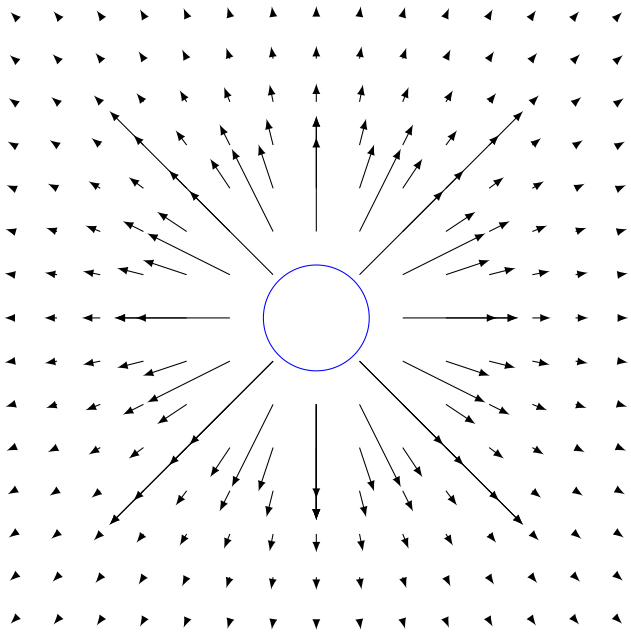


FIGURE 11.13 – Le champ de gravitation de la Terre.

L'application

$$\begin{aligned} F\colon \mathbb{R}^3 &\to \mathbb{R}^3 \\ r &\mapsto F(r) \end{aligned} \tag{11.1139}$$

est le champ gravitationnel de la Terre.

△

11.43.1 Matrice jacobienne

La **matrice jacobienne** de la fonction $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ au point $a \in \mathbb{R}^3$ est la matrice dont les colonnes sont les vecteurs $\frac{\partial f}{\partial x}(a)$, $\frac{\partial f}{\partial y}(a)$ et $\frac{\partial f}{\partial z}(a)$, c'est-à-dire

$$J_f(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x}(a) & \frac{\partial f_1}{\partial y}(a) & \frac{\partial f_1}{\partial z}(a) \\ \frac{\partial f_2}{\partial x}(a) & \frac{\partial f_2}{\partial y}(a) & \frac{\partial f_2}{\partial z}(a) \\ \frac{\partial f_3}{\partial x}(a) & \frac{\partial f_3}{\partial y}(a) & \frac{\partial f_3}{\partial z}(a) \end{pmatrix}. \tag{11.1140}$$

Exemple 11.416.
Si

$$f(x,y,z) = \begin{pmatrix} xye^z \\ x^2 + \cos(yz) \\ xyz \end{pmatrix}, \tag{11.1141}$$

alors

$$J_f(x,y,z) = \begin{pmatrix} ye^z & xe^z & xye^z \\ 2x & -z\sin(yz) & -y\sin(yz) \\ yz & xz & xy \end{pmatrix}. \tag{11.1142}$$

△

11.44 Divergence, rotationnel et l'opérateur nabla

Nous avons déjà vu le gradient d'une fonction $f\colon \mathbb{R}^3 \to \mathbb{R}$

$$\nabla f(x,y,z) = \begin{pmatrix} \partial_x f(x,y,z) \\ \partial_y f(x,y,z) \\ \partial_z f(x,y,z) \end{pmatrix} \tag{11.1143}$$

Afin de définir la divergence et le rotationnel, nous introduisons ∇ sous une forme un peu plus abstraite comme le « vecteur »

$$\nabla = \begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \end{pmatrix}. \tag{11.1144}$$

Vue comme ça, la formule (11.1143) est claire.

Si F est un champ de vecteurs, nous introduisons la **divergence** de F par

$$\nabla \cdot F = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}. \tag{11.1145}$$

Cela est une fonction. Et nous introduisons le rotationnel du champ de vecteur F par

$$\begin{aligned} \nabla \times F &= \begin{vmatrix} e_x & e_y & e_z \\ \partial_x & \partial_y & \partial_z \\ F_x & F_y & F_z \end{vmatrix} \\ &= \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right) e_x - \left(\frac{\partial F_z}{\partial x} - \frac{\partial F_x}{\partial z}\right) e_y + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right) e_z. \end{aligned} \tag{11.1146}$$

Cela est un champ de vecteur. En utilisant le symbole complètement antisymétrique ϵ_{ijk}, le rotationnel d'un champ de vecteur peut s'écrire

$$\nabla \times F = \sum_{ijk} \epsilon_{ijk} \partial_i F_j e_k. \tag{11.1147}$$

Le gradient, la divergence et le rotationnel consistent à appliquer simplement à ∇ est trois produits qu'on peut effectuer sur un vecteur :

(1) Le produit d'un vecteur par un scalaire multiplie chacune des composantes :

$$\begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \end{pmatrix} f = \begin{pmatrix} \partial_x f \\ \partial_y f \\ \partial_z f \end{pmatrix}. \tag{11.1148}$$

(2) Le produit scalaire d'un vecteur avec un autre vecteur donne lieu à la divergence :

$$\begin{pmatrix} \partial_x \\ \partial_y \\ \partial_z \end{pmatrix} \cdot \begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}. \tag{11.1149}$$

(3) Le produit vectoriel de deux vecteurs :

$$\begin{pmatrix}\partial_x\\ \partial_y\\ \partial_z\end{pmatrix}\times\begin{pmatrix}F_x\\ F_y\\ F_z\end{pmatrix}=\begin{vmatrix}e_x & e_y & e_z\\ \partial_x & \partial_y & \partial_z\\ F_x & F_y & F_z\end{vmatrix}. \tag{11.1150}$$

Ces trois opérations joueront un rôle central en électromagnétisme dans les équations de Maxwell.

Exemple 11.417.
Soit $F(x,y,z)=xe_x+xye_y+e_z$, c'est-à-dire

$$F(x,y,z)=\begin{pmatrix}x\\ xy\\ 1\end{pmatrix}. \tag{11.1151}$$

Son rotationnel est donné par

$$\nabla\times F=\begin{vmatrix}e_x & e_y & e_z\\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial y}\\ x & xy & 1\end{vmatrix}=(0-0)e_x-(0-0)e_y+(y-0)e_z=ye_z=\begin{pmatrix}0\\ 0\\ y\end{pmatrix}. \tag{11.1152}$$

△

Afin d'étudier comment se comporte la composition de ces opérateurs, nous aurons besoin de ce lemme que nous n'énoncerons pas précisément.

Lemme 11.418.
Si $f\colon\mathbb{R}^3\to\mathbb{R}$ est une fonction de classe C^2, alors on peut permuter l'ordre des dérivées :

$$\begin{aligned}\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)&=\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\\ \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial z}\right)&=\frac{\partial}{\partial z}\left(\frac{\partial f}{\partial x}\right)\\ \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial y}\right)&=\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial z}\right)\end{aligned} \tag{11.1153}$$

La fonction

$$(x,y,z)\mapsto\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right)(x,y,z) \tag{11.1154}$$

sera notée

$$\frac{\partial^2 f}{\partial x\partial y}. \tag{11.1155}$$

Il y a deux propriétés importantes :

Théorème 11.419.
Soit $f\colon\mathbb{R}^3\to\mathbb{R}$ une fonction de classe C^2. Alors

$$\nabla\times(\nabla f)=0. \tag{11.1156}$$

Si $F\colon\mathbb{R}^3\to\mathbb{R}^3$ est un champ de vecteurs de classe C^2, alors

$$\nabla\cdot(\nabla\times F)=0. \tag{11.1157}$$

Démonstration. Ce sont seulement deux calculs qui manipulent les définitions. Pour le premier, la divergence de f est le champ de vecteurs

$$\nabla f=\frac{\partial f}{\partial x}e_x+\frac{\partial f}{\partial y}e_y+\frac{\partial f}{\partial z}e_z. \tag{11.1158}$$

En mettant ce champ dans la définition du rotationnel,

$$\begin{aligned}\nabla \times (\nabla f) = \begin{vmatrix} e_x & e_y & e_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{vmatrix} &= \left[\frac{\partial}{\partial y}\left(\frac{\partial f}{\partial z}\right) - \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial y}\right)\right] e_x \\ &- \left[\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial z}\right) - \frac{\partial}{\partial z}\left(\frac{\partial f}{\partial x}\right)\right] e_y \\ &+ \left[\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y}\right) - \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\right] e_z.\end{aligned} \tag{11.1159}$$

En utilisant le lemme 11.418, chacun des termes fait zéro.

La seconde propriété se démontre en utilisant le même type de calcul. □

Remarque 11.420.
Il n'y a pas de propriétés du même style pour la combinaison $\nabla \times (\nabla \cdot F)$ pour le rotationnel de la divergence. En effet la divergence d'un champ de vecteur est une fonction, et il n'y a pas de rotationnel pour une fonction.

11.45 Interprétation géométrique et physique de la divergence

En physique, on dit qu'un champ de vecteurs à divergence nulle est **incompressible**. Nous allons essayer de comprendre pourquoi. Lorsqu'un fluide incompressible se déplace, il faut qu'en chaque point il y ait autant de fluide qui rentre que de fluide qui sort. Nous allons voir sur quelques exemples que la divergence d'un champ de vecteurs est le « bilan de masse » d'un fluide qui se déplace selon le champ de vecteurs.

Si en un point la divergence est positive, cela signifie qu'il y a une perte de masse et si la divergence est négative, cela signifie qu'il y a une accumulation de masse.

Prenons par exemple un fluide qui se déplace selon le champ de vitesse montré à figure 11.14.

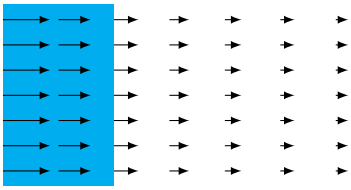


FIGURE 11.14 – Le champ de vecteurs $F(x, y) = \frac{1}{x}(1, 0)$.

Étant donné que la vitesse diminue lorsque x avance, il y a une accumulation de fluide. Regardez en effet la quantité de fluide qui rentre dans le rectangle par rapport à la quantité de fluide qui en sort. Ce champ de vecteurs a pour équation :

$$F(x, y) = \frac{1}{x}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1/x\\0\end{pmatrix}. \tag{11.1160}$$

Sa divergence vaut donc

$$(\nabla \cdot F)(x, y) = \frac{\partial F_x}{\partial x}(x, y) + \underbrace{\frac{\partial F_y}{\partial y}(x, y)}_{=0} = -\frac{1}{x^2}. \tag{11.1161}$$

Cette divergence étant négative, il y a bien accumulation de fluide en tout point, et d'autant plus que x est petit.

Exemple 11.421.
Prenons le champ de vecteurs tournant

$$F(x,y) = \frac{1}{\sqrt{x^2+y^2}}\begin{pmatrix} y \\ -x \end{pmatrix} \tag{11.1162}$$

représenté à la figure 11.15. Cela est un vecteur qui est constamment perpendiculaire au rayon.

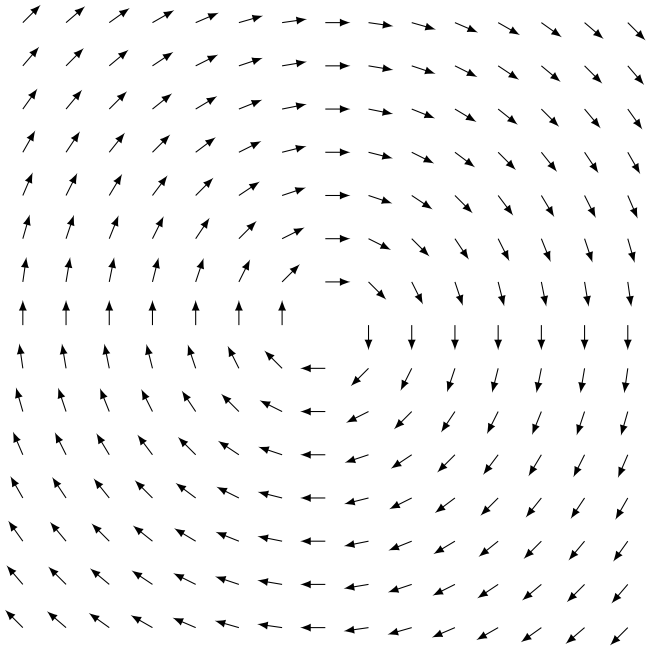


FIGURE 11.15 – Le champ de vecteurs $F(x,y) = (y,-x)$.

Un fluide dont la vitesse serait donné par ce champ de vecteur se contente de tourner. Intuitivement il ne devrait pas y avoir de divergence parce qu'il n'y a aucune accumulation de fluide. En effet,

$$\nabla \cdot F(x,y) = \frac{-2xy}{(x^2+y^2)^2} + \frac{2xy}{(x^2+y^2)^2} = 0. \tag{11.1163}$$

△

Exemple 11.422.
Prenons le cas du champ de force de gravitation :

$$F(x,y,z) = \frac{1}{(x^2+y^2+z^2)^{3/2}}\begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{11.1164}$$

Nous pouvons rapidement remarquer que $\nabla \cdot F = 0$. Est-ce que cela peut se comprendre sur le dessin de la figure 11.16 ?

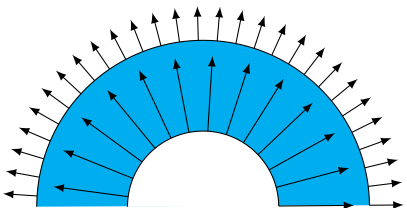


FIGURE 11.16 – Le champ de vecteur de la gravité. Nous avons tracé, sur les deux cercles la même densité de vecteurs, c'est-à-dire le même nombre de vecteurs par unité de surface.

Essayons de voir combien de fluide entre dans la zone bleue et combien en sort. D'abord, il est certain que les vecteurs qui sortent sont plus courts que ceux qui rentrent, ce qui voudrait dire qu'il y a plus de fluide qui rentre. Mais on voit également que le *nombre* de vecteurs qui sortent est plus grand parce que la seconde sphère est plus grande et qu'il y a un vecteur en chaque point de la sphère.

Intuitivement nous pouvons dire que la quantité qui rentre dans la sphère de rayon r_1 donnée par la taille des vecteurs entrants multiplié par la surface de la sphère, c'est-à-dire

$$4\pi r_1^2 \|F(x,y,z)\|, \tag{11.1165}$$

mais $\|F(x,y,z)\| = \frac{1}{r_1^2}$, donc la quantité de fluide entrant est 4π. La quantité de fluide sortant sera la même.

Cela explique deux choses

(1) Pourquoi les forces de gravitation et électromagnétiques sont en $1/r^2$; c'est parce que nous vivons dans un monde avec trois dimensions d'espace. En étudiant très précisément le champ de gravitation, certains physiciens espèrent trouver des déviations expérimentales par rapport à la règle du $1/r^2$; cela *pourrait* être un signe que l'espace contient des dimensions supplémentaires.

(2) Pourquoi il y a un 4π comme coefficient dans beaucoup d'équations en électromagnétisme ; en particulier dans certaines anciennes unités de flux.

△

Remarque 11.423.
Nous allons voir plus loin comment s'assurer que l'équation (11.1165) représente bien la « quantité de fluide » qui rentre dans la zone délimitée

11.46 Quelques formules de Leibnitz

La divergence étant une combinaison de dérivées, il n'est pas tellement étonnant que la divergence de produits donne lieux à des formules en deux termes. Si f est une fonction et si F et G sont des champs de vecteurs, nous avons la proposition suivante.

Proposition 11.424.
Si F et G sont des champs de vecteurs dont toutes les dérivées partielles existent, alors

(1) $\nabla \cdot (fF) f\nabla \cdot F + F \cdot \nabla f$

(2) $\nabla \cdot (F \times G) = G \cdot \nabla \times F - F \cdot \nabla \times G$

(3) $\nabla \times (fF) = f\nabla \times F + \nabla f \times F$.

Chapitre 12

Analyse sur des groupes

12.1 Action de groupe et connexité

Sources : [260] et wikipédia.

Théorème 12.1.
Soit G un groupe topologique localement compact et dénombrable à l'infini[1] *agissant continument et transitivement sur un espace topologique localement compact E. Alors l'application*

$$\begin{aligned} \varphi\colon G/G_x &\to E \\ [g] &\mapsto g \cdot x \end{aligned} \tag{12.1}$$

est un homéomorphisme.

Lemme 12.2.
Si G et H sont des groupes topologiques tels que G/H et H sont connexes[2]*, alors G est connexe.*

Démonstration. Soit $f\colon G \to \{0,1\}$ une fonction continue. Considérons l'application

$$\begin{aligned} \tilde{f}\colon G/H &\to \{0,1\} \\ [g] &\mapsto f(g). \end{aligned} \tag{12.2}$$

D'abord nous montrons qu'elle est bien définie. En effet si $h \in H$ nous aurions $\tilde{f}([gh]) = f(gh)$, mais étant donné que H est connexe, l'ensemble gH est également connexe ; la fonction continue f est donc constante sur gH. Nous avons donc $f(gh) = f(g)$.

Étant donné que G/H est également connexe, la fonction $\tilde{f}$ doit être constante. Si g_1 et g_2 sont deux éléments du groupe, nous avons $f(g_1) = \tilde{f}([g_1]) = \tilde{f}([g_2]) = f(g_2)$. Nous en déduisons que f est constante et que G est connexe. □

Théorème 12.3.
Pour tout $n \geqslant 2$, le groupe $\mathrm{SO}(n)$ est connexe, le groupe $\mathrm{O}(n)$ a deux composantes connexes.

Démonstration. La seconde assertion découle de la première parce que les matrices de déterminant 1 et celles de déterminant -1 ne peuvent pas être reliées par un chemin continu tandis que l'application

$$M \mapsto \begin{pmatrix} -1 & & \\ & 1 & \\ & & 1 \end{pmatrix} M \tag{12.3}$$

est un homéomorphisme entre les matrices de déterminant 1 et celles de déterminants -1. Montrons donc que $G = \mathrm{SO}(n)$ est connexe par arcs pour $n \geqslant 2$ en procédant par récurrence sur la dimension.

1. Cela signifie qu'il est une réunion dénombrable de compacts
2. Définition 7.42.

Nous acceptons le résultat pour $G = \mathrm{SO}(2)$. Notons que nous en avons besoin pour prouver que la sphère S^{n-1} est connexe.

Le groupe $\mathrm{SO}(n)$ agit, par définition, de façon transitive sur la sphère S^{n-1}. Soit $a \in S^{n-1}$, nous avons

$$G \cdot a = S^{n-1} \tag{12.4a}$$
$$G_a \simeq \mathrm{SO}(n-1) \tag{12.4b}$$

où G_a est le fixateur de a dans G. Pour montrer le second point, nous considérons $\{e_i\}$, la base canonique de $\mathbb{R}^n$ et $M \in G$ telle que $Ma = e_1$. Le fixateur de e_1 est évidemment isomorphe à $\mathrm{SO}(n-1)$ parce qu'il est constitué des matrices de la forme

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & a_{11} & \dots & a_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n-1,1} & \dots & a_{n-1,n-1} \end{pmatrix} \tag{12.5}$$

où $(a_{ij}) \in \mathrm{SO}(n-1)$. L'application

$$\begin{aligned} \alpha\colon G_{e_1} &\to G_a \\ A &\mapsto M^{-1}AM \end{aligned} \tag{12.6}$$

est un isomorphisme entre G_a et $\mathrm{SO}(n-1)$. Le théorème 12.1 nous montre alors que, en tant qu'espaces topologiques,

$$G/G_a = S^{n-1}. \tag{12.7}$$

L'hypothèse de récurrence montre que $G_a = \mathrm{SO}(n-1)$ est connexe tandis que nous savons que S^{n-1} est connexe. Le lemme 12.2 conclut que $G = \mathrm{SO}(n)$ est connexe. □

Lemme 12.4.
Une bijection continue entre un espace compact et un espace séparé est un homéomorphisme.

Proposition 12.5.
Les groupes $\mathrm{U}(n)$ *et* $\mathrm{SU}(n)$ *sont connexes.*

Démonstration. Soit $G(n)$ le groupe $\mathrm{SU}(n)$ ou $\mathrm{U}(n)$. Ce groupe opère transitivement sur la sphère complexe

$$S_{\mathbb{C}}^{n-1} = \{z \in \mathbb{C}^n \text{ tel que } \langle z, z\rangle = \sum_k |z_k|^2 = 1\}. \tag{12.8}$$

Cet ensemble est le même que S^{2n-1} parce que $|z_k| = x_k^2 + y_k^2$. Nous avons une bijection continue entre S^{n-1} et $S_{\mathbb{C}}^{n-1}$ et donc un homéomorphisme (lemme 12.4). Soit $a \in S_{\mathbb{C}}^{n-1}$, nous avons

$$G \cdot a = S_{\mathbb{C}}^{n-1} \tag{12.9a}$$
$$G_a \simeq G(n-1). \tag{12.9b}$$

La seconde ligne est un isomorphisme de groupe et un homéomorphisme. Il est donné de la façon suivante. D'abord le fixateur de e_1 dans $G(n)$ est donné par les matrices de la forme

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & a_{11} & \dots & a_{1,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n-1,1} & \dots & a_{n-1,n-1} \end{pmatrix} \tag{12.10}$$

où $(a_{ij}) \in G(n-1)$. Par ailleurs si M est une matrice de $G(n)$ telle que $Ma = e_1$, nous avons l'homéomorphisme

$$\begin{aligned} \alpha\colon G_{e_1} &\to G_a \\ A &\mapsto M^{-1}AM. \end{aligned} \tag{12.11}$$

Encore une fois, cela est un homéomorphisme par le lemme 12.4. Par composition nous avons $G_a \simeq G(n-1)$ et un homéomorphisme

$$G(n)/G_a = S_{\mathbb{C}}^{n-1}. \tag{12.12}$$

Le groupe G_a et l'ensemble $S_{\mathbb{C}}^{n-1}$ étant connexes, le groupe $G(n)$ est connexe par le lemme 12.2. □

Lemme 12.6 ([202]).
Si G est un sous-groupe connexe de $\mathrm{GL}(n,\mathbb{C})$ alors son groupe dérivé[3] *l'est également.*

Démonstration. Soit S_m l'ensemble des produits de m commutateurs de G :

$$S_m = \{g_1, \dots, g_m \text{ où les } g_i \text{ sont des commutateurs}\}. \tag{12.13}$$

La partie S_m est l'image de G par l'application continue

$$\begin{aligned} \underbrace{G \times \dots \times G}_{2m \text{ facteurs}} &\to G \\ (g_1, h_1, g_2, h_2, \dots, g_m, h_m) &\mapsto [g_1, h_1] \dots [g_m, h_m] \end{aligned} \tag{12.14}$$

En tant qu'image d'un connexe par une application continue, S_m est connexe par la proposition 7.90. Vu que les S_m ont l'identité en commun, le groupe dérivé

$$D(G) = \bigcup_{m=1}^{\infty} S_m \tag{12.15}$$

est également connexe. □

12.2 Espaces de matrices

L'ensemble des matrices est un espace vectoriel. Nous identifions $\mathbb{M}(n,\mathbb{R})$ avec $\mathbb{R}^{n^2}$; plus précisément, nous identifions une matrice

$$A = (a_{i,j})_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant n} \tag{12.16}$$

avec le vecteur $x = (x_1, x_2, \dots, x_{n^2}) \in \mathbb{R}^{n^2}$, où $a_{i,j} = x_{(n-1)i+j}$.

12.2.1 Dilatations et transvections

Soit un corps commutatif $\mathbb{K}$ et $n \geqslant 2$.

Théorème-définition 12.7 ([202]).
Soit une application linéaire $u\colon E \to E$ dont les points fixes forment un hyperplan noté H d'équation $H = \ker(f)$ avec $f \in E^*$.

(1) Les affirmations suivantes sont équivalentes :

(1a) $\det(u) \neq 1$

(1b) L'application u est diagonalisable et a une valeur propre qui vaut $\det(u) \neq 1$.

(1c) $\mathrm{Image}(u - \mathrm{Id}) \nsubseteq H$.

(1d) Il existe une base de E dans laquelle la matrice de u est $\mathrm{diag}(1, \dots, 1, \lambda)$ *avec* $\lambda \neq 1$.

(2) Les affirmation suivantes sont équivalentes :

(2a) Il existe $a \in H$ tel que pour tout $x \in E$, $u(x) = x + f(x)a$.

3. Définition 2.24.

(2b) Dans une base adaptée, la matrice de u est donnée par

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix}. \tag{12.17}$$

(3) Les conditions (1a)-(1d) sont respectées si et seulement si les conditions (2a)-(2b) ne sont pas respectées (elles sont les négations l'une de l'autre.).

Un endomorphisme qui est soit l'identité soit respecte les conditions (1) est une ***dilatation****. Un endomorphisme qui est soit l'identité soit qui vérifie les conditions (2) est une* ***transvection*** *(dans les deux cas il faut que les points fixes forment un hyperplan).*

Notons que selon cette terminologie, l'application $x \mapsto \lambda x$ n'est pas une dilatation mais un produit de dilatations.

Démonstration. Nous allons prouver plein d'implications ...

(1a) implique (1b) Le théorème de la base incomplète (voir remarque 4.12) permet de considérer une base $\{e_1, \dots, e_n\}$ de E telle que $\{e_1, \dots, e_{n-1}\}$ soit une base de H. Dans cette base, la matrice de u est de la forme suivante (les cases non remplies sont nulles et les étoiles correspondent à des valeurs inconnues mais pas spécialement nulles) :

$$\begin{pmatrix} 1 & & & * \\ & \ddots & & \vdots \\ & & 1 & * \\ & & & \lambda \end{pmatrix} \tag{12.18}$$

Le fait que le déterminant de u ne soit pas 1 implique que $\lambda \neq 1$. Par conséquent le polynôme caractéristique

$$\chi_u(X) = (1 - X)^{n-1}(\lambda - X) \tag{12.19}$$

possède une racine $\lambda \neq 1$, et donc u possède un vecteur propre v pour cette valeur[4]. Le vecteur v est linéairement indépendant de $\{e_1, \dots, e_{n-1}\}$ (parce que vecteur propre de valeur propre différente). Par conséquent l'ensemble $\{e_1, \dots, e_{n-1}, v\}$ est une base par la proposition 4.16. Cela est une base de vecteurs propres et donc une base de diagonalisation[5].

(1b) implique (1c) Nous nommons maintenant $\{e_1, \dots, e_n\}$ la base de diagonalisation. Nous avons $u(e_n) = \lambda e_n$ avec $\det(u) = \lambda \neq 1$. Nous avons

$$(u - \mathrm{Id})(e_n) = (\lambda - 1)e_n \notin H, \tag{12.20}$$

ce qui prouver que l'image de e_n par $u - \mathrm{Id}$ n'est pas dans H.

(1c) implique (1d) Reprenons une base $\{e_1, \dots, , e_n\}$ donnant la matrice (12.18). Il existe $x \in E$ tel que $u(x) - x$ n'est pas dans H, c'est-à-dire tel que $u\big(u(x) - x\big) \neq u(x) - x$. Nous en déduisons que

$$u^2(x) - 2u(x) + x \neq 0 \tag{12.21}$$

ou encore que

$$(X - 1)^2(u)x \neq 0. \tag{12.22}$$

C'est-à-dire que $(X - 1)^2$ n'est pas un polynôme annulateur de u. Or ce serait le cas si $X - 1$ était le polynôme minimal (proposition 9.96). Le polynôme caractéristique étant $(X -$

4. Proposition 9.118.

5. Nous pourrions en dire à peine plus et prouver le point (1d), mais cela ne servirait à rien parce que nous voulons prouver les équivalences et qu'il faudra quand même prouver que (1c) implique (1d).

$1)^{n-1}(X-\lambda)$ (et étant annulateur[6]), le polynôme minimal est de la forme

$$\mu_u(X) = \begin{cases} (X-1)(X-\lambda) & \text{si } \lambda \neq 1 \\ X-1 & \text{si } \lambda = 1. \end{cases} \tag{12.23}$$

Dans notre cas nous venons de voir que ce n'est pas $X-1$ et donc c'est $(X-1)(X-\lambda)$ avec $\lambda \neq 1$.

Nous devons trouver une base de diagonalisation ... Supposons

$$u(e_n) = \sum_{k=1}^{n-1} a_k e_k + \lambda e_n, \tag{12.24}$$

dans lequel nous venons de prouver que $\lambda \neq 1$, et cherchons

$$e'_n = \sum_{j=1}^{n} p_j e_j \tag{12.25}$$

de telle sorte à avoir $u(e'_n) = \lambda e_n$. Nous avons

$$u(e'_n) = \sum_{j=1}^{n-1} p_j u(e_j) + p_n u(e_n) = \sum_{j=1}^{n-1} (p_j + p_n a_j) e_j + p_n \lambda e_n. \tag{12.26}$$

En égalisant à $\lambda \sum_{j=1}^{n} p_j e_j$, il vient

$$p_j + p_n a_j = \lambda p_j \tag{12.27}$$

pour tout $j = 1, \dots, n-1$ et la condition triviale $p_n \lambda = \lambda p_n$ pour $j = n$. Nous en déduisons que le choix

$$p_j = \frac{p_n a_j}{\lambda - 1} \tag{12.28}$$

fonctionne (parce que $\lambda \neq 1$ comme nous l'avons démontré plus haut). En bref, il suffit de poser

$$e'_n = \sum_{j=1}^{n-1} \frac{p_n a_j}{\lambda - 1} e_j + p_n e_n \tag{12.29}$$

avec p_n au choix pour avoir une base $\{e_1, \dots, e_{n-1}, e'_n\}$ de diagonalisation de u avec $\lambda \neq 1$ comme dernière valeur propre.

(1d) implique (1a) Évident ... encore qu'il faut invoquer l'invariance du déterminant par changement de base.

Nous avons terminé la première série d'équivalences. Nous continuons avec la seconde.

(2a) implique (2b) Nous prenons $e_{n-1} = a$ et nous complétons en une base de H. Pour e_n il suffit de prendre n'importe quel vecteur v tel que $f(v) \neq 0$ (qui existe parce que $f = 0$ est seulement un hyperplan), et de le normaliser.

Dans cette base, la matrice de u a la forme désirée parce que $u(e_n) = e_n + f(e_n)a = e_n + e_{n-1}$ du fait que $e_{n-1} = a$ et $f(e_n) = 1$.

(2b) implique (2a) Soit $\{e_1, \dots, e_n\}$ cette base. En prenant $a = e_{n-1}$ et en posant $x = \sum_k x_k e_k$ nous avons

$$u(x) = \sum_{k=1}^{n-1} x_k e_k + x_n(e_{n-1} + e_n) = x + x_n e_{n-1} = x_n a. \tag{12.30}$$

Mais vu que $f(x) = \sum_i f_i x_i$, et que $f(e_i) = 0$ pour tout $i = 1, \dots, n-1$ nous avons $f(x) = f_n x_n$. Il n'y a cependant pas de raisons d'avoir $f_n = 1$. Cependant en définissant

$$e'_i = \frac{1}{f_n} e_i \tag{12.31}$$

6. Théorème de Cayley-Hamilton 9.115.

nous avons bien $u(e'_n) = \frac{1}{f_n}(e_{n-1} + e_n) = e'_{n-1} + e'_n$. Donc dans cette base nous avons encore la matrice de u de la forme

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix}, \tag{12.32}$$

mais cette fois avec $f(e'_n) = 1$.

Nous avons terminé avec la seconde série d'équivalences. Il nous reste à prouver que la première est équivalente à la négation de la seconde.

non (1c) implique (2a) Considérons $x_0 \in E$ tel que $f(x_0) = 1$ et posons $a = u(x_0) - x_0 \in \text{Image}(u - \text{Id})$. Par la négation de (1c) nous avons $a \in H$. De plus $x_0 \notin H$ (sinon $f(x_0) = 0$) donc $u(x_0) \neq x_0$ et $a \neq 0$.

Nous montrons que ce choix de a fonctionne : $u(x) = x + f(x)a$ pour tout $x \in E$. Nous faisons cela séparément pour $x \in H$ et pour $x = x_0$.

Si $h \in H$ alors $u(h) = h$ et $f(h) = 0$ donc $h + f(h)a = h = u(h)$. Si $x = x_0$ alors $u(x_0) = a + x_0$ (cela est la définition de a) et$x_0 + f(x_0)a = x_0 + a$.

(2b) implique non (1a) Dans une base adaptée nous avons

$$\begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & 1 \\ & & & 1 \end{pmatrix}, \tag{12.33}$$

et donc $\det(u) = 1$, ce qui contredit (1a).

□

Remarque 12.8.
Nous notons E_{ij} la matrice qui possède uniquement 1 en position (i, j). C'est-à-dire que $\big(E_{ij}\big)_{kl} = \delta_{ik}\delta_{jl}$. Soit H l'hyperplan des points fixes de f. Dans une base contenant une base de H, la matrice d'une transvection a pour forme type :

$$T_{ij}(\lambda) = \mathbb{1} + \lambda E_{ij} \tag{12.34}$$

avec $i \neq j$ et $\lambda \in \mathbb{K}$, et une dilatation a pour forme type la matrice diagonale

$$D_i(\alpha) = \mathbb{1} + (\alpha - 1)E_{ii} \tag{12.35}$$

avec $\alpha \in \mathbb{K}^*$.

Bien entendu, en choisissant une base quelconque, les matrices des dilatations et des translations peuvent avoir des formes différentes.

Lemme 12.9.
Quelques manipulations de lignes et de colonnes pour les matrices.

(1) La multiplication à gauche par $T_{ij}(\lambda)$ revient à effectuer le remplacement de ligne

$$L_i \to L_i + \lambda L_j. \tag{12.36}$$

(2) La multiplication à droite par $T_{ij}(\lambda)$ revient à effectuer le remplacement de colonne

$$C_j \to C_j + \lambda C_i. \tag{12.37}$$

(3) La multiplication à gauche par $T_{ij}(1)T_{ji}(-1)T_{ij}(1)$ revient à la substitution de lignes

$$\begin{cases} L_i \to L_j & \text{(12.38a)} \\ L_j \to -L_i. & \text{(12.38b)} \end{cases}$$

Note qu'il n'est pas possible d'inverser deux lignes à l'aide de transvections sans changer un signe parce que les transvections sont de déterminant 1 alors que l'inversion de lignes change le signe du déterminant.

Démonstration. Point par point.

Pour (1) Nous devons prouver que

$$\big(T_{ij}(\lambda)A\big)_{kl} = \begin{cases} A_{kl} & \text{si } k \neq i \\ A_{il} + \lambda A_{jl} & \text{si } k = i. \end{cases} \tag{12.39}$$

Un peu de calcul matriciel avec utilisation modérée des indices donner :

$$\big(T_{ij}(\lambda)A\big)_{kl} = \sum_s \big(T_{ij}(\lambda)\big)_{ks} A_{sl} \tag{12.40a}$$

$$= \sum_s \delta_{ks} A_{sl} + \lambda \delta_{ik}\delta_{js} A_{sl} \tag{12.40b}$$

$$= A_{kl} + \lambda \delta_{ik} A_{jl}. \tag{12.40c}$$

Pour (2) C'est la même chose.

Pour (3) Si nous appliquons successivement ces trois matrices (de droite à gauche) nous effectuons les substitutions :

$$\begin{cases} L'_i = L_i + L_j \\ L'_j = L_j \end{cases} \quad \text{suivit de} \quad \begin{cases} L''_i = L'_i \\ L''_j = L'_j - L'_i \end{cases} \quad \text{et de} \begin{cases} L'''_i = L''_i + L''_j \\ L'''_j = L''_j. \end{cases} \tag{12.41}$$

En effectuant ces substitutions,

$$L'''_i = L''_i + L''_j = L'_i + (L'_j - L'_i) = L'_j = L_j \tag{12.42}$$

et

$$L'''_j = L''_j = L'_j - L'_i = L_j - (L_i + L_j) = -L_i, \tag{12.43}$$

ce qu'il fallait.

□

Proposition 12.10 ([261]).
Soient $n \geqslant 2$ et $\mathbb{K}$ un corps commutatif.

(1) Si $A \in \mathrm{GL}(n, \mathbb{K})$, il existe des transvections $U_1, \dots, U_r$, $V_1, \dots, V_s$ telles que

$$A = U_1 \dots U_r \, D_n\big(\det(A)\big) \, V_1 \dots V_s. \tag{12.44}$$

(2) L'ensemble des transvections engendre le groupe spécial linéaire $\mathrm{SL}(n, \mathbb{K})$.

(3) L'ensemble des transvections et des dilatations engendre le groupe linéaire $\mathrm{GL}(n, \mathbb{K})$.

Démonstration. Nous allons montrer que toutes les matrices de $\mathrm{SL}(n, \mathbb{K})$ peuvent être écrites comme produits de matrices de la forme (12.34). Cela montrera qu'étant donné un endomorphisme f et une base *pas spécialement liée à* f, il est possible décrire la matrice de f comme produit de transvections dont les hyperplans invariants sont « contenus » dans cette base. Cela suffit à prouver que les transvections engendrent $\mathrm{SL}(n, \mathbb{K})$ grâce au lemme 2.9.

Toutes les transvections ont un déterminant égal à 1. Donc le groupe engendré par les transvections est inclus dans $\mathrm{SL}(2, \mathbb{K})$. Soit $A \in \mathrm{GL}(n, \mathbb{K})$; nous allons utiliser le pivot de Gauss pour la diagonaliser. Étant donné que A est inversible, sa première colonne n'est pas nulle. Si $A_{i1} \neq 0$ alors une multiplication à gauche par $L_{1i}\big((A_{11} - 1)/A_{i1}\big)$ effectue la substitution

$$L_1 \to L_1 - \frac{A_{11} - 1}{A_{i1}} L_i \tag{12.45}$$

qui met un 1 en la position $(1,1)$. Notons que si la première colonne est de la forme

$$\begin{pmatrix} s \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{12.46}$$

avec $s \neq 0$ alors il faut plutôt faire les substitutions $L_2 \to L_2 + L_1$ et ensuite $L_1 \to L_1 - \frac{1}{s}L_2$ pour obtenir le même résultat. En effectuant le pivot avec A_{11}, une suite d'opérations sur les lignes et les colonnes donnent

$$M_1 \dots M_p A N_1 \dots N_q = \begin{pmatrix} 1 & 0 \\ 0 & A_1 \end{pmatrix} \tag{12.47}$$

où $A_1 \in \mathrm{GL}(n-1,\mathbb{K})$ et $\det(A_1) = \det(A)$. En continuant de la sorte nous arrivons sur une matrice diagonale [7]

$$M_1 \dots M_{p'} A N_1 \dots N_{q'} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \alpha \end{pmatrix} \tag{12.48}$$

avec $\alpha = \det(A)$. En d'autres termes nous avons prouvé qu'il existe des transvections $U_1, \dots, U_r$ et $V_1, \dots, V_s$ telles que

$$A = U_1 \dots U_r \, D_n\big(\det(A)\big) \, V_1 \dots V_s. \tag{12.49}$$

Cela prouve que les transvections et les translations engendrent $\mathrm{GL}(n,\mathbb{K})$. Si $A \in \mathrm{SL}(n,\mathbb{K})$ alors $D_n\big(\det(A)\big) = 1$ et l'équation (12.49) est un produit de transvections. □

Proposition 12.11.
Le groupe $\mathrm{GL}(n,\mathbb{R})$ *est engendré par les endomorphismes inversibles diagonalisables.*

Démonstration. Par la proposition 12.10, le groupe $\mathrm{GL}(n,\mathbb{R})$ est engendré par les dilatations et les transvections. Il suffit donc de montrer qu'à leur tour, ces deux types d'endomorphismes sont engendrés par les endomorphismes inversibles et diagonalisables.

Les dilatations sont diagonalisables et inversibles. C'est bon pour elles.

Soit une transvection u, et une base $\{e_i\}_{i=1,\dots,n}$ dans laquelle u est de la forme (12.17). Nous considérons l'endomorphisme $d\colon E \to E$ défini par $d(e_k) = ke_k$. Cet endomorphisme est diagonalisable parce que son polynôme minimal, $\mu_d = \prod_{k=1}^{n}(X-k)$, est scindé à racines simples (voir le théorème 9.128).

Nous avons évidemment $u = d^{-1} \circ (d \circ u)$ où d^{-1} est diagonalisable et inversible. Voyons que $d \circ u$ est également diagonalisable en montrant que μ_d est son polynôme minimal (qui est scindé à racines simples).

Il suffit de montrer que $\mu_d(d \circ u)(e_k) = 0$ pour tout k. Ainsi μ_d sera un polynôme annulateur de $d \circ u$ de degré n, et donc minimal.

Si $k \leqslant n-1$ Alors $u(e_k) = e_k$ et $(d \circ u - n)e_k = (k-n)e_k$. En tout :

$$\mu_d(d\circ u)(e_k) = (d\circ u - 1)(d\circ u - 2)\dots(d\circ u - n)e_k = (k-1)(k-2)\dots(k-n)e_k = 0 \tag{12.50}$$

parce que dans le produit des $k-i$, il y en a forcément un de nul.

Si $k = n$ Dans un premier temps,

$$(d \circ u - n)e_n = d(e_n + e_{n-1}) - ne_n = ne_n + (n-1)e_{n-1} - ne_n = (n-1)e_{n-1}. \tag{12.51}$$

7. Attention : les opérations sur les lignes et le colonnes ne sont pas des opérations de similitude. Il n'est pas question de prétendre ici que toutes les matrices de $\mathrm{GL}(n,\mathbb{K})$ sont diagonales, voir la définition 4.110.

Ensuite

$$\begin{aligned}
\big(d\circ u-(n-1)\big)e_{n-1} &= d(e_{n-1})-(n-1)e_{n-1} && (12.52a)\\
&= d(e_{n-1})-(n-1)e_{n-1} && (12.52b)\\
&= (n-1)e_{n-1}-(n-1)e_{n-1} && (12.52c)\\
&= 0 && (12.52d)
\end{aligned}$$

Le polynôme μ_d est donc un polynôme scindé à n racines simples annulateur de $d\circ u$, qui est alors diagonalisable et inversible (parce que u et d le sont).

Donc sous la forme $u=d^{-1}(du)$, la transvection u est écrite comme produit de diagonalisables inversibles. □

Proposition 12.12 ([91]).
Soient $n\geqslant 3$ et $\mathbb{K}$ un corps de caractéristique différente de 2. Alors

(1) le groupe dérivé de $D(\mathrm{GL}(n,\mathbb{K}))$ est $\mathrm{SL}(n,\mathbb{K})$;

(2) le groupe dérivé de $\mathrm{SL}(n,\mathbb{K})$ est $\mathrm{SL}(n,\mathbb{K})$.

La preuve utilise le fait que les transvections engendrent $\mathrm{SL}(n,\mathbb{K})$ et que les transvections avec les dilatations engendrent $\mathrm{GL}(n,\mathbb{K})$. Voir la proposition 12.10.

12.2.2 Connexité de certains groupes

Lemme 12.13.
Le groupe $\mathrm{O}(n,\mathbb{R})$ n'est pas connexe.

Démonstration. La non connexité par arcs est facile parce que les éléments de déterminant 1 ne peuvent pas être reliés aux éléments de déterminant -1 par un chemin continu restant dans $\mathrm{O}(n)$ à cause du théorème des valeurs intermédiaires 11.50.

En ce qui concerne la connexité, il faut en dire un peu plus.

Les éléments de $\mathrm{O}(n,\mathbb{R})$ ont des déterminants égaux à 1 ou à -1. Ces deux parties sont des ouverts (pour la topologie induite de $\mathbb{M}(n,\mathbb{R})$). En effet soit $A\in\mathrm{SO}(n,\mathbb{R})$ (la partie contenant les déterminants 1 ; ce que l'on va dire tient pour l'autre partie). Alors, vu que le déterminant est une fonction continue sur $\mathbb{M}(n,\mathbb{R})$ il existe un voisinage $\mathcal{O}$ de A dans $\mathbb{M}(n,\mathbb{R})$ dans lequel le déterminant reste entre $\frac{1}{2}$ et $\frac{3}{2}$ (c'est la définition de la continuité avec $\epsilon=1/2$). L'ensemble $\mathcal{O}\cap\mathrm{O}(n,\mathbb{R})$ est par définition un ouvert de $\mathrm{O}(n,\mathbb{R})$ et ne contient que des éléments de déterminant 1.

La partie $\mathrm{O}(n,\mathbb{R})$ de $\mathbb{M}(n,\mathbb{R})$ est donc non-connexe selon la définition 7.42. □

Lemme 12.14.
Les groupes $\mathrm{U}(n)$ et $\mathrm{SU}(n)$ sont connexes par arcs.

Démonstration. Soient A, une matrice unitaire et Q une matrice unitaire qui diagonalise A. Étant donné que les valeurs propres arrivent par paires complexes conjuguées,

$$QAQ^{-1}=\begin{pmatrix} e^{i\theta_1} & & & & \\ & e^{-i\theta_1} & & & \\ & & \ddots & & \\ & & & e^{i\theta_r} & \\ & & & & e^{-i\theta_r}\end{pmatrix}. \tag{12.53}$$

Le chemin $U(t)$ obtenu en remplaçant θ_i par $t\theta_i$ avec $t\in[0,1]$ joint QAQ^{-1} à l'identité. Par conséquent $Q^{-1}U(t)Q$ joint A à l'unité. □

Théorème 12.15.
Les matrices normales[8] *forment un espace connexe par arc.*

8. Définition 9.146.

Démonstration. Soient A une matrice normale et U une matrice unitaire qui diagonalise A. Nous considérons $U(t)$, un chemin qui joint $\mathbb{1}$ à U dans $\mathrm{U}(n)$. Pour chaque t, la matrice

$$A(t) = U(t)^{-1}AU(t) \tag{12.54}$$

est normale. Nous avons donc trouvé un chemin dans les matrices normales qui joint A à une matrice diagonale. Il est à présent facile de la joindre à l'identité.

Toutes les matrices normales étant connexes à l'identité, l'ensemble des matrices normales est connexe. □

Proposition 12.16.
Le groupe $\mathrm{SL}(n,\mathbb{K})$ *est connexe par arcs.*

Démonstration. Soit $A \in \mathrm{SL}(n,\mathbb{K})$; par la proposition 12.10(2) nous pouvons écrire

$$A = \prod_{c\in X} T_c(\lambda_c) \tag{12.55}$$

où X est une partie de l'ensemble des couples (i,j) dans $\{1,\dots,n\}$. En posant

$$\begin{aligned}\varphi\colon [0,1] &\to \mathrm{SL}(n,\mathbb{K})\\ t &\mapsto \prod_{c\in X} T_c(t\lambda_c)\end{aligned} \tag{12.56}$$

nous avons une application continue de A vers $\mathbb{1}$, dont pour tout t la matrice $\varphi(t)$ est inversible de déterminant1.

Donc tous les éléments de $\mathrm{SL}(n,\mathbb{K})$ peuvent être reliés à $\mathbb{1}$. Donc $\mathrm{SL}(n,\mathbb{K})$ est connexe par arcs. □

Proposition 12.17 ([262])**.**
Le groupe $\mathrm{GL}(n,\mathbb{C})$ *est connexe par arcs.*

Démonstration. Soient $A \in \mathrm{GL}(n,\mathbb{C})$ et sa décomposition (12.44). Comme fait précédemment, chacune des transvections peut être reliée à $\mathbb{1}$ par un chemin continu dans $\mathrm{SL}(n,\mathbb{C})$. En ce qui concerne le facteur de translation, nous ne pouvons pas simplement prendre le chemin donné par $t \mapsto D_n\big(t\det(A)\big)$ parce que le résultat n'est pas inversible en $t=0$.

Vu que C^* il existe une application continue $\alpha\colon [0,1] \to \mathbb{C}^*$ telle que $\alpha(0) = \det(A) \in \mathbb{C}^*$ et $\alpha(1) = 1$. Il suffit alors de prendre $D_n\big(\alpha(t)\big)$ et nous avons un chemin continu de A vers $\mathbb{1}$ restant dans $\mathrm{GL}(n,\mathbb{C})$. □

Proposition 12.18.
Le groupe $\mathrm{GL}(n,\mathbb{R})$ *a exactement deux composantes connexes par arcs.*

Démonstration. Nous notons $\mathrm{GL}^+(n,\mathbb{R})$ et $\mathrm{GL}^-(n,\mathbb{R})$ les parties de $\mathrm{GL}(n,\mathbb{R})$ formées des applications de déterminant ± 1 respectivement. Vu le théorème des valeurs intermédiaires (théorème 11.50), il n'existe pas d'applications continues dans $\mathrm{GL}(n,\mathbb{R})$ reliant $\mathrm{GL}^+(n,\mathbb{R})$ à $\mathrm{GL}^-(n,\mathbb{R})$ tout en restant dans les applications de déterminant non nul [9].

Montrons que $\mathrm{GL}^\pm(n,\mathbb{R})$ sont connexes par arcs. Si $A \in \mathrm{GL}^+(n,\mathbb{R})$ alors grâce à la décomposition (12.44), il existe un chemin continu de A vers $D_n\big(\det(A)\big)$. Vu que $\mathbb{R}^\pm$ sont connexes par arc, il est possible de relier $D_n\big(\det(A)\big)$ à $D_n(\pm 1)$ par un chemin continu. □

9. Si $\varphi\colon [0,1] \to \mathrm{GL}(n,\mathbb{R})$ est le chemin, la fonction à mettre dans le théorème des valeurs intermédiaires est la fonction $f\colon [0,1] \to \mathbb{R}$ $t \mapsto \det\big(\varphi(t)\big)$.

12.2.3 Densité

Proposition 12.19.
Les matrices diagonalisables sont denses dans $\mathbb{M}(n,\mathbb{C})$.

Démonstration. D'après le lemme de Schur 9.137, une matrice de $\mathbb{M}(n,\mathbb{C})$ est de la forme

$$A = Q\begin{pmatrix}\lambda_1 & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_n\end{pmatrix}Q^{-1}. \tag{12.57}$$

Les valeurs propres sont sur la diagonale. La matrice est diagonalisable si les éléments de la diagonales sont tous différents. Il suffit maintenant de considérer n suites $(\epsilon_k^{(r)})_{k\in\mathbb{N}}$ convergentes vers zéro telles que pour chaque k les nombres $\lambda_r+\epsilon_k^{(r)}$ soient tous différents. La suite de matrices

$$A_k = Q\begin{pmatrix}\lambda_1+\epsilon_k^{(1)} & * & * \\ 0 & \ddots & * \\ 0 & 0 & \lambda_n+\epsilon_k^{(n)}\end{pmatrix}Q^{-1}. \tag{12.58}$$

est alors diagonalisable pour tout k et nous avons $\lim_{k\to\infty}A_k = A$. □

Proposition 12.20.
Les matrices inversibles sont denses dans l'ensemble des matrices. C'est-à-dire que $\mathrm{GL}(n,\mathbb{R})$ *est dense dans* $\mathbb{M}(n,\mathbb{R})$.

Démonstration. Soit $A\in\mathbb{M}(n,\mathbb{R})$; le lemme de Schur réel 9.149 nous permet d'écrire

$$A = Q\begin{pmatrix}\lambda_1 & & & & \\ & \ddots & & & \\ & & \lambda_r & & \\ & & & \begin{pmatrix}a & b\\ c & d\end{pmatrix} & \\ & & & & \ddots\end{pmatrix}Q^{-1} \tag{12.59}$$

avec Q orthogonale.

Pour définir A_k nous remplaçons λ_i par $\lambda_i+\epsilon^{(i)_k}$ de façon à avoir $\epsilon_k^{(i)}\to 0$ et $\lambda_i+\epsilon_k^{(i)}\neq 0$. En ce qui concerne les blocs, ceux dont le déterminant est non nul, nous n'y touchons pas, et ceux dont le déterminant est nul, nous remplaçons a par $a+\epsilon_k$.

Avec cela, QA_kA^{-1} est une suite dans $\mathrm{GL}(n,\mathbb{R})$ qui converge vers A. □

Proposition 12.21.
Si $A\in\mathbb{M}(n,\mathbb{C})$ *alors*

$$e^{\mathrm{Tr}(A)} = \det(e^A). \tag{12.60}$$

Démonstration. Ici, e^A est l'exponentielle soit d'endomorphisme soit de matrice définie par la proposition 10.154.

Le résultat est un simple calcul pour les matrices diagonalisable. Si A n'est pas diagonalisable, nous considérons une suite de matrices diagonalisables A_k dont la limite est A (proposition 12.19). La suite

$$a_k = e^{\mathrm{Tr}(A_k)} \tag{12.61}$$

converge vers $e^{\mathrm{Tr}(A)}$ tandis que la suite

$$b_k = \det(e^{A_k}) \tag{12.62}$$

converge vers $\det(e^A)$. Mais nous avons $a_k=b_k$ pour tout k ; les limites sont donc égales. □

Corolaire 12.22.
Si $A \in \mathbb{M}(n, \mathbb{C})$ alors

$$\frac{d}{dt}\Big[\det(e^{tX})\Big]_{t=0} = \mathrm{Tr}(X). \tag{12.63}$$

Démonstration. Nous écrivons la proposition 12.21 pour tX au lieu de X ; pour chaque t nous avons

$$\det\left(e^{tA}\right) = e^{\mathrm{Tr}(tA)} = e^{t\,\mathrm{Tr}(A)}. \tag{12.64}$$

La dérivation par rapport à t en $t = 0$ donne le résultat attendu. □

Théorème 12.23 (Cayley-Hamilton[263, 264])**.**
Tout endomorphisme d'un espace vectoriel de dimension finie sur un corps commutatif quelconque annule son propre polynôme caractéristique

Une autre démonstration est donnée en le théorème 9.115.

Démonstration. La preuve est divisée en plusieurs étapes.

Endomorphisme diagonalisable Soit u un endomorphisme sur un espace vectoriel V de dimension n sur un corps $\mathbb{K}$ et χ_u sont polynôme caractéristique. Nous savons que si λ est une valeur propre de u alors $\chi_u(\lambda) = 0$ le théorème 9.111(2). En combinant avec le lemme 9.90, si x est vecteur propre pour la valeur propre λ de u nous avons

$$\chi_u(u)x = \chi_u(\lambda)x = 0. \tag{12.65}$$

Donc tant que u possède une base de vecteurs propres nous avons $\chi_u(u) = 0$.

Le cas complexe Nous nous restreignons à présent (et provisoirement) au cas $\mathbb{K} = \mathbb{C}$, ce qui nous donne $u \in \mathbb{M}(n, \mathbb{C})$. Les matrices diagonalisables sont denses dans $\mathbb{M}(n, \mathbb{C})$ par la proposition 12.19. Si $A \in \mathbb{M}(n, \mathbb{C})$ nous considérons une suite de matrices diagonalisables $A_k \xrightarrow{\mathbb{M}(n,\mathbb{C})} A$. Pour chaque k nous avons par le point précédent

$$\chi_{u_k}(u_k) = 0. \tag{12.66}$$

Chacune des composantes de $\chi_{u_k}(u_k)$ est un polynôme en les composantes de u_k, ce qui légitime le passage à la limite :

$$\chi_u(u) = 0. \tag{12.67}$$

Le théorème est établi pour toutes les matrices de $\mathbb{M}(n, \mathbb{C})$ et donc aussi pour tous les sous-corps de $\mathbb{C}$ comme $\mathbb{R}$ ou $\mathbb{Z}$.

La cas général Par définition, $\chi_u(X) = \det(u - X\mathbb{1})$; les coefficients de X sont des polynômes à coefficients entiers en les composantes de u. En substituant u à X nous obtenons une matrice dont chacune des entrées est un polynôme à coefficients entiers en les coefficients de u. Pour chaque i et j entre 1 et n il existe donc un polynôme $P_{ij} \in \mathbb{Z}(X_1, \dots, X_{n^2})$ tel que

$$\chi_u(u)_{ij} = P(u_{11}, \dots, u_{nn}). \tag{12.68}$$

Ces polynômes ne dépendent pas de u ni du corps sur lequel on travaille. Notre but est maintenant de prouver que $P_{ij} = 0$.

Étant donné que le cas complexe (et a fortiori entier) est déjà prouvé nous savons que pour tout $u \in \mathbb{M}(n, \mathbb{Z})$ nous avons $P(u_{11}, \dots, u_{nn}) = 0$. La proposition 6.170 nous donne effectivement $P = 0$, en conséquence de quoi l'endomorphisme $\chi_u(u)$ est nul.

□

Exemple 12.24.
Pour montrer que chaque composante $\chi_u(u)$ est bien un polynôme à coefficients entiers en les coefficients de u, voyons l'exemple 2×2 : $u = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. D'abord

$$\chi_u(X) = \det \begin{pmatrix} a - X & b \\ c & d - X \end{pmatrix} = X^2 - (a + d)X + ad - cb. \tag{12.69}$$

Le coefficient de X^2 est 1, celui de X est $-a - d$ et le terme indépendant est $ad - cb$; tout trois sont des polynômes à coefficients entiers en a, b, c, d. Après substitution de X par u,

$$\chi_u(u)_{ij} = (u^2)_{ij} - (a + d)u_{ij} + ad - cb. \tag{12.70}$$

Cela est bien un polynôme à coefficients entiers en les entrées de la matrice u. △

12.2.4 Racine carrée d'une matrice hermitienne positive

Proposition-définition 12.25.
Si $A \in \mathbb{M}(n, \mathbb{C})$ est une matrice hermitienne[10] *positive, alors il existe une unique matrice hermitienne positive R telle que $A = R^2$. De plus R est un polynôme (de $\mathbb{R}[X]$) en A.*

La matrice R ainsi définie est la ***racine carrée de*** *de A, et est notée $\sqrt{A}$.*

Démonstration. **Existence** Étant donné que A est hermitienne, elle est diagonalisable par une matrice unitaire (proposition 9.147), et ses valeurs propres sont réelles et positives (parce que A est positive). Soit donc P une matrice unitaire telle que

$$P^*AP = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix} \tag{12.71}$$

avec $\alpha_i > 0$. Si on pose

$$R = P \begin{pmatrix} \sqrt{\alpha_1} & & \\ & \ddots & \\ & & \sqrt{\alpha_n} \end{pmatrix} P^*, \tag{12.72}$$

alors $R^2 = A$ parce que $P^*P = \mathbb{1}$.

Hermitienne positive La matrice R est hermitienne parce que, avec un peu de notation raccourcie, $R = P^*\sqrt{\alpha}P$ et $R^* = P^*\sqrt{\alpha}P$. D'autre part, elle est positive parce que ses valeurs propres sont les $\sqrt{\alpha_i}$ qui sont positives.

Polynôme Nous montrons maintenant que la matrice R est un polynôme en A. Pour cela nous considérons un polynôme Q tel que $A(\alpha_i) = \sqrt{\alpha_i}$ pour tout i. Soit $\{e_i\}$ une base de diagonalisation de A : $Ae_i = \alpha_i e_i$. Alors c'est encore une base de diagonalisation de $Q(A)$. En effet si $Q = \sum_k a_k X^k$, alors

$$Q(A)e_i = (\sum_k a_k A^k)e_i = (\sum_k a_k \alpha_i^k)e_i = Q(\alpha_i)e_i = \sqrt{\alpha_i}e_i. \tag{12.73}$$

Les valeurs propres de $Q(A)$ sont donc $\sqrt{\alpha_i}$. Nous savons maintenant que $Q(A)$ a la même base de diagonalisation de A (et donc la même matrice unitaire P qui diagonalise), c'est-à-dire que

$$Q(A) = P^* \begin{pmatrix} \sqrt{\alpha_1} & & \\ & \ddots & \\ & & \sqrt{\alpha_n} \end{pmatrix} = R. \tag{12.74}$$

Donc oui, R est un polynôme en A.

Notons que ce Q n'est pas du tout unique ; il existe une infinité de polynômes envoyant n nombres donnés sur n nombres donnés.

10. Définition 9.35.

Unicité Soit S une matrice hermitienne positive telle que $R^2 = S^2 = A$. D'abord S commute avec A parce que

$$SA = S^3 = S^2S = AS. \tag{12.75}$$

Donc S commute aussi avec $Q(A) = R$. Étant donné que S et R commutent et sont diagonalisables, ils sont simultanément diagonalisables par le corolaire 9.129. Soient $D_R = PRP^*$ et $D_S = PSP^*$ les formes diagonales de R et S dans une base de simultanée diagonalisation. Les carrés des valeurs propres de R et S étant identiques (ce sont les valeurs propres de A) et les valeurs propres de R et S étant positives, nous déduisons que $D_R = D_S$ et donc que $R = P^*D_RP = P^*D_SP = S$.

□

Une des applications usuelles de cette proposition est la décomposition polaire.

12.2.5 Racine carrée d'une matrice symétrique positive

Lemme 12.26 ([265]).
Le groupe orthogonal $\mathrm{O}(n, \mathbb{R})$ *est compact.*

Démonstration. Nous avons $O(n) = f^{-1}\big(\{\mathbb{1}_n\}\big)$ où f est l'application continue $A \mapsto A^tA$. En tant qu'image inverse d'un fermé par une application continue, le groupe $O(n)$ est fermé.

De plus il est borné parce que tous les coefficients d'une matrice orthogonale sont $\leqslant 1$, donc $\|A\|_\infty$ pour tout $A \in O(n)$. □

Proposition 12.27.
Une matrice symétrique semi (ou pas) définie positive admet une unique racine carrée symétrique. Le spectre de la racine carrée est la racine carrée du spectre de la matrice de départ.

Démonstration. Ceci est une phrase pour que les titres se mettent bien.

Existence Soit T une matrice symétrique et Q une matrice orthogonale qui diagonalise[11] T : $QTQ^{-1} = D$ avec $D = \mathrm{diag}(\lambda_i)$ et $\lambda_i \geqslant 0$. En posant $R = Q^{-1}\sqrt{D}Q$, il est vite vérifié que $R^2 = T$ et que R est symétrique. En ce qui concerne le spectre, R a pour valeurs propres les $\sqrt{\lambda_i}$.

Unicité Soit R une matrice symétrique de T : $R^2 = T$. Du coup R et T commutent : $RT = R^3 = TR$. Par conséquent les espaces propres de T sont stables sous R. Soit E_λ l'un d'eux de dimension d, et T_F, R_F les restrictions de T et R à E_λ. L'application T_F est une homothétie et $R_F^2 = T_F = \lambda\mathbb{1}$. Mais R_F est encore une matrice symétrique définie positive, donc nous pouvons considérer une base $\{e_1, \dots, e_d\}$ de E_λ qui diagonalise R_F avec les valeurs propres μ_i ; nous avons donc en même temps

$$R_f^2(e_i) = \mu_i^2 e_i \tag{12.76a}$$
$$T_F(e_i) = \lambda e_i, \tag{12.76b}$$

de telle sorte que $\mu_i^2 = \lambda$. Mais les valeurs propres de R_F sont positives, sont $\mu_i = \sqrt{\lambda}$ pour tout i. En conclusion R_F est univoquement déterminé par la donnée de T. Vu que cela est valable pour tous les espaces propres de T et que ces espaces propres engendrent tout E, l'opérateur R est déterminé de façon univoque par T.

□

Notons que nous n'avons démontré l'unicité qu'au sein des matrices symétriques.

11. Théorème 9.150.

12.2.6 Décomposition polaires : cas réel

Nous rappelons que $S^{++}(n,\mathbb{R})$ est l'ensemble des matrices symétriques strictement définies positives définies en 9.153.

Lemme 12.28.
La partie $S^{+}(n,\mathbb{R})$ est fermée dans $\mathbb{M}(n,\mathbb{R})$.

Démonstration. En effet si S_k est une suite de matrices symétriques convergeant dans $\mathbb{M}(n,\mathbb{R})$ vers la matrice A, les suites $(S_k)_{ij}$ et $(S_k)_{ji}$ des composantes ij et ji sont des suites égales, et donc leurs limites sont égales [12]. Donc la limite est symétrique.

En ce qui concerne le spectre, le théorème 9.150 nous permet de diagonaliser : $S_k = Q_k D_k Q_k^{-1}$ où les D_k sont des matrices diagonales remplies de nombres positifs ou nuls. Vu que $O(n)$ est compact [13], nous avons une sous-suite $Q_{\varphi(k)}$ convergente : $Q_{\varphi(k)} \to Q$. Pour chaque k, nous avons

$$S_{\varphi(k)} = Q_{\varphi(k)} D_{\varphi(k)} Q_{\varphi(k)}^{-1}, \tag{12.77}$$

dont la limite existe et vaut A. Vu que pour tout k, $D_{\varphi(k)} = Q_{\varphi(k)}^{-1} S_{\varphi(k)} Q_{\varphi(k)}$ et que le produit matriciel est continu, la suite $k \mapsto D_{\varphi(k)}$ est une suite convergente dans $\mathbb{M}(n,\mathbb{R})$. Nous notons D sa limite qui est encore une matrice diagonale contenant des nombres positifs ou nuls sur la diagonale.

$$A = \lim_{k\to\infty} S_{\varphi(k)} = QDQ^{-1}, \tag{12.78}$$

et donc le spectre de A est la limite de ceux des matrices $D_{\varphi(k)}$. Chacun étant positif, la limite est positive. Donc $A \in S^{+}(n,\mathbb{R})$. □

Lemme 12.29.
La fermeture de l'ensemble des matrices symétriques strictement définies positives est l'ensemble des matrices définies positives : $\overline{S^{++}(n,\mathbb{R})} = S^{+}(n,\mathbb{R})$.

Démonstration. Le lemme 12.28 nous a à peine dit que $S^{+}(n,\mathbb{R})$ était fermé. Nous devons prouver que pour tout élément de $S^{+}(n,\mathbb{R})$, il existe une suite (S_k) dans $S^{++}(n,\mathbb{R})$ convergeant vers S.

Si $S \in S^{+}(n,\mathbb{R})$ alors nous avons la diagonalisation

$$S = QDQ^{-1} = Q \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} Q^{-1} \tag{12.79}$$

où $\lambda_i \geqslant 0$ pour tout i. Nous définissons

$$D_k = \begin{pmatrix} \lambda_1 + \epsilon_k^{(1)} & & \\ & \ddots & \\ & & \lambda_n + \epsilon_k^{(n)} \end{pmatrix} \tag{12.80}$$

où ϵ_k^{i} est une suite convergent vers 0 telle que $\lambda_i + \epsilon_n^{(i)} > 0$ pour tout n. Typiquement si $\lambda_i > 0$ alors $\epsilon_k^{(i)} = 0$ et sinon $\epsilon_k^{(i)} = 1/k$.

Pour tout k nous avons $QD_kQ^{-1} \in S^{++}(n,\mathbb{R})$ et de plus $QD_kQ^{-1} \to QDQ = S$. □

Théorème 12.30 (Décomposition polaire de matrices symétriques définies positives[265, 266, 267])**.**
En ce qui concerne les matrices inversibles :

$$\begin{aligned} f\colon O(n,\mathbb{R}) \times S^{++}(n,\mathbb{R}) &\to \mathrm{GL}(n,\mathbb{R}) \\ (Q,S) &\mapsto SQ \end{aligned} \tag{12.81}$$

12. Ici nous utilisons le critère de convergence composante par composante et le fait que nous ne sommes pas trop inquiétés par la norme que nous choisissons parce que toutes les normes sont équivalentes par le théorème 7.224.
13. Lemme 12.26.

est un homéomorphisme[14].

En ce qui concerne les matrices en général :

$$\begin{aligned} g\colon O(n,\mathbb{R}) \times S^+(n,\mathbb{R}) &\to \mathbb{M}(n,\mathbb{R}) \\ (Q,S) &\mapsto SQ \end{aligned} \tag{12.82}$$

est une surjection mais pas une injection.

De plus les mêmes conclusions tiennent si nous regardons $(Q,S) \mapsto QS$ au lieu de SQ.

Démonstration. Nous commençons par prouver les résultats concernant les matrices inversibles.

Existence et unicité Si $M = SQ$, alors $MM^t = SQQ^tS^t = S^2$, donc S doit être une racine carrée symétrique de la matrice définie positive MM^t. La proposition 12.27 nous dit que ça existe et que c'est unique. Donc S est univoquement déterminé par M. Maintenant avoir $Q = MS^{-1}$ est obligatoire (unicité) et fonctionne :

$$Q^tQ = (S^{-1})^tM^tMS^{-1} = S^{-1}S^2S^{-1} = \mathbb{1}, \tag{12.83}$$

donc Q ainsi défini est orthogonale.

Notons que ceci ne fonctionne pas lorsque M n'est pas inversible parce qu'alors S n'est pas inversible.

Homéomorphisme Le fait que f soit continue n'est pas un problème : c'est un produit de matrices. Nous devons vérifier que f^{-1} est continue. Soit une suite convergente $M_k \to M$ dans $\mathrm{GL}(n,\mathbb{R})$. Si nous nommons (Q_k,S_k) la décomposition polaire de M_k et (Q,S) celle de M, nous devons prouver que $Q_k \to Q$ et $S_k \to S$. En effet dans ce cas nous aurions

$$\lim_{k\to\infty} f^{-1}(M_k) = \lim_{k\to\infty}(Q_k,S_k) = (Q,S) = f^{-1}(M). \tag{12.84}$$

Étant donné que $O(n)$ est compact (lemme 12.26), la suite (Q_k) admet une sous-suite convergente (Bolzano-Weierstrass, théorème 7.103) que nous nommons

$$Q_{\varphi(k)} \to F \in O(n). \tag{12.85}$$

Vu que la suite (M_k) converge, sa sous-suite converge vers la même limite : $M_{\varphi(k)} \to M$ et vu que pour tout k nous avons $S_k = M_kQ_k^{-1}$,

$$S_{\varphi(k)} \to G = MF^{-1}. \tag{12.86}$$

Vu que chacune des matrices $S_{\varphi(k)}$ est symétrique définie positive, la limite est symétrique et semi-définie positive[15]. Donc $G \in S^+(n,\mathbb{R}) \cap \mathrm{GL}(n,\mathbb{R})$ parce que de plus M et F étant inversibles, G est inversible. En ce qui concerne la sous-suite nous avons

$$M_{\varphi(k)} = S_{\varphi(k)}Q_{\varphi(k)} \to GF = M \tag{12.87}$$

où $F \in O(n)$ et $G \in S^+(n,\mathbb{R})$. Par unicité de la décomposition polaire de M (partie déjà démontrée), nous avons $G = S$ et $F = Q$.

Nous avons prouvé que toute sous-suite convergente de Q_k a Q pour limite. Donc la suite elle-même converge[16] vers Q. Donc $Q_k \to Q$. Du coup vu que $S_k = M_kQ_k^{-1}$ est un produit de suites convergentes, S_k converge également, vers S : $S_k \to S$.

Au final l'application f^{-1} est bien continue parce que les égalités (12.84) ont bien lieu.

14. Cela est en réalité en difféomorphisme, voir la remarque 12.31.
15. Lemme 12.29
16. Proposition 7.311, pas difficile.

Nous passons maintenant à la preuve dans le cas des matrices en général.

Soit $A \in \mathbb{M}(n,\mathbb{R})$; par densité (lemme 12.20), il existe une suite (A_k) dans $\mathrm{GL}(n,\mathbb{R})$ telle que $A_k \to A$. Pour chacun des k nous appliquons la décomposition polaire déjà prouvée : $A_k = Q_k S_k$. D'abord (Q_k) est une suite dans le compact [17] $\mathrm{O}(n,\mathbb{R})$ et accepte donc une sous-suite convergente. Quitte à redéfinir la suite de départ, nous supposons pour alléger les notations que $Q_k \to Q \in \mathrm{O}(n,\mathbb{R})$. Vu que Q_k est inversible,

$$S_k = Q_k^{-1} A_k \tag{12.88}$$

Le produit matriciel étant continu nous avons $S_k \to S$ dans $\mathbb{M}(n,\mathbb{R})$. Mais $S^+(n,\mathbb{R})$ étant fermé (lemme 12.28) nous avons aussi $S \in S^+(n,\mathbb{R})$. □

Remarque 12.31.
Pour démontrer que f est différentiable, nous devons utiliser le théorème d'inversion locale 16.48 ; cela est fait dans la proposition 16.58.

Corolaire 12.32.
Toute matrice peut être écrite sous la forme $Q_1 D Q_2$ où Q_1 et Q_2 sont orthogonales et D est diagonale.

Démonstration. Si $A \in \mathbb{M}(n,\mathbb{R})$ alors la décomposition polaire 12.30 nous donne $A = SQ$ où S est symétrique définie positive et Q est orthogonale. La matrice S peut ensuite être diagonalisée par le théorème 9.150 : $S = RDR^{-1}$ où D est diagonale et R est orthogonale. Avec ces deux décompositions en main, $A = SQ = RDR^{-1}Q$. La matrice $R^{-1}Q$ est orthogonale. □

12.2.7 Enveloppe convexe

Définition 12.33.
*Sur C est un ensemble convexe, un point $x \in C$ est un **point extrémal** si $C \backslash \{x\}$ est encore convexe.*

Théorème 12.34 ([54]).
Soit E un espace euclidien de dimension $n \geqslant 1$ et $\mathcal{L}(E)$ l'espace des opérateurs linéaires sur E sur lequel nous considérons la norme subordonnée [18] à celle sur E. L'ensemble des points extrémaux de la boule unité fermée de $\mathcal{L}(E)$ est le groupe orthogonal $O(n,\mathbb{R})$.

Démonstration. Nous notons $\mathcal{B}$ la boule unité fermée de $\mathcal{L}(E)$. Montrons pour commencer que les éléments de $O(n)$ sont extrémaux dans $\mathcal{B}$. D'abord si $A \in O(E)$ alors $\|A\| = 1$ parce que $\|Ax\| = \|x\|$. Supposons maintenant que A n'est pas extrémal, c'est-à-dire qu'il est le milieu d'un segment joignant deux points (distincts) de la boule unité de $\mathcal{L}(E)$. Soient donc $T, U \in \mathcal{B}$ tels que $A = \frac{1}{2}(T+U)$. Pour tout $x \in E$ tel que $\|x\| = 1$ nous avons

$$1 = \|x\| = \|Ax\| = \frac{1}{2}\|Tx + Ux\| \leqslant \frac{1}{2}\big(\|Tx\| + \|Ux\|\big) \leqslant \frac{1}{2}\big(\|T\| + |U|\big) \leqslant 1 \tag{12.89}$$

Toutes les inégalités sont en réalité des égalités. En particulier nous avons

$$\|Tx + Ux\| = \|Tx\| + \|Ux\|, \tag{12.90}$$

mais alors nous sommes dans un cas d'égalité dans l'inégalité de Cauchy-Schwarz (théorème 7.191) et donc il existe $\lambda \geqslant 0$ tel que $Tx = \lambda Ux$. Mais de plus les ~~inégalité~~ égalités (12.89) nous donnent

$$\frac{1}{2}\big(\|Tx\| + \|Ux\|\big) = 1 \tag{12.91}$$

alors que nous savons que $\|Tx\|, \|Ux\| \leqslant 1$, donc $\|Tx\| = \|Ux\| = 1$. La seule possibilité est d'avoir $\lambda = 1$ et donc que $U = T$ parce que nous avons $Tx = Ux$ pour tout x de norme 1. Au final A n'est pas le milieu d'un segment dans $\mathcal{B}$.

17. Lemme 12.26.
18. Définition 10.3.

Nous passons donc à l'inclusion inverse : nous prouvons que les points extrémaux de $\mathcal{B}$ sont dans $O(E)$. Pour cela nous prenons $U \in \mathcal{B}\backslash O(E)$ et nous allons montrer que U n'est pas un point extrémal : nous allons l'écrire comme milieu d'un segment dans $\mathcal{B}$.

Par la seconde partie du théorème de décomposition polaire 12.30, il existe $Q \in O(n,\mathbb{R})$ et $S \in S^+(n,\mathbb{R})$ tels que $U = QS$. Nous diagonalisons S à l'aide de la matrice orthogonale P :

$$S = PDP^{-1} \tag{12.92}$$

avec $D = \operatorname{diag}(\lambda_i)$. En termes de normes, nous avons

$$\|U\| = \|S\| = \|S\|. \tag{12.93}$$

En effet vu que Q est orthogonale, $\|Ux\| = \|QSx\| = \|Sx\|$ pour tout x, donc $\|U\| = \|S\|$. De plus pour tout x nous avons

$$\|Sx\| = \|PDP^{-1}x\| = \|DP^{-1}x\|. \tag{12.94}$$

Étant donné que P^{-1} est une bijection, le supremum des $\|Sx\|$ sera le même que celui des $\|Dx\|$ et donc $\|S\| = \|D\|$. Étant donné que par définition $\|U\| \leqslant 1$, nous avons aussi $\|D\| \leqslant 1$ et donc $0 \leqslant \lambda_i \leqslant 1$ (pour rappel, les valeurs propres de D sont positives ou nulles parce que S est ainsi).

Comme $U \notin O(E)$, au moins une des valeurs propres n'est pas 1, supposons que ce soit λ_1. Alors nous avons $\alpha, \beta \in [-1,1]$ avec $-1 \leqslant \alpha < \beta \leqslant 1$ et $\lambda_1 = \frac{1}{2}(\alpha+\beta)$. Nous posons alors

$$D_1 = \operatorname{diag}(\alpha, \lambda_2, \dots, \lambda_n) \tag{12.95a}$$
$$D_2 = \operatorname{diag}(\beta, \lambda_2, \dots, \lambda_n). \tag{12.95b}$$

Nous avons bien $D_1 \neq D2$ et $D_1 + D_2 = D$. Par conséquent

$$U = \frac{1}{2}\big(QPD_1P^{-1} + QPD_2P^{-1}\big) \tag{12.96}$$

avec $QPD_1P^{-1} \neq QPD_2^{-1}$. La matrice U est donc le milieu d'un segment. Reste à montrer que ce segment est dans $\mathcal{B}$. Pour ce faire, prenons $x \in E$ et calculons :

$$\|QPD_iP^{-1}x\| = \|D_iP^{-1}x\| \leqslant \|P^{-1}x\| = \|x\| \tag{12.97}$$

parce que $\|D_i\| \leqslant 1$ et P^{-1} est orthogonale. Au final la norme de QPD_iP est plus petite que 1 et donc U est bien le milieu d'un segment dans $\mathcal{B}$, et donc non extrémal. □

Théorème 12.35 ([268]).
L'enveloppe convexe de $O(n)$ dans $\mathbb{M}_n(\mathbb{R})$ est la boule unité pour la norme induite de $\|.\|_2$ sur $\mathbb{R}^n$.

Démonstration. Nous notons $\mathcal{B}$ la boule unité fermée de $\mathbb{M}(n,\mathbb{R})$ et $\mathcal{C}\mathrm{onv}\big(O(n,\mathbb{R})\big)$ l'enveloppe convexe de $O(n,\mathbb{R})$. Vu que $\mathcal{B}$ est convexe nous avons $\mathcal{C}\mathrm{onv}\big(O(n)\big) \subset \mathcal{B}$.

Maintenant nous devons prouver l'inclusion inverse. Pour ce faire nous supposons avoir un élément $A \in \mathcal{B}\backslash\mathcal{C}\mathrm{onv}\big(O(n)\big)$ et nous allons dériver une contradiction.

Remarquons que $O(n)$ est compact par le lemme 12.26 et que par conséquent $\mathcal{C}\mathrm{onv}(O(n))$ est compacte par le corolaire 8.40 et donc fermée. Nous considérons un produit scalaire $(X,Y) \mapsto X \cdot Y$ sur $\mathbb{M}$. Vu que $\mathcal{C}\mathrm{onv}\big(O(n)\big)$ est un fermé convexe nous pouvons considérer la projection [19] sur $\mathcal{C}\mathrm{onv}(A)$ relativement au produit scalaire choisi.

Nous notons $P = \mathtt{proj}_{\mathcal{C}\mathrm{onv}\left(O(n)\right)}(A)$. En vertu du théorème de projection, nous avons

$$(A-P)\cdot(M-P) \leqslant 0 \tag{12.98}$$

pour tout $M \in \mathcal{C}\mathrm{onv}\,O(n)$. Notons $B = A - P$ pour alléger les notations. L'équation (12.98) s'écrit

$$B \cdot M \leqslant B \cdot P. \tag{12.99}$$

19. Le théorème de projection : théorème 11.232.

D'autre par vu que $B \neq 0$ nous avons $B \cdot B > 0$, c'est-à-dire $B \cdot (A - P) > 0$ et donc

$$B \cdot A > B \cdot P. \tag{12.100}$$

En combinant avec (12.99),

$$B \cdot M \leqslant B \cdot P < B \cdot A. \tag{12.101}$$

Nous utilisons maintenant la décomposition polaire, théorème 12.30, pour écrire $B = QS$ avec $Q \in O(n)$ et $S \in S^+(n, \mathbb{R})$. Vu que l'inégalité (12.101) tient pour tout $M \in \mathcal{C}\mathrm{onv}(O(n))$, elle tient en particulier pour $Q \in O(n)$. Donc

$$B \cdot Q = B \cdot A. \tag{12.102}$$

Nous nous particularisons à présent au produit scalaire $(X, Y) \mapsto \mathrm{Tr}(X^t Y)$ de la proposition 10.23. D'abord

$$B \cdot Q = \mathrm{Tr}(B^t Q) = \mathrm{Tr}(S^t Q^t Q) = \mathrm{Tr}(S^t) = \mathrm{Tr}(S), \tag{12.103}$$

et ensuite l'inégalité (12.103) devient

$$\mathrm{Tr}(S) < B \cdot A = \mathrm{Tr}(S^t Q^t A). \tag{12.104}$$

Nous choisissons une basse $\{e_i\}$ diagonalisant S : $Se_i = \lambda_i e_i$ vérifiant automatiquement $\lambda_i \geqslant 0$ parce que S est semi-définie positive [20]. Alors

$$\begin{aligned}
\mathrm{Tr}(S) &< \mathrm{Tr}(S^t Q^t A) && \text{(12.105a)}\\
&= \sum_i \langle S^t Q^t A e_i, e_i \rangle && \text{(12.105b)}\\
&= \sum_i \langle A e_i, Q S e_i \rangle && \text{(12.105c)}\\
&\leqslant \sum_i \| A e_i \| |\lambda_i| \underbrace{\| Q e_i \|}_{=1} && \text{(12.105d)}\\
&\leqslant \sum_i \lambda_i \qquad\qquad A \in \mathcal{B} \Rightarrow \| A e_i \| \leqslant 1 && \text{(12.105e)}\\
&= \mathrm{Tr}(S). && \text{(12.105f)}
\end{aligned}$$

Il faut noter que la première inégalité est stricte, et donc nous avons une contradiction. □

12.2.8 Décomposition de Bruhat

Théorème 12.36 (Décomposition de Bruhat)**.**
Soit $\mathbb{K}$ un corps ; un élément $M \in \mathrm{GL}(n, \mathbb{R})$ s'écrit sous la forme

$$M = T_1 P_\sigma T_2 \tag{12.106}$$

où T_1 et T_2 sont des matrices triangulaires supérieures inversibles et où P_σ est une matrice de permutations $\sigma \in S_n$. De plus il y a unicité de σ.

Démonstration. Afin de rendre les choses plus visuelles, nous nous permettons de donner des exemples au fur et à mesure de la preuve. Nous prenons l'exemple de la matrice

$$\begin{pmatrix} 1 & 3 & 4 \\ 2 & 5 & 6 \\ 0 & 7 & 8 \end{pmatrix}. \tag{12.107}$$

20. Définition 9.152.

Existence Soit $M \in \mathrm{GL}(n,\mathbb{R})$; vu qu'elle est inversible, on a un indice i_1 maximum tel que $M_{i_1,1} \neq 0$. Nous changeons toutes les lignes jusque là, c'est-à-dire que nous faisons, pour $1 \leqslant i < i_1$,

$$L_i \to L_i - \frac{M_{i1}}{M_{i_1 1}} L_{i_1}. \tag{12.108}$$

Voir le lemme 12.9(3).

Nous avons donc obtenu une matrice dont la première colonne est nulle sauf la case numéro i_1. L'opération (12.108) revient à considérer la multiplication par la matrice de transvection

$$T_1^{(i)} = T_{ii_1}\left(-\frac{M_{i1}}{M_{i_1 1}}\right) \tag{12.109}$$

pour tout $i < i_1$. Pour rappel nous ne changeons que les lignes *au-dessus* de la i_1. Du coup les matrices $T_1^{(i)}$ sont triangulaires supérieures. Nous avons donc la nouvelle matrice $M_1 = \left(\prod_{i<i_1} T_1^{(i)}\right) M$ pour laquelle toute la première colonne est nulle sauf un élément.

Dans le cas de l'exemple, le « pivot » sera la ligne $(2,5,6)$ et la matrice se transforme à l'aide de la matrice $T_1 = T_{12}(-1/2)$:

$$\begin{pmatrix} 1 & -1/2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 & 4 \\ 2 & 5 & 6 \\ 0 & 7 & 8 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 & 1 \\ 2 & 5 & 6 \\ 0 & 7 & 8 \end{pmatrix}. \tag{12.110}$$

Maintenant nous faisons de même avec les colonnes (en renommant M la matrice obtenue à l'étape précédente) :

$$C_j \to C_j - \frac{M_{i_1 j}}{M_{i_1 1}} C_1, \tag{12.111}$$

qui revient à multiplier à droite par les matrices $T_{1j}(\frac{M_{i_1 i}}{M_{i_1 1}})$ avec $j > 1$. Encore une fois ce sont des matrices triangulaires supérieures.

Dans l'exemple, pour traiter la seconde colonne, nous multiplions (12.110) à droite par la matrice $T_{12}(-5/2)$:

$$\begin{pmatrix} 0 & 1/2 & 1 \\ 2 & 5 & 6 \\ 0 & 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & -5/2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1/2 & 1 \\ 2 & 0 & 6 \\ 0 & 7 & 8 \end{pmatrix}. \tag{12.112}$$

Appliquer encore la matrice $T_{13}(-6/2)$ apporte la matrice

$$\begin{pmatrix} 0 & 1/2 & 1 \\ 2 & 0 & 0 \\ 0 & 7 & 8 \end{pmatrix}. \tag{12.113}$$

Enfin nous multiplions la matrice obtenue par $\frac{1}{M_{i_1 1}}\mathbb{1}$ pour normaliser à 1 l'élément « pivot » que nous avions choisit. Dans notre exemple nous multiplions par $1/2$ pour trouver

$$\begin{pmatrix} 0 & 1/4 & 1/2 \\ 1 & 0 & 0 \\ 0 & 7/2 & 4 \end{pmatrix}. \tag{12.114}$$

La matrice obtenue jusqu'ici possède une ligne et une colonne de zéros avec un 1 à leur intersection, et elle est de la forme

$$M' = T_1 M T_2 \tag{12.115}$$

où T_1 et T_2 sont triangulaires supérieures et inversibles, produits de matrices de transvection (et d'une matrice scalaire pour la normalisation).

Il reste à recommencer l'opération avec la seconde colonne (qui n'est pas toute nulle parce que le déterminant est encore non nul) puis la suivante etc. Dans notre exemple de l'équation (12.114), nous éliminerions le 1/4 et le 4 en utilisant le 7/2.

Encore une fois tout cela se fait à l'aide de matrice supérieures parce qu'à chaque étape, les colonnes précédent le pivot sont déjà nulles (saut un 1) et ne doivent donc pas être touchées.

À la fin de ce processus, ce qui reste est une matrice TMT' qui ne contient plus que un seul 1 sur chaque ligne et chaque colonne, c'est-à-dire une matrice de permutations : $P_\sigma = TMT'$ et donc

$$M = T_\sigma^{-1}(T')^{-1}. \tag{12.116}$$

Unicité Soient $\sigma, \sigma \in S'_n$ tels que $T_1 P_\sigma T_2 = S_1 P_\tau S_2$ avec T_i et S_i triangulaires supérieures et inversibles. En posant $T = T_2 S_2^{-1}$ et $S = T_1^{-1} S_1$, nous avons

$$P_\sigma T = S P_\tau \tag{12.117}$$

où S et T sont des matrices triangulaires supérieures et inversibles. Par les calculs de la preuve du lemme 4.102,

$$\begin{cases} (P_\sigma T)_{kl} = T_{\sigma^{-1}(k)l} & (12.118a) \\ (SP_\tau)_{kl} = S_{k\tau(l)}, & (12.118b) \end{cases}$$

et donc

$$T_{\sigma^{-1}(k)l} = S_{k\tau(l)}. \tag{12.119}$$

En écrivant cette équation avec $k = \sigma(i)$ (nous rappelons que σ est bijective),

$$T_{il} = S_{\sigma(i)\tau(l)}. \tag{12.120}$$

Nous savons que les termes diagonaux de T sont non nuls parce que T est triangulaire supérieure et inversible (donc pas de colonnes entières nulles). Nous avons donc, en prenant $i = l = k$,

$$0 \neq T_{kk} = S_{\sigma(k)\tau(k)}. \tag{12.121}$$

La matrice étant triangulaire supérieure, cela implique

$$\sigma(k) \leqslant \tau(k). \tag{12.122}$$

De la même manière en écrivant (12.119) avec $l = \tau^{-1}(i)$,

$$S_{ki} = T_{\sigma^{-1}(k)\tau^{-1}(i)} \tag{12.123}$$

et donc

$$\sigma^{-1}(k) \leqslant \tau^{-1}(k). \tag{12.124}$$

En écrivant cela avec $k = \sigma(j)$, nous avons $j \leqslant \tau^{-1}\sigma(j)$ et en appliquant enfin τ,

$$\tau(j) \leqslant \sigma(j). \tag{12.125}$$

En comparant avec (12.122), nous avons $\sigma = \tau$.

□

12.3 Sous-groupes du groupe linéaire

Lemme 12.37 ([54]).
Soit V un espace vectoriel de dimension finie muni d'une norme euclidienne $\|.\|$. Soit K un compact convexe de V et G, un sous-groupe compact de $\mathrm{GL}(V)$ tel que

$$u(K) \subset K \tag{12.126}$$

pour tout $u \in G$. Alors il existe $a \in K$ tel que $u(a) = a$ pour tout $u \in G$.

Démonstration. Avant de nous lancer dans la preuve, nous avons besoin d'un petit résultat.

Un pré-résultat Nous commençons par prouver que si $v \in \mathcal{L}(V)$ vérifie $v(K) \subset K$, alors v a un point fixe dans K. Pour cela nous considérons $x_0 \in K$ et la suite

$$x_k = \frac{1}{k+1} \sum_{i=0}^{k} v^i(x_0). \tag{12.127}$$

Étant donné que K est convexe et stable par v, la suite (x_k) est contenue dans K et accepte une sous-suite convergente[21] que nous allons noter $x_{\varphi(n)}$ avec $\varphi\colon \mathbb{N} \to \mathbb{N}$ strictement croissante. Soit $a \in K$ la limite :

$$\lim_{n\to\infty} x_{\varphi(n)} = a. \tag{12.128}$$

Tant que nous y sommes nous pouvons aussi calculer $v(x_k)$:

$$v(x_k) = v\left(\frac{1}{k+1}\sum_{i=1}^{k} v^i(x_0)\right) \tag{12.129a}$$

$$= \frac{1}{k+1}\sum_{i=0}^{k} v^{i+1}(x_0) \tag{12.129b}$$

$$= x_k + \frac{1}{k+1}\Big(v^{k+1}(x_0) - x_0\Big). \tag{12.129c}$$

La norme $\|v^{k+1}(x_0) - x_0\|$ est bornée par le diamètre de K, donc en prenant la limite $k \to \infty$ le second terme de (12.129c) tend vers zéro. En prenant ces égalités en $k = \varphi(n)$ et en prenant $n \to \infty$, nous trouvons

$$v(a) = a, \tag{12.130}$$

c'est-à-dire le résultat que nous voulions dans un premier temps.

Une norme sur V Nous passons maintenant à la preuve du lemme. D'abord nous remarquons que le groupe G agit sur V par $u \cdot x = u(x)$ et de plus, considérant la fonction continue

$$\begin{aligned} \alpha\colon G &\to V \\ u &\mapsto u(x), \end{aligned} \tag{12.131}$$

nous voyons que les orbites de cette action sont compactes en tant qu'image par α du compact G (théorème 7.92). Nous posons

$$\begin{aligned} \nu\colon V &\to \mathbb{R}^+ \\ x &\mapsto \max_{u\in G} \|u(x)\|. \end{aligned} \tag{12.132}$$

Cette définition a un sens parce que l'orbite $\{u(x) \text{ tel que } u \in G\}$ est compacte dans V et donc l'ensemble des normes est compact dans $\mathbb{R}$ et admet un maximum. De plus cela donne une norme sur V parce que nous vérifions les conditions de la définition 7.113 :

(1) Pour tout $x, y \in V$ nous avons :

$$\nu(x+y) = \max_{u\in G} \|u(x) + u(y)\| \leqslant \max_{u\in G} \big(\|u(x)\| + \|u(y)\|\big) \leqslant \nu(x) + \nu(y). \tag{12.133}$$

(2) Si $\nu(x) = 0$, alors l'égalité $\max_{u\in G} \|u(x)\| = 0$ nous enseigne que $\|u(x)\| = 0$ pour tout $u \in G$ et donc en particulier avec $u = \mathrm{Id}$ nous trouvons $x = 0$.

(3) Pour tout $\lambda \in \mathbb{R}$ et $x \in V$,

$$\nu(\lambda x) = \max_{u\in G} \|u(\lambda x)\| = \max \|\lambda u(x)\| = \max |\lambda| \|u(x)\| = |\lambda| \nu(x). \tag{12.134}$$

21. C'est Bolzano-Weierstrass, théorème 7.103.

De plus la fonction ν est constante sur les orbites de G.

Un point fixe Pour tout $u \in G$ nous posons

$$F_u = \{x \in K \text{ tel que } u(x) = x\}; \tag{12.135}$$

par le pré-résultat, aucun de ces ensembles n'est vide. Ils sont de plus tous fermés par continuité de u (le complémentaire est ouvert). Nous devons prouver que $\bigcap_{u\in G} F_u \neq \varnothing$ parce qu'une intersection serait un point fixe de tous les éléments de G. Supposons donc que $\bigcap_{u\in G} F_u = \varnothing$. Alors les complémentaires des F_u forment un recouvrement ouvert de K et nous pouvons en extraire un sous-recouvrement fini par compacité. Soient $\{u_i\}_{i=1,\dots,p}$ les éléments qui réalisent ce recouvrement. Alors

$$\bigcap_{i=1}^{p} F_{u_i} = \varnothing. \tag{12.136}$$

Nous considérons l'opérateur

$$v = \frac{1}{p}\sum_{i=1}^{p} u_i \in \mathcal{L}(V). \tag{12.137}$$

Vu que K est convexe et stable sous chacun des u_i, nous avons aussi $v(K) \subset K$ et donc il existe $a \in K$ tel que $v(a) = a$. Pour ce a, nous avons

$$\nu\big(v(a)\big) = \nu\left(\frac{1}{p}\sum_{i=1}^{p} u_i(a)\right) \tag{12.138a}$$

$$\leqslant \frac{1}{p}\sum_{i=1}^{p} \nu\left(u_i(a)\right) \tag{12.138b}$$

$$= \frac{1}{p}\sum_{i=1}^{p} \nu(a) \tag{12.138c}$$

$$= \nu(a) \tag{12.138d}$$

où nous avons utilisé la constance de ν sur les orbites de G. Par ailleurs nous savons que $v(a) = a$, donc en réalité à gauche dans (12.138a) nous avons $\nu(a)$ et toutes les inégalités sont des égalités. Nous avons en particulier

$$\nu\left(\sum_{i=1}^{p} u_i(a)\right) = \sum_{i=1}^{p} \nu\left(u_i(a)\right). \tag{12.139}$$

Notons $u_0 \in G$ l'élément qui réalise le maximum de la définition de ν pour le vecteur $\sum_i u_i(a)$:

$$\nu\left(\sum_i u_i(a)\right) = \|u_0\left(\sum_i u_i(a)\right)\| \leqslant \sum_i \|u_0 u_i(a)\| \leqslant \sum_i \nu\big(u_i(a)\big). \tag{12.140}$$

Mais nous venons de voir (équation (12.139)) que l'expression de gauche est égale à celle de droite. Donc les inégalités sont des égalités et en particulier la première inégalité devient l'égalité

$$\|\sum_i u_0 u_i(a)\| = \sum_i \|u_0 u_i(a)\|. \tag{12.141}$$

En vertu du lemme 7.195, il existe des nombres positifs λ_i tels que

$$u_0 u_1(a) = \lambda_2 u_0 u_2(a) = \dots = \lambda_p u_0 u_p(a). \tag{12.142}$$

Du fait que u_0 est inversible nous avons aussi

$$u_1(a) = \lambda_2 u_2(a) = \dots = \lambda_p u_p(a). \tag{12.143}$$

Mais par constance de ν sur les orbites nous avons $\nu(u_i(a)) = \nu(u_j(a))$ pour tout i et j ; en appliquant ν à la série d'égalités (12.143), nous trouvons que tous les λ_i doivent être égaux à 1. En particulier

$$u_1(a) = u_2(a) = \ldots = u_p(a). \tag{12.144}$$

Nous récrivons maintenant l'équation $v(a) = a$ avec la définition de v :

$$a = v(a) = \frac{1}{p}\sum_{i=1}^{p} u_i(a) = u_j(a) \tag{12.145}$$

pour n'importe quel j. Donc

$$a \in \bigcap_{i=1}^{p} F_{u_i}, \tag{12.146}$$

ce qui contredit notre hypothèse de départ.

□

Proposition 12.38 ([268, 54, 269])**.**
Soit G un sous-groupe compact de $\mathrm{GL}(n,\mathbb{R})$. Alors

(1) Il existe une forme quadratique définie positive q sur $\mathbb{R}^n$ telle que $G \subset \mathrm{O}(q)$.

(2) Le groupe G est conjugué à un sous-groupe de $\mathrm{O}(n,\mathbb{R})$.

Démonstration. Nous considérons le (pas tout à fait) morphisme de groupe

$$\begin{aligned} \rho\colon G &\to \mathrm{GL}\big(\mathrm{S}(n,\mathbb{R})\big) \\ u &\mapsto \rho_u\colon s \to u^t s u, \end{aligned} \tag{12.147}$$

et tant que nous y sommes à considérer, nous considérons l'ensemble

$$H = \{M^t M \text{ tel que } M \in G\} \subset \mathrm{S}(n,\mathbb{R}). \tag{12.148}$$

Cet ensemble est constitué de matrices définies positives parce que si $\langle M^t M x, x\rangle = 0$, alors $0 = \langle Mx, Mx\rangle = \|Mx\|$, mais M étant inversible, cela implique que $x = 0$. Qui plus est cet ensemble est compact dans $\mathrm{GL}(n,\mathbb{R})$ en tant qu'image du compact G par l'application continue $M \mapsto M^t M$. L'enveloppe convexe $K = \mathcal{C}\mathrm{onv}(H)$ est alors également compacte par le théorème 8.40. Enfin nous considérons $L = \rho(G)$, qui est un sous-groupe compact de $\mathrm{GL}\big(\mathrm{S}(n,\mathbb{R})\big)$ parce que $\rho_u\rho_v = \rho_{vu} \in \rho(G)$. Nous remarquons que ρ_u étant linéaire, elle préserve les combinaisons convexes et donc pour tout $u \in G$, $\rho_u(K) \subset K$.

Bref, L est un sous-groupe compact de $\mathrm{GL}(n,\mathbb{R})$ préservant le compact K de $\mathrm{S}(n,\mathbb{R})$. Par le lemme 12.37, il existe $s \in K$ tel que $\rho_u(s) = s$ pour tout $u \in G$. Ou encore :

$$u^t s u = s \tag{12.149}$$

pour tout $u \in G$. Fort de ce s bien particulier, nous considérons la forme quadratique associée : $q(x) = x^t s x$. Cette forme est définie positive parce que s l'est. Nous avons $G \subset \mathrm{O}(q)$ parce que si $u \in G$ alors

$$q\big(ux\big) = (ux)^t s u x = x^t \underbrace{u^t s u}_{=s} x = q(x). \tag{12.150}$$

Le premier point est prouvé.

La matrice s est symétrique et définie positive. Le théorème 9.150 nous permet donc de la diagonaliser en $\mathrm{diag}(\lambda_1,\ldots,\lambda_n)$ avec $\lambda_i > 0$, et ensuite transformée en la matrice $\mathbb{1}_n$ par la matrice $\mathrm{diag}(1/\sqrt{\lambda_i})$. Nous avons donc une matrice $a \in \mathrm{GL}(n,\mathbb{R})$ telle que $a^t s a = \mathbb{1}_n$. Avec ça, si $u \in G$, nous avons

$$(a^{-1}ua)^t(a^{-1}ua) = (a^{-1}ua)^t\mathbb{1}_n(a^{-1}ua) = a^t u^t (a^t)^{-1} a^t s a a^{-1} u a = a^t u^t s u a = a^t s a = \mathbb{1}, \tag{12.151}$$

ce qui prouve que $a^{-1}ua$ est dans $\mathrm{O}(n,\mathbb{R})$, et donc que $a^{-1}Ga \subset \mathrm{O}(n,\mathbb{R})$. □

Chapitre 13

Tribus, théorie de la mesure, intégration

13.1 Tribus

Vous pouvez voir le thème 2 pour voir plus vite où sont les définitions associées.

13.1.1 Généralités

Définition 13.1 (Tribu, espace mesurable[270])**.**
Si Ω est un ensemble, un ensemble $\mathcal{A}$ de sous-ensembles de Ω est une ***tribu*** *si*

(1) $\Omega \in \mathcal{A}$ *;*

(2) $A^c \in \mathcal{A}$ *pour tout* $A \in \mathcal{A}$ *;*

(3) si $(A_i)_{i\in\mathbb{N}}$ *une suite dénombrable d'éléments de* $\mathcal{A}$*, alors* $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$*.*

Le couple $(\Omega, \mathcal{A})$ *est alors un* ***espace mesurable****.*

Remarque 13.2.
Nous trouvons parfois la notation

$$\bigcup_{k\in\mathbb{N}} A_k = \sup_{k\geqslant 0} A_k. \tag{13.1}$$

Lemme 13.3.
Opérations ensemblistes sur les tribus.

(1) Une tribu est stable par intersections au plus dénombrables.

(2) Une tribu est stable par différence ensembliste.

Démonstration. Soit $(A_i)_{i\in I}$ une famille au plus dénombrable d'éléments de la tribu $\mathcal{A}$. Nous devons prouver que $\bigcap_{i\in I} A_i$ est également un élément de $\mathcal{A}$. Pour cela nous passons au complémentaire :

$$\complement\left(\bigcap_{i\in I} A_i\right) = \bigcup_{i\in I} \complement A_i. \tag{13.2}$$

La définition d'une tribu implique que le membre de droite est un élément de la tribu. Par stabilité d'une tribu par complémentaire, l'ensemble $\bigcap_{i\in I} A_i$ est également un élément de la tribu.

La seconde assertion est immédiate à partir de la première parce que $A\backslash B = A \cap \complement B$. □

Si $(\mathcal{A}_i)_{i\in I}$ est un ensemble de tribus (indexé par un ensemble I quelconque) alors

$$\mathcal{A} = \bigcap_{i\in I} \mathcal{A}_i \tag{13.3}$$

est également une tribu.

Définition 13.4.
*Soit $\mathcal{D}$ un ensemble de parties de Ω. La **tribu engendrée** par $\mathcal{D}$ est l'intersection de toutes les tribus de Ω contenant $\mathcal{D}$. C'est la plus petite tribu contenant $\mathcal{D}$. Nous la noterons le plus souvent $\sigma(\mathcal{A})$*

Note : une tribu engendrée par une application sera la définition 13.46.

Proposition 13.5 ([271])**.**
Soit S un ensemble et $\mathcal{F}$ une tribu de S. Soit une classe $\mathcal{N} \subset \mathcal{P}(S)$ telle que

(1) Si $A \in \mathcal{N}$ alors il existe $Y \in \mathcal{F} \cap \mathcal{N}$ tel que $A \subset Y$.

(2) Si $A \in \mathcal{N}$ et $B \subset A$ alors $B \in \mathcal{N}$.

(3) La classe $\mathcal{N}$ est stable par union dénombrable.

Alors la classe

$$\mathcal{T} = \{X \cup A \text{ avec } A \in \mathcal{N} \text{ et } X \in \mathcal{F}\} \tag{13.4}$$

est une tribu.

Démonstration. L'ensemble $\mathcal{T}$ est stable par union dénombrable parce que $\mathcal{F}$ et $\mathcal{N}$ le sont. De plus S et $\emptyset$ sont dans $\mathcal{F}$ et donc dans $\mathcal{T}$. Nous devons voir que $\mathcal{T}$ est stable par complémentarité.

Soit donc $A \in \mathcal{N}$ et $X \in \mathcal{F}$; nous savons que $(A \cup X)^c = A^c \cap X^c$. De plus il existe $Y \in \mathcal{F} \cap \mathcal{N}$ tel que $A \subset Y$ et nous pouvons exprimer A^c en termes de Y : $A^c = Y^c \cup (Y \backslash A)$. Donc

$$(A \cup X)^c = \big(Y^c \cup (Y \backslash A)\big) \cap X^c = \underbrace{(Y^c \cap X^c)}_{\in \mathcal{F}} \cup \underbrace{\big((Y \backslash A) \cap X^c\big)}_{\in \mathcal{N}}. \tag{13.5}$$

Le fait que la seconde partie soit dans $\mathcal{N}$ est due au fait que ce soit une partie de $Y \in \mathcal{N}$. Nous avons donc bien $(A \cup X)^c \in \mathcal{T}$. □

13.1.2 Tribu induite

Proposition-définition 13.6 (Tribu induite, tribu-trace[272])**.**
Soit un espace mesurable $(S, \mathcal{F})$ et une partie $R \subset S$. L'ensemble

$$\mathcal{F}_R = \{A \cap R \text{ tel que } A \in \mathcal{F}\} \tag{13.6}$$

*est une tribu. Elle est la **tribu induite** de R depuis S. Elle est aussi nommée « tribu trace ».*

Démonstration. D'abord R et $\emptyset$ dont dans $\mathcal{F}_R$. Si $C \in \mathcal{F}_R$ alors $C = A \cap R$ pour un certain $A \in \mathcal{F}$ et nous devons prouver que $R \cap C^c$ est dans $\mathcal{F}_R$ (le complémentaire de C dans R). Nous avons

$$R \cap C^c = R \cap (A \cap R)^c = R \cap A^c \in \mathcal{F}_R \tag{13.7}$$

parce que $A^c \in \mathcal{F}$. Enfin si $C_i \in \mathcal{F}_R$ alors $C_i = R \cap A_i$ pour des A_i dans $\mathcal{F}$. Nous avons

$$\bigcup_{i \in \mathbb{N}} C_i = \bigcup_{i \in \mathbb{N}} (R \cap A_i) = R \cap \big(\bigcup_{i \in \mathbb{N}} A_i\big), \tag{13.8}$$

mais $\bigcup_i A_i \in \mathcal{F}$ donc $\bigcup_i C_i \in \mathcal{F}_R$. □

Proposition 13.7.
Si R est mesurable dans $(\Omega, \mathcal{A})$ alors

$$\mathcal{F}_R = \{A \cap R \text{ tel que } A \in \mathcal{F}\} = \{S \in \mathcal{A} \text{ tel que } S \subset R\}, \tag{13.9}$$

où $\mathcal{F}_R$ est la tribu induite de $\mathcal{A}$ sur R.

Démonstration. La première égalité est simplement la définition (13.6) de la tribu induite. Pour le reste, nous notons $\mathcal{A}_R = \{S \in \mathcal{A} \text{ tel que } S \subset R\}$.

Si $S \subset R$ et $S \in \mathcal{A}$ alors $S = S \cap R \in \mathcal{A}_R$.

Dans l'autre sens, si $S \in \mathcal{A}_R$, alors il existe $A \in \mathcal{A}$ tel que $S = A \cap R$. Donc $S \subset A$ et $S \in \mathcal{A}$ parce que R et A sont des éléments de $\mathcal{A}$ (stable par intersection). □

13.1.3 Tribu de Baire

Définition 13.8.
*Une partie d'un espace topologique est **rare** si elle est contenue dans un fermé d'intérieur vide.*
*Une partie est **maigre** si elle est réunion finie ou dénombrable de parties rares.*

Exemple 13.9.
L'ensemble $\mathbb{Q}$ dans $\mathbb{R}$ est maigre mais n'est pas rare parce que $\bar{\mathbb{Q}} = \mathbb{R}$. △

Proposition 13.10 ([271])**.**
Soit X un espace topologique. L'ensemble de parties

$$\mathcal{B}a(X) = \{B \cup A \text{ avec } B \text{ borélien et } A \text{ maigre}\} \tag{13.10}$$

*est une tribu. Elle est appelée la **tribu de Baire** de l'espace X.*

Démonstration. Nous allons montrer que les boréliens et les maigres vérifient les conditions de la proposition 13.5.

(1) Si A est maigre, il s'écrit comme $A = \bigcup_{i\in\mathbb{N}} R_i$ où les R_i sont rares. Il existe donc des fermés d'intérieur vide F_i tels que $R_i \subset F_i$; en particulier $A \subset \bigcup_i F_i$. En tant que fermés, $F_i \in \mathcal{B}\mathrm{or}(X)$; de plus chaque F_i est rare, donc $\bigcup_i F_i$ est maigre. L'ensemble A est donc bien contenu dans un ensemble maigre et borélien.

(2) Soi A maigre et $B \subset A$. Nous avons, avec les mêmes notations, $A = \bigcup_i R_i$ et $B = \bigcup_i (R_i \cap B)$. Les ensembles $R_i \cap B$ sont encore rares, donc B est une union dénombrable d'ensembles rares. L'ensemble B est donc maigre.

(3) Si les ensembles (A_i) sont maigres, alors ils sont unions dénombrables de rares : $A_i = \bigcup_k R_k^{(i)}$. Nous avons alors

$$\bigcup_i A_i = \bigcup_{(i,k)\in\mathbb{N}^2} R_k^{(i)}, \tag{13.11}$$

et donc $\bigcup_i A_i$ est encore une union dénombrable d'ensembles rares.

□

Proposition 13.11 ([271])**.**
Une partie B de l'espace topologique X est dans la tribu de Baire de X si et seulement s'il existe un ouvert U tel que $B\Delta U$ est maigre.

Démonstration. Nous définissons la relation d'équivalence[1] suivante sur $\mathcal{P}(X)$: nous disons que $A \sim B$ si et seulement si $A\Delta B$ est maigre.

Réflexive Nous avons $A\Delta A = \emptyset$, donc $A \sim A$.

symétrique Nous avons $A\Delta B = B\Delta A$, donc $\sim$ est symétrique.

transitive Si A, B, C sont des parties de X alors nous avons toujours

$$A\Delta C \subset (A \cup B \cup C)\backslash(A \cap B \cap C) = (A\Delta B) \cup (B\Delta C). \tag{13.12}$$

Donc si $A \sim B$ et $B \sim C$ alors $A\Delta C$ est contenu dans une union de maigres et est donc maigre.

Autres propriétés de $\sim$ De plus la relation d'équivalence $\sim$ vérifie $A \sim B$ si et seulement si $A^c \sim B^c$, par le lemme 1.19(1).

Pour compléter les propriétés de $\sim$ mentionnons encore le fait que si F est fermé alors $F \sim \mathrm{Int}(F)$. En effet $F \cup \mathrm{Int}(F) = F$ et $F \cap \mathrm{Int}(F) = \mathrm{Int}(F)$, de telle sorte que $F\Delta\,\mathrm{Int}(F) = F\backslash\,\mathrm{Int}(F)$. Cela est un fermé parce que son complémentaire est $F^c \cup \mathrm{Int}(F)$ qui est une union d'ouverts. De plus $F \subset \mathrm{Int}(F)$ est d'intérieur vide, de telle sorte qu'il est rare et donc maigre.

1. Définition 1.20

Pour la suite de la preuve nous posons

$$\mathcal{F} = \{A \subset X \text{ tel que il existe ouvert } U \text{ avec } U \sim A\}, \tag{13.13}$$

et nous devons prouver que $\mathcal{F} = \mathcal{B}\mathrm{a}(X)$.

<u>$\mathcal{F} \subset \mathcal{B}\mathbf{a}(X)$</u> Soit $A \in \mathcal{F}$ et un ouvert U tel que $U \sim A$. Alors nous posons $M = U\Delta A$ qui est maigre. En vertu du lemme 1.19(2), nous avons

$$A = M\Delta U = (M \cup U)\backslash(M \cap U), \tag{13.14}$$

ce qui prouve que A est dans la tribu engendrée par les ouverts et les maigres, laquelle tribu est contenue dans $\mathcal{B}\mathrm{a}(X)$.

<u>$\mathcal{B}\mathbf{a}(X) \subset \mathcal{F}$</u> Nous allons montrer que $\mathcal{F}$ est une tribu contenant tous les ouverts et tous les maigres. Alors en particulier $\mathcal{F}$ contiendra $\mathcal{B}\mathrm{a}(X)$. Si U est ouvert, $U \sim U$ et donc $U \in \mathcal{F}$. Si M est maigre, alors $M \sim \emptyset$ et donc $M \in \mathcal{F}$. Il reste à prouver que $\mathcal{F}$ est une tribu.

<u>**Vide et ensemble**</u> Cela est facile : X et $\emptyset$ sont dans $\mathcal{F}$.

<u>**Comlémentaire**</u> Commençons par nous souvenir que $F \sim \mathrm{Int}(F)$ dès que F est fermé. Si $A \in \mathcal{F}$ alors il existe un ouvert U tel que $A \sim U$ et donc aussi $A^c \sim U^c$. D'autre part U^c est fermé, donc $U^c \sim \mathrm{Int}(U^c)$, donc

$$A^c \sim U^c \sim \mathrm{Int}(U^c), \tag{13.15}$$

ce qui implique que $A^c \in \mathcal{F}$.

<u>**Union dénombrable**</u> Soit $A_n \in \mathcal{F}$ et $M_n = A_n\Delta U_n$ avec M_n maigre et U_n ouvert. Nous allons prouver que

$$\bigcup_n A_n \sim \bigcup_n U_n. \tag{13.16}$$

Pour cela il faut remarque que

$$\Big(\bigcup_n A_n\Big)\Delta\Big(\bigcup_n U_n\Big) \subset \bigcup_n (A_n\Delta U_n) = \bigcup_n M_n. \tag{13.17}$$

Le terme le plus à droite est maigre, ce qui signifie que celui le plus à gauche est contenu dans un maigre et donc est maigre lui-même.

□

Proposition 13.12.
Si B est un borélien de X, alors il existe un ouvert U et un maigre M tels que

(1) $B\Delta U$ est maigre,

(2) $M\Delta U = B$,

(3) $D\Delta M$ est ouvert.

Démonstration. Vu que B est borélien, il est aussi dans la tribu de Baire et il existe par la proposition 13.11 un ouvert U tel que $M = B\Delta U$ est maigre. En prenant ce U et ce M, les trois conditions sont vérifiées parce que

$$M\Delta U = (B\Delta U)\Delta U = B \tag{13.18}$$

et

$$B\Delta M = M\Delta B = (U\Delta B)\Delta B = U. \tag{13.19}$$

Tout par le lemme 1.19(2). □

13.2 Théorie de la mesure

Définition 13.13 ([273]).
*Une **mesure extérieure** sur un ensemble S est une application $m^*\colon \mathcal{P}(S) \to [0,\infty]$ telle que*

(1) $m^*(\emptyset) = 0$,

(2) Si $A \subset B$ *dans* S *alors* $m^*(A) \leqslant m^*(B)$

(3) Si les A_n *sont des parties de* S *alors*

$$m^*\big(\bigcup_{n\in\mathbb{N}} A_n\big) \leqslant \sum_{n\in\mathbb{N}} m^*(A_n). \tag{13.20}$$

La différence avec une mesure est que nous ne demandons pas que (13.20) soit une égalité lorsque les A_n sont disjoints.

13.2.1 Mesure sur un ensemble de parties

Définition 13.14 (Mesure sur un ensemle de parties[274]).
*Soit S un ensemble et $\mathcal{C}$ un ensemble de parties de S contenant $\emptyset$. Une **mesure positive** sur $(S,\mathcal{C})$ est une application $\mu\colon \mathcal{C} \to [0,\infty]$ telle que*

(1) $\mu(\emptyset) = 0$,

(2) Si $A_n \in \mathcal{C}$ *sont des éléments deux à deux disjoints dans* $\mathcal{C}$ *et tels que* $\bigcup_n A_n \in \mathcal{C}$ *alors*

$$\mu\big(\bigcup_{n=0}^{\infty} A_n\big) = \sum_{n=0}^{\infty} \mu(A_n). \tag{13.21}$$

*La mesure est **finie** si $\mu(S) < \infty$ et σ**-finie** s'il existe une suite (S_n) dans $\mathcal{C}$ telle que $S = \bigcup_n S_n$ et $\mu(S_n) < \infty$.*

Remarque 13.15.
La condition $\mu(\emptyset) = 0$ est nécessaire. Certes, si $A \in \mathcal{A}$ nous avons

$$\mu(A) = \mu(A \cup \emptyset) = \mu(A) + \mu(\emptyset) \tag{13.22}$$

parce que A et $\emptyset$ sont disjoints. Cela semble indiquer que $\mu(\emptyset) = 0$, mais pas tout à fait : il est encore possible d'avoir $\mu(B) = \infty$ pour tout $B \in \mathcal{A}$, y compris $\mu(\emptyset) = \infty$. À cause de cette exception, la relation (13.22) n'implique pas $\mu(\emptyset) = 0$.

Sans hypothèses sur l'ensemble de parties, nous ne pouvons pas dire grand chose de plus.

13.2.2 Mesure sur une algèbre de parties

Définition 13.16 (Algèbre de parties[273]).
*Soit S, un ensemble. Une classe $\mathcal{D}$ de parties de S est une **algèbre de parties** de S si*

(1) $S \in \mathcal{D}$ *et* $\emptyset \in \mathcal{D}$,

(2) si $A \in \mathcal{D}$ *alors* $A^c \in \mathcal{D}$,

(3) si $A, B \in \mathcal{D}$ *alors* $A \cup B \in \mathcal{D}$.

Les algèbre de parties ne sont pas des classes si sauvages que ça ; en témoigne le lemme suivant.

Lemme 13.17.
Une algèbre de partie est stable par intersection (finie) et par différence ensembliste.

Démonstration. Il suffit de remarquer que $A \cap B = \big(A^c \cup B^c\big)^c$ et que $A \backslash B = A \cap B^c$. □

Lemme 13.18 ([273]).
Si $\mathcal{D}$ est une algèbre de parties de S et si μ est une mesure sur $(S,\mathcal{D})$ alors

(1) si $A, B \in \mathcal{D}$ avec $A \subset B$ alors $\mu(A) \leqslant \mu(B)$

(2) si $A_n \in \mathcal{D}$ et $\bigcup_n A_n \in \mathcal{D}$ alors

$$\mu(\bigcup_n A_n) \leqslant \sum_n \mu(A_n). \tag{13.23}$$

La propriété (2) est la **σ-sous-additivité**.

Démonstration. Si $A \subset B$ alors $B = A \cup (B \backslash A)$ avec A et $B \backslash A$ disjoints donc

$$\mu(B) = \mu(A) + \mu(B \backslash A) \geqslant \mu(A). \tag{13.24}$$

Pour la seconde, on passe par les compléments deux à deux : nous posons

$$\begin{cases} B_0 = \varnothing & (13.25a) \\ B_n = A_n \backslash \bigcup_{k<n} B_k. & (13.25b) \end{cases}$$

Ces ensembles sont deux à deux disjoints et $\bigcup_n B_n = \bigcup_n A_n \in \mathcal{D}$, donc

$$\mu(\bigcup_n A_n) = \sum_n \mu(A_n \backslash \bigcup_{k<n} B_k) \leqslant \sum_n \mu(A_n), \tag{13.26}$$

où nous avons utilisé la première partie du lemme. □

Proposition 13.19 (Mesure extérieure à partir d'une algèbre de parties[273])**.**
Soit $\mathcal{D}$ une algèbre de partie sur l'ensemble S et μ une mesure sur $(S, \mathcal{D})$. Alors l'application

$$\begin{aligned} \mu^* \colon \mathcal{P}(S) &\to [0, +\infty] \\ X &\mapsto \inf\{\sum_n \mu(A_n) \text{ tel que } A_n \in \mathcal{D}, X \subset \bigcup_n A_n\} \end{aligned} \tag{13.27}$$

est une mesure extérieure[2] *sur S et pour tout $A \in \mathcal{D}$ nous avons $\mu^*(A) = \mu(A)$.*

Démonstration. Ceci est une phrase servant à aligner correctement le `description` qui suit.

La définition est bonne Notons d'abord que la définition est bonne : l'ensemble sur lequel l'infimum est pris n'est pas vide : il suffit de prendre $A_1 = S$ et $A_{n\geqslant 2} = \varnothing$.

Le vide D'abord $\mu^*(\varnothing) = 0$ parce que $\varnothing \in \mathcal{D}$. Prendre $A_n = \varnothing$.

Inégalité d'inclusion Soient $X \subset Y$ dans $\mathcal{P}(S)$. Si $Y \subset A$ alors $X \subset A$, donc

$$\{\sum_n \mu(A_n) \text{ tel que } A_n \in \mathcal{D}, Y \subset \bigcup_n A_n\} \subset \{\sum_n \mu(A_n) \text{ tel que } A_n \in \mathcal{D}, X \subset \bigcup_n A_n\}, \tag{13.28}$$

ce qui prouve que $\mu^*(X) \leqslant \mu^*(Y)$.

Inégalité par union dénombrable Soit $(X_n)_{n\in\mathbb{N}}$ une suite de parties de S. S'il existe n_0 tel que $\mu^*(X_{n_0}) = \infty$ alors nous avons automatiquement $\sum_n \mu^*(X_n) = \infty$ et l'inégalité demandée est évidente parce que n'importe quoi est plus petit ou égal à ∞. Nous supposons donc que $\mu^*(X_n) < \infty$ pour tout n.

Soit $\epsilon > 0$. Pour tout $n \geqslant 1$ il existe une suite $(B_k^{(n)})_{k\in\mathbb{N}}$ dans $\mathcal{D}$ tells que $X_n \subset \bigcup_k B_k^{(n)}$ et

$$\mu^*(X_n) + \frac{\epsilon}{2^n} \geqslant \sum_k \mu(B_k^{(n)}). \tag{13.29}$$

Étant donné que

$$\bigcup_n X_n \subset \bigcup_n (\bigcup_k B_k^{(n)}), \tag{13.30}$$

2. Définition 13.13.

nous avons

$$\mu^*\big(\bigcup_n X_n\big) \leqslant \sum_n \sum_k \mu^*(B_k^{(n)}) \leqslant \sum_n \Big(\mu^*(X_n) + \frac{\epsilon}{2^n}\Big) = \sum_n \mu^*(X_n) + \epsilon. \tag{13.31}$$

Cette inégalité étant valable pour tout ϵ, nous avons bien

$$\mu^*\big(\bigcup_n X_n\big) = \sum_n \mu^*(X_n). \tag{13.32}$$

Restriction Soit $A \in \mathcal{D}$. Nous avons automatiquement $\mu^*(A) \leqslant \mu(A)$ parce que $\mu(A)$ est dans l'ensemble dont nous prenons l'infimum (prendre $A_1 = A$ et $A_{n\geqslant 2} = \varnothing$).

En ce qui concerne l'inégalité inverse nous considérons une suite A_n dans $\mathcal{D}$ telle que $A \subset \bigcup_n A_n$. Vu que $A \in \mathcal{D}$ et que $\mathcal{D}$ est une algèbre de parties nous avons $A \cap A_n \in \mathcal{D}$ et $\bigcup_n (A \cap A_n) = A \in \mathcal{D}$. Par conséquent

$$\mu(A) = \mu\big(\bigcup_n (A \cap A_n)\big) \leqslant \sum_n \mu(A \cap A_n) \leqslant \sum_n \mu(A_n). \tag{13.33}$$

Donc tous les éléments de l'ensemble sur lequel nous prenons l'infimum sont plus grands que $\mu(A)$. Nous en déduisons que $\mu^*(A) \geqslant \mu(A)$.

□

13.2.3 Mesure sur une tribu, espace mesuré

La définition suivante est un simple recopiage de la définition générale 13.14 d'une mesure sur un ensemble de parties. La seule différence est que dans le cas d'une tribu, l'union est automatiquement dans la tribu et la condition (2) ne demande pas de le préciser.

Définition 13.20 (Espace mesuré[274]).
*Une **mesure positive** sur l'espace mesurable*[3] $(\Omega, \mathcal{A})$ *est une application* $\mu\colon \mathcal{A} \to [0, \infty]$ *telle que*

(1) $\mu(\varnothing) = 0$,

(2) $\mu\big(\bigcup_{i=0}^{\infty} A_i\big) = \sum_{i=0}^{\infty} \mu(A_i)$ *si les* A_i *sont des éléments de* $\mathcal{A}$ *deux à deux disjoints.*

Le triple $(\Omega, \mathcal{A}, \mu)$ *est alors un **espace mesuré**.*

Une mesure est σ***-finie*** *s'il existe un recouvrement dénombrable de* Ω *par des ensembles de mesure finie. Si la mesure est* σ*-finie, nous disons que l'espace* $(\Omega, \mathcal{A}, \mu)$ *est un espace mesuré* σ*-fini.*

La mesure μ *sur* Ω *est **finie** si* $\mu(\Omega) < \infty$.

Si $(\Omega, \mathcal{A}, \mu)$ et $(S, \mathcal{F}, \nu)$ sont deux espaces mesurés, alors nous notons

$$(\Omega, \mathcal{A}, \mu) \subset (S, \mathcal{F}, \nu) \tag{13.34}$$

lorsque $\Omega \subset S$, $\mathcal{A} \subset \mathcal{F}$ et pour tout $A \in \mathcal{A}$, $\mu(A) = \nu(A)$.

Définition 13.21 (Ensemble mesurable).
Les éléments de $\mathcal{A}$ *sont les ensembles **mesurables** pour la mesure* μ.

Si la mesure des σ-finie, nous pouvons choisir le recouvrement croissant pour l'inclusion. En effet si $(E_n)_{n\in\mathbb{N}}$ est le recouvrement, il suffit de considérer $F_n = \bigcup_{k\leqslant n} E_k$. Ces ensembles F_n forment tout autant un recouvrement dénombrable, mais il est évidemment croissant.

Le lemme suivant complète la propriété 13.20(2) lorsque les ensembles ne sont pas disjoints.

Lemme 13.22.
À propos de mesure d'unions et de différences dans un espace mesurable.

3. Les définitions de tribus et d'espaces mesurables sont en 13.1.

(1) Si $A \subset B$ sont deux ensembles μ-mesurables de mesure finie alors

$$\mu(B\backslash A) = \mu(B) - \mu(A) \tag{13.35}$$

(2) Si $A \subset B$ sont deux ensembles μ-mesurables de mesure finie alors

$$\mu(B) \geqslant \mu(A). \tag{13.36}$$

(3) Si (M_n) est une suite d'éléments de $\mathcal{A}$ pas spécialement disjoints, alors

$$\mu\big(\bigcup_k M_k\big) \leqslant \sum_k \mu(M_k). \tag{13.37}$$

Démonstration. Vu que les ensembles $B\backslash A$ et A sont disjoints par la propriété (2) de la définition de mesure nous avons

$$\mu\big((B\backslash A) \cup A\big) = \mu(B\backslash A) + \mu(A) \tag{13.38}$$

et donc

$$\mu(B) = \mu(B\backslash A) + \mu(A) \tag{13.39}$$

comme demandé.

Pour la seconde partie nous considérons la suite disjointe

$$\begin{cases} M'_0 = \varnothing & \text{(13.40a)} \\ M'_k = M_k\backslash M'_{k-1}. & \text{(13.40b)} \end{cases}$$

Nous avons $\bigcup_k M'_k = \bigcup_k M_k$. Le calcul suivant est alors immédiat :

$$\mu\big(\bigcup_k M_k\big) = \mu\big(\bigcup_k M'_k\big) = \sum_k \mu(M'_k) = \sum_k \mu(M_k\backslash M'_{k-1}) \leqslant \sum_k \mu(M_k). \tag{13.41}$$

□

Lemme 13.23 ([1])**.**
Résultats sur les unions croissantes d'ensembles mesurables dans $(S, \mathcal{A}, \mu)$.

(1) Si (A_k) est une suite croissante d'ensembles μ-mesurables dont l'union est mesurable, alors

$$\lim_{n\to\infty} \mu(A_k) = \mu(\bigcup_k A_k). \tag{13.42}$$

(2) Soit K_n, une suite emboîtée d'éléments de $\mathcal{A}$ tels que $K_n \to S$. Si $A \in \mathcal{A}$ alors

$$\lim_{n\to\infty} \mu(A \cap K_n) = \mu(A). \tag{13.43}$$

Démonstration. Pour prouver (1), nous faisons le coup de l'union télescopique, en posant $A_0 = \varnothing$:

$$\bigcup_{k=1}^{\infty} A_k = \bigcup_{k=1}^{\infty} (A_k\backslash A_{k-1}). \tag{13.44}$$

Les ensembles $A_k\backslash A_{k-1}$ sont deux à deux disjoints, donc la propriété (2) de la définition d'une mesure donne

$$\begin{aligned} \mu(\bigcup_{k=1}^{\infty} A_k) &= \mu\left(\bigcup_{k=1}^{\infty} (A_k\backslash A_{k-1})\right) && \text{(13.45a)} \\ &= \sum_{k=1}^{\infty} \mu(A_k\backslash A_{k-1}) && \text{(13.45b)} \\ &= \sum_{k=1}^{\infty} \big(\mu(A_k) - \mu(A_{k-1})\big) && \text{(13.45c)} \\ &= \lim_{k\to\infty} \mu(A_k) - \mu(A_0) && \text{(13.45d)} \\ &= \lim_{k\to\infty} \mu(A_k). && \text{(13.45e)} \end{aligned}$$

où pour obtenir 13.45c, nous avons utilisé le lemme 13.22.

Le point (2) est une application du point (1). □

Exemple 13.24.
L'intégration à la Riemann n'est pas dans la théorie des espaces mesurés. En effet l'ensemble

$$\mathcal{A} = \{A \subset [0,1] \text{ tel que } \mathbb{1}_A \text{ est intégrable au sens de Riemann}\} \tag{13.46}$$

n'est pas une tribu. Par exemple les singletons en font partie tandis que $[0,1] \cap \mathbb{Q}$ n'en fait pas partie malgé que ce soit une union dénombrable de singletons. △

Définition 13.25.
*Si μ est une mesure nous disons qu'une propriété est vraie μ-**presque partout** si elle est fausse seulement sur un ensemble de mesure nulle.*

Par exemple la fonction de Dirichlet est presque partout égale à la fonction 1 (pour la mesure de Lebesgue).

Définition 13.26 (fonction mesurable)**.**
Une application entre espace mesurés

$$f\colon (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}') \tag{13.47}$$

*est **mesurable** si pour tout $B \in \mathcal{A}'$, l'ensemble $f^{-1}(B)$ est dans $\mathcal{A}$.*

Lemme 13.27.
Une union dénombrable d'ensemble de mesure nulle est de mesure nulle.

Démonstration. C'est une conséquence immédiate du point (2) de la définition d'une mesure : si les A_i sont de mesure nulle,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \mu(A_i) = 0 \tag{13.48}$$

□

Définition 13.28.
*Si (A_n) est une suite croissante d'ensembles alors la **limite** est*

$$\lim_n A_n = \bigcup_{i=0}^{\infty} A_i. \tag{13.49}$$

Si la suite est décroissante alors la limite est

$$\lim_n A_n = \bigcap_{i=0}^{\infty} A_i. \tag{13.50}$$

Proposition 13.29 ([275])**.**
Soit μ une mesure sur Ω et (S_n) une suite croissante d'ensembles μ-mesurables de Ω. Nous notons

$$S = \lim_n S_n. \tag{13.51}$$

Alors pour tout ensemble mesurable[4] *$A \subset \Omega$ nous avons*

$$\mu(A \cap S) = \lim_{n\to\infty} \mu(A \cap S_n). \tag{13.52}$$

Note : dans la référence le résultat fonctionne pour tout ensemble A (et non seulement les mesurables) parce que la définition de la mesurabilité est un peu différente.

4. Définition 13.21

Démonstration. L'inégalité $\lim \mu(A \cap S_n) \leqslant \mu(A \cap S)$ est simple à prouver. En effet pour tout n nous avons $A \cap S_n \subset A \cap S$ et donc par le lemme 13.22 nous avons

$$\mu(A \cap S_n) \leqslant \mu(A \cap S). \tag{13.53}$$

En passant à la limite (qui respecte les inégalités) nous avons l'inégalité.

Nous passons à l'inégalité dans l'autre sens. D'abord si $\mu(A \cap S_n) = \infty$ pour un certain n, alors il cela vaut encore ∞ pour tous les n suivants et la limite est ∞ sans problèmes. Donc nous supposons que $\mu(A \cap S_n) < \infty$ pour tout $n \in \mathbb{N}$. De plus, quitte à renommer les indices, nous pouvons supposer que $S_0 = \emptyset$.

Un élément x est dans S si et seulement s'il existe $n \geqslant 0$ tel que $x \in S_{n+1}$. En prenant le plus petit de ces n nous avons $x \neq S_n$ (éventuellement $n = 0$) et donc

$$S = \bigcup_{n=0}^{\infty} \big(S_{n+1} \backslash S_n\big). \tag{13.54}$$

Par conséquent

$$A \cap S = A \cap \bigcup_{n=0}^{\infty} (S_{n+1} \backslash S_n) = \bigcup_{n=0}^{\infty} A \cap (S_{n+1} \backslash S_n) \tag{13.55}$$

Étant donné que les ensembles $A \cap (S_{n+1} \backslash S_n)$ sont disjoints,

$$\begin{aligned}
\mu(A \cap S) &= \sum_{n=0}^{\infty} \mu\big(A \cap (S_{n+1} \backslash S_n)\big) && (13.56a)\\
&= \sum_{n=0}^{\infty} \mu\Big((A \cap S_{n+1}) \backslash (A \cap S_n)\Big) && (13.56b)\\
&= \sum_{n=0}^{\infty} \big[\mu(A \cap S_{n+1}) - \mu(A \cap S_n)\big] && (13.56c)\\
&= \lim_{n \to \infty} \mu(A \cap S_{n+1}) - \underbrace{\mu(A \cap S_0)}_{=0} && (13.56d)\\
&= \lim_{n \to \infty} \mu(A \cap S_n). && (13.56e)
\end{aligned}$$

Dans ce calcul nous avons utilisé plusieurs fois le fait que les S_n et A étaient mesurables (et la propriété de tribu qui dit que $A \cap S_n$ est également mesurable) ainsi que le lemme 13.22. Nous avons aussi utilisé la série télescopique dans $\mathbb{R}$ pour obtenir (13.56d). □

Définition 13.30 (λ-système[276])**.**
Soit E un ensemble. Un ensemble $\mathcal{D}$ de parties de E est un ***λ-système*** *lorsqu'il vérifie les conditions suivantes :*

(1) si $A, B \in \mathcal{D}$ avec $A \subset B$ alors $B \backslash A \in \mathcal{D}$,

(2) si $(A_k)_{k \geqslant 1}$ est une suite croissante d'éléments de $\mathcal{D}$ alors $\bigcup_k A_k \in \mathcal{D}$.

Note : une tribu est un λ-système.

Lemme 13.31 ([276])**.**
Une intersection quelconque de λ-systèmes dans E est un λ-système dans E.

Démonstration. Soient $\{\mathcal{D}_l\}_{l \in L}$ des λ-systèmes indicés par un ensemble L. Si $A, B \in \bigcap_{l \in L} \mathcal{D}_l$ alors $B \backslash A \in \mathcal{D}_l$ pour tout $l \in L$ et donc $A \backslash B \in \bigcap_{l \in L} \mathcal{D}_l$. De la même façon si (A_k) est une suite croissante dans $\bigcap_{l \in L} \mathcal{D}_l$ alors pour tout $l \in L$ nous avons $\bigcup_k A_k \in \mathcal{D}_l$. Donc $\bigcup_k A_k \in \bigcap_l \mathcal{D}_l$. □

Ce lemme est ce qui permet de définir le λ-système **engendré** par une classe $\mathcal{A}$ de parties de E : c'est l'intersection de tous les λ-systèmes de E contenant $\mathcal{A}$.

Lemme 13.32 ([276]).
Soit $\mathcal{C}$ une classe de parties de E (contenant E lui-même) qui soit stable par intersection finie. Alors le λ-système engendré par $\mathcal{C}$ coïncide avec la tribu engendrée par $\mathcal{C}$.

Démonstration. Nous notons $\mathcal{E}$ le λ-système engendré par $\mathcal{C}$ et $\mathcal{F}$ la tribu engendrée par $\mathcal{C}$. Étant donné que $\mathcal{F}$ est un λ-système nous avons $\mathcal{E} \subset \mathcal{F}$. Pour montrer l'inclusion inverse nous allons prouver que $\mathcal{E}$ est une tribu.

D'abord pour $C \in \mathcal{C}$ nous posons

$$\mathcal{G}_C = \{A \subset \mathcal{E} \text{ tel que } A \cap C \in \mathcal{E}\}. \tag{13.57}$$

et pour $F \in \mathcal{E}$,

$$\mathcal{H}_F = \{A \in \mathcal{E} \text{ tel que } A \cap F \in \mathcal{E}\}. \tag{13.58}$$

Nous allons montrer que $\mathcal{G}_C$ et $\mathcal{H}_F$ sont des λ-systèmes et que $\mathcal{G}_C = \mathcal{H}_F = \mathcal{E}$.

Nous commençons par $\mathcal{G}_C$. Si $A, B \in \mathcal{G}_C$ avec $A \subset B$ alors

$$(B \backslash A) \cap C = \underbrace{(B \cap C)}_{\in \mathcal{E}} \backslash \underbrace{(A \cap C)}_{\in \mathcal{E}}. \tag{13.59}$$

Vu que $\mathcal{E}$ est un λ-système et que $(A \cap C) \subset (B \cap C)$ nous avons bien $(B \backslash A) \cap C \in \mathcal{E}$ et donc $B \backslash A \in \mathcal{G}_C$. Soit maintenant (A_k) une suite croissante dans $\mathcal{G}_C$. Nous avons

$$\Big(\bigcup_{k=1}^{\infty} A_k\Big) \cap C = \bigcup_{k=1}^{\infty} (A_k \cap C) \tag{13.60}$$

qui est une union d'une suite croissante d'éléments de $\mathcal{E}$. Donc $\bigcup_{k=1}^{\infty} (A_k \cap C) \in \mathcal{E}$, ce qui signifie que $\bigcup_{k=1}^{\infty} A_k \in \mathcal{G}_C$. Cela termine la preuve du fait que $\mathcal{G}_C$ soit une λ-système.

Étant donné que $\mathcal{C}$ est stable par intersection finie, si $K \in \mathcal{C}$ nous avons $C \cap K \in \mathcal{C}$, ce qui signifie que $K \in \mathcal{G}_C$. Nous avons donc $\mathcal{C} \subset \mathcal{G}_C$. Donc $\mathcal{G}_C$ est un λ-système vérifiant $\mathcal{C} \subset \mathcal{G}_C \subset \mathcal{E}$. Mais comme $\mathcal{E}$ est le plus petit λ-système contenant $\mathcal{C}$ nous avons en fait $\mathcal{G}_C = \mathcal{E}$.

Nous montrons à présent que $\mathcal{H}_F$ est un λ-système. Si $A, B \in \mathcal{H}_F$ avec $A \subset B$ alors $(B \backslash A) \cap F = (B \cap F) \backslash (A \cap F)$. Vu que $\mathcal{E}$ est une λ-système et que $A \cap F$ et $B \cap F$ sont dans $\mathcal{E}$ avec $A \cap F \subset B \cap F$, nous avons

$$(B \cap F) \backslash (A \cap F) \in \mathcal{H}_F. \tag{13.61}$$

Soit maintenant $(A_k)_{k \geqslant 1}$ une suite croissante dans $\mathcal{H}_F$. Pour tout k nous avons $A_k \cap F \in \mathcal{E}$, ce qui donne

$$\Big(\bigcap_{k=1}^{\infty} A_k\Big) \cap F = \bigcap_{k=1}^{\infty} (A_k \cap F) \in \mathcal{E}. \tag{13.62}$$

Donc $\mathcal{H}_F$ est un λ-système vérifiant $\mathcal{C} \subset \mathcal{H}_F \subset \mathcal{E}$. Nous en concluons que pour tout $C \in \mathcal{C}$ et pour tout $F \in \mathcal{E}$,

$$\mathcal{G}_C = \mathcal{H}_F = \mathcal{E}. \tag{13.63}$$

Nous allons maintenant prouver que $\mathcal{E}$ est une tribu [5].

(1) Si $F \in \mathcal{E}$ alors $E \cap F = F \in \mathcal{E}$, ce qui signifie que $E \in \mathcal{H}_F = \mathcal{E}$.

(2) Si $A \in \mathcal{E}$ alors $E \backslash A \in \mathcal{E}$ parce que $\mathcal{E}$ est un λ-système et $E \in \mathcal{E}$. Donc $\complement A \in \mathcal{E}$.

(3) Montrons que $\mathcal{E}$ est stable par union finie en considérant $A, B \in \mathcal{E}$. Vu que E est également un élément de $\mathcal{E}$ nous avons

$$E \backslash (A \cup B) = (E \backslash A) \cap (E \backslash B) \in \mathcal{E}. \tag{13.64}$$

Cela prouve que $\complement(A \cup B) \in \mathcal{E}$. Par complémentarité nous avons aussi $A \cup B \in \mathcal{E}$.

Soient $A_k \in \mathcal{E}$, et nommons $B_p = A_1 \cup \ldots \cup A_p$. Les ensembles B_p forment une suite croissante d'éléments de $\mathcal{E}$. L'union est donc dans $\mathcal{E}$ et ce dernier est au final stable par union dénombrable.

5. Définition 13.1.

Maintenant que $\mathcal{E}$ est une tribu nous avons $\mathcal{F} \subset \mathcal{E}$ parce que $\mathcal{F}$ est la plus petite tribu contenant $\mathcal{C}$. Nous en déduisons que $\mathcal{E} = \mathcal{F}$, ce qu'il fallait démontrer. □

Le théorème suivant permet de prouver que la mesure de Lebesgue est l'unique mesure possible ayant les bonnes valeurs sur les intervalles (théorème 13.126).

Théorème 13.33 (Unicité des mesures[276])**.**
Soient μ et ν, deux mesures sur $(E, \mathcal{A})$ et une classe $\mathcal{E}$ de parties de E telles que

(1) La tribu engendrée par $\mathcal{E}$ soit $\mathcal{A}$.

(2) si $A, B \in \mathcal{E}$ alors $A \cap B \in \mathcal{E}$

(3) il existe une suite croissante (E_n) dans $\mathcal{E}$ telle que $E = \lim E_n$.

Alors si les mesures μ et ν coïncident sur $\mathcal{E}$, elles coïncident sur $\mathcal{A}$ en entier.

Démonstration. Soit (E_n) la suite des hypothèses ; nous considérons μ_n et ν_n, les restrictions de μ et ν à E_n, c'est-à-dire

$$\mu_n(A) = \mu(A \cap E_n) \tag{13.65a}$$
$$\nu_n(A) = \nu(A \cap E_n). \tag{13.65b}$$

Vu que les E_n sont dans $\mathcal{E} \subset \mathcal{A}$ ils sont mesurables au sens de μ et ν. Par la proposition 13.29, pour tout $A \in \mathcal{E}$ nous avons alors

$$\lim_{n\to\infty} \mu_n(A) = \mu(A) \tag{13.66a}$$
$$\lim_{n\to\infty} \nu_n(A) = \nu(A) \tag{13.66b}$$

Nous devons donc seulement montrer que pour tout $A \in \mathcal{A}$ et pour tout $n \in \mathbb{N}$, $\mu_n(A) = \nu_n(A)$. Pour cela nous nous fixons un n et nous considérons la classe

$$\mathcal{D} = \{A \in \mathcal{A} \text{ tel que } \mu_n(A) = \nu_n(A)\}. \tag{13.67}$$

Le but sera de prouver que $\mathcal{D} = \mathcal{A}$.

Par hypothèse $A \cap E_n \in \mathcal{E}$ et donc

$$\mu(A \cap E_n) = \nu(A \cap E_n) < \infty, \tag{13.68}$$

c'est-à-dire que $\mu_n = \nu_n$ sur $\mathcal{E}$. Par ailleurs, $E \cap E_n = E_n \in \mathcal{E}$, donc

$$\mu_n(E) = \nu_n(E) < \infty. \tag{13.69}$$

Par conséquent $\mu_n = \nu_n$ sur la classe $\mathcal{E}' = \mathcal{E} \cup \{E\}$: $\mathcal{E}' \subset \mathcal{D}$.

Montrons que $\mathcal{D}$ est un λ-système. Soient $A, B \in \mathcal{D}$ avec $A \subset B$. Alors, étant donné que les mesures μ_n et ν_n sont finies, le lemme 13.22 nous donne

$$\mu_n(B\backslash A) = \mu_n(B) - \mu_n(A) \tag{13.70a}$$
$$\nu_n(B\backslash A) = \nu_n(B) - \nu_n(A). \tag{13.70b}$$

Donc $\mu_n(B\backslash A) = \nu_n(B\backslash A)$ et $B\backslash A \in \mathcal{D}$.

Soit par ailleurs une suite croissante $(A_k)_{k\geqslant 1}$ d'éléments de $\mathcal{D}$. En posant $B_p = \bigcup_{k=1}^{p} A_k$, le lemme 13.23(1) nous donne

$$\mu_n(\bigcup_{k=1}^{\infty} A_k) = \lim_{p\to\infty} \mu_n(A_p). \tag{13.71}$$

Mais vu que pour chaque p nous avons $\mu_n(A_p) = \nu_n(A_p)$, nous avons aussi

$$\mu_n(\bigcup_{p=1}^{\infty} A_p) = \nu_n(\bigcup_{p=1}^{\infty} A_p). \tag{13.72}$$

Donc $\mathcal{D}$ est bel et bien un λ-système contenant $\mathcal{E}'$. Par le lemme 13.32, le λ-système engendré par $\mathcal{E}'$ est égal à la tribu engendrée par $\mathcal{E}'$, mais par hypothèse la tribu engendrée par $\mathcal{E}$ est $\mathcal{A}$, donc le λ-système engendré par $\mathcal{E}'$ est $\mathcal{A}$. Vu que $\mathcal{D}$ est une λ-système contenant $\mathcal{E}'$, nous avons alors $\mathcal{A} \subset \mathcal{D}$ et donc $\mathcal{A} = \mathcal{D}$, ce qu'il fallait. □

Exemple 13.34.
La partie $\mathcal{E}$ des intervalles de $\mathbb{R}$ de la forme $]a, b[$ engendre les boréliens par la proposition 7.97. Par conséquent pour vérifier que deux mesures sont égales sur les boréliens de $\mathbb{R}$ il suffit de prouver qu'elles sont égales sur les intervalles ouverts. △

13.2.4 Mesure extérieure

Nous avons déjà défini la notion de mesure extérieure en la définition 13.13.

Lemme 13.35 ([273])**.**
Soit $(S, \mathcal{F}, \mu)$ un espace mesuré et $X \subset S$. Alors

$$\inf_{\substack{A \in \mathcal{F} \\ X \in A}} \mu(A) = \inf\{\sum_n \mu(A_n) \text{ tel que } A_n \in \mathcal{F}, X \subset \bigcup_k A_k\}. \tag{13.73}$$

Démonstration. Pour montrer l'inégalité $\geqslant$, nous remarquons qu'il y a plus d'éléments dans l'ensemble du second membre que dans le premier. En effet si $A \in \mathcal{F}$ avec $X \subset A$ alors dans le membre de gauche nous pouvons prendre $A_1 = A$ et $A_{n \geqslant 1} = \varnothing$.

Pour l'inégalité dans l'autre sens, nous montrons que tout élément de

$$\{\sum_n \mu(A_n) \text{ tel que } A_n \in \mathcal{F}, X \subset \bigcup_k A_k\} \tag{13.74}$$

est plus grand qu'un élément de

$$\{\mu(A) \text{ tel que } A \in \mathcal{F}, X \subset A\}. \tag{13.75}$$

En effet si $A_n \in \mathcal{F}$ avec $X \subset \bigcup_k A_k$ alors en posant $A = \bigcup_k A_k$ nous avons $A \in \mathcal{F}$ avec $X \subset A$ ainsi que $\mu(A) \leqslant \sum_n \mu(A_n)$. Cela prouve que l'élément $\sum_n \mu(A_n)$ de (13.74) est plus grand que l'élément $\mu(A)$ de (13.75). □

Proposition 13.36 ([273])**.**
Soit un espace mesuré $(S, \mathcal{F}, \mu)$ et l'application

$$\begin{aligned} \mu^* \colon \mathcal{P}(S) &\to [0, \infty] \\ X &\mapsto \inf\{\mu(A) \text{ tel que } A \in \mathcal{F}, X \subset A\}. \end{aligned} \tag{13.76}$$

Alors μ^* est une mesure extérieure sur S et sa restriction à $\mathcal{F}$ est égale à μ.

Cela est un cas particulier de 13.19 en utilisant 13.35. Nous en donnons cependant une preuve directe, qui est presque identique à celle de 13.19, mais avec une ou deux simplifications.

Démonstration. Notons que la définition est bonne parce que l'ensemble sur lequel l'infimum est pris n'est pas vide : prendre $A = S$.

Le vide D'abord $\mu^*(\varnothing) = O$ parce que $\varnothing \in \mathcal{F}$.

Inégalité d'inclusion Soient $X \subset Y$ dans $\mathcal{P}(S)$. Si $Y \subset A$ alors $X \subset A$, donc

$$\inf\{\mu(A) \text{ tel que } A \in \mathcal{F}, X \subset A\} \leqslant \inf\{\mu(A) \text{ tel que } A \in \mathcal{F}, Y \subset A\}, \tag{13.77}$$

ce qui signifie que $\mu^*(X) \leqslant \mu^*(Y)$.

Inégalité par union dénombrable Soit $(X_n)_{n\in\mathbb{N}}$ une suite de parties de S. S'il existe n_0 tel que $\mu^*(X_{n_0}) = \infty$ alors nous avons automatiquement $\sum_n \mu^*(X_n) = \infty$ et l'inégalité demandée est évidente parce que n'importe quoi est plus petit ou égal à ∞. Nous supposons donc que $\mu^*(X_n) < \infty$ pour tout n.

Soit $\epsilon >$ et par définition pour chaque n, il existe un $A_n \in \mathcal{F}$ tel que $X_n \subset A_n$ et $\mu(A_n) \leqslant \mu^*(X_n) + \frac{\epsilon}{2^n}$. Bien entendu nous avons

$$\bigcup_n X_n \subset \bigcup_n A_n \in \mathcal{F}. \tag{13.78}$$

Nous en déduisons que

$$\mu^*(\bigcup_n X_n) \leqslant \mu(\bigcup_n A_n). \tag{13.79}$$

Mais $(S, \mathcal{F}, \mu)$ étant un espace mesuré,

$$\mu(\bigcup_n A_n) \leqslant \sum_n \mu(A_n). \tag{13.80}$$

Au final nous avons les inégalités

$$\mu^*(\bigcup_n X_n) \leqslant \mu(\bigcup_n A_n) \leqslant \sum_n \mu(A_n) \leqslant \sum_n \mu^*(X_n) + \epsilon \sum_n \frac{1}{2^n} = \sum_n \mu^*(X_n) + \epsilon. \tag{13.81}$$

Cela étant vrai pour tout ϵ,

$$\mu^*(\bigcup_n X_n) \leqslant \sum_n \mu^*(X_n), \tag{13.82}$$

ce qui prouve que μ^* est une mesure extérieure.

Restriction Supposons que $X \in \mathcal{F}$. Alors si $X \subset A$ nous avons $\mu(X) \leqslant \mu(A)$; mais en même temps, $\mu(X)$ est dans l'infimum qui définit $\mu^*(X)$ donc

$$\mu^*(X) \leqslant \mu(X) \leqslant \inf\{\mu(A) \text{ tel que } A \in \mathcal{F}, X \subset A\} \leqslant \mu(X) \leqslant \mu^*(X). \tag{13.83}$$

Donc nous avons égalité de tous les éléments de cette chaine d'inégalité.

□

Définition 13.37.
*Soit S un ensemble et m^* une mesure extérieure sur S. Une partie $A \subset X$ est m^*-**mesurable** si pour tout $X \subset S$,*

$$m^*(X) = m^*(X \cap A) + m^*(X \cap A^c). \tag{13.84}$$

Remarque 13.38.
L'inégalité

$$m^*(X) \leqslant m^*(X \cap A) + m^*(X \cap A^c) \tag{13.85}$$

étant toujours vraie, pour prouver qu'un ensemble est m^*-mesurable, il est suffisant de prouver l'inégalité inverse :

$$m^*(X) \geqslant m^*(X \cap A) + m^*(X \cap A^c) \tag{13.86}$$

La définition 13.37 est motivée par la proposition suivante.

Proposition 13.39.
Soit un espace mesuré $(S, \mathcal{F}, \mu)$ et μ^* la mesure extérieure qui va avec. Alors tous les éléments de $\mathcal{F}$ sont μ^*-mesurables.

En d'autres termes, pour tout $A \in \mathcal{F}$ et tout $X \subset S$ nous avons

$$\mu^*(X) = \mu^*(X \cap A) + \mu^*(X \cap A^c). \tag{13.87}$$

Démonstration. Vu que $X = (X \cap A) \cup (X \cap A^c)$, et que μ^* est une mesure extérieure,

$$\mu^*(X) \leqslant \mu^*(X \cap A) + \mu^*(X \cap A^c). \tag{13.88}$$

Nous devons montrer l'inégalité inverse.

Soit $B \in \mathcal{F}$ tel que $X \subset B$. D'une part nous avons $X \cap A \subset B \cap A \in \mathcal{F}$, donc

$$\mu^*(X \cap A) \leqslant \mu^*(B \cap A) = \mu(B \cap A). \tag{13.89}$$

Et d'autre part, $X \cap A^c \subset B \cap A^c \in \mathcal{F}$, donc

$$\mu^*(X \cap A^c) \leqslant \mu(B \cap A^c). \tag{13.90}$$

En remettant ensemble,

$$\mu^*(X \cap A) + \mu^*(X \cap A^c) \leqslant \mu(B \cap A) + \mu(B \cap A^c) = \mu(B). \tag{13.91}$$

La dernière égalité est le fait que $B \cap A$ et $B \cap A^c$ sont disjoints et que μ est une mesure. L'inégalité (13.91) étant vraie pour tout $B \in \mathcal{F}$ tel que $X \subset B$, elle est encore vraie pour l'infimum :

$$\mu^*(X \cap A) + \mu^*(X \cap A^c) \leqslant \inf\{\mu(B) \text{ tel que } B \in \mathcal{F}, X \subset B\} = \mu^*(X). \tag{13.92}$$

Nous avons donc prouvé que

$$\mu^*(X \cap A) + \mu^*(X \cap A^c) \leqslant \mu^*(X). \tag{13.93}$$

□

Remarque 13.40.
Notons la duplicité du vocabulaire. Les ensembles μ-mesurables sont les éléments de $\mathcal{F}$, qui sont à priori les seuls sur lesquels μ est calculable[6], alors que les μ^*-mesurables sont les parties de S qui vérifient une certaine propriété (et μ^* est calculable sur toutes les parties de S).

13.3 Applications mesurables

13.3.1 Propriétés

Définition 13.41 (Fonction mesurable)**.**
Soient $(E, \mathcal{A})$ et $(F, \mathcal{F})$ deux espaces mesurés. Une fonction $f\colon E \to F$ est ***mesurable*** *si pour tout $\mathcal{O} \in \mathcal{F}$, l'ensemble $f^{-1}(\mathcal{O})$ est dans $\mathcal{A}$.*

Définition 13.42 (Fonction borélienne)**.**
Une application $f\colon (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d))$ est ***borélienne*** *si elle est mesurable, c'est-à-dire si pour tout $B \in \mathcal{B}or(\mathbb{R}^d)$ nous avons $f^{-1}(B) \in \mathcal{A}$.*

Si rien n'est précisé, une application entre deux espaces topologiques est borélienne lorsqu'elle est mesurable en considérant la tribu borélienne sur les deux *espace.*

Si $\mathcal{A}$ est une tribu sur un ensemble E, nous notons $m(\mathcal{A})$ l'ensemble des fonctions qui sont $\mathcal{A}$-mesurables.

Le plus souvent lorsque nous parlerons de fonctions $f\colon X \to Y$ où Y est un espace topologique, nous considérons la tribu borélienne sur Y. Ce sera en particulier le cas dans la théorie de l'intégration.

Proposition 13.43.
Soient $(S_i, \mathcal{F}_i)$ $(i = 1, 2, 3)$ des espaces mesurables et des fonctions mesurables $f\colon S_1 \to S_2$ et $g\colon S_2 \to S_3$. Alors la fonction $g \circ f\colon S_1 \to S_3$ est mesurable.

Démonstration. Soit $B \in \mathcal{F}_3$. Alors

$$(g \circ f)^{-1}(B) = f^{-1}\big(g^{-1}(B)\big) \in f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1. \tag{13.94}$$

□

6. « calculable » au sens où μ y vaut un nombre bien définit ; après, que ce soit facile ou pas à calculer dans la pratique, c'est une autre histoire.

13.3.2 D'une tribu à l'autre

Lemme 13.44 ([277]).
Soit une application $f\colon S_1 \to S_2$ et une tribu $\mathcal{F}_2$ sur S_2. Alors $f^{-1}(\mathcal{F}_2)$ est une tribu sur S_1

Démonstration. Il faut prouver les trois propriétés de la définition 13.1 d'une tribu.

(1) D'abord f est définit sur tout S_1, donc $f^{-1}(S_2) = S_1$ alors que $S_2 \in \mathcal{F}_2$.

(2) Soit $A \in f^{-1}(\mathcal{F}_2)$, c'est-à-dire $A = f^{-1}(B)$ pour un certain $B \in \mathcal{F}_2$. En ce qui concerne le complémentaire :

$$A^c = f^{-1}(B)^c = S_1 \backslash f^{-1}(B) = f^{-1}(S_2 \backslash B) = f^{-1}(B^c). \tag{13.95}$$

(3) Si $(A_i)_{i\in\mathbb{N}}$ sont des éléments de $f^{-1}(\mathcal{F}_2)$ avec $A_i = f^{-1}(B_i)$ alors

$$\bigcup_i A_i = \bigcup_i f^{-1}(B_i) = f^{-1}\big(\bigcup_i B_i\big). \tag{13.96}$$

Ce qui est dans la dernière parenthèse est dans $\mathcal{F}_2$ parce que cette dernière est une tribu.

□

Lemme 13.45 ([277]).
Soit une application $f\colon S_1 \to S_2$ et $\mathcal{F}$ une tribu de S_1. Alors

(1) L'ensemble

$$\mathcal{F}_f = \{B \subset S_2 \text{ tel que } f^{-1}(B) \in \mathcal{F}\} \tag{13.97}$$

est une tribu sur S_2.

(2) C'est la plus grande tribu de S_2 pour laquelle f est mesurable.

Démonstration. Encore les trois propriétés à vérifier.

(1) $S_2 \in \mathcal{F}$, sont $S_1 = f^{-1}(S_2) \in \mathcal{F}_f$.

(2) Si $A \in \mathcal{F}_f$ alors $A = f^{-1}(B)$ pour un certain $B \in \mathcal{F}$. Nous avons alors aussi $B^c \in \mathcal{F}$ et donc

$$f^{-1}(B^c) = f^{-1}(B)^c = A^c. \tag{13.98}$$

Par conséquent A^c est dans $\mathcal{F}_f$.

(3) Si (A_i) sont des éléments de $\mathcal{F}_f$ avec $A_i = f^{-1}(B_i)$ pour $B_i \in \mathcal{F}$ alors $\bigcup_i B_i \in \mathcal{F}$ et

$$f^{-1}\big(\bigcup_i B_i\big) = \bigcup_i f^{-1}(B_i) \in \mathcal{F}_f. \tag{13.99}$$

En ce qui concerne la maximalité, si $R \subset S_2$ n'est pas dans $\mathcal{F}_f$ alors $f^{-1}(R)$ n'est pas dans $\mathcal{F}$ et donc f ne serait pas mesurable. □

Définition 13.46 (Tribu engendrée).
Soit une application $f\colon S_1 \to S_2$ et $\mathcal{F}$ une tribu de S_1. Alors conformément au lemme 13.45 l'ensemble

$$\mathcal{F}_f = \{B \subset S_2 \text{ tel que } f^{-1}(B) \in \mathcal{F}\} \tag{13.100}$$

*est la **tribu engendrée**.*

Le lemme suivant est également nommé « lemme de transfert ».

Lemme 13.47 (Lemme de transport).
Soit $f\colon S_1 \to S_2$ une application et une classe $\mathcal{C}$ de parties de S_2. Alors

$$\sigma\big(f^{-1}(\mathcal{C})\big) = f^{-1}\big(\sigma(\mathcal{C})\big). \tag{13.101}$$

Démonstration. Vu que $\sigma(\mathcal{C})$ es tune tribu, dans S_2 alors le lemme 13.45 dit que $f^{-1}\big(\sigma(\mathcal{C})\big)$ est une tribu qui contient en particulier $f^{-1}(\mathcal{C})$. Nous en déduisons que $\sigma\big(f^{-1}(\mathcal{C})\big) \subset f^{-1}\big(\sigma(\mathcal{C})\big)$.

Réciproquement. Dans S_1 nous avons la tribu $\sigma\big(f^{-1}(\mathcal{C})\big)$. Nous pouvons alors considérer la tribu

$$\mathcal{F}_f = \{B \subset S_2 \text{ tel que } f^{-1}(B) \in \sigma\big(f^{-1}(\mathcal{C})\big)\}. \tag{13.102}$$

Montrons que $\mathcal{C} \subset \mathcal{F}_f$. Lorsque $B \in \mathcal{C}$ nous avons $f^{-1}(B) \in f^{-1}(\mathcal{C}) \subset \sigma\big(f^{-1}(\mathcal{C})\big)$. Du coup $B \in \mathcal{F}_f$. Nous avons alors, en passant aux tribus engendrées :

$$\sigma(\mathcal{C}) \subset \sigma(\mathcal{F}_f) = \mathcal{F}_f. \tag{13.103}$$

Si maintenant $B \in \sigma(\mathcal{C})$, nous avons $f^{-1}(B) \in \sigma\big(f^{-1}(\mathcal{C})\big)$, ce qui signifie que

$$f^{-1}\big(\sigma(\mathcal{C})\big) \subset \sigma\big(f^{-1}(\mathcal{C})\big). \tag{13.104}$$

□

Le théorème suivant est important pour prouver qu'une application est mesurable. En effet, il permet de ne tester si une application est mesurable uniquement sur une partie génératrice de la tribu d'arrivé[7].

Théorème 13.48.
Soient des espaces mesurables $(S_1, \mathcal{F}_1)$ et $(S_2, \mathcal{F}_2)$ ainsi qu'une application $f\colon S_1 \to S_2$. S'il existe un ensemble de parties $\mathcal{C}$ de S_2 tel que

- $\sigma(\mathcal{C}) = \mathcal{F}_2$
- $f^{-1}(B) \in \mathcal{F}_1$ *pour tout* $B \in \mathcal{C}$

alors f est mesurable.

Démonstration. Par hypothèse, $\sigma(\mathcal{C}) = \mathcal{F}_2$ et $f^{-1}(\mathcal{C}) \subset \mathcal{F}_1$ et nous pouvons utiliser le lemme de transfert 13.47 :

$$\sigma\big(f^{-1}(\mathcal{C})\big) = f^{-1}\big(\sigma(\mathcal{C})\big) \tag{13.105}$$

qui s'écrit ici

$$\sigma\big(f^{-1}(\mathcal{C})\big) = f^{-1}(\mathcal{F}_2). \tag{13.106}$$

Mais vu que $f^{-1}(\mathcal{C}) \subset \mathcal{F}_1$, nous avons aussi $\sigma\big(f^{-1}(\mathcal{C})\big) \subset \mathcal{F}_1$, ce qui signifie que

$$f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1. \tag{13.107}$$

Cela est exactement le fait que f soit mesurable. □

Le théorème suivant est très important parce qu'en pratique c'est souvent lui, en conjonction avec la proposition 13.110 qui permet de déduire qu'une fonction est borélienne.

Théorème 13.49 ([277])**.**
Soient X et Y deux espaces topologiques. Alors toute application continue $f\colon X \to Y$ est borélienne[8].

Démonstration. Pour vérifier que f est borélienne, nous devons prouver que $f^{-1}(B)$ est borélien pour tout borélien B de Y. Heureusement, le théorème 13.48 nous permet de limiter la vérification aux B appartenant à une classe engendrant les boréliens de Y.

La classe en question est toute trouvée : ce sont les ouverts. Si $\mathcal{O}$ est un ouvert de Y alors $f^{-1}(\mathcal{O})$ est un ouvert de X et donc un borélien de X. □

Le théorème suivant donne une importante compatibilité entre l'induction de tribu et l'induction de topologie : la tribu induite à partir des boréliens sur un sous-espace topologique est la tribu des boréliens pour la topologie induite.

7. Typiquement les ouverts pour les boréliens.
8. Définition 13.42.

Théorème 13.50 ([277]).
Soit X, un espace topologique et $Y \subset X$ une partie munie de la topologie induite. Alors

$$\mathcal{B}or(Y) = \mathcal{B}or(X)_Y \tag{13.108}$$

où $\mathcal{B}or(X)_Y$ est la tribu sur Y induite de $\mathcal{B}or(X)$ par la définition 13.6.

Démonstration. Nous notons τ_X et τ_Y les topologies de X et Y.

$\mathcal{B}$or$(Y) \subset \mathcal{B}$or$(X)_Y$ Si $A \in \tau_Y$ alors $A = Y \cap \Omega$ pour un $\Omega \in \tau_X$. Mais vu que Ω est un ouvert de X, il est un borélien de X, ce qui donne que $Y \cap \Omega$ est un élément de $\mathcal{B}\text{or}(X)_Y$. Cela prouve que $\tau_Y \subset \mathcal{B}\text{or}(X)_Y$, c'est-à-dire que $\mathcal{B}\text{or}(X)_Y$ est une tribu sur Y contenant les ouverts de Y. Nous avons donc

$$\mathcal{B}\text{or}(X) \subset \mathcal{B}\text{or}(X)_Y. \tag{13.109}$$

Réciproquement L'application $\text{Id}\colon (Y, \tau_Y) \to (X, \tau_X)$ est continue parce que si Ω est ouvert de X alors $\text{Id}^{-1}(\Omega) = \Omega \cap Y \in \tau_Y$. Par conséquent l'identité est une application borélienne (théorème 13.49), ce qui signifie que $\text{Id}^{-1}\big(\mathcal{B}\text{or}(X)\big) \subset \mathcal{B}\text{or}(Y)$, ou encore que si $B \in \mathcal{B}\text{or}(X)$, alors $\text{Id}^{-1}(B) = B \cap Y \in \mathcal{B}\text{or}(Y)$. Cela signifie que

$$\mathcal{B}\text{or}(X)_Y \subset \mathcal{B}\text{or}(Y). \tag{13.110}$$

□

Corolaire 13.51.
Si U est un borélien de l'espace topologique X, alors les boréliens de U sont les boréliens de X inclus dans U :

$$\mathcal{B}or(U) = \{B \in \mathcal{B}or(X) \text{ tel que } B \subset U\}. \tag{13.111}$$

Démonstration. Si $B' \in \mathcal{B}\text{or}(U)$, le théorème 13.50 donne un borélien $B \in \mathcal{B}\text{or}(X)$ tel que $B' = B \cap U$. Mais U étant borélien de X, l'intersection $B \cap U$ est encore un borélien de X. □

Ce corolaire s'applique en particulier lorsque U est un ouvert.

La proposition suivante montre comment il est possible de construire un espace mesuré à partir d'une bijection avec un espace mesuré déjà connu. Attention cependant : la mesure construite dans cette proposition n'est pas celle qui est le plus adapté. Voir la proposition 13.257 et l'exemple 13.257.

Proposition 13.52.
Soient un espace mesuré $(\Omega, \mathcal{A}, \mu)$, un ensemble Ω' et une bijection $\varphi\colon \Omega \to \Omega'$. Nous posons

(1) $\mathcal{A}' = \varphi(\mathcal{A})$,

(2) $\mu'(B) = \mu\big(\varphi^{-1}(B)\big)$ *pour tout* $B \in \mathcal{A}'$.

Alors $(\Omega', \mathcal{A}', \mu')$ est un espace mesuré.

Démonstration. En plusieurs points.

$\mathcal{A}'$ est une tribu Il faut vérifier les différents points de la définition 13.1. D'abord, vu que $\Omega \in \mathcal{A}$, nous avons $\Omega' = \varphi(\Omega) \in \mathcal{A}'$. Pour le complémentaire, si $B \in \mathcal{A}'$ alors $B = \varphi(A)$ pour un certain $A \in \mathcal{A}$. Vu que $\mathcal{A}$ est une tribu nous avons alors $\Omega \backslash A \in \mathcal{A}$ et donc $\varphi(\Omega \backslash A) \in \mathcal{A}'$. Mais comme φ est bijective,

$$\varphi(\Omega \backslash A) = \Omega' \backslash \varphi(A) = \Omega' \backslash B. \tag{13.112}$$

Le complémentaire de B est donc bien dans $\mathcal{A}'$. Pour la troisième condition soient $B_i \in \mathcal{A}'$. Pour chauqe i, il existe $A_i \in \mathcal{A}$ tel que $B_i = \varphi(A_i)$. Nous avons $\bigcup_i A_i \in \mathcal{A}$, donc

$$\bigcup_i B_i = \bigcup_i \varphi(A_i) = \varphi\big(\bigcup_i A_i\big) \in \mathcal{A}'. \tag{13.113}$$

Nous avons fini de prouver que $(\Omega', \mathcal{A}')$ était un espace mesurable.

μ' est une mesure positive D'abord $\mu'(\emptyset) = \mu\big(\varphi^{-1}(\emptyset)\big) = \mu(\emptyset) = 0$. Ensuite si les A_i sont disjoints dans $\mathcal{A}'$ nous avons

$$\mu\big(\bigcup_{i=0}^{\infty} A_i\big) = \mu\left(\varphi^{-1}\big(\bigcup_{i=0}^{\infty} A_i\big)\right) = \mu\left(\bigcup_i \varphi^{-1}(A_i)\right) = \sum_{i=0}^{\infty} \mu\big(\varphi^{-1}(A_i)\big) = \sum_i \mu'(A_i). \quad (13.114)$$

□

Proposition 13.53.
Soit une bijection continue d'inverse continue $\varphi\colon \Omega \to \Omega'$. Alors

$$\varphi\big(\mathcal{B}or(\Omega)\big) = \mathcal{B}or(\Omega'). \quad (13.115)$$

Démonstration. Si $A \in \mathcal{B}\mathrm{or}(\Omega')$, alors $A = \varphi\big(\varphi(A)\big) \in \varphi\big(\mathcal{B}\mathrm{or}(\Omega)\big)$ parce que φ est continue et donc borélienne (proposition 13.49). Le même raisonnement fonctionne dans l'autre sens parce que nous avons supposé que φ est continue et d'inverse continu. □

13.4 Espace mesuré complet

13.4.1 Partie négligeable

Définition 13.54.
*Soit un espace mesuré $(X, \mathcal{A}, \mu)$. Une partie N de X est **négligeable** pour μ s'il existe $Y \in \mathcal{A}$ tel que $N \subset Y$ et $\mu(Y) = 0$.*

Lemme 13.55.
L'ensemble des parties négligeables est stable par union dénombrable.

Démonstration. Si les ensembles N_i sont négligeables, alors pour chaque i nous avons $Y_i \in \mathcal{A}$ tel que $N_i \subset Y_i$ et $\mu(Y_i) = 0$. Alors bien entendu $\bigcup_i N_i \subset \bigcup_i Y_i$ et en utilisant (13.37),

$$\mu\big(\bigcup_i Y_i\big) \leqslant \sum_i \mu(Y_i) = 0. \quad (13.116)$$

□

Définition 13.56.
*L'espace mesuré $(X, \mathcal{F}, \mu)$ est **complet** si tout ensemble μ-négligeable est dans $\mathcal{F}$.*

Notons que la proposition 13.5 s'applique si $(X, \mathcal{F}, \mu)$ est un espace mesuré et $\mathcal{N}$ est l'ensemble des parties μ-négligeables. C'est ce qui permet de donner le théorème suivant, que nous redémontrons de façon indépendante de la proposition 13.5.

Théorème 13.57 (Complétion d'espace mesuré[273, 278, 279])**.**
Soit un espace mesuré $(X, \mathcal{F}, \mu)$ et $\mathcal{N}$ l'ensemble des parties μ-négligeables de X.

(1) Les ensembles suivants sont égaux :

$$\mathcal{A} = \{A \subset X \text{ tel que } \exists B, C \in \mathcal{F} \text{ tel que } B \subset A \subset C, \mu(C \backslash B) = 0\} \quad (13.117a)$$
$$\mathcal{B} = \{B \cup N \text{ tel que } B \in \mathcal{F}, N \in \mathcal{N}\} \quad (13.117b)$$
$$\mathcal{C} = \{A \subset X \text{ tel que } \exists B \in \mathcal{F} \text{ tel que } A \Delta B \in \mathcal{N}\}. \quad (13.117c)$$

Ici $A\Delta B$ est la différence symétrique de A et B, définition 1.18.

(2) L'ensemble $\hat{\mathcal{F}} = \mathcal{A} = \mathcal{B} = \mathcal{C}$ est une tribu.

(3) La définition

$$\begin{aligned} \mu'\colon \mathcal{B} &\to [0, \infty] \\ A \cup N &\mapsto \mu(A) \end{aligned} \quad (13.118)$$

est cohérente.

(4) L'application μ' ainsi définie est une mesure sur $(X, \mathcal{A})$.
(5) L'espace $(X, \mathcal{A}, \mu')$ est complet.
(6) La mesure μ' prolonge μ.
(7) La mesure μ' est minimale au sens où toute mesure complète prolongeant μ prolonge μ'.

Démonstration. Commençons par prouver que les trois ensembles $\mathcal{A}$, $\mathcal{B}$ et $\mathcal{C}$ sont égaux.

$\underline{\mathcal{A} \subset \mathcal{B}.}$ Soit $A \in \mathcal{A}$. Alors nous avons des ensembles $B, C \in \mathcal{F}$ tels que $B \subset A \subset V$ avec $\mu(C \backslash B) = 0$. Alors nous avons aussi $A = B \cup (C \backslash B)$, ce qui prouve que $A \in \mathcal{B}$.

$\underline{\mathcal{B} \subset \mathcal{C}.}$ Soit $A \in \mathcal{B}$, c'est-à-dire que $A = B \cup N$ avec $B \in \mathcal{F}$ et $N \in \mathcal{N}$. Nous avons évidemment $A \cup B = A$ et donc

$$A \Delta B = (A \cup B) \backslash (A \cap B) = A \backslash (A \cap B) = (B \cup N) \backslash (A \cap B) \subset N. \tag{13.119}$$

Pour comprendre la dernière inclusion, si x appartient à $A = B \cup N$ sans être dans N alors $x \in B$ et donc $x \in A \cap B$. Par conséquent nous avons $A \Delta B \subset N$ et donc $A \Delta B \in \mathcal{N}$.

$\underline{\mathcal{C} \subset \mathcal{A}}$ Soit donc $A \in \mathcal{C}$; il existe $B \in \mathcal{F}$ tel que $A \Delta B \in \mathcal{N}$ ou encore, il existe $D \in \mathcal{F}$ tel que $A \Delta B \subset D$ avec $\mu(D) = 0$. Si nous posons $B' = B \cap D^c$ et $C' = B \cup D$ alors nous prétendons avoir

$$B' \subset A \subset C'. \tag{13.120}$$

Et nous le prouvons. En effet si $x \in B \cap D^c$ alors en remarquant que B se divise en

$$B = (B \cap A) \cup \big(B \cap (A \Delta B)\big), \tag{13.121}$$

et en nous souvenant que $B \cap (A \Delta B) \subset D$, il vient que $B \cap D^c \subset B \cap A$. Et en particulier $x \in A$. D'autre part

$$A \subset B \cup (A \Delta B) \subset B \cup D. \tag{13.122}$$

Nous avons donc bien $B' \subset A \subset C'$. Par stabilité de la tribu $\mathcal{F}$ sous les intersections et complémentaires nous avons aussi $B', C' \in \mathcal{F}$. De plus

$$C' \backslash B' = (B \cup D) \backslash (B \cap D^c) \subset D, \tag{13.123}$$

et donc

$$\mu(C' \backslash B') \leqslant \mu(D) = 0. \tag{13.124}$$

Nous avons donc prouvé que $\mathcal{A} \subset \mathcal{B} \subset \mathcal{C} \subset \mathcal{A}$, et donc que $\mathcal{A} = \mathcal{B} = \mathcal{C}$. Nous allons donc maintenant noter $\mathcal{A}$ indifféremment les trois ensembles. Nous prouvons à présent que c'est une tribu.

Tribu : le vide Pas de problèmes à $\emptyset \in \mathcal{A}$

Tribu : complémentaire Soit $A \in \mathcal{A}$. Alors il existe $B, C \in \mathcal{F}$ tels que $B \subset A \subset C$ avec $\mu(C \backslash B) = 0$. En passant au complémentaire,

$$C^c \subset A^c \subset B^c. \tag{13.125}$$

Mais $B^c \backslash C^c = C \backslash B$, donc $\mu(B^c \backslash C^c) = 0$.

Tribu : union dénombrable Soit (A_n) des éléments de $\mathcal{A}$. Pour chaque n nous avons des ensembles $B_n, C_n \in \mathcal{F}$ tels que$B_n \subset A_n \subset C_n$ avec $\mu(C_n \backslash B_n) = 0$. En ce qui concerne les unions nous avons

$$\bigcup_n B_n \subset \bigcup_n A_n \subset \bigcup_n C_n, \tag{13.126}$$

et

$$(\bigcup_n C_n) \backslash (\bigcup_n B_n) \subset \bigcup_n (C_n \backslash B_n). \tag{13.127}$$

Par conséquent, en utilisant (13.37),

$$\mu\left((\bigcup_n C_n) \backslash (\bigcup_n B_n)\right) \leqslant \mu\left(\bigcup_n (C_n \backslash B_n)\right) \leqslant \sum_n \mu(C_n \backslash B_n) = 0. \tag{13.128}$$

Cela prouve que $\bigcup_n A_n \in \mathcal{A}$, et donc que $\mathcal{A}$ est une tribu.

Définition cohérente Soient $A, A' \in \mathcal{F}$ et $N, N' \in \mathcal{N}$ tels que $A \cup N = A' \cup N'$. Nous considérons $Y, Y' \in \mathcal{F}$ tel que $N \subset Y$, $N' \subset Y'$ et $\mu(Y) = \mu(Y') = 0$. En vertu de (13.37) nous avons

$$\mu(A) \leqslant \mu(A \cup Y) \leqslant \mu(A' \cup Y \cup Y') \leqslant \mu(A') + \mu(Y) + \mu(Y') = \mu(A'). \tag{13.129}$$

En écrivant la même chose en échangeant les primes nous prouvons également $\mu(A') \leqslant \mu(A)$. Au final $\mu(A) = \mu(A')$, c'est-à-dire

$$\mu'(A \cup N) = \mu'(A' \cup N'). \tag{13.130}$$

La définition de μ' est donc cohérente.

μ' est une mesure Le fait que μ' soit positive et que $\mu'(\emptyset)$ soit nul ne pose pas de problèmes. Il faut voir l'union dénombrable disjointe. Si les ensembles $A_i = B_i \cup N_i$ sont disjoints, alors les B_i et le N_i sont tous disjoints deux à deux. De plus l'ensemble $\bigcup_i N_i$ est négligeable parce que nous avons déjà vu que $\mathcal{N}$ était stable par union dénombrable (13.37). Donc

$$\mu'\Big(\bigcup_i B_i \cup N_i\Big) = \mu'\Big(\big(\bigcup_i B_i\big) \cup \underbrace{\big(\bigcup_i N_i\big)}_{\in\mathcal{N}}\Big) = \mu\big(\bigcup_i B_i\big) = \sum_u \mu(B_i) = \sum_i \mu'(B_i \cup N_i). \tag{13.131}$$

Espace complet Un ensemble μ'-négligeable est automatiquement μ-négligeable. En effet si H est μ'-négligeable, il existe $B \in \mathcal{F}$ et $N \in \mathcal{N}$ tels que $H \subset B \cup N$ avec $\mu(B) = 0$. Vu que N est μ-négligeable, il existe $Y \in \mathcal{F}$ tel que $N \subset Y$ et $\mu(Y) = 0$. Donc $H \subset B \cup N \subset B \cup Y$ avec $\mu(B \cup Y) = 0$.

Tous les ensembles μ-négligeables faisant partie de $\mathcal{B}$, tous les ensembles μ'-négligeables font partie de $\mathcal{A}$.

Prolongement La mesure μ' prolonge μ. En effet si $A \in \mathcal{F}$ alors $A = A \cup \emptyset \in \mathcal{B}$ et A est m'-mesurable. De plus $\mu'(A) = \mu'(A \cup \emptyset) = \mu(A)$.

Minimalité Soit un espace mesuré complet $(X, \mathcal{M}, \nu)$ prolongeant $(X, \mathcal{F}, \mu)$. Pour $A \in \mathcal{A}$ nous devons prouver que $A \in \mathcal{M}$ et que $\mu'(A) = \nu(A)$. Il existe $B \in \mathcal{F}$ et $N \in \mathcal{N}$ tels que $A = B \cup N$. Vu que N est μ-négligeable, il est également ν-négligeable et donc ν-mesurable parce que ν est complète : $A \in \mathcal{M}$. En ce qui concerne l'égalité $\mu'(A) = \nu(A)$ nous avons

$$\nu(B) \leqslant \nu(B \cup N) \leqslant \nu(B) + \nu(N) = \nu(B), \tag{13.132}$$

donc $\nu(A) = \nu(B \cup N) = \nu(B) = \mu(B)$. La dernière égalité est le fait que ν prolonge μ. Mais par définition de μ' nous avons aussi $\mu'(A) = \mu'(B \cup N) = \mu(B)$. Au final $\mu'(A) = \nu(A) = \mu(B)$.

□

Définition 13.58.
*L'espace mesuré complet $(X, \mathcal{A}, \mu')$ défini par le théorème 13.57 est l'**espace mesuré complétée** de $(X, \mathcal{F}, \mu)$.*

Nous noterons le complété de $(S, \mathcal{F}, \mu)$ par $(S, \hat{\mathcal{F}}, \hat{\mu})$

Théorème 13.59 (Carathéodory[273])**.**
Soit S un ensemble et m^* une mesure extérieure sur S. Alors

(1) l'ensemble $\mathcal{M}$ des parties m^*-mesurables est une tribu,

(2) la restriction de m^* est une mesure sur $(S, \mathcal{M})$,

(3) l'espace mesuré $(S, \mathcal{M}, m^*)$ est complet[9]*.*

Démonstration. Une grosse partie de la preuve sera de prouver la stabilité de $\mathcal{M}$ par union dénombrable quelconque ; cela sera divisé en plusieurs parties.

9. Définition 13.56.

Tribu : le vide L'ensemble vide est m^*-mesurable.

Tribu : complémentaire Soit $A \in \mathcal{M}$ et $X \in S$. La condition qui dirait $A^c \in \mathcal{M}$ est :

$$m^*(X) = m^*(X \cap A^c) + m^*(X \cap A), \tag{13.133}$$

qui est la même que celle qui dit que A est dans $\mathcal{M}$.

Tribu : union finie Soient $A, B \in \mathcal{M}$ et $X \subset S$. Alors, vu que m^* est une mesure extérieure,

$$\begin{aligned} m^*(X) &\leqslant m^*\big(X \cap (A \cup B)\big) + m^*\big(X \cap (A \cup B)^x\big) && (13.134a)\\ &= m^*\big((X \cap A) \cup (X \cap B)\big) + m^*\big(X \cap A^c \cap B^c\big). && (13.134b) \end{aligned}$$

Mais nous pouvons écrire la première union sous forme d'une union disjointe de la façon suivante :

$$(X \cap A) \cup (X \cap B) = (X \cap A) \cup (X \cap B \cap A^c), \tag{13.135}$$

ce qui donne

$$\begin{aligned} m^*(X) &\leqslant m^*(X \cap A) + m^*(X \cap B \cap A^c) + m^*(X \cap A^c \cap B^c) && (13.136a)\\ &= m^*(X \cap A) + m^*(X \cap A^c) && (13.136b)\\ &= m^*(X) && (13.136c) \end{aligned}$$

parce que les deux derniers termes de (13.136a) se somment à $m^*(X \cap A^c)$ parce que $B \in \mathcal{M}$. La dernière ligne est le fait que A soit m^*-mesurable.

Union finie disjointe Soient $\{A_1, \dots, A_n\}$ des éléments deux à deux disjoints de $\mathcal{M}$. Nous allons maintenant prouver par récurrence que

$$m^*\Big(X \cap \big(\bigcup_{k=1}^{n} A_k\big)\Big) = \sum_{k=1}^{n} m^*(X \cap A_k). \tag{13.137}$$

Si $n = 1$ le résultat est évident. Sinon, le fait que A_{n+1} soit m^*-mesurable donne

$$m^*\Big(X \cap \big(\bigcup_{k=1}^{n+1} A_k\big)\Big) = m^*\Big(X \cap \big(\bigcup_{k=1}^{n+1} A_k\big) \cap A_{n+1}\Big) + m^*\Big(X \cap \big(\bigcup_{k=1}^{n+1} A_k\big) \cap A_{n+1}^c\Big). \tag{13.138}$$

Le fait que les A_k soient disjoints implique aussi que

$$X \cap \big(\bigcup_{k=1}^{n+1} A_k\big) \cap A_{n+1} = X \cap A_{n+1} \tag{13.139}$$

et

$$X \cap \big(\bigcup_{k=1}^{n+1} A_k\big) \cap A_{n+1}^c = X \cap \big(\bigcup_{k=1}^{n} A_k\big) \tag{13.140}$$

et donc

$$\begin{aligned} m^*\Big(X \cap \big(\bigcup_{k=1}^{n+1} A_k\big)\Big) &= m^*(X \cap A_{n+1}) + m^*\Big(X \cap \big(\bigcup_{k=1}^{n} A_k\big)\Big) && (13.141a)\\ &\overset{rec.}{=} m^*(X \cap A_{n+1}) + \sum_{k=1}^{n} m^*(X \cap A_k) && (13.141b)\\ &= \sum_{k=1}^{n+1} m^*(X \cap A_k). && (13.141c) \end{aligned}$$

La relation (13.137) est prouvée.

Notons qu'en particularisant à $X = S$ nous avons

$$m^*\big(\bigcup_{k=1}^{n} A_k\big) = \sum_{k=1}^{n} m^*(A_k) \tag{13.142}$$

dès que les A_k sont des éléments deux à deux disjoints de $\mathcal{M}$.

Union dénombrable disjointe Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'éléments deux à deux disjoints dans $\mathcal{M}$. Nous allons prouver les choses suivantes :

— $\bigcup_n A_n \in \mathcal{M}$

— $m^*\big(\bigcup_n A_n\big) = \sum_n m^*(A_n)$

où toutes les sommes et union sur n sont entre 1 et ∞.

Nous posons $A = \bigcup_k A_k$ et $B_n = \bigcup_{k=1}^n A_k$. Nous savons que $B_n \in \mathcal{M}$ pour tout n par le point précédent. Donc si $X \in S$ nous avons

$$m^*(X) = m^*(X \cap B_n) + m^*(X \cap B_n^c) \tag{13.143a}$$

$$= \sum_{k=1}^n m^*(X \cap A_k) + m^*(X \cap B_n^x) \tag{13.143b}$$

$$\geqslant \sum_{k=1}^n m^*(X \cap A_k) + m^*(X \cap A^c) \tag{13.143c}$$

où nous avons utilisé la relation (13.137) sur les B_n ainsi que le fait que $A^c \subset B_n^c$ (parce que $B_n \subset A$). L'inégalité (13.143a) étant vraie pour tout n, elle est vraie à la limite :

$$m^*(X) \geqslant \sum_{k=1}^\infty m^*(A \cap A_k) + m^*(X \cap A^c) \tag{13.144a}$$

$$\geqslant m^*\Big(\bigcup_k (X \cap A_k)\Big) + m^*(X \cap A^c) \tag{13.144b}$$

$$= m^*\Big(X \cap \big(\bigcup_k A_k\big)\Big) + m^*(X \cap A^c) \tag{13.144c}$$

$$= m^*(X \cap A) + m^*(X \cap A^c), \tag{13.144d}$$

ce qui signifie que $A \in \mathcal{M}$. La première des deux choses que nous voulions montrer est faite. En la particularisant à $X = A$ et en tenant compte des faits que $A \cap A_k = A_k$ et $A \cap A^c = \varnothing$,

$$m^*(A) \geqslant \sum_{k=1}^\infty m^*(A \cap A_k) + m^*(A \cap A^c), \tag{13.145}$$

c'est-à-dire que pour tout n nous avons

$$m^*\big(\bigcup_{k\in\mathbb{N}} A_k\big) \geqslant \sum_{k=1}^n m^*(A_k). \tag{13.146}$$

L'inégalité est encore vraie à la limite, et l'inégalité inverse étant toujours vraie pour une mesure extérieure,

$$m^*\big(\bigcup_{k\in\mathbb{N}} A_k\big) = \sum_{k=1}^\infty m^*(A_k). \tag{13.147}$$

Union dénombrable quelconque Soit maintenant une suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{M}$ que nous ne supposons plus être disjoints. Nous nous ramenons au cas disjoint en posant

$$\begin{cases} B_1 = A_1 & \text{(13.148a)} \\ B_n = A_n \cap \big(\bigcup_{k=1}^{n-1} A_k\big)^c, & \text{(13.148b)} \end{cases}$$

c'est-à-dire que nous mettons dans B_n les éléments de A_n qui ne sont dans aucun des A_k précédents. Autrement dit, nous posons $B_0 = \varnothing$ et $B_n = A_n \backslash B_{n-1}$. L'ensemble $\mathcal{M}$ étant stable par réunion finie, par complément et par intersection finie nous avons $B_n \in \mathcal{M}$. De plus les B_n sont disjoints, donc

$$\bigcup_{k=1}^\infty A_k = \bigcup_{k=1}^\infty B_k \in \mathcal{M}. \tag{13.149}$$

La première égalité se justifie de la façon suivante : si $x \in \bigcup_{k=1}^{\infty} A_k$ alors nous notons n_0 le plus petit n tel que $x \in A_n$ et alors $x \in B_{n_0}$.

Espace complet Nous prouvons à présent que $(S, \mathcal{M}, m^*)$ est un espace mesuré complet. Soit N une partie m^*-négligeable de S et $Y \in \mathcal{M}$ tel que $m^*(Y) = 0$ et $N \subset Y$. D'abord $m^*(N) = 0$ parce que

$$m^*(N) \leqslant m^*(Y) = 0. \tag{13.150}$$

Si $X \subset S$ nous avons

$$X \cap N \subset N \Rightarrow m^*(X \cap N) = 0 \tag{13.151a}$$
$$X \cap N^c \subset X \Rightarrow m^*(X \cap N^c) \leqslant m^*(X). \tag{13.151b}$$

Donc

$$m^*(X \cap N) + m^*(X \cap N^c) \leqslant m^*(X), \tag{13.152}$$

ce qui montre que N est est m^*-mesurable.

□

13.60.
Ce théorème nous pousse à adopter de la notation. Lorsqu'un espace mesuré $(S, \mathcal{F}, \mu)$ est donné, nous noterons

$$(S, \mathcal{M}, \mu^*) \tag{13.153}$$

l'espace mesuré construit de la façon suivante. D'abord μ^* est la mesure extérieure associée à μ par la proposition 13.36. Ensuite $\mathcal{M}$ est la tribu des parties μ^*-mesurables, qui est bien une tribu parce que μ^* est une mesure extérieure (13.59). La proposition (13.39) dit alors que $\mathcal{F} \subset \mathcal{M}$. De plus 13.59 nous explique que si $A \in \mathcal{F}$ alors $\mu(A) = \mu^*(A)$. Tout cela pour dire que

$$(S, \mathcal{F}, \mu) \subset (S, \mathcal{M}, \mu^*). \tag{13.154}$$

Et enfin, 13.59 nous dit que l'espace mesuré $(S, \mathcal{M}, \mu^*)$ est complet.

Exemple 13.61.
Montrons un cas dans lequel $(S, \mathcal{M}, \mu^*)$ n'est pas σ-fini. Soit S un ensemble non dénombrable et $\mathcal{F}$ la tribu des parties de S qui sont soit fini ou dénombrables soit de complémentaire fini ou dénombrable. Nous y mettons la mesure

$$\mu(A) = \begin{cases} 0 & \text{si } A \text{ est au plus dénombrable} \\ \infty & \text{sinon.} \end{cases} \tag{13.155}$$

Cette mesure n'est pas σ-finie parce qu'aucune union de dénombrables est non dénombrable. De plus $(S, \mathcal{F}, \mu)$ est complet parce que toute partie contenue dans un ensemble fini ou dénombrable est fini ou dénombrable (1.48).

$\mathcal{F}$ n'est pas $\mathcal{P}(S)$ La tribu $\mathcal{F}$ est différente de $\mathcal{P}(S)$. En effet S étant infini, il existe par 1.59 une bijection $\varphi\colon \{1, 2\} \times S \to S$. Alors l'ensemble $\varphi\big(\{1\} \times S\big)$ est non dénombrable et son complémentaire

$$\varphi\big(\{1\} \times S\big)^c = \varphi\big(\{2\} \times S\big) \tag{13.156}$$

n'est as dénombrable non plus. Cet ensemble n'est donc pas de $\mathcal{F}$.

$\mathcal{M}$ est $\mathcal{P}(S)$ En effet, soit $A \subset S$; il faut prouver que pour tout $X \subset S$ nous avons

$$\mu^*(X) = \mu^*(X \cap A) + \mu^*(X \cap A^c). \tag{13.157}$$

Nous prouvons cela en séparant les cas suivant que X est dénombrable ou non.

Si X est fini ou dénombrable, alors $X \cap A$ et $X \cap A^c$ le sont également et nous avons $\mu^*(X) = \mu(X) = 0$ ainsi que $\mu^*(X \cap A) = \mu^*(X \cap A^c) = 0$.

Si au contraire X n'est pas dénombrable,

$$\mu^*(X) = \inf_{\substack{A\in\mathcal{F}\\ X\subset A}} \mu(A) = \infty, \tag{13.158}$$

parce que X n'étant pas dénombrable, l'ensemble A ne l'est pas non plus et $\mu(A) = \infty$. Mais comme X n'est pas dénombrable, soit $X\cap A$ soit $X\cap A^c$ (soit les deux) n'est pas dénombrable non plus ; par conséquent

$$\mu^*(X\cap A) + \mu^*(X\cap A^c) = \infty. \tag{13.159}$$

Par conséquent $(S,\mathcal{F},\mu) \neq (S,\mathcal{M},\mu^*)$. Mais vu que $(S,\mathcal{F},\mu)$ est complété nous devons avoir $(S,\mathcal{F},\mu) = (S,\hat{\mathcal{F}},\hat{\mu})$. Tout cela pour dire que nous avons un exemple avec

$$(S,\mathcal{M},\mu^*) \neq (S,\hat{\mathcal{F}},\hat{\mu}). \tag{13.160}$$

△

Nous avons deux façons de créer un espace complet à partir de $(S,\mathcal{F},\mu)$.

(1) Partir de la mesure extérieure μ^* et construire $(S,\mathcal{M},\mu^*)$.

(2) Partir des ensembles μ-négligeables, construire $\hat{\mathcal{F}}$ et ensuite $(S,\hat{\mathcal{F}},\hat{\mu})$.

Ces deux façons ne sont pas équivalentes en général comme le montre l'exemple 13.61. Mais il sera montré par la proposition 13.65 que si $(S,\mathcal{F},\mu)$ est σ-fini alors les deux sont équivalent.

Lemme 13.62.
Soit $(S,\mathcal{F},\mu)$ un espace mesuré. Alors pour tout $X\subset S$ tel que $\mu^*(X) < \infty$ il existe $A\in\mathcal{F}$ tel que $X\subset A$ et $\mu^*(X) = \mu(A)$.

C'est-à-dire que μ^* a beau être défini sur toutes les parties de S, ce qu'il faut rajouter pour être μ-mesurable, c'est pas grand chose.

Démonstration. Par définition de la mesure extérieure associée à μ en tant qu'infimum, pour tout $n\geqslant 1$, il existe $A_n\in\mathcal{F}$ tel que $X\subset A_n$ et $\mu(A_n)\leqslant \mu^*(X) + \frac{1}{2^n}$. Nous posons $A = \bigcap_{n\geqslant 1} A_n$ et nous vérifions que ce A fait l'affaire.

D'abord $A\in\mathcal{F}$ parce qu'une tribu est stable par union dénombrable. Ensuite pour tout $n\geqslant 1$ nous avons

$$\mu(A)\leqslant \mu(A_n)\leqslant \mu^*(X) + \frac{1}{2^n}, \tag{13.161}$$

et à la limite $\mu(A)\leqslant \mu^*(X)$. Mais $X\subset A$ implique $\mu^*(X)\leqslant \mu(A)$ parce que $\mu^*(X)$ l'infimum d'un ensemble contenant $\mu(A)$. □

Corolaire 13.63.
Soit une mesure μ et la mesure extérieure μ^* associée[10]*. Une partie N de X est négligeable si et seulement si $\mu^*(N) = 0$.*

Démonstration. Si μ^* est la mesure extérieure associée à μ et si N est μ-négligeable alors $\mu^*(N) = 0$ parce que

$$\mu^*(N)\leqslant \mu^*(Y) = \mu(Y) = 0 \tag{13.162}$$

pour un certain Y mesurable de mesure nulle contenant N.

D'autre part si $\mu^*(N) = 0$ alors le lemme 13.62 donne une partie mesurable A telle que $N\subset A$ et $\mu(A) = 0$, c'est-à-dire que N est négligeable. □

Lemme 13.64.
Si l'espace mesuré $(S,\mathcal{F},\mu)$ est σ-fini alors l'espace mesuré $(S,\mathcal{M},\mu^*)$ est également σ-fini.

10. Par la proposition 13.36.

Démonstration. Vu que $(S, \mathcal{F}, \mu)$ est σ-fini, nous avons une suite croissante A_n d'éléments de $\mathcal{F}$ tels que $\bigcup_n A_n = S$ et telle que $\mu(A_n) < \infty$ pour tout n. Étant donné que $\mathcal{F} \subset \mathcal{M}$, cette suite convient également pour montrer que $(S, \mathcal{M}, \mu^*)$ est σ-fini parce que $\mu^*(A_n) = \mu(A_n) < \infty$. □

La proposition suivante montre que si $(S, \mathcal{F}, \mu)$ est σ-finie alors nous avons l'égalité.

Proposition 13.65.
Soit $(S, \mathcal{F}, \mu)$ un espace mesuré σ-fini, μ^* la mesure extérieure associée et $\mathcal{M}$ la tribu des ensembles μ^*-mesurables[11]*. Alors*

$$(S, \mathcal{M}, \mu^*) = (S, \hat{\mathcal{F}}, \hat{\mu}). \tag{13.163}$$

Démonstration. La proposition 13.39 indique que tous les éléments de $\mathcal{F}$ sont μ^*-mesurables, c'est-à-dire que $\mathcal{F} \subset \mathcal{M}$. Mais l'espace $(S, \mathcal{M}, \mu^*)$ est complet par le théorème de Carathéodory 13.59, donc par minimalité du complété (13.57(7)),

$$(S, \hat{\mathcal{F}}, \hat{\mu}) \subset (S, \mathcal{M}, \mu^*) \tag{13.164}$$

au sens où $\hat{\mathcal{F}} \subset \mathcal{M}$ et si $A \in \hat{\mathcal{F}}$ alors $\hat{\mu}(A) = \mu^*(A)$. Notons que cette inclusion est vraie même si la mesure n'est pas σ-finie.

Nous passons à l'inclusion inverse. Soit $A \in \mathcal{M}$, c'est-à-dire que pour tout $Y \subset S$ nous avons

$$\mu^*(Y) = \mu^*(Y \cap A) + \mu^*(Y \cap A^c). \tag{13.165}$$

Nous allons montrer que $A \in \hat{\mathcal{F}}$ en séparant les cas suivant que $\mu^*(A) = \infty$ ou non.

Si $\mu^*(A) < \infty$ Par le lemme 13.62, il existe $X \in \mathcal{F}$ tel que $A \subset X$ et $\mu^*(A) = \mu(X)$. Vu que $(S, \mathcal{F}, \mu) \subset (S, \mathcal{M}, \mu^*)$ nous avons alors

$$\mu^*(A) = \mu(X) = \mu^*(X). \tag{13.166}$$

Nous écrivons la relation (13.165) avec ce X en guise de Y, et en nous souvenant que $X \cap A = A$ et $X \cap A^c = X \backslash A$:

$$\mu^*(X) = \mu^*(A) + \mu^*(X \backslash A). \tag{13.167}$$

En tenant compte de (13.166) et du fait que $\mu^*(A) < \infty$, nous pouvons simplifier et trouver $\mu^*(X \backslash A) = 0$. Le lemme 13.62 nous donne alors $B \in \mathcal{F}$ tel que $X \backslash A \subset B$ et $\mu(B) = \mu^*(X \backslash A) = 0$, c'est-à-dire que $X \backslash A$ est μ-négligeable. Par conséquent $X \backslash A \in \hat{\mathcal{F}}$. En écrivant

$$A = X \backslash (X \backslash A), \tag{13.168}$$

nous avons écrit A comme différence de deux éléments de $\hat{\mathcal{F}}$ et nous concluons que $A \in \hat{\mathcal{F}}$.

Si $\mu^*(A) < \infty$ Le lemme 13.64 nous indique que $(S, \mathcal{M}, \mu^*)$ est σ-fini et il existe donc une suite $(S_n)_{n \geqslant 1}$ dans $\mathcal{M}$ telle que $\bigcup_n S_n = S$ et $\mu^*(S_n) < \infty$. L'ensemble $A \cap S_n$ est un élément de $\mathcal{M}$ vérifiant

$$\mu^*(A \cap S_n) \leqslant \mu^*(A) < \infty, \tag{13.169}$$

ce qui implique que $A \cap S_n \in \hat{\mathcal{F}}$ par la première partie. Maintenant $A = \bigcup_n (A \cap S_n) \in \hat{\mathcal{F}}$ par union dénombrable d'éléments de la tribu $\hat{\mathcal{F}}$.

□

Proposition 13.66 ([1])**.**
Soit un espace mesuré $(\Omega, \mathcal{F}, \mu)$. Nous considérons un mesurable $M \in \mathcal{F}$ ainsi que

— *la tribu induite $\mathcal{F}_M = \{A \cap M$ tel que $A \in \mathcal{F}\}$,*

— *la tribu complétée $\hat{\mathcal{F}}$ de $\mathcal{F}$ dans Ω,*

11. C'est bien une tribu par 13.59(1).

— *la tribu complétée* $\widehat{\mathcal{F}_M}$ *de* $\mathcal{F}_M$ *dans* M *(où nous avons considéré la mesure restreinte*[12] *de* μ*).*

— *la tribu induite* $(\hat{\mathcal{F}})_M = \{A \cap M \text{ tel que } A \in \hat{\mathcal{F}}\}$ *de* $\hat{\mathcal{F}}$ *sur* M*.*

Alors

$$(\hat{\mathcal{F}})_M = \widehat{\mathcal{F}_M}. \tag{13.170}$$

Démonstration. L'utilisation de la proposition 13.7 nous donne déjà les expressions alternatives

$$(\hat{\mathcal{F}})_M = \{A \cap M \text{ tel que } A \in \hat{\mathcal{F}}\} = \{A \subset M \text{ tel que } A \in \hat{\mathcal{F}}\} \tag{13.171}$$

et

$$\mathcal{F}_M = \{A \cap M \text{ tel que } A \in \mathcal{F}\} = \{A \subset M \text{ tel que } A \in \mathcal{F}\}. \tag{13.172}$$

Pour prouver $(\hat{\mathcal{F}})_M = \widehat{\mathcal{F}_M}$ il faudra faire deux inclusions, et nous avons l'embarras du choix.

Première : $\widehat{\mathcal{F}_M} \subset \{M \cap A \textbf{ tel que } A \in \hat{\mathcal{F}}\}$ Un élément de $\widehat{\mathcal{F}_M}$ est de la forme $B \cup N$ où $B \in \mathcal{F}_M$ et où N est négligeable[13] dans M. Vu que $B \in \mathcal{F}_M$, il existe $A \in \mathcal{F}$ tel que $B = A \cap M$. Vu que B et N sont dans M nous pouvons « factoriser » l'intersection :

$$B \cup N = M \cap (A \cup N) \tag{13.173}$$

avec N négligeable dans M et donc également négligeable dans Ω. Donc $A \cup N \in \hat{\mathcal{F}}$.

Deuxième : $\{M \cap A \textbf{ tel que } A \in \hat{\mathcal{F}}\} \subset \widehat{\mathcal{F}_M}$ Soit $A \in \hat{\mathcal{F}}$. Nous avons une partie négligeable N de Ω et un élément $B \in \mathcal{F}$ tels que $A = B \cup N$. Nous avons la décomposition

$$M \cap (B \cup N) = (M \cap B) \cap (M \cap N). \tag{13.174}$$

Il s'agit maintenant de nous assurer que cette décomposition implique que $M \cap (B \cup N) \in \widehat{\mathcal{F}_M}$. Soit $N_1 \in \mathcal{F}$ tel que $\mu(N_1) = 0$ et $N \subset N_1$. Vu que $M \cap N_1 \in \mathcal{F}$ (intersections dans une tribu), nous pouvons écrire

$$M \cap N \subset M \cap N_1 \tag{13.175}$$

avec $\mu(M \cap N_1) = 0$. Cela pour dire que $M \cap N$ est négligeable dans M. La décomposition (13.174) est donc bien une union d'un élément de $\mathcal{F}_M$ avec un négligeable de M, et donc bien un élément de $\widehat{\mathcal{F}_M}$.

□

13.67.
La principale application de la proposition 13.66 est le cas où $\mathcal{F} = \mathcal{B}\text{or}(\mathbb{R}^n)$ et M est un borélien B de $\mathbb{R}^n$. Dans ce cas, la proposition explique que la tribu de Lebesgue sur B (complétée depuis les boréliens de la topologie induite) est donnée directement par l'intersection entre B et la tribu de Lebesgue de $\mathbb{R}^n$. Donc sans devoir passer par la topologie induite, les boréliens et la completion :

$$\mathcal{L}\text{eb}(\mathbb{R}^n)_M = \widehat{\mathcal{B}\text{or}(\mathbb{R}^n)_M}. \tag{13.176}$$

Exemple dans la proposition 17.52 qui donne une structure d'espace mesuré dans S^1 à partir de la mesure de Lebesgue sur $\mathbb{C}$.

12. Ce n'est pas ce qu'il se passe dans le cas de S^1 par rapport à $\mathbb{C}$, voir la proposition 17.54(3) malgré que S^1 est un borélien de $\mathbb{C}$.

13. Pour rappel, une partie est négligeable quand elle est inclue à une partie de mesure nulle.

13.4.2 Prolongement

Le théorème suivant est parfois nommé théorème d'extension de Carathéodory, par exemple sur Wikipédia. Le théorème de Carathéodory en étant un des ingrédients principaux, on comprend.

Théorème 13.68 (Prolongement de Hahn[273])**.**
Soit $\mathcal{A}$ une algèbre de parties d'un ensemble S et μ une mesure sur $(S,\mathcal{A})$. Soit $\mathcal{F}=\sigma(\mathcal{A})$ la tribu engendrée par $\mathcal{A}$. Alors

(1) La mesure μ se prolonge en une mesure m sur $\mathcal{F}$.
(2) Si μ est σ-finie alors le prolongement est unique et m est σ-finie.
(3) Si μ est finie, alors m l'est aussi.

Démonstration. La proposition 13.19 nous donne une mesure extérieure μ^* sur S dont la restriction à $\mathcal{A}$ est μ. Si $\mathcal{M}$ est la tribu des parties μ^*-mesurables de S alors le théorème de Carathéodory 13.59 nous dit que $(S,\mathcal{M},\mu^*)$ est un espace mesuré.

<u>$\mathcal{A}\subset\mathcal{M}$</u> Cette partie est une adaptation de ce qui a déjà été fait dans la preuve de la proposition 13.39. Soit $A\in\mathcal{A}$ et $X\in S$; nous devons prouver la relation de la définition 13.37. Vu que μ^* est une mesure extérieure nous avons automatiquement

$$\mu^*(A)\leqslant\mu^*(X\cap A)+\mu^*(X\cap A^c). \tag{13.177}$$

Il reste à prouver l'inégalité inverse. Soit une suite B_k d'éléments de $\mathcal{A}$ telle que $X\subset\bigcup_k B_k$; nous avons alors

$$\mu^*(X\cap A)\leqslant\mu^*\big(\bigcup_{k=1}^{\infty}B_k\cap A\big)\leqslant\sum_{k=1}^{\infty}\mu^*(B_k\cap A)=\sum_k\mu(B_k\cap A) \tag{13.178}$$

où nous avons utilisé la définition 13.13(3) ainsi que le lemme 13.17. De la même façon,

$$\mu^*(X\cap A^c)\leqslant\sum_k\mu(B_k\cap A^c). \tag{13.179}$$

Mettant les deux bouts ensemble, en remarquant que $B_k\cap A\in\mathcal{A}$ et donc que $\mu^*(B_k\cap A)=\mu(B_k\cap A)$,

$$\mu^*(X\cap A)+\mu^*(X\cap A^c)\leqslant\sum_k\mu(B_k\cap A)+\mu(B_k\cap A^c)=\sum_k\mu(B_k). \tag{13.180}$$

La somme $\mu^*(X\cap A)+\mu^*(X\cap A^c)$ est donc inférieure à chacun des éléments de l'ensemble sur lequel on prend l'infimum pour définir [14] $\mu^*(X)$, donc

$$\mu^*(X\cap A)+\mu^*(X\cap A^c)\leqslant\mu^*(X). \tag{13.181}$$

A fortiori nous avons $\sigma(\mathcal{A})\subset\mathcal{M}$ et donc $(S,\sigma(\mathcal{A}),\mu^*)$ est un espace mesuré. Cela prouve l'existence d'une mesure prolongeant μ à $\sigma(\mathcal{A})$.

<u>**Unicité**</u> Nous supposons à présent que μ est σ-finie. Soient m_1 et m_2 deux mesures prolongeant μ et définies sur une tribu contenant $\mathcal{A}$. Nous posons

$$\mathcal{C}=\{A\in\mathcal{A}\text{ tel que }\mu(A)<\infty\}. \tag{13.182}$$

Dans l'optique d'utiliser le théorème d'unicité des mesures 13.33, nous prouvons que $\sigma(\mathcal{A})=\sigma(\mathcal{C})$. Vu que μ est σ-finie, il existe une suite croissante (S_n) d'éléments de $\mathcal{A}$ telle que $S=\bigcup_n S_n$ et $\mu(S_n)\in\mathcal{C}$. Alors si $A\in\mathcal{A}$ nous avons $A=\bigcup_n(A\cap S_n)$, et donc $A\in\sigma(\mathcal{C})$. Donc $\mathcal{A}\subset\sigma(\mathcal{C})$. Mais étant donné que $\mathcal{C}\subset\mathcal{A}$ nous avons aussi $\sigma(\mathcal{C})\subset\sigma(\mathcal{A})$. Au final $\sigma(\mathcal{A})=\sigma(\mathcal{C})$. Les mesures m_1 et m_2 sont des mesures sur $\sigma(\mathcal{C})$ coïncidant sur $\mathcal{C}$ (parce que $\mathcal{C}\subset\mathcal{A}$). De plus la classe $\mathcal{C}$ est stable par intersection finie et contient une suite croissante dont l'union est S (parce que μ est σ-finie).

Le théorème 13.33 nous dit alors que m_1 et m_2 coïncident sur $\sigma(\mathcal{C})=\sigma(\mathcal{A})$.

14. Définition 13.27.

Extension finie et σ-finie Enfin si μ est σ-finie il existe $S_n \in \mathcal{A}$ avec $\mu(S_n) < \infty$ et $\bigcup_n S_n = S$. Ces ensembles vérifient tout autant $m(S_n) = \mu(S_n) < \infty$ pour tout prolongement m de μ.

Idem si μ est finie, tout prolongement est fini.

□

Exemple 13.69 ([273]).
Soit $\mathcal{A}$, l'algèbre de parties de $\mathbb{R}$ formée par les réunions finies d'intervalles de la forme $]-\infty, a[$, $[a, b[$ et $[b, +\infty[$ avec $-\infty < a \leqslant b < +\infty$. Notons que les singletons ne font pas partie de $\mathcal{A}$ parce que $[a, a[= \varnothing$. Nous posons

$$\mu(A) = \begin{cases} 0 & \text{si } A = \varnothing \\ \infty & \text{sinon.} \end{cases} \tag{13.183}$$

Cela donne une mesure (non σ-finie) sur $(\mathbb{R}, \mathcal{A})$.

Nous allons prouver que la tribu engendrée par $\mathcal{A}$ est la tribu des boréliens et que μ accepte (au moins) deux prolongements distincts à $\sigma(\mathcal{A})$.

D'abord nous avons

$$]a, b[= \big(-\infty, a[\cup [b, +\infty[\big) \cap [a, b[, \tag{13.184}$$

donc toutes les boules ouvertes appartiennent à $\sigma(\mathcal{A})$. Ces dernières comprenant une base dénombrable de la topologie de $\mathbb{R}$ (par la proposition 7.97), tous les ouverts de $\mathbb{R}$ sont dans $\sigma(\mathcal{A})$. Par conséquent $\mathcal{B}\mathrm{or}(\mathbb{R}) \subset (\mathbb{R}^d)$. Mais en même temps tous les éléments de $\mathcal{A}$ sont des boréliens, donc $\mathcal{B}\mathrm{or}(\mathbb{R}) = \sigma(\mathcal{A})$ parce que la fermeture en tant qu'algèbre de parties est plus petite que la fermeture en tant que tribu.

La mesure de comptage prolonge μ parce qu'à part l'ensemble vide, tous les éléments de $\mathcal{A}$ sont infinis. Notons que les singletons sont dans $\sigma(\mathcal{A})$, donc la mesure de comptage prend d'autres valeurs que 0 et $+\infty$.

Par ailleurs la mesure

$$\mu'(A) = \begin{cases} 0 & \text{si } A = \varnothing \\ +\infty & \text{sinon} \end{cases} \tag{13.185}$$

est également une mesure prolongeant μ à $\sigma(\mathcal{A}) = \mathcal{B}\mathrm{or}(\mathbb{R})$.

La mesure de comptage et μ' sont deux prolongements distincts de μ. △

Exemple 13.70 ([273]).
Nous montrons maintenant une mesure non σ-finie qui se prolonge en deux mesures distinctes, toutes deux σ-finies.

Nous considérons la même algèbre $\mathcal{A}$ de parties que celle donnée dans l'exemple 13.69, mais cette fois vue sur $\mathbb{Q}$ uniquement. La mesure de comptage m sur $(\mathbb{Q}, \mathcal{A})$ n'est pas σ-finie.

Vu que les singletons sont des boréliens nous avons $\sigma(\mathcal{A}) = \mathcal{P}(\mathbb{Q})$, ce qui fait que $(\mathbb{Q}, \sigma(\mathcal{A}), m)$ est un prolongement σ-fini de m. L'espace mesuré $(\mathbb{Q}, \sigma(\mathcal{A}), 2m)$ est également σ-fini et est un prolongement distinct de $(\mathbb{Q}, \mathcal{A}, m)$. △

Proposition 13.71.
Soient des espaces mesurés $(S_1, \mathcal{F}_1, \mu_1)$ et $(S_2, \mathcal{F}_2, \mu_2)$ ainsi qu'une application $\varphi\colon S_1 \to S_2$ avec les hypothèses suivantes :

(1) φ est une bijection,

(2) φ est mesurable d'inverse mesurable,

(3) si $\mu_1(A) = 0$ alors $\mu_2\big(\varphi(A)\big) = 0$,

(4) si $\mu_2(A) = 0$ alors $\mu_1\big(\varphi^{-1}(A)\big) = 0$.

Alors

$$\hat{\mathcal{F}}_2 = \varphi(\hat{\mathcal{F}}_1). \tag{13.186}$$

Démonstration. Nous prouvons que $\hat{\mathcal{F}}_1 \subset \varphi(\hat{\mathcal{F}}_1)$. Vu la symétrie des hypothèses, l'inclusion inverse se fera de même.

Soit $A \in \hat{\mathcal{F}}_2$. Nous avons $A = B \cup N$ avec $B \in \mathcal{F}_2$ et N, une partie μ_2-négligeable. Nous considérons $N_1 \in \mathcal{F}_2$ tel que $\mu_2(N_1) = 0$ et $N \subset N_1$. Notre but est maintenant de prouver que $\varphi^{-1}(B \cup N) \in \hat{\mathcal{F}}_1$.

Vu que φ est une bijection, nous avons

$$\varphi^{-1}(B \cup N) = \varphi^{-1}(B) \cup \varphi^{-1}(N). \tag{13.187}$$

Là-dedans, $\varphi^{-1}(B) \in \mathcal{F}_1$ parce que φ est borélienne. Il nous reste à voir que $\varphi^{-1}(N)$ est μ_1-négligeable. Vu que $N \subset N_1$, nous avons $\varphi^{-1}(N) \subset \varphi^{-1}(N_1)$ où $\varphi^{-1}(N_1) \in \mathcal{F}_1$.

Par construction, $\mu_2(N_1) = 0$ et par hypothèse, $\mu_1\big(\varphi^{-1}(N_1)\big) = 0$.

Au total,

$$\varphi^{-1}(B \cup N) = \underbrace{\varphi^{-1}(B)}_{\in \mathcal{F}_1} \cup \underbrace{\varphi^{-1}(N)}_{\mu_1\text{-négligeable}} \in \hat{\mathcal{F}}_1. \tag{13.188}$$

□

13.5 Tribu borélienne

13.5.0.1 Définition

Définition 13.72 (Tribu borélienne)**.**
La tribu des ***boréliens****, notée $\mathcal{B}or(\mathbb{R}^d)$ est la tribu engendrée par les ouverts de $\mathbb{R}^d$. Plus généralement si Y est un espace topologique, la tribu des boréliens est la tribu engendrée par les ouverts de Y.*

Proposition 13.73.
La tribu engendrée par une base dénombrable de la topologie est celle des boréliens.

Démonstration. Si une base de topologie est donnée, tout ouvert peut être écrit comme union d'élément de la base, proposition 7.60. Dans le cas d'une base dénombrable, cette union sera forcément dénombrable. Une tribu étant stable par union dénombrable, tout ouvert est dans la tribu engendrée par la base de topologie. Les autres boréliens suivent automatiquement.

Dit avec plus de lettres et moins de phrases, si $\mathcal{D}$ est une base dénombrable de la topologie de X, et si $\mathcal{O}$ est un ouvert de X, nous avons $\mathcal{O} = \bigcup_{i=1}^{\infty} A_i$ avec $A_i \in \mathcal{D}$. Vu qu'une tribu est stable par union dénombrable[15], nous avons $\mathcal{O} \in \sigma(\mathcal{D})$. En conséquence de quoi $\mathcal{B}\mathrm{or}(X) \subset \sigma(\mathcal{D})$.

Mais comme $\mathcal{D} \subset \mathcal{B}\mathrm{or}(X)$ l'inclusion inverse est automatique. D'où l'égalité $\mathcal{B}\mathrm{or}(X) = \sigma(\mathcal{D})$. □

13.5.0.2 Les boréliens de $\mathbb{R}$

Nous rappelons que la topologie de $\mathbb{R}$ est celle des boules donnée par le théorème 7.94. Nous rappelons (voir la proposition 7.97 et sa preuve) que les boules ouvertes de la forme $B(q, r)$ avec $q, r \in \mathbb{Q}$ forment une base dénombrable de la topologique de $\mathbb{R}$.

Lemme 13.74.
Soit $\{q_i\}$ une énumération des rationnels. La tribu engendrée par les ouverts $\sigma_i =]q_i, \infty[$ est la tribu des boréliens.

Démonstration. Si $a < b$ dans $\mathbb{Q}$ alors $\sigma_a \backslash \sigma_b =]a, b]$. Ensuite

$$\bigcup_{n \in \mathbb{N}^*} \sigma_a \backslash \sigma_{b-\frac{1}{n}} = \bigcup_{n \in \mathbb{N}^*}]a, b - \frac{1}{n}] =]a, b[. \tag{13.189}$$

15. Définition 13.1(3)

Par union dénombrable, tous les intervalles $]a,b[$ avec $a,b\in\mathbb{Q}$ sont dans la tribu engendrée par les σ_i.

Ces boules ouvertes forment une base de la topologie de $\mathbb{R}$ par la proposition 7.97 et la proposition 13.73 conclu. □

Exemple 13.75.
Les singletons sont des boréliens de $\mathbb{R}$ parce que

$$\{x\}=\Big(]-\infty,x[\cup]x,+\infty[\Big)^c. \tag{13.190}$$

Vu qu'une tribu est stable par union dénombrable, l'ensemble $\mathbb{Q}$ est un borélien de $\mathbb{R}$. Et comme les tribus sont stables par différence ensembliste (13.3(2)), l'ensemble des irrationnels est un borélien de $\mathbb{R}$. △

13.5.0.3 Diverses expressions

Lemme 13.76.
Soient un espace topologique X et un borélien B de X. Nous considérons sur B la topologie induite[16] *de X et les boréliens $\mathcal{B}or(B)$ correspondants. Nous avons :*

$$\mathcal{B}or(B)=\{A\in\mathcal{B}or(X)\text{ tel que }A\subset B\}=\{B\cap A\text{ tel que }A\in\mathcal{B}or(X)\}. \tag{13.191}$$

En particulier,

$$\mathcal{B}or(B)=\mathcal{B}or(X)_B. \tag{13.192}$$

Démonstration. L'égalité

$$\{A\in\mathcal{B}\mathrm{or}(X)\text{ tel que }A\subset B\}=\{B\cap A\text{ tel que }A\in\mathcal{B}\mathrm{or}(X)\} \tag{13.193}$$

est déjà dans la proposition 13.7.

Nous démontrons maintenant que

$$\mathcal{B}\mathrm{or}(B)=\{A\cap B\text{ tel que }A\in\mathcal{B}\mathrm{or}(X)\}. \tag{13.194}$$

Pour ce faire, nous nous rappelons du lemme de transport 13.47. Soit l'injection canonique $f\colon B\to X$; pour tout $A\subset X$ nous avons $f^{-1}(A)=A\cap B$.

Nous considérons la classe $\mathcal{T}$ des ouverts de X. Par définition de la topologie induite, les ouverts de B sont les éléments de $f^{-1}(\mathcal{T})$. Donc

$$\sigma\big(f^{-1}(\mathcal{T})\big)=\mathcal{B}\mathrm{or}(B). \tag{13.195}$$

Mais d'autre part,

$$f^{-1}\big(\sigma(\mathcal{T})\big)=\{A\cap B\text{ tel que }A\in\mathcal{B}\mathrm{or}(X)\}. \tag{13.196}$$

Donc le lemme de transport 13.47 nous dit que

$$\mathcal{B}\mathrm{or}(B)=\sigma\big(f^{-1}(\mathcal{T})\big)=f^{-1}\big(\sigma(\mathcal{T})\big)=\{A\cap B\text{ tel que }A\in\mathcal{B}\mathrm{or}(X)\}. \tag{13.197}$$

Pour finir, l'égalité (13.192) n'est autre que le fait que

$$\mathcal{B}\mathrm{or}(X)_B=\{B\cap A\text{ tel que }A\in\mathcal{B}\mathrm{or}(X)\}, \tag{13.198}$$

alors que nous venons de prouver que le membre de droite est $\mathcal{B}\mathrm{or}(B)$. □

16. Définition 7.15.

13.5.1 Mesure image

Le produit d'une mesure par une fonction est définit par la propriété 13.189.

Proposition-définition 13.77 (Mesure image[277]).
Soient $(S_1, \mathcal{F}_1)$ et $(S_2, \mathcal{F}_2)$ des espaces mesurables. Soit $\varphi\colon S_1 \to S_2$ une application mesurable. Si m_1 est une mesure positive sur S_1 alors l'application définie par

$$m_2(A_2) = m_1\big(\varphi^{-1}(A_2)\big) \tag{13.199}$$

est une mesure positive sur $(S_2, \mathcal{F}_2)$.

La mesure m_2 ainsi définie est la ***mesure image*** *de m_1 par l'application φ. Elle est notée $\varphi(m_1)$.*

Démonstration. Il y a deux choses à vérifier pour avoir une mesure positive[17]. D'abord pour l'ensemble vide :

$$m_2(\emptyset) = m_1\big(\varphi^{-1}(\emptyset)\big) = m_1(\emptyset) = 0. \tag{13.200}$$

Ensuite pour l'additivité. Soient A_n dans $\mathcal{F}_2$ des parties deux à deux disjointes et telles que $\bigcup_n A_n \in \mathcal{F}_2$. Alors nous avons

$$m_2\big(\bigcup_n A_n\big) = m_1\Big(\varphi^{-1}\big(\bigcup_n A_n\big)\Big) \tag{13.201a}$$

$$= m_1\big(\bigcup_n \varphi^{-1}(A_n)\big) \tag{13.201b}$$

$$= \sum_n m_1\big(\varphi(A_n)\big) \tag{13.201c}$$

$$= \sum_n m_2(A_n). \tag{13.201d}$$

□

Lemme 13.78.
Soient deux espaces mesurables $(S_1, \mathcal{F}_1)$ et $(S_2, \mathcal{F}_2)$ ainsi que deux mesures μ et ν sur $(S_1, \mathcal{F}_1)$. Si $\varphi\colon S_1 \to S_2$ est mesurable et si $\mu \leqslant \nu$ alors $\varphi(\mu) \leqslant \varphi(\nu)$.

Démonstration. Soit B mesurable dans $(S_2, \mathcal{F}_2)$ (c'est-à-dire $B \in \mathcal{F}_2$). Alors

$$\varphi(\mu)(B) = \mu\big(\varphi^{-1}(B)\big) \leqslant \nu\big(\varphi^{-1}(B)\big) = \varphi(\nu)(B). \tag{13.202}$$

□

Il est naturel de se demander comment il faut intégrer par rapport à une mesure image. La réponse sera dans le théorème 13.196.

13.5.2 Régularité d'une mesure

Certaines mesures ont de la compatibilité avec la topologie. Nous allons étudier ça.

Théorème 13.79 ([277]).
Soit X un espace métrique et m une mesure positive bornée sur $\big(X, \mathcal{B}or(X)\big)$. Alors si B est un borélien,

(1) Régularité extérieure : $m(B) = \inf\{m(\Omega) \text{où } \Omega \text{ est un ouvert contenant } B\}$

(2) Régularité intérieure : $m(B) = \sup\{m(F) \text{où } F \text{ est un fermé, } F \subset B\}$.

Démonstration. Soit $\mathcal{F}$ l'ensemble des $B \in \mathcal{B}\mathrm{or}(X)$ tels que pour tout $\epsilon > 0$, il existe Ω_ϵ ouvert et F_ϵ fermé tels que $F_\epsilon \subset B \subset \Omega_\epsilon$ et $m(\Omega_\epsilon \backslash F_\epsilon) \leqslant \epsilon$. Nous allons montrer que cela est une tribu contenant les ouverts. Comme cela est inclus dans la tribu borélienne, nous en déduirons que $\mathcal{F} = \mathcal{B}\mathrm{or}(X)$.

17. Définition 13.20

$\mathcal{F}$ contient les ouverts Soit Ω un ouvert de X. Alors Ω^c est fermé et $d(x,\Omega^c)=0$ si et seulement si $x\in\Omega^c$ par la proposition 7.305. Nous pouvons donc écrire

$$\Omega^c=\bigcap_{n\geqslant 1}\{x\in X \text{ tel que } d(x,\Omega^c)<\frac{1}{n}\}. \tag{13.203}$$

En passant au complémentaire et en posant $F_n=\{x\in X \text{ tel que } d(x,\Omega^c)\geqslant\frac{1}{n}\}$ nous avons

$$\Omega=\bigcup_{n\geqslant 1}F_n. \tag{13.204}$$

Chacun des F_n est fermé parce que F_n est l'image réciproque du fermé $[\frac{1}{n},\infty[$ par l'application $x\mapsto d(x,\Omega^c)$ qui est continue. De plus les F_n forment une suite croissante, donc le lemme 13.23 nous assure que $m(\Omega)=\lim_{n\to\infty}m(F_n)$. Et le lemme 13.22 que $m(\Omega\backslash F_n)=m(\Omega)-m(F_n)$.
Soit $\epsilon>0$. Il existe alors $n_\epsilon\geqslant 1$ tel que

$$m(\Omega\backslash F_n)=m(\Omega)-m(F_n)\leqslant\epsilon. \tag{13.205}$$

Bref si Ω est ouvert nous considérons $\Omega_\epsilon=\Omega$ et $F_\epsilon=F_{n_\epsilon}$ et nous avons

$$F_\epsilon\subset\Omega\subset\Omega_\epsilon \tag{13.206}$$

avec $m(\Omega_\epsilon\backslash F_\epsilon)\leqslant\epsilon$.
L'ensemble $\mathcal{F}$ contient les ouverts.

$\mathcal{F}$ est une tribu Il y a à vérifier les trois conditions de la définition 13.1.

Les ensembles faciles Les ensembles X et $\varnothing$ sont dans $\mathcal{F}$ parce qu'ils sont ouverts et fermés.

Complémentaire Soit $B\in\mathcal{F}$, soit $\epsilon>0$ et les ensembles F_ϵ et Ω_ϵ qui vont avec. Alors en passant au complémentaire nous avons

$$\Omega_\epsilon^c\subset B^c\subset F_\epsilon^c \tag{13.207}$$

De plus

$$F_\epsilon^c\backslash\Omega_\epsilon^c=F_\epsilon^c\cap(\Omega_\epsilon^c)^c=F_\epsilon^c\cap\Omega_\epsilon=\Omega_\epsilon\backslash F_\epsilon. \tag{13.208}$$

Par conséquent

$$m(F_\epsilon^c\backslash\Omega_\epsilon^c)=m(\Omega_\epsilon\backslash F_\epsilon)\leqslant\epsilon. \tag{13.209}$$

Cela montre que $B^c\in\mathcal{F}$.

Union dénombrable Soient (B_n) une suite d'éléments de $\mathcal{F}$ et $\epsilon>0$. Pour chaque n nous choisissons un ouvert Ω_n et un fermé F_n tels que $F_n\subset B_n\subset\Omega_n$ et

$$m(\Omega_n\backslash F_n)\leqslant\frac{\epsilon}{2^{n+2}}. \tag{13.210}$$

Vu que $\Omega_n\backslash B_n\subset\Omega_n\backslash F_n$ nous avons aussi

$$m(\Omega_n\backslash B_n)\leqslant m(\Omega_n\backslash F_n)\leqslant\frac{\epsilon}{2^{n+2}}. \tag{13.211}$$

Nous posons $\Omega=\bigcup_{n\geqslant 1}\Omega_n$ (un ouvert) et $B=\bigcup_{n\geqslant 1}B_n$ ainsi que $A=\bigcup_{n\geqslant 1}F_n$ (qui n'est pas spécialement fermé).
Le but est de majorer $m(\Omega\backslash F)$ où F est un fermé qui est encore à déterminer. Calculons déjà ceci :

$$\Omega\backslash B=\bigcup_n\Omega_n\cap\big(\bigcup_k B_k\big)^c \tag{13.212a}$$

$$=\bigcup_n\Big(\Omega_n\cap\big(\bigcap_k B_k^c\big)\Big) \tag{13.212b}$$

$$\subset\bigcup_n\big(\Omega_n\cap B_n^c\big) \tag{13.212c}$$

$$=\bigcup_n(\Omega_n\backslash B_n) \tag{13.212d}$$

où l'union n'est pas spécialement disjointe. Par conséquent,

$$m(\Omega\backslash B) \leqslant \sum_{n=1}^{\infty} m(\Omega_n\backslash B_n) \leqslant \sum_{n=1}^{\infty} \frac{\epsilon}{2^{n+2}} = \frac{\epsilon}{4}. \tag{13.213}$$

De la même façon nous avons

$$B\backslash A = \big(\bigcup_{n=1}^{\infty} B_n\big) \cap \big(\bigcup_{k=1}^{\infty} F_n\big)^c \subset \bigcup_{n=1}^{\infty} B_n\backslash F_n. \tag{13.214}$$

Nous avons alors le inégalités de mesures

$$m(B\backslash A) \leqslant \sum_{n=1}^{\infty} m(B_n\backslash F_n) \tag{13.215a}$$

$$\leqslant \sum_{n=1}^{\infty} m(\Omega_n\backslash F_n) \tag{13.215b}$$

$$\leqslant \frac{\epsilon}{4}. \tag{13.215c}$$

C'est vraiment dommage que A ne soit pas en générale un fermé, sinon il répondrait à la question. Nous posons $F'_1 = F_1$ et $F'_n = \bigcup_{k=1}^{n} F_k$. En tant qu'unions finies de fermés, les F'_n sont des fermés (lemme 7.3(2)). De plus la suite (F'_n) est croissante et l'union est A. Par le lemme 13.23(1) nous avons

$$m(A) = m\big(\bigcup_n F'_n\big) = \lim_{n\to\infty} m(F'_n). \tag{13.216}$$

Il existe donc n_ϵ tel que

$$m(A) - m(F'_n) \leqslant \epsilon \tag{13.217}$$

Nous posons $F = F'_{n_\epsilon}$. Vu que $F \subset A$ nous avons aussi $m(A\backslash F) = m(A) - m(F) \leqslant \epsilon$. Et en plus $F \subset A \subset B \subset \Omega$, ce qui donne bien la propriété voulue $F \subset B \subset \Omega$. Il reste à nous assurer de $m(\Omega\backslash F)$. Nous avons d'abord

$$m(B\backslash F) = m\big((B\backslash A) \cup (A\backslash F)\big) = m(B\backslash A) + m(A\backslash F) \leqslant \frac{5\epsilon}{4}. \tag{13.218}$$

Et enfin :

$$m(\Omega\backslash F) = m\big((\Omega\backslash B) \cup (B\backslash F)\big) = m(\Omega\backslash B) + m(B\backslash F) \leqslant \frac{6\epsilon}{4}. \tag{13.219}$$

Et donc à redéfinition près de ϵ c'est d'accord.

Il est donc établi que $\mathcal{F}$ est une tribu. Qui plus est, l'ensemble $\mathcal{F}$ est une tribu incluse aux boréliens et contenant les ouverts. Ergo $\mathcal{F} = \mathcal{B}\mathrm{or}(X)$.

Régularité extérieure Soit B un borélien et $\epsilon > 0$. Alors il existe F_ϵ fermé et Ω_ϵ ouvert tels que $F_\epsilon \subset B \subset \Omega_\epsilon$ et $m(\Omega_\epsilon\backslash F_\epsilon) \leqslant \epsilon$. Vu que $B \subset \Omega_\epsilon$ pour tout ϵ, nous avons aussi

$$m(B) \leqslant \inf_\epsilon m(\Omega_\epsilon). \tag{13.220}$$

Mais comme $\mu(\Omega_\epsilon) \geqslant m(B)$ pour tout ϵ, nous avons en réalité $m(B) = \inf_\epsilon m(\Omega_\epsilon)$.

Soit maintenant un ouvert Ω tel que $B \subset \Omega$. Nous devons prouver l'existence d'un $\epsilon > 0$ tel que $m(\Omega_\epsilon) \leqslant m(\Omega)$. Cela permettra de conclure que l'infimum sur tous les ouverts contenant B est égal à l'infimum sur les ouverts de la forme Ω_ϵ.

Nous posons $m(\Omega) = m(B) + \delta$ et avec $\epsilon \leqslant \delta$ nous avons

$$m(\Omega_\epsilon\backslash B) \leqslant m(\Omega_\epsilon\backslash F_\epsilon) \leqslant \epsilon \tag{13.221}$$

et donc aussi

$$m(\Omega_\epsilon) \leqslant m(B) + \epsilon \leqslant m(B) + \delta = m(\Omega). \tag{13.222}$$

Régularité intérieure Elle se fait de même.

□

Définition 13.80.
Soit X un espace topologique et m une mesure positive sur $\big(X, \mathcal{B}or(X)\big)$.

*(1) m est une **mesure de Borel** si elle est finie sur tout compact.*

*(2) m est **régulière extérieurement** si $\forall B \in \mathcal{B}or(X)$,*

$$m(B) = \inf\{m(\Omega) \text{ tel que } \Omega \text{ est ouvert et } B \subset \Omega\} \tag{13.223}$$

*(3) m est **régulière intérieurement** si $\forall B \in \mathcal{B}or(X)$,*

$$m(B) = \sup\{m(K) \text{ tel que } K \text{ est compact et } K \subset B\} \tag{13.224}$$

*(4) m est une mesure **régulière** si elle est régulière dans les deux sens.*

*(5) m est une **mesure de Radon** si elle est de Borel et régulière.*

Proposition 13.81.
Soit X un espace localement compact et dénombrable à l'infini[18] *Alors toute mesure de Borel sur $\big(X, \mathcal{B}or(X)\big)$ est de Radon.*

Démonstration. Nous avons une suite exhaustive[19] de compacts X_k tels que

$$X = \bigcup_{k\geqslant 1} X_k = \bigcup_{k\geqslant 1} \mathrm{Int}(X_k). \tag{13.225}$$

Régularité intérieure Soit B, un borélien de X ; nous avons $B = \bigcup_{k\geqslant 1}(B \cap X_k)$ et comme cette union est croissante,

$$m(B) = \lim_{k\to\infty} m(B \cap X_k) \tag{13.226}$$

par le lemme 13.23(1). Dans la suite, il va y avoir beaucoup de considérations sur les topologies induites. Nous nommons τ_k la topologie de X_k induite depuis celle de X. Il ne faudra pas confondre les expressions « un compact *de* X_k » et « un compact *dans* X_k ». La première parle d'un compact pour la topologie τ_k. La seconde parle d'un compact pour la topologie de X, inclus dans X_k.

Si $a < m(B)$ alors il existe $k \geqslant 1$ tel que $a < m(B \cap X_k)$, c'est-à-dire

$$a < m(B \cap X_k) \leqslant m(B). \tag{13.227}$$

Mais (X_k, m) est un espace mesuré borné parce que m est de Borel et X_k est compact. Par conséquent la (restriction de la) mesure m est régulière sur l'espace mesuré $\big(X_k, \mathcal{B}\mathrm{or}(X_k)\big)$ par le théorème 13.79. De plus l'ensemble $B \cap X_k$ est un borélien de (X_k, τ_k) parce que

$$B \cap X_k \in \mathcal{B}\mathrm{or}(X)_{X_k} = \mathcal{B}\mathrm{or}(X_k) \tag{13.228}$$

où nous avons utilisé la propriété de compatibilité entre topologie induite et tribu des borélien du théorème 13.50. Il existe donc un fermé F_ϵ de (X_k, τ_k) tel que

$$\begin{cases} F_\epsilon \subset B \cap X_k & \text{(13.229a)} \\ m(B \cap X_k) \leqslant m(F_\epsilon) + \epsilon. & \text{(13.229b)} \end{cases}$$

En mettant bout à bout les inégalités nous avons trouvé

$$a < m(B \cap X_k) \leqslant m(F_\epsilon) + \epsilon < m(F_\epsilon), \tag{13.230}$$

18. Définitions 7.53 et 7.55.
19. Définition 7.312.

et donc en particulier $a < m(F_\epsilon)$. L'ensemble F_ϵ est en plus un compact de (X, τ_X). En effet X_k étant fermé de (X, τ_X), le lemme 7.16 nous dit que F_ϵ est un fermé de (X, τ_X). Mais X_k étant compact, F_ϵ est un fermé inclus dans un compact, il est donc compact (lemme 7.64).

Pour tout $a < m(B)$ nous avons trouvé un compact F_ϵ inclus dans B dont la mesure est plus grande que a. Cela prouve la régularité intérieure de la mesure m.

Régularité extérieure Soit un borélien B de X. Si $m(B) = \infty$ alors tous les ouverts contenant B ont mesure infinie et $m(B)$ en est évidemment le supremum. Nous supposons donc que $m(B) < \infty$.

Nous notons τ_k la topologie induite de X sur $\mathrm{Int}(X_k)$. Nous posons $B_k = B \cap \mathrm{Int}(X_k)$. L'espace $\big(\mathrm{Int}(X_k), m\big)$ est un espace mesuré borné et $B_k \in \mathcal{B}\mathrm{or}\Big(\mathrm{Int}(X_k)\Big)$. Il existe donc un ouvert Ω_k de $\big(\mathrm{Int}(X_k), \tau_k\big)$ tel que $B_k \subset \Omega_k$ et

$$m(\Omega_k \backslash B_k) \leqslant \frac{\epsilon}{2^k}. \tag{13.231}$$

De plus $\mathrm{Int}(X_k)$ est un ouvert de (X, τ_X), donc en réalité Ω_k est un ouvert de X. Nous posons

$$\Omega = \bigcup_{k=1}^{\infty} \Omega_k \tag{13.232}$$

qui est encore un ouvert de (X, τ_X).

Il est temps de voir que Ω vérifie $m(\Omega \backslash B) \leqslant \epsilon$. Pour cela,

$$\begin{aligned}
\Omega \backslash B &= \big(\bigcup_k \Omega_k\big) \cap \big(\bigcup_l B_l\big)^c && \text{(13.233a)}\\
&= \big(\bigcup_k \Omega_k\big) \cap \big(\bigcap B_l^c\big) && \text{(13.233b)}\\
&\subset \bigcup_k (\Omega_k \cap B_k^c) && \text{(13.233c)}\\
&= \bigcup_k (\Omega_k \backslash B_k), && \text{(13.233d)}
\end{aligned}$$

ce qui donne au niveau des mesures :

$$m(\Omega \backslash B) \leqslant \sum_{k=1}^{\infty} m(\Omega_k \backslash B_k) \leqslant \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = \epsilon. \tag{13.234}$$

□

Remarque 13.82.
Exprimé sur $\mathbb{R}^N$, la proposition 13.81 s'exprime en disant que toute mesure de Borel sur $\mathbb{R}^N$ est régulière. Typiquement, l'espace X dont il est question est un ouvert de $\mathbb{R}^N$.

13.5.3 Théorème de récurrence

Soit X un espace mesurable, μ une mesure finie sur X et $\phi\colon X \to X$ une application mesurable[20] préservant la mesure, c'est-à-dire que pour tout ensemble mesurable $A \subset X$,

$$\mu\big(\phi^{-1}(A)\big) = \mu(A). \tag{13.235}$$

Si $A \subset X$ est un ensemble mesurable, un point $x \in A$ est dit **récurrent** par rapport à A si et seulement si pour tout $p \in \mathbb{N}$, il existe $k \geqslant p$ tel que $\phi^k(x) \in A$.

Théorème 13.83 (Théorème de récurrence de Poincaré.)**.**
Si A est mesurable dans X, alors presque tous les points de A sont récurrents par rapport à A.

20. Définition 13.41.

Démonstration. Soit $p \in \mathbb{N}$ et l'ensemble

$$U_p = \bigcup_{k=p}^{\infty} \phi^{-k}(A) \tag{13.236}$$

des points qui repasseront encore dans A après p itérations de ϕ. C'est un ensemble mesurable en tant que union d'ensembles mesurables (pour rappel, les tribus sont stables par union dénombrable, comme demandé à la définition 13.1), et nous avons donc

$$\mu(U_p) \leqslant \mu(X) < \infty. \tag{13.237}$$

De plus $U_p = \phi^{-p}(U_0)$, donc $\mu(U_p) = \mu(U_0)$. Vu que $U_p \subset U_p$, nous avons

$$\mu(U_0 \backslash U_p) = 0. \tag{13.238}$$

Étant donné que $A \subset U_0$ nous avons a fortiori que

$$\{x \in A \text{ tel que } x \notin U_p\} \subset U_0 \backslash U_p, \tag{13.239}$$

et donc

$$\mu\{x \in A \text{ tel que } x \notin U_p\} = 0. \tag{13.240}$$

Cela signifie exactement que l'ensemble des points x de A tels que aucun des $\phi^k(x)$ avec $k \geqslant p$ n'est dans A est de mesure nulle. □

13.6 Mesurabilité des fonctions à valeurs réelles

Nous allons parler de la mesurabilité de fonctions

$$f\colon (S, \mathcal{F}) \to \big(\bar{\mathbb{R}}, \mathcal{B}\mathrm{or}(\bar{\mathbb{R}})\big) \tag{13.241}$$

où $\bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$.

13.84.
Nous convenons que $0 \times \pm\infty = 0$ parce que nous voulons qu'une droite (qui est un rectangle dont une mesure est 0 et l'autre ∞) soit de mesure nulle dans $\mathbb{R}^2$.

Les produits et sommes $\pm\infty \pm \pm\infty$ et $\pm\infty \times \pm\infty$ sont ceux que l'on croit. Sauf bien entendu $+\infty - \infty$ et $1/0$ qui ne sont toujours pas définis.

Lemme 13.85.
L'ensemble B est un borélien de $\bar{\mathbb{R}}$ si et seulement s'il existe un borélien B_0 de $\mathbb{R}$ tel que B soit B_0 ou $B_0 \cup \{-\infty\}$ ou $B_0 \cup \{-\infty\}$ ou $B_0 \cup \{+\infty, -\infty\}$.

Démonstration. Vu que la topologie usuelle sur $\mathbb{R}$ est la topologie induite de celle sur $\bar{\mathbb{R}}$, la tribu induite l'est aussi par le théorème 13.49. Donc si B est un borélien de $\bar{\mathbb{R}}$, l'ensemble $B \cap \mathbb{R}$ est un borélien de $\mathbb{R}$. □

Lemme 13.86 ([277])**.**
Si $\mathcal{S}_0$ est l'ensemble des intervalles du type

$$]\alpha, \beta[, \quad [-\infty, \beta[, \quad]\alpha, +\infty] \tag{13.242}$$

avec $-\infty < \alpha < \beta < +\infty$ alors $\sigma(\mathcal{S}_0) = \mathcal{B}or(\bar{\mathbb{R}})$.

Démonstration. Les intervalles $]\alpha, \beta[$ engendrent la topologie de $\mathbb{R}$ [21], donc $\mathcal{B}\mathrm{or}(\mathbb{R}) \subset \sigma(\mathcal{S}_0)$. De plus le lemme 13.3 nous autorise à dire que

$$\bigcap_{n \geqslant 1} [n, +\infty] = \{+\infty\} \in \sigma(\mathcal{S}_0). \tag{13.243}$$

21. Parce toutes les boules sont des intervalles de ce type et que les boules forment une base de topologie, proposition 7.97.

Par conséquent tous les ensembles énumérés dans le lemme 13.85 font partie de $\sigma(\mathcal{S}_0)$. Cela implique que $\mathcal{B}\mathrm{or}(\bar{\mathbb{R}}) \subset \sigma(\mathcal{S}_0)$.

Pour l'inclusion inverse, $\sigma(\mathcal{S}_0)$ est engendré par des parties qui font parie de $\mathcal{B}\mathrm{or}(\bar{\mathbb{R}})$, donc $\sigma(\mathcal{S}_0) \subset \mathcal{B}\mathrm{or}(\bar{\mathbb{R}})$. □

13.6.1 Fonctions à valeurs réelles sur un espace mesurable

Théorème 13.87.
Soit un espace mesurable $(S, \mathcal{F})$ et une fonction $f\colon S \to \bar{\mathbb{R}}$. Les faits suivants sont équivalents.

(1) La fonction f est mesurable.

(2) L'ensemble $\{f < a\}$ est dans $\mathcal{F}$ pour tout $a \in \mathbb{R}$

(3) L'ensemble $\{f \leqslant a\}$ est dans $\mathcal{F}$ pour tout $a \in \mathbb{R}$

Démonstration. Plusieurs implications à prouver.

(1)⇒(2) Vu que f est mesurable et que $[-\infty, a[\in \mathcal{B}\mathrm{or}(\bar{\mathbb{R}})$, nous avons $f^{-1}\big([-\infty, a[\big) \in \mathcal{F}$.

(2)⇒(1) Nous posons $\mathcal{A} = \{[-\infty, a[\text{ tel que } a \in \mathbb{R}\}$.

Nous avons $\mathcal{A} \subset \mathcal{S}_0$ (le $\mathcal{S}_0$ du lemme 13.86). Et de plus,

$$]\alpha, \beta[= [-\infty, \beta[\backslash[-\infty, \alpha] = [-\infty, \beta[\backslash \bigcap_{n\geqslant 1}[-\infty, \alpha + \frac{1}{n}[. \tag{13.244}$$

Donc $]\alpha, \beta[\in \sigma(\mathcal{A})$.

Et aussi :

$$]\alpha, +\infty] = \bar{\mathbb{R}}\backslash[-\infty, \alpha + \frac{1}{n}[, \tag{13.245}$$

ce qui donne $]\alpha, +\infty] \in \sigma(\mathcal{A})$.

Au final, $\mathcal{S}_0 \subset \sigma(\mathcal{A})$ et donc $\sigma(\mathcal{S}_0) \subset \sigma(\mathcal{A})$. Le lemme 13.86 nous dit que $\sigma(\mathcal{S}_0) = \mathcal{B}\mathrm{or}(\bar{\mathbb{R}})$. Nous avons donc bien $\sigma(\mathcal{S}_0) = \sigma(\mathcal{A}) = \mathcal{B}\mathrm{or}(\bar{\mathbb{R}})$.

par ailleurs, nous savons que $f^{-1}(\mathcal{A}) \subset \mathcal{F}$ parce que les éléments de $\mathcal{A}$ sont de la forme $\{f < a\}$. Cela donne $\sigma\big(f^{-1}(\mathcal{A})\big) = \mathcal{F}$. Mais $\sigma\big(f^{-1}(\mathcal{A})\big)$ peut aussi s'exprimer par le lemme de transport 13.47 : $\sigma\big(f^{-1}(\mathcal{A})\big) = f^{-1}\big(\sigma(\mathcal{A})\big)$. En combinant les deux,

$$f^{-1}\big(\sigma(\mathcal{A})\big) = \mathcal{F}, \tag{13.246}$$

et en remplaçant $\sigma(\mathcal{A})$ par $\mathcal{B}\mathrm{or}(\bar{\mathbb{R}})$ nous avons ce que nous voulions :

$$f^{-1}\big(\mathcal{B}\mathrm{or}(\bar{\mathbb{R}})\big) \in \mathcal{F}, \tag{13.247}$$

ce qui signifie que f est mesurable.

(3)⇒(2) Nous avons

$$\{f < a\} = \bigcup_{n\geqslant 1}\{f \leqslant a - \frac{1}{n}\}. \tag{13.248}$$

donc cela est une union dénombrable d'éléments de $\mathcal{F}$. Donc $\{f < a\}$ est dans $\mathcal{F}$.

(1)⇒(3) Nous avons

$$\{f \leqslant a\} = \{f < a\} \cup f^{-1}\big([-\infty, a]\big). \tag{13.249}$$

Le premier ensemble est dans $\mathcal{F}$ par (2). Ensuite $[-\infty, a]$ est un fermé de $\bar{\mathbb{R}}$ et donc un borélien de $\bar{\mathbb{R}}$. Son image réciproque est donc un élément de $\mathcal{F}$ parce que f est mesurable. Au final nous avons bien $\{f \leqslant a\} \in \mathcal{F}$.

□

Lemme 13.88 ([280])**.**
Une fonction $f\colon X \to \mathbb{R}$ est mesurable si et seulement si $f^{-1}(I)$ est mesurable pour tout I de la forme $]a, \infty[$.

Démonstration. Nous devons prouver que $f^{-1}(A)$ est mesurable dans X pour tout borélien A de $\mathbb{R}$. Nous posons

$$S = \{A \subset \mathbb{R} \text{ tel que } f^{-1}(A) \text{ est mesurable dans } X\} \tag{13.250}$$

et nous prouvons que cela est une tribu. D'abord $f^{-1}(\mathbb{R}) = X$, et X est mesurable, donc $\mathbb{R} \in S$. Ensuite si $A \in S$ alors $f^{-1}(A^c) = f^{-1}(A)^c$. En tant que complémentaire d'un mesurable de X, l'ensemble $f^{-1}(A)^c$ est mesurable dans X. Et enfin si $A_n \in S$ alors $f^{-1}(\bigcup_n A_n) = \bigcup_n f^{-1}(A_n)$ qui est encore mesurable dans X en tant qu'union de mesurables.

Donc S est une tribu qui contient tous les ensembles de la forme $]a, \infty]$. Le lemme 13.74 conclu que S contient tous les boréliens de $\mathbb{R}$. □

Lemme 13.89 ([280]).
Soit $f_n \colon X \to \mathbb{R}$ une suite de fonctions mesurables[22]*. Alors $\sup_n f_n$ est mesurable.*

Démonstration. Nous avons

$$(\sup f_n)^{-1}\big(]a, \infty]\big) = \{x \in X \text{ tel que } (\sup f_n)(x) > a\} \tag{13.251a}$$

$$= \bigcup_n \{x \in X \text{ tel que } f_n(x) > a\} \tag{13.251b}$$

$$= \bigcup_n f_n^{-1}\big(]a, \infty]\big). \tag{13.251c}$$

Étant donné que f_n est mesurable et que $]a, \infty]$ est mesurable, chacun des $f_n^{-1}\big(]a, \infty]\big)$ est mesurable dans X. L'ensemble $(\sup f_n)^{-1}\big(]a, \infty]\big)$ donc une union dénombrable de parties mesurables. Il est donc mesurable.

Le lemme 13.88 conclu que $\sup f_n$ est mesurable. □

Proposition 13.90.
Si $f_n \colon X \to \mathbb{R}$ est une suite de fonctions mesurables et positives, alors la fonction[23] *$\sum_n f_n$ est mesurable.*

Démonstration. Nous considérons les fonctions $s_k(x) = \sum_{n=0}^{k} f_n(x)$ qui vaut éventuellement ∞ en certains points. Nous avons

$$\sum_n f_n(x) = \sup_k s_k(x), \tag{13.252}$$

donc le lemme 13.89 nous donne la mesurabilité de la somme de f_n. □

Définition 13.91.
Soit $(S, \mathcal{F})$ un espace mesurable. Une ***partition mesurable dénombrable*** *de e S est une suite $(S_n)_{n \geqslant 1}$ de parties de S telles que*

(1) $S_n \in \mathcal{F}$ pour tout n,
(2) $S_N \cap S_k = \varnothing$ si $n \neq k$,
(3) $S = \bigcup_{n \geqslant 1} S_n$.

Lemme 13.92 (Lemme de recollement).
Soit (S_n) une partition mesurable dénombrable de l'espace mesurable $(S, \mathcal{F})$. Soit $(S', \mathcal{F}')$ un autre espace mesurable et des fonctions mesurables

$$f_n \colon (S_n, \mathcal{F}_{S_n}) \to (S', \mathcal{F}') \tag{13.253}$$

où $\mathcal{F}_{S_n}$ est la tribu induite[24]*. Alors la fonction*

$$\begin{aligned} f \colon (S, \mathcal{F}) &\to (S', \mathcal{F}') \\ x &\mapsto f_n(x) \text{ si } x \in S_n \end{aligned} \tag{13.254}$$

est mesurable.

22. Ici X est un espace mesuré et $\mathbb{R}$ est muni des boréliens.
23. Définition 11.301 pour la série de fonctions.
24. Définition 13.6.

Démonstration. Soit $A' \in \mathcal{F}'$; nous devons prouver que $f^{-1}(A') \in \mathcal{F}$. Nous savons que

$$f^{-1}(A') = \bigcup_{n\geqslant 1} f_n^{-1}(A'), \tag{13.255}$$

qui est une union dénombrable d'éléments $f_n^{-1}(A') \in \mathcal{F}_{S_n}$.

Vu que $S_n \in \mathcal{F}$ nous avons $\mathcal{F}_{S_n} \subset \mathcal{F}$ parce qu'un élément de $\mathcal{F}_{S_n}$ est de la forme $S_n \cap B$ avec $B \in \mathcal{F}$. Du coup pour chaque n nous avons

$$f_n^{-1}(A') \in \mathcal{F}_{S_n} \subset \mathcal{F}. \tag{13.256}$$

Au final l'égalité (13.255) écrit $f^{-1}(A')$ comme une union d'éléments de $\mathcal{F}$ et est donc un élément de $\mathcal{F}$. □

Proposition 13.93.
Soit $(S, \mathcal{F})$ un espace mesurable et des applications mesurables $f, g\colon S \to \bar{\mathbb{R}}$. Alors les fonctions suivantes sont mesurables :

(1) λf pour tout $\lambda \in \mathbb{R}$

(2) $f + g$ si elle existe.

(3) $1/f$ si elle existe.

(4) fg.

Démonstration. Commençons par clarifier « si elle existe ». La fonction $f+g$ n'existe pas au point $x \in S$ si $f(x) = +\infty$ et $g(x) = -\infty$. La fonction $1/f$ n'existe pas au point $x \in S$ si $f(x) = 0$. Voir le point 13.84.

La partie où $f + g$ existe est mesurable La partie de S sur laquelle $f + g$ existe est

$$\{x \in S \text{ tel que } \big(f(x), g(x)\big) \neq (+\infty, -\infty), \big(f(x), g(x)\big) \neq (-\infty, +\infty)\}. \tag{13.257}$$

Nous avons

$$\{(f, g) = (+\infty, -\infty)\} = \{f = \infty\} \cap \{g = -\infty\} \tag{13.258}$$

qui est un ensemble mesurable parce que, par exemple,

$$\{+\infty\} = \bigcap_{n\geqslant 1} [n, +\infty]. \tag{13.259}$$

La cas $(-\infty, +\infty)$ est identique, et au final la partie de S sur laquelle $f + g$ n'existe pas est mesurable. Par complémentarité la partie sur laquelle $f+g$ existe est également mesurable [25].

Idem pour la partie sur laquelle $1/f$ existe Idem.

Mesurabilité de λf Si $\lambda = 0$, nous avons une fonction constante dont la mesurabilité est évidente [26]. Nous supposons $\lambda > 0$. Alors

$$\{\lambda f < a\} = \{f < a/\lambda\} \in \mathcal{F}. \tag{13.260}$$

.Pour $\lambda < 0$ nous avons de la même manière

$$\{\lambda f < a\} = \{f > a/\lambda\} \in \mathcal{F}. \tag{13.261}$$

Ce dernier point est suffisant pour que λf soit mesurable par la théorème 13.87(3) et par complémentarité.

25. Parfois on a envie de dire que l'affirmation « A est mesurable » ne passe pas le test de Popper.
26. Prenez quand même le temps d'y penser.

Mesurabilité de $f+g$ Soit $a \in \mathbb{R}$; le théorème 13.87 nous demande d'avoir envie de prouver que $\{f+g<a\} \in \mathcal{F}$. Nous avons

$$f(x)+g(x)<a \tag{13.262}$$

si et seulement si

$$f(x)<a-g(x) \tag{13.263}$$

si et seulement si

$$\exists q \in \mathbb{Q} \text{ tel que } f(x)<q<a-g(x). \tag{13.264}$$

Donc

$$\{f+g<a\} = \bigcup_{q\in\mathbb{Q}} \Big(\{f<q\} \cap \{g<a-r\}\Big), \tag{13.265}$$

qui est une union dénombrable d'éléments de $\mathcal{F}$. Donc $\{f+g<a\} \in \mathcal{F}$ et $f+g$ est mesurable. Note qu'en toute rigueur il faudrait « $\cap$là où $f+g$ est définie » un peu partout, mais cela ne change rien parce que l'intersection de deux parties mesurables est mesurable.

Mesurabilité de $1/f$ Soit $a \in \mathbb{R}$. Si $a>0$ alors

$$\{1/f<a\} = \{f<0\} \cup \{f>\frac{1}{a}\} \in \mathcal{F}. \tag{13.266}$$

et si $a<0$ alors

$$\{1/f<a\} = \{f<0\} \cap \{f>\frac{1}{a}\} \in \mathcal{F}. \tag{13.267}$$

Mesurabilité de fg Nous allons la prouver en plusieurs fois.

Si f est mesurable alors f^2 est mesurable Si $a \leqslant 0$ alors $\{f^2<a\}=\emptyset$. Si $a>0$ nous avons

$$\{f^2<a\} = \{-\sqrt{a}<f<\sqrt{a}\} \in \mathcal{F}. \tag{13.268}$$

$f\mathbb{1}_A$ est mesurable Soit $A \in \mathcal{F}$, et prouvons que $f\mathbb{1}_A$ est mesurable. Par définition,

$$(f\mathbb{1}_A)(x) = \begin{cases} f(x) & \text{si } x \in A \\ 0 & \text{si } x \notin A. \end{cases} \tag{13.269}$$

Nous posons

$$\begin{aligned} f_1 \colon A^c &\to \bar{\mathbb{R}} \\ x &\mapsto 0 \end{aligned} \tag{13.270}$$

et

$$\begin{aligned} f_2 \colon A &\to \bar{\mathbb{R}} \\ x &\mapsto f(x). \end{aligned} \tag{13.271}$$

Alors nous avons

$$(\mathbb{1}_A f)(x) = \begin{cases} f_1(x) & \text{si } x \in A^c \\ f_2(x) & \text{si } x \in A. \end{cases} \tag{13.272}$$

Les ensembles A et A^c forment une partition mesurable dénombrable de S. La fonction f_1 est mesurable ; pour prouver que f_2 est mesurable, nous l'écrivons $f_2 = f \circ j_A$ où $j_A \colon A \to S$ est l'injection canonique. L'application

$$j_A \colon (A, \mathcal{F}_A) \to (S, \mathcal{F}) \tag{13.273}$$

est mesurables parce que si $B \in \mathcal{F}$ alors $j_A^{-1}(B) = A \cap B \in \mathcal{F}_A$. D'autre part l'application

$$f \colon (S, \mathcal{F}) \to \big(\bar{\mathbb{R}}, \mathcal{B}\mathrm{or}(\bar{\mathbb{R}})\big) \tag{13.274}$$

est mesurable par hypothèse. La composée $f_2 = f \circ j_A$ est alors mesurable par la proposition 13.43. Le lemme de recollement 13.92 nous donne alors la mesurabilité de $f\mathbb{1}_A$.

Le produit fg est mesurable Nous posons

$$F = \{x \in S \text{ tel que } |f(x)| < +\infty, |g(x)| < \infty\}. \tag{13.275}$$

En tant qu'intersection de deux ensembles mesurables, F est mesurable. Par la partie précédente, les applications $f_1 = g\mathbb{1}_F$ et $g_1 = g\mathbb{1}_F$ sont mesurables. L'application $f_1 + g_1 \colon S \to \mathbb{R}$ est encore mesurable. Par conséquent l'application

$$f_1 g_1 = \frac{1}{2}\big((f_1 + g_1)^2 - f_1^2 - g_1^2\big) \tag{13.276}$$

est mesurable.

Voyons maintenant ce qui se passe en dehors de F. Nous allons utiliser le lemme de recollement sur la fonction

$$(fg)(x) = \begin{cases} (f_1 f_2)(x) & \text{si } x \in F \\ -\infty & \text{si } x \in \mathcal{U} \\ 0 & \text{si } x \in \mathcal{V} \\ +\infty & \text{si } x \in \mathcal{W} \end{cases} \tag{13.277}$$

où $F, \mathcal{U}, \mathcal{V}, \mathcal{W}$ forment une partition mesurable dénombrable [27] de S. Pour le sport nous montrons que $\mathcal{U}$ est mesurable :

$$\mathcal{U} = \big(\{f = -\infty\} \cap \{g > 0\}\big) \tag{13.278a}$$
$$\cup \big(\{f = +\infty\} \cap \{g < 0\}\big) \tag{13.278b}$$
$$\cup \big(\{g = -\infty\} \cap \{f > 0\}\big) \tag{13.278c}$$
$$\cup \big(\{g = +\infty\} \cap \{f < 0\}\big). \tag{13.278d}$$

□

Proposition 13.94.
Si $f_n \colon S \to \bar{\mathbb{R}}$ est une suite de fonctions mesurables, alors les fonctions $\inf_n f_n$ et $\sup_n f_n$ sont mesurables.

Démonstration. Nous avons les découpages

$$\{\inf_n f_n < a\} = \bigcup_n \{f_n < a\} \in \mathcal{F} \tag{13.279}$$

et

$$\{\sup_n f_n \leqslant a\} = \bigcap_n \{f_n \leqslant a\} \in \mathcal{F}. \tag{13.280}$$

Le théorème 13.87 permet de conclure. □

Note que pour (13.280) nous ne pouvions pas utiliser les inégalités strictes parce que $\{\sup_n f_n < a\}$ n'est pas spécialement égal à $\bigcap_n \{f_n < a\}$.

13.95.
La proposition 13.94 nous permet de définir les parties positives et négatives de f par $f^+ = \sup(f, 0)$ et $f^- = \sup(-f, 0)$. Ce sont des applications mesurables. Nous avons les décompositions

$$f = f^+ - f^- \tag{13.281a}$$
$$|f| = f^+ + f^-. \tag{13.281b}$$

Corolaire 13.96.
Si $f \colon S \to \bar{\mathbb{R}}$ est mesurable alors les applications f^+, f^- et $|f|$ sont mesurables en tant qu'applications $S \to \bar{\mathbb{R}}^+$.

27. Définition 13.91.

Démonstration. Nous faisons la preuve pour f^+. Nous savons que $f^+ \colon S \to \bar{\mathbb{R}}$ est mesurable par la proposition 13.94. Nous considérons l'injection canonique $f \colon \bar{\mathbb{R}}^+ \to \bar{\mathbb{R}}$ et

$$\begin{aligned} f_1^+ \colon S &\to \bar{\mathbb{R}}^+ \\ x &\mapsto f^+(x). \end{aligned} \tag{13.282}$$

Alors $f_1^+ = j \circ f^+$ est mesurable. Et c'est bien cela que nous voulions.

□

Note : f^+ et f_1^+ sont exactement les mêmes fonctions. Elles ne diffèrent que par la tribu que nous considérons sur l'espace d'arrivée. Nous allons à partir de maintenant les noter toutes deux f^+.

Remarque 13.97.
L'application $|f|$ peut être mesurable sans que f le soit. Soit en effet une partie $A \notin \mathcal{F}$, et posons

$$f(x) = \begin{cases} 1 & \text{si } x \in A \\ -1 & \text{si } x \in A^c. \end{cases} \tag{13.283}$$

Alors $f^{-1}(\{1\}) = A$ n'est pas mesurable alors que $|f|(x) = 1$ pour tout x.

Il est temps d'aller relire les définitions 7.145.

Proposition 13.98.
Si les fonctions $f_n \colon S \to \bar{\mathbb{R}}$ sont mesurables alors les fonctions $\limsup f_n$ et $\liminf f_n$ sont mesurables.

Démonstration. Par le lemme 7.147 nous écrivons $\limsup_n f_n(x) = \inf_{n\geqslant 1} \sup_{k\geqslant n} f_k(x)$. Pour chaque k nous considérons la fonction $g_k = \sup_{n\geqslant k} f_n$. Par la proposition 13.94, les fonctions g_k sont mesurables. En utilisant encore la même proposition, $\inf_{n\geqslant 1} g_k$ est encore mesurable. □

Proposition 13.99 ([281])**.**
Si $f_n \colon S \to \bar{\mathbb{R}}$ est une suite de fonctions mesurables telle dont la limite ponctuelle existe, alors la limite est mesurable.

Démonstration. Si la limite existe, elle est égale à la limite supérieure par le lemme 7.148. Or la limite supérieure est mesurable par la proposition 13.98. □

13.6.2 Fonction étagée

Définition 13.100 ([282])**.**
Soit $(S, \mathcal{F})$ un espace mesurable et une fonction $f \colon S \to \big(\bar{\mathbb{R}}, \mathcal{B}or(\bar{\mathbb{R}})\big)$. Il serait dommage de confondre les trois concepts suivants.

- *Une **fonction simple** est une fonction dont l'image est constituée d'un nombre fini de valeurs.*
- *Une **fonction étagée** est une fonction simple qui est elle-même une fonction mesurable.*
- *Une **fonction en escalier** est une fonction étagée dont les valeurs sont constantes sur des intervalles : ce sont donc des fonctions constantes par morceaux.*

Dans les trois cas, la fonction f peut être écrite comme somme de fonctions caractéristiques :

$$f(x) = \sum_{j=1}^{p} \alpha_j \mathbb{1}_{A_j}(x) \tag{13.284}$$

où $A_j = f^{-1}(\alpha_j)$. Ce qui change est la nature des A_j.

- Si f est simple, les A_j sont quelconques.

— Si f est étagée, les A_i peuvent être choisis mesurables parce que $\{\alpha_i\}$ est un borélien, ce qui fait que $A_i = f^{-1}(\alpha_i)$ un choix mesurable.

— Si f est en escalier, les A_i sont des intervalles.

Définition 13.101.
*La **forme canonique** d'une fonction simple f est la suivante. Soit $\{\alpha_i\}_{i=1,\dots,l}$ les valeurs distinctes prises par f et $A_i = f^{-1}(\alpha_i)$. La forme canonique de f est alors*

$$f = \sum_{i=1}^{l} \alpha_i \mathbb{1}_{A_i}. \tag{13.285}$$

Lemme 13.102.
Si f est une fonction simple dont la représentation canonique est

$$f = \sum_{i=1}^{l} \alpha_i \mathbb{1}_{A_i}, \tag{13.286}$$

alors

(1) les A_i sont disjoints,

(2) l'union fait S : $S = \bigcup_i A_i$.

Problèmes et choses à faire

Le lemme 13.103 et le théorème 13.105 disent la même chose alors que la preuve du théorème 13.105 est beaucoup plus compliquée.La démonstration du lemme serait fausse ?

M'est avis que ce que le théorème donne en plus est la convergence uniforme en cas de fonction bornée. La suite (13.287) ne va pas converger uniformément.

Lemme 13.103 (Limite croissante de fonctions étagées[1])**.**
Soit $f\colon (S,\mathcal{F}) \to \bar{\mathbb{R}}$ une fonction positive mesurable. Il existe une suite $f_n\colon S \to \mathbb{R}$ de fonctions étagées positives telles que $f_n \to f$ ponctuellement et $f_n \leqslant f$.

Démonstration. Nous considérons (q_n) une suite parcourant tous les rationnels positifs[28] avec $q_0 = 0$ pour être sûr. Pour $n \in \mathbb{N}$ nous définissons la fonction

$$f_n(x) = \max\{q_i \text{ tel que } i \leqslant n,\ q_i \leqslant f(x)\}. \tag{13.287}$$

L'ensemble sur lequel le maximum est pris n'est pas vide parce que $q_0 = 0$. La fonction f_n est simple parce qu'elle ne prend que n valeurs différentes. Nous avons aussi par construction que $f_n(x) \leqslant f(x)$. Nous avons aussi pour tout $x \in S$ que $f_n(x) \to f(x)$ parce que $\mathbb{Q}$ est dense dans $\mathbb{R}$.

En ce qui concerne le fait que f_n est mesurable, nous notons $\{r_0,\dots,r_n\}$ l'ensemble des $\{q_0,\dots,q_n\}$ classés dans l'ordre croissant. Nous posons en plus $r_{n+1} = +\infty$. Nous avons alors

$$f_n^{-1}(r_k) = \{x \in S \text{ tel que } f(x) \geqslant r_k, f(x) < r_{k+1}\} = \{f \geqslant r_k\} \cap \{f < r_{k+1}\}. \tag{13.288}$$

En tant qu'intersection de deux ensembles mesurables, le théorème 13.87 dit que $f_n^{-1}(r_k)$ est mesurable. □

Remarque 13.104.
Pour avoir $f_n < |f|$ nous pouvons poser

$$f_n(x) = \begin{cases} \max\{q_i \text{ tel que } i \leqslant n,\ q_i \leqslant f(x)\} & \text{si } f(x) \geqslant 0 \\ \min\{q_i \text{ tel que } i \leqslant n,\ q_i \geqslant f(x)\} & \text{si } f(x) < 0. \end{cases} \tag{13.289}$$

Théorème 13.105 (Théorème fondamental d'approximation, thème 3[277, 283, 284])**.**
Soit un espace mesuré $(S,\mathcal{A},\mu)$.

28. Nous rappelons que $\mathbb{Q}$ est dénombrable et dense dans $\mathbb{R}$ par la proposition 7.121.

(1) Soit une fonction mesurable $f\colon S \to [0,+\infty]$. Alors il existe une suite croissante de fonctions $\varphi_n\colon S \to [0,+\infty[$ étagées positives dont la limite ponctuelle est f.

(2) Si de plus f est bornée, la convergence est uniforme.

(3) Idem pour f à valeurs dans $\bar{\mathbb{R}}$ ou $\mathbb{C}$.

Démonstration. Nous découpons l'intervalle $[0,n]$ en plusieurs morceaux.

$$I_{n,k} = \begin{cases} [\frac{k}{2^n}, \frac{k+1}{2^n}[& \text{si } 0 \leqslant k \leqslant n2^n - 1 \\ [n,\infty] & \text{si } k = n2^n. \end{cases} \tag{13.290}$$

Nous posons $S_{n,k} = f^{-1}(I_{n,k})$. Ce sont des ensembles mesurables parce que f est mesurable. Et de plus pour chaque n, la suite $(S_{n,k})_{k\geqslant 0}$ est une partition mesurable finie de S. Nous posons

$$\varphi_n = \sum_{k=0}^{n2^n} \frac{k}{2^n} \mathbb{1}_{S_{n,k}}. \tag{13.291}$$

C'est-à-dire que sur chaque $S_{n,k}$ nous approximons f par le bas. La fonction φ_n est étagée et positive : $0 \leqslant \varphi_n(x) \leqslant f(x)$ par construction.

Croissance Nous allons voir que $\varphi_n \leqslant \varphi_{n+1}$. Soit $k \neq n2^n$. Si $x \in S_{n,k}$ alors $\varphi_n(x) = \frac{k}{2^n}$ et nous avons aussi la décomposition

$$S_{n,k} = S_{n+1,2k} \cup S_{n+,2k+1}. \tag{13.292}$$

Si $x \in S_{n+1,2k}$ alors $\varphi_{n+1}(x) = \frac{2k}{2^{n+1}} = \frac{k}{2^n} = \varphi_n(x)$. Et si $x \in S_{n+1,2k+1}$ alors

$$\varphi_{n+1}(x) = \frac{2k+1}{2^{n+1}} = \frac{k+\frac{1}{2}}{2^n} > \varphi_n(x). \tag{13.293}$$

Il reste à traiter le cas $x \in \{f \geqslant n\}$. Dans ce cas nous avons $\varphi_n(x) = n$. Il y a encore deux cas à traiter :

$$\{f \geqslant n\} = \{f \in [n,n+1[\} \cup \{f \in [n+1,\infty]\}. \tag{13.294}$$

Pour plus de simplicité dans les notations, nous notons $\bar{n} = n2^n$, c'est-à-dire que $I_{n,\bar{n}}$ est le $I_{n,k}$ avec le k le plus grand possible. Nous avons

$$I_{n,\bar{n}} = [n,n+1[\cup [n+1,\infty]. \tag{13.295}$$

Le premier élément se décompose en $I_{n+1,k}$ avec $k < n+1$ (nous préciserons plus tard exactement les valeurs de k) tandis que le second est $[n+1,\infty] = I_{n+1,\overline{n+1}}$.
Pour $x \in S_{n+1,\overline{n+1}}$ nous avons

$$\varphi_{n+1}(x) = \frac{(n+1)2^{n+1}}{2^{n+1}} = n+1 > \varphi_n(x). \tag{13.296}$$

Si au contraire $f(x) \in [n,n+1[$ nous devons précisément voir quels sont les k qui font en sorte que $I_{n+1,k}$ recouvre $[n,n+1[$. Le plus petit k est donné par $\frac{k}{2^{n+1}} = n$, c'est-à-dire $k = n2^{n+1}$ et le plus grand k est donné par $\frac{k}{2^{n+1}} < n+1$, c'est-à-dire $k = 2^{n+1}(n+1) - 1$. Donc si $f(x) \in [n,n+1[$ alors $x \in S_{n+1,k}$ avec

$$n2^{n+1} \leqslant k \leqslant (n+1)2^{n+1} - 1 \tag{13.297}$$

Dans ce cas

$$\varphi_{n+1}(x) = \frac{k}{2^{n+1}} \geqslant \frac{n2^{n+1}}{2^{n+1}} = n = \varphi_n(x). \tag{13.298}$$

Convergence ponctuelle Si $f(x) < \infty$ alors il existe[29] $n_0 \in \mathbb{N}$ tel que $f(x) < n_0$. Pour $bn \geqslant n_0$ nous avons $f(x) < n$ et donc $\varphi_n(x)$ se calcule à partir d'un des intervalles de taille $1/2^n$:

$$\varphi_n(x) = \frac{k}{2^n} \leqslant f(x) < \frac{k+1}{2^n}. \tag{13.299}$$

Donc

$$|\varphi_n(x) - f(x)| \leqslant \frac{1}{2^n}, \tag{13.300}$$

ce qui signifie que $\lim_{n\to\infty} \varphi_n(x) = f(x)$.

Si $f(x) = +\infty$ alors $f(x) > n$ pour tout n. Et alors $\varphi_n(x) = n$ pour tout n, ce qui donne bien $\varphi_n(x) \to \infty$.

Convergence uniforme Soit f bornée : $0 \leqslant f(x) < M$ pour tout $x \in S$. Soit aussi $\epsilon > 0$. Nous prenons $n_0 > M$ tel que $\frac{1}{2^{n_0} < \epsilon}$. Alors pour tout $n \geqslant n_0$ nous avons

$$0 \leqslant f(x) - \varphi_n(x) \leqslant \frac{1}{2^n} \leqslant \frac{1}{2^{n_0}} \leqslant \epsilon. \tag{13.301}$$

Note qu'il n'y a pas de valeurs absolues parce que nous savons déjà que la limite est croissante.

□

13.6.3 Fonctions réelles à variables réelles

Nous nous particularisons à présent au cas de fonctions

$$f\colon \big(\mathbb{R}, \mathcal{B}\mathrm{or}(\mathbb{R})\big) \to \big(\bar{\mathbb{R}}, \mathcal{B}\mathrm{or}(\bar{\mathbb{R}})\big). \tag{13.302}$$

13.106 ([1]).
Anticipons un peu pour expliquer pourquoi ce que nous allons faire maintenant est suffisant pour ce que nous avons en tête[30]. Toutes les fonctions mesurables

$$f\colon \big(\mathbb{R}, \mathcal{B}\mathrm{or}(\mathbb{R})\big) \to \big(\bar{\mathbb{R}}, \mathcal{B}\mathrm{or}(\bar{\mathbb{R}})\big) \tag{13.303}$$

seront a fortiori mesurables au sens de

$$f\colon \big(\mathbb{R}, \mathcal{L}\mathrm{eb}(\mathbb{R})\big) \to \big(\bar{\mathbb{R}}, \mathcal{B}\mathrm{or}(\bar{\mathbb{R}})\big) \tag{13.304}$$

où $\mathcal{L}\mathrm{eb}(\mathbb{R})$ est la tribu de Lebesgue sur $\mathbb{R}$, c'est-à-dire la tribu complétée de celle des boréliens (définition 13.127).

13.107.
Nous allons maintenant donner quelques conditions pour que des fonctions soient mesurables au sens de la tribu des boréliens sur l'espace d'arrivée et de départ. Ces résultats seront donc immédiatement applicables à la théorie de l'intégration où nous considérons la tribu de Lebesgue sur l'espace de départ.

Autrement dit, les résultats présentés ici sont un peu plus forts que ce dont nous avons réellement besoin ... ou alors ce sont les hypothèses que nous allons nous mettre en théorie de l'intégration qui seront un peu plus fortes que nécessaires. C'est une question de point de vue.

Corolaire 13.108.
Si I est un intervalle de $\mathbb{R}$, alors toute application monotone $f\colon I \to \mathbb{R}$ est borélienne.

Démonstration. Vu que f est monotone, l'ensemble $\{f < a\}$ est un intervalle. Or tous les intervalles sont boréliens, donc f est mesurable par le théorème 13.87. □

29. Le vrai snob citera ici le lemme 1.141.
30. Pour rappel, nous avons en tête de définir une théorie de la mesure afin d'y définir des intégrales. En particulier nous allons étudier l'inttégrale de Lebesgue et en ce qui concerne $\mathbb{R}^n$, nous aurons la tribu de Lebesgue.

Définition 13.109.
Si I est un intervalle de $\mathbb{R}$, une fonction $f\colon I \to \mathbb{R}$ a une propriété (monotone, mesurable, continue, etc.) ***par morceaux*** *s'il existe une suite strictement croissante de points $(x_I)_{i\in\mathbb{Z}}$ dans I telle que f ait la propriété sur chacun des ouverts $]x_j, x_{j+1}[$..*

Dans cette définition, les points sont numérotés par $\mathbb{Z}$ et non par $\mathbb{N}$ parce que nous nous laissons la liberté d'avoir une infinité de points de chacun des deux côtés.

Proposition 13.110.
Soit I un intervalle de $\mathbb{R}$ et une fonction $f\colon I \to \mathbb{R}$. Si f est continue ou monotone par morceaux sur I alors elle y est borélienne.

Démonstration. L'ensemble $\{]x_j, x_{j+1}[\}_{j\in\mathbb{Z}} \cup \{x_i\}_{i\in\mathbb{Z}}$ forme une partition mesurable dénombrable de I (les singletons sont des boréliens). À une belle redéfinition près de la numérotation (deux fois $\mathbb{Z}$ va dans $\mathbb{N}$), nous les appelons $(I_n)_{n\in\mathbb{N}}$, et nous définissons les fonctions f_n comme étant les restrictions de f aux intervalles I_k.

Toute fonction sur un singleton est mesurable. Toute fonction continue sur un ouvert est mesurable (théorème 13.49). Toute fonction monotone sur un ouvert est mesurable (corolaire 13.108).

Le lemme de recollement 13.92 donne alors la mesurabilité de f. □

13.111.
Toutes les fonctions que nous pouvons écrire explicitement sont mesurables ... en tout cas toutes celles que l'on trouve en pratique. En effet nous avons déjà toutes les fonctions continues par morceaux via la proposition 13.110 et ensuite toutes les limites par la proposition 13.99. Cela donne les séries, les dérivées, les primitives, etc.

13.7 Tribu produit, mesure produit

13.7.1 Produit d'espaces mesurables

Définition 13.112.
Si $\mathcal{A}_1$ et $\mathcal{A}_2$ sont deux tribus sur deux ensembles Ω_1 et Ω_2, nous définissons la ***tribu produit*** *$\mathcal{A}_1 \otimes \mathcal{A}_2$ comme étant la tribu engendrée par*

$$\{X \times Y \text{ tel que } X \in \mathcal{A}_1, Y \in \mathcal{A}_2\}. \tag{13.305}$$

Ces ensembles sont appelés ***rectangles*** *de $(\Omega_1, \mathcal{A}_1) \otimes (\Omega_2, \mathcal{A}_2)$.*

Proposition 13.113 ([285]).
Soient deux espaces mesurables $(S_1, \mathcal{F}_1)$ et $(S_2, \mathcal{F}_2)$. Si $\mathcal{C}_i$ est une classe de parties de S_i avec $\mathcal{F}_i = \sigma(\mathcal{C}_i)$ et $S_i \in \mathcal{C}_i$. Alors

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{C}_1 \times \mathcal{C}_2). \tag{13.306}$$

Démonstration. Nous notons p_1 et p_2 les projections de $S_1 \times S_2$ vers S_1 et S_2. Nous commençons par prouver que

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma\big(\mathrm{P}_1^{-1}(\mathcal{F}_1) \cup p_2^{-1}(\mathcal{F}_2)\big). \tag{13.307}$$

En effet cette union est dans $\mathcal{F}_1 \otimes \mathcal{F}_2$ parce que ce sont tous des produits de la forme $A_1 \times S_2$ et $S_1 \times A_2$ où $A_i \in \mathcal{F}_i$. Inversement, tous les produits de la forme $A_1 \times A_2$ sont dans la tribu engendrée par l'union parce que

$$A_1 \cup A_2 = (A_1 \times S_2) \cap (S_1 \times A_2). \tag{13.308}$$

Par conséquent, la partie $p_1^{-1}(\mathcal{F}_1) \cup p_2^{-1}(\mathcal{F}_2)$ engendre tous les produits qui engendrent la tribu $\mathcal{F}_1 \otimes \mathcal{F}_2$. L'égalité (13.307) est donc correcte.

Si $C_1 \in \mathcal{C}_1$ alors

$$p_1^{-1}(C_1) = C_1 \times S_2 \in \mathcal{C}_1 \times \mathcal{C}_2 \tag{13.309}$$

et donc $p_1^{-1}(\mathcal{C}_1) \subset \mathcal{C}_1 \times \mathcal{C}_2$. En utilisant le lemme de transport 13.47 nous avons alors

$$p_1^{-1}(\mathcal{F}_1) = p_1^{-1}\big(\sigma(\mathcal{C}_1)\big) = \sigma\big(p_1^{-1}\mathcal{C}_1\big) \subset \sigma(\mathcal{C}_1 \times \mathcal{C}_1) \tag{13.310}$$

et au bout de la même façon,

$$p_2^{-1}(\mathcal{F}_1) \subset \sigma(\mathcal{C}_1 \times \mathcal{C}_2). \tag{13.311}$$

Vu les relations (13.310), (13.311) et (13.307) nous avons

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma\big(\mathrm{P}_1^{-1}(\mathcal{F}_1) \cup p_2^{-1}(\mathcal{F}_2)\big) \subset \sigma(\mathcal{C}_1 \times \mathcal{C}_2). \tag{13.312}$$

Réciproquement, si $C_1 \in \mathcal{C}_1$ et $C_2 \in \mathcal{C}_2$ alors

$$C_1 \times C_2 = (C_1 \times S_1) \cap (S_1 \times C_2) = p_1^{-1}(C_1) \cap p_2^{-1}(C_2) \in \mathcal{F}_1 \otimes \mathcal{F}_2. \tag{13.313}$$

□

13.7.2 Le cas des boréliens

Si X_1 et X_2 sont des espaces topologiques et si nous notons $\mathcal{O}_i$ l'ensemble de leurs ouverts, par définition $\mathcal{B}\mathrm{or}(X_i) = \sigma(\mathcal{O}_i)$. De plus par la proposition 13.113 nous savons que

$$\sigma(\mathcal{O}_1 \times \mathcal{O}_2) = \mathcal{B}\mathrm{or}(X_1) \otimes \mathcal{B}\mathrm{or}(X_2). \tag{13.314}$$

Lemme 13.114.
Si $(X_i, \mathcal{O}_i)$ sont des espaces topologiques, alors

$$\mathcal{B}or(X_1) \otimes \mathcal{B}or(X_2) \subset \mathcal{B}or(X_1 \times X_2) \tag{13.315}$$

Démonstration. Si $A_i \in \mathcal{O}_i$ alors $A_1 \times A_2$ est un ouvert de $X_1 \times X_2$ (voir la définition 7.12). Par conséquent, $\mathcal{O}_1 \times \mathcal{O}_2$ est contenu dans l'ensemble des ouverts de $X_1 \times X_2$ ou encore

$$\mathcal{O}_1 \times \mathcal{O}_2 \subset \mathcal{B}\mathrm{or}(X_1 \times X_2), \tag{13.316}$$

et donc

$$\sigma(\mathcal{O}_1 \times \mathcal{O}_2) \subset \sigma\big(\mathcal{B}\mathrm{or}(X_1 \times X_2)\big) \tag{13.317}$$

finalement, par (13.314)

$$\mathcal{B}\mathrm{or}(X_1) \otimes \mathcal{B}\mathrm{or}(X_2) \subset \mathcal{B}\mathrm{or}(X_1 \times X_2). \tag{13.318}$$

□

Il n'y a en général pas égalité, mais nous allons immédiatement voir que dans (presque) tous les cas raisonnables, les boréliens sur un produit sont le produit des boréliens.

Proposition 13.115 ([285])**.**
Soient (X_1, d_1) et (X_2, d_2) des espaces métriques séparables. Alors

$$\mathcal{B}or(X_1 \times X_2) = \mathcal{B}or(X_1) \otimes \mathcal{B}or(X_2). \tag{13.319}$$

Démonstration. Nous savons par le lemme 7.255 que tout ouvert de $X_1 \times X_2$ est une réunion dénombrable d'éléments de $\mathcal{O}_1 \times \mathcal{O}_2$. Donc tout ouvert de $X_1 \times X_2$ est dans $\mathcal{B}\mathrm{or}(X_1) \otimes \mathcal{B}\mathrm{or}(X_2)$. Par conséquent

$$\mathcal{B}\mathrm{or}(X_1 \times X_2) \subset \mathcal{B}\mathrm{or}(X_1) \otimes \mathcal{B}\mathrm{or}(X_2). \tag{13.320}$$

L'inclusion inverse étant déjà acquise par le lemme 13.114, nous avons l'égalité. □

Proposition 13.116.
Les boréliens sur $\mathbb{R}^N$ sont ceux qu'on croit.

(1) $\mathcal{B}or(\mathbb{R}^2) = \mathcal{B}or(\mathbb{R}) \otimes \mathcal{B}or(\mathbb{R})$

(2) $\mathcal{B}or(\mathbb{R}^{N+1}) = \mathcal{B}or(\mathbb{R}^N) \otimes \mathcal{B}or(\mathbb{R})$

Démonstration. Cela n'est rien d'autre que la proposition 13.115. □

Proposition 13.117.
Soit un espace mesurable $(S, \mathcal{F})$ et des applications $f_k \colon S \to \mathbb{R}$ $(k = 1, \dots, N)$. Alors l'application

$$\begin{aligned} f \colon (S, \mathcal{F}) &\to (\mathbb{R}^N, \mathcal{B}or(\mathbb{R}^N)) \\ x &\mapsto \big(f_1(x), \dots, f_N(x)\big) \end{aligned} \tag{13.321}$$

est mesurable si et seulement si chacun des f_i est mesurable.

Démonstration. Division en deux.

Condition nécessaire Nous supposons que les f_i sont mesurables. Nous avons

$$f^{-1}\Big(\prod_{k=1}^{N}]a_k, b_k[\Big) = \{x \in S \text{ tel que } f_1(x) \in]a_1, b_1[, \cdots f_N(x) \in]a_N, b_N[\} \tag{13.322a}$$

$$= \bigcap_{k=1}^{N} f_k^{-1}\big(]a_k, b_k[\big). \tag{13.322b}$$

Cela est une intersection finie d'éléments de $\mathcal{F}$ et est donc un élément de $\mathcal{F}$. Mais les pavés ouverts engendrent $\mathcal{B}\mathrm{or}(\mathbb{R}^N)$ parce qu'ils sont une base dénombrable de la topologie (proposition 13.73). Le théorème 13.48 nous assure alors que f est mesurable parce que l'image inverse d'une base de la tribu est mesurable.

Condition suffisante Si f est mesurable alors en particulier

$$f_k^{-1}\big(]a, b[\big) = f^{-1}\big(\mathbb{R} \times \dots \times]a, b[\times \mathbb{R} \times \dots \times \mathbb{R}\big) \in \mathcal{F}. \tag{13.323}$$

Pour cela nous avons utilisé la proposition 13.116 qui nous indique que le produit dans la parenthèse est un borélien de $\mathbb{R}^N$ en tant que produit de boréliens de $\mathbb{R}$.

Encore une fois f_k^{-1} tombe dans $\mathcal{F}$ pour une base dénombrable de la topologie de $\mathbb{R}$ et est donc mesurable.

□

13.8 Mesure de Lebesgue sur $\mathbb{R}$

Nous notons $\mathcal{S}$ l'ensemble des intervalles[31] de $\mathbb{R}$.

Proposition 13.118.
L'ensemble réunions finies d'éléments de $\mathcal{S}$ est une algèbre de parties de $\mathbb{R}$ que nous allons noter $\mathcal{A}_{\mathcal{S}}$.

Démonstration. Nous devons vérifier la définition 13.16. Les ensembles $\mathbb{R}$ et $\emptyset$ sont des intervalles et font donc partie de $\mathcal{A}_{\mathcal{S}}$.

Si $A \in \mathcal{A}_{\mathcal{S}}$ se décompose en union d'intervalles de la forme (a_k, b_k) avec $k = 1, \dots, n$ (ici nous mettons des parenthèses au lieu de crochets parce qu'à priori nous ne savons pas). Alors

$$A^c = \bigcup_{k=0}^{k} (b_k, a_{k+1}) \tag{13.324}$$

où nous avons posé $b_0 = -\infty$ et $a_{n+1} = +\infty$. Ici encore les parenthèses sont soit fermées soit ouvertes en fonction de ce qu'étaient celles dans la décomposition de A. Quoi qu'il en soit, cette décomposition de A^c montre que $A^c \in \mathcal{A}_{\mathcal{S}}$.

Enfin si $A, B \in \mathcal{A}_{\mathcal{S}}$ alors $A \cup B \in \mathcal{A}_{\mathcal{S}}$. □

31. Définition 1.13.

Lemme 13.119.
Tout élément de $\mathcal{A}_{\mathcal{S}}$ admet une décomposition minimale unique en réunion finie d'intervalles. Cette décomposition est formée d'intervalles deux à deux disjoints.

Démonstration. Nous allons montrer que si $A \in \mathcal{A}_{\mathcal{S}}$, alors la décomposition minimale consiste en les composantes connexes[32] de A. Pour cela nous rappelons que la proposition 7.153 dit qu'une partie de $\mathbb{R}$ est connexe si et seulement si elle est un intervalle. D'abord cela nous dit immédiatement que les composantes connexes de A forment une décomposition de A en intervalles. Nous devons prouver qu'elle est minimale.

Soit $\{C_k\}_{k=1,\dots,n}$ les composantes connexes de A. Aucun connexe de $\mathbb{R}$ contenu dans A ne peut intersecter plus d'un des C_k, et par conséquent nous ne pouvons pas décomposer A en moins de n intervalles.

Pour l'unicité, soit $\{I_k\}_{k=1,\dots,n}$ un ensemble de n intervalles tels que $\bigcup_{k=1}^n I_k = A$. Chacun des I_k intersecte un et un seul des C_k. En effet si $x \in I_k \cap C_i$ et $y \in I_k \cap C_j$, alors $[x,y] \subset I_k$ parce que I_k est un intervalle. Mais C_i étant le plus grand connexe contenant x, $[x,y] \subset C_i$ et de la même façon, $[x,y] \subset C_j$. Par conséquent C_i et C_j sont tous deux la composante connexe de x et y. Nous en déduisons que $C_i = C_j$, c'est-à-dire $i = j$.

Par ailleurs nous avons $I_k \cap I_l = \emptyset$ dès que $k \neq l$ parce que sinon l'ensemble $I_k \cap I_l$ serait connexe et la décomposition des $\{I_k\}_{k=1,\dots,n}$ ne serait pas minimale : en remplaçant I_k et I_l par $I_k \cup I_l$ on aurait eu une décomposition contenant moins d'éléments. Donc à renumérotation près nous pouvons supposer que I_k intersecte C_l si et seulement si $k = l$.

Dans ce cas nous devons avoir $I_k = C_k$, sinon les éléments de $C_k \backslash I_k$ ne seraient pas dans $\bigcup_{i=1}^n I_i$. □

Définition 13.120 (longueur d'intervalle[273])**.**
Si I est un intervalle d'extrémités a et b avec $-\infty \leqslant a \leqslant b \leqslant +\infty$ alors nous définissons la ***longueur*** *de I par*

$$\ell(I) = \begin{cases} b - a & \text{si } -\infty < a \leqslant b < +\infty \\ \infty & \text{si } a \text{ ou } b \text{ est infini} \end{cases} \tag{13.325}$$

Si $A \in \mathcal{A}_{\mathcal{S}}$ et si sa décomposition minimale est $A = \bigcup_{k=1}^n I_k$, alors on définit

$$\ell(A) = \sum_{k=1}^n \ell(I_k). \tag{13.326}$$

Le lemme suivant nous indique que nous pouvons calculer la longueur d'un élément de $\mathcal{A}_{\mathcal{S}}$ sans savoir la décomposition minimale, pourvu que l'on connaisse une décomposition disjointe.

Lemme 13.121 ([273])**.**
Si

$$B = \bigcup_{r=1}^p J_r \tag{13.327}$$

est une décomposition de $B \in \mathcal{A}_{\mathcal{S}}$ en intervalles deux à deux disjoints alors

$$\ell(B) = \sum_{r=1}^p \ell(J_r). \tag{13.328}$$

Démonstration. Nous prouvons dans un premier temps le résultat dans le cas où $B = I$ est un intervalle. Soit I un intervalle et une décomposition en intervalles disjoints $I = \bigcup_{r=1}^p J_r$. Nous montrons qu'alors $\ell(I) = \sum_{r=1}^p \ell(J_r)$. Nous verrons ensuite comment passer au cas où B est un élément générique de $\mathcal{A}_{\mathcal{S}}$.

<u>Si $B = I$ est un intervalle infini</u> Si I est infini alors un des J_r soit l'être et donc $\sum_{r=1}^p \ell(J_r) = \infty = \ell(I)$.

32. Définition 7.43.

Si $B = I$ est un intervalle ininfini Pour chaque $r = 1, \dots, p$ nous notons a_r et b_r les extrémités de J_r. Vu que les J_r sont connexes et disjoints, si $a_k \leqslant a_l$ alors $b_k \leqslant a_l$, sinon l'ensemble (non vide) $]a_l, b_k[$ serait dans l'intersection $I_k \cap I_l$ qui, elle, est vide. Plus généralement, si $x \in J_k$ et $y \in J_l$ avec $x < y$ alors pour tout $x' \in J_k$ et tout $y' \in J_l$ nous avons $x' < y'$. Vu qu'il y a un nombre fini d'ensembles J_r, nous pouvons les classer dans l'ordre croissant :

$$a_1 \leqslant b_1 \leqslant a_2 \leqslant b_2 \leqslant \dots \leqslant b_{p-1} \leqslant a_p \leqslant b_p. \tag{13.329}$$

Vu que les J_r sont disjoints et que leur union est connexe nous avons en réalité

$$a = a_1 \leqslant b_1 = a_2 \leqslant b_2 = a_3 \leqslant \dots \leqslant b_{p-1} = a_p \leqslant b_p, \tag{13.330}$$

donc une somme télescopique donne

$$\ell(I) = b - a = \sum_{r=1}^{p} (b_r - a_r) = \sum_{r=1}^{p} \ell(J_r). \tag{13.331}$$

Si B n'est pas un intervalle Soit $\{I_k\}_{k=1,\dots,n}$ la décomposition minimale de B. Alors

$$\spadesuit = \ell(B) = \sum_{k=1}^{n} \ell(I_k) = \sum_{k=1}^{n} \ell\big(\bigcup_{r=1}^{p} (I_k \cap J_r)\big). \tag{13.332}$$

Mais I_k est un intervalle et s'écrit comme union disjointe $I_k = \bigcup_{r=1}^{p} (I_k \cap J_r)$, donc par la première partie

$$\spadesuit = \sum_{k=1}^{n} \sum_{r=1}^{p} \ell(I_k \cap J_r) = \sum_{r=1}^{p} \sum_{k=1}^{n} \ell(I_k \cap J_r). \tag{13.333}$$

Ici J_r est un intervalle qui se décompose en $J_r = \bigcup_{k=1}^{n} (I_k \cap J_r)$, donc nous pouvons encore utiliser la première partie :

$$\spadesuit = \sum_{r=1}^{p} \ell(J_r), \tag{13.334}$$

ce qu'il fallait.

□

Lemme 13.122.
Si $A, B \in \mathcal{A}_{\mathcal{S}}$ avec $A \subset B$ alors $\ell(A) \leqslant \ell(B)$.

Démonstration. Nous avons évidemment $B = A \cup B \backslash A$. Notons que $B \backslash A \in \mathcal{A}_{\mathcal{S}}$ par le lemme 13.17. Si $\{I_k\}$ est une décomposition disjointe de A et $\{J_i\}$ une de $B \backslash A$ alors $\{I_k\} \cup \{J_i\}$ est une décomposition disjointe de $A \cup B \backslash A$ et le lemme 13.121 nous dit que

$$\ell(B) = \ell(A \cup B \backslash A) = \ell(A) + \ell(B \backslash A). \tag{13.335}$$

Par conséquent $\ell(B) \geqslant \ell(A)$. □

Lemme 13.123.
Si I est un intervalle et s'il se décompose en

$$I = \bigcup_{n \in \mathbb{N}} I_n \tag{13.336}$$

où les I_n sont des intervalles disjoints, alors

$$\ell(I) = \sum_{n=1}^{\infty} \ell(I_n). \tag{13.337}$$

Démonstration. Nous allons encore diviser la preuve en deux parties suivant que I soit de longueur finie ou pas.

Si I est de longueur finie Soient a et b les extrémités de I : $-\infty < a \leqslant b < +\infty$. Pour tout $N \geqslant 1$ nous avons

$$\sum_{n=1}^{N} \ell(I_n) = \ell\big(\bigcup_{n=1}^{n} I_n\big) \leqslant \ell(I). \tag{13.338}$$

La première égalité est le lemme dans le cas d'une union finie 13.121. L'inégalité est le lemme 13.122. Cela étant vrai pour tout N, à la limite $N \to \infty$ nous conservons l'inégalité :

$$\sum_{n=1}^{\infty} \ell(I_n) \leqslant \ell(I). \tag{13.339}$$

Nous devons encore voir l'inégalité inverse. Pour cela nous supposons que $a < b$. Sinon $\ell(I) = 0$ et tous les I_n doivent être vide sauf un qui contiendra seulement $\{a\}$ (si I le contient).

Soit $\epsilon > 0$ avec $\epsilon < b - a$ et l'intervalle

$$[a + \frac{\epsilon}{4}, b - \frac{\epsilon}{4}] = [a', b'] \subset I. \tag{13.340}$$

Si les a_n et le b_n sont le extrémités des I_n alors

$$[a', b'] \subset I = \bigcup_{n\geqslant 1} I_n \subset \bigcup_{n\geqslant 1}]a_n - \frac{\epsilon}{2^{n+2}}, b_n + \frac{\epsilon}{2^{n+2}}[= \bigcup_{n\geqslant 1}]a'_n, b'_n[\tag{13.341}$$

où nous avons posé $a'_n = a_n - \epsilon/2^{n+2}$ et $b'_n = b_n + \epsilon/2^{n+2}$. Nous avons donc recouvert le compact [33] $[a', b']$ par des ouverts. Nous pouvons donc en extraire un sous-recouvrement fini (c'est la définition de la compacité), c'est-à-dire une partie finie F de $\mathbb{N}$ telle que

$$[a', b'] \subset \bigcup_{n\in F}]a'_n, b'_n[. \tag{13.342}$$

Le lemme 13.122 nous dit alors que

$$\heartsuit = b' - a' \leqslant \ell\big(\bigcup_{n\in F}]a'_n, b'_n[\big) \leqslant \sum_{n\in F} (b'_n - a'_n). \tag{13.343}$$

La seconde inégalité se prouve en recopiant [34] la preuve de 13.18. Nous continuons le calcul :

$$\heartsuit \leqslant \sum_{n\in F} (b_n - a_n) + \sum_{n\in F} \frac{\epsilon}{2^{n+1}} \leqslant \sum_{n\in F} (b_n - a_n) + \frac{\epsilon}{2}. \tag{13.344}$$

Mais $b' - a' = (b - a) - \frac{\epsilon}{2}$, donc

$$b - a - \frac{\epsilon}{2} \leqslant \sum_{n\in F} (b_n - a_n) + \frac{\epsilon}{2}. \tag{13.345}$$

D'où nous déduisons que

$$\ell(I) = b - a \leqslant \sum_{n\in F} (b_n - a_n) + \epsilon \leqslant \sum_{n\in \mathbb{N}} (b_n - a_n) + \epsilon = \sum_{n\in \mathbb{N}} \ell(I_n) + \epsilon. \tag{13.346}$$

Cela étant valable pour tout ϵ nous déduisons que

$$\ell(I) \leqslant \sum_{n\in \mathbb{N}} \ell(I_n). \tag{13.347}$$

33. Lemme 7.125.

34. Nous ne pouvons pas invoquer directement le lemme 13.18 parce que nous n'avons pas encore prouvé que ℓ était une mesure sur $(\mathbb{R}, \mathcal{A}_{\mathcal{S}})$.

Si I est de longueur infinie Étant donné que I est un intervalle de longueur infinie, il doit au moins contenir un ensemble du type $]-\infty, a]$ ou $[a, +\infty[$; donc pour tout $M > 0$, il existe $N \geqslant 1$ tel que

$$\ell\big(I \cap [-N, N]\big) \geqslant M. \tag{13.348}$$

Mais $I \cap [-N, N]$ est un intervalle et

$$I \cap [-N, N] = \bigcup_{n\in\mathbb{N}} I_n \cap [-N, N] \tag{13.349}$$

qui est une union disjointe. Par conséquent,

$$M \leqslant \ell\big(I \cap [-N, N]\big) = \sum_n \ell\big(I_n \cap [-N, N]\big) \leqslant \sum_n \ell(I_n). \tag{13.350}$$

Cela étant vrai pour tout $M > 0$, nous concluons que

$$\sum_{n\in\mathbb{N}} \ell(I_n) = \infty. \tag{13.351}$$

□

Remarque 13.124.
Pour la preuve de 13.123 nous ne pouvons pas classer les I_n en ordre croissant comme nous l'avons fait dans la preuve de 13.121. En effet si $I = [0, 1]$ et que nous recouvrons $[0, \frac{1}{2}[$ et $]\frac{1}{2}, 1]$ par une infinité d'intervalles chacun, nous ne pouvons plus les classer par ordre croissant.

Proposition 13.125 ([273])**.**
La fonction ℓ ainsi définie est une mesure σ-finie sur l'algèbre de parties $\mathcal{A}_{\mathcal{S}}$.

Démonstration. Le fait que ℓ soit σ-finie provient par exemple du fait que $\ell\big(]-n, n[\big) = 2n$ tandis que $\bigcup_n]-n, n[= \mathbb{R}$.

Nous devons à présent prouver que ℓ est additive. Soient $(A_i)_{i\in\mathbb{N}}$ des éléments disjoints de $\mathcal{A}_{\mathcal{S}}$, avec leurs décomposition minimales

$$A_i = \bigcup_{k=1}^{n} I_k^{(i)}. \tag{13.352}$$

Pour chaque $i \in \mathbb{N}$, le lemme 13.123 nous indique que

$$\ell(A_i) = \sum_{k\in\mathbb{N}} \ell(I_k^{(i)}). \tag{13.353}$$

L'ensemble $\mathbb{N} \times \mathbb{N}$ est dénombrable et nous pouvons considérer la décomposition

$$\bigcup_{i\in\mathbb{N}} A_i = \bigcup_{(i,k)\in\mathbb{N}\times\mathbb{N}} I_k^{(i)}. \tag{13.354}$$

Cette décomposition n'est pas spécialement minimale[35] mais elle est disjointe. Le lemme 13.123 donne

$$\ell(\bigcup_i A_i) = \sum_{(i,k)\in\mathbb{N}\times\mathbb{N}} \ell(I_k^{(i)}) = \sum_{i\in\mathbb{N}} \left(\sum_{k\in\mathbb{N}} \ell(I_k^{(i)}) \right) = \sum_{i\in\mathbb{N}} \ell(A_i). \tag{13.355}$$

La décomposition de la somme sur $\mathbb{N}^2$ en deux sommes sur $\mathbb{N}$ est faite en vertu de la proposition 10.98. □

35. A_1 pourrait contenir $[0, 1]$ et A_2 contenir $]1, 2]$.

13.8.1 Mesure et tribu de Lebesgue

Théorème 13.126.
Il existe une unique mesure λ sur $\big(\mathbb{R}, \mathcal{B}or(\mathbb{R})\big)$ telle que

$$\lambda\big(]a,b[\big) = b - a \tag{13.356}$$

pour tout $a \leqslant b$ dans $\mathbb{R}$.

Démonstration. L'existence provient du théorème de prolongement de Hahn 13.68 : la mesure ℓ sur $(\mathcal{A}_{\mathcal{S}})$ se prolonge à $\sigma(\mathcal{A}_{\mathcal{S}}) = \mathcal{B}\mathrm{or}(\mathbb{R})$.

Nous ne pouvons pas prouver l'unicité en invoquant la partie unicité de Hahn (c'est tentant parce que ℓ est σ-finie) parce que dans ce théorème nous ne fixons la valeur de λ que sur une toute petite partie de $\mathcal{A}_{\mathcal{S}}$. Nous allons cependant voir que cette petite partie suffit à garantir l'unicité.

La classe

$$\mathcal{D} = \{]a,b[\text{ tel que } -\infty < a \leqslant b < +\infty\} \tag{13.357}$$

est stable par intersection finie et engendre la tribu borélienne. En effet $\mathcal{D}$ contient toutes les boules et donc une base dénombrable de la topologie de $\mathbb{R}$ (proposition 7.97). Donc tous les ouverts de $\mathbb{R}$ sont dans $\sigma(\mathcal{D})$ et $\sigma(\mathcal{D}) = \mathcal{B}\mathrm{or}(\mathbb{R})$. Nous pouvons donc dire grâce au théorème 13.33 qu'il y a unicité de la mesure sur $\mathcal{B}\mathrm{or}(\mathbb{R})$ lorsque les valeurs sur $\mathcal{D}$ sont fixées. □

Définition 13.127.
*La mesure de l'espace mesuré $\big(\mathbb{R}, \mathcal{B}or(\mathbb{R}), \lambda\big)$ donné par le théorème 13.126 est la **mesure de Lebesgue** sur $\big(\mathbb{R}, \mathcal{B}or(\mathbb{R})\big)$.*

*Nous définissons aussi la **tribu de Lebesgue** par la proposition 13.65 : $\big(\mathbb{R}, \mathcal{L}eb(\mathbb{R}), \lambda\big)$ est l'espace mesuré complété de $\big(\mathbb{R}, \mathcal{B}or(\mathbb{R}), \lambda\big)$.*

Remarque 13.128.
Il n'est pas évident que la tribu de Lebesgue soit plus grande que celle des boréliens, ni que la tribu des parties soit plus grande que celle de Lebesgue. Nous mentionnons cependant les faits suivants.

(1) Il existe des ensembles mesurables non-boréliens, et cela ne nécessite pas l'axiome du choix. Un argument classique de cardinalité est donné dans [271]. La construction la plus explicite que j'aie trouvée est dans [286], mais ça a l'air de demander des connaissances précises sur les ordinaux.

(2) Vu que l'ensemble de Cantor C est mesurable de mesure nulle (proposition 13.147), tout sous-ensemble de Cantor est mesurable de mesure nulle parce que la tribu de Lebesgue est complète par définition. Le cardinal de $\mathcal{P}(C)$ est strictement supérieur à la puissance du continu, alors que le cardinal de l'ensemble des boréliens est au plus égal à la puissance du continu. Donc il existe des non boréliens contenus dans Cantor ; de tels non boréliens sont alors mesurables au sens de Lebesgue.

(3) Si nous admettons l'axiome du choix alors il existe des ensembles non mesurables au sens de Lebesgue. Nous en verrons un dans l'exemple 13.141.

Exemple 13.129 (Un ouvert contenant tous les rationnels et de mesure arbitrairement petite)**.**
Il est possible de construire un ouvert de $\mathbb{R}$ contenant $\mathbb{Q}$ et de mesure de Lebesgue plus petite que ϵ. Pour cela si (q_i) est une énumération des rationnels, il suffit de prendre

$$\mathcal{O} = \bigcup_{n=1}^{\infty} B(q_n, \frac{\epsilon}{2^{n+1}}). \tag{13.358}$$

Cela est un ouvert comme union d'ouverts, ça contient tous les rationnels, et sa mesure se majore. En effet le théorème 13.126 donne $\lambda\big(B(q_n, \frac{\epsilon}{2^n})\big) = \frac{\epsilon}{2^n}$. Vu que ces boules ne sont à priori pas disjointes, le lemme 13.22 donne

$$\lambda(\mathcal{O}) \leqslant \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon \tag{13.359}$$

par (10.127) avec $q = \frac{1}{2}$.

Par complémentarité, nous pouvons construire un ensemble fermé de mesure non nulle et ne contenant aucun rationnel. Et même un fermé dans $[0, 1]$, de mesure $1 - \epsilon$ ne contenant aucun rationnel.

Cela peut surprendre parce qu'il existe des tonnes de suites d'irrationnels qui convergent vers des rationnels[36], et il semble difficile de créer un ensemble contenant beaucoup d'irrationnels tout en préservant la propriété de fermeture vis à vis des suites convergentes. △

Exemple 13.130 (Mesure finie, non borné)**.**
Il existe des parties de $\mathbb{R}$ qui sont de mesure finie sans être bornés. Par exemple en posant

$$A = \bigcup_{n=1}^{\infty} B(n, \frac{1}{2^n}). \tag{13.360}$$

La partie A n'est pas bornée parce que que $\mathbb{N} \subset A$. Mais en termes de mesure,

$$\lambda(A) \leqslant \sum_{n=1}^{\infty} \frac{1}{2^n} < \infty \tag{13.361}$$

en vertu de la somme de la série géométrique, exemple 10.67. △

13.8.2 Propriétés de la mesure de Lebesgue

Proposition 13.131.
Tout ensemble dénombrable de $\mathbb{R}$ est mesurable de mesure nulle.

Démonstration. Un point de $\mathbb{R}$ est un intervalle de mesure nulle. Si D est dénombrable, il est union disjointes et dénombrable de points. Le lemme 13.123 nous dit alors que sa mesure est $\lambda(D) = \sum_{i=1}^{\infty} \lambda(\{a_i\}) = 0$. □

Remarque 13.132.
Il existe cependant des ensembles non dénombrables et tout de même de mesure nulle. Par exemple l'ensemble de Cantor (voir la proposition 13.147).

Proposition 13.133.
La mesure de Lebesgue est invariante par translation, c'est-à-dire que si A est mesurable alors $\lambda(A) = \lambda(A + \alpha)$ pour tout réel α.

Démonstration. Nous commençons par les intervalles ouverts :

$$\lambda\big(]a, b[+ \alpha\big) = \lambda\big(]a + \alpha, b + \alpha[\big) = (b + \alpha) - (a + \alpha) = b - a = \lambda\big(]a, b[\big). \tag{13.362}$$

D'après ce qui est dit dans l'exemple 13.34, la mesure de Lebesgue sur les boréliens est invariante par translation.

Si A est mesurable alors il existe un borélien B et un ensemble négligeable N tels que $A = B \cup N$ par la caractérisation 13.117b de la complétion. Alors $A + \alpha = B + \alpha \cup N + \alpha$ et $N + \alpha$ est encore un ensemble négligeable. Donc $\lambda(A + \alpha) = \alpha(B + \alpha) = \lambda(B)$. □

Le mesure ℓ définie sur l'algèbre de parties $\mathcal{A}_{\mathcal{S}}$ (voir proposition 13.125). La proposition 13.19 nous donne donc une mesure extérieure par

$$\lambda^*(X) = \inf\{\sum_n \ell(A_n); A_n \in \mathcal{A}_{\mathcal{S}}, X \subset \bigcup_n A_n\}. \tag{13.363}$$

La proposition suivante montre que cette mesure extérieure peut être exprimée seulement avec des intervalles ouverts.

36. Si $q \in \mathbb{Q}$ et $r \in \mathbb{R}\backslash\mathbb{Q}$ alors la suite $(q + r/10^k)_k$ est une suite d'irrationnels convergente vers le rationnel q.

Proposition 13.134.
Nous avons

$$\lambda^*(X) = \inf\{\sum_{n\geqslant 1} \ell(I_n); I_n \text{ sont des intervalles ouverts et } X \subset \bigcup_n I_n\}. \tag{13.364}$$

Démonstration. Nous savons que dans la définition (13.363), chacun des A_n est une réunion disjointe d'intervalles (pas spécialement ouverts) deux à deux disjoints ; donc

$$\lambda^*(X) = \inf\{\sum_n \ell(I_n); I_n \in \mathcal{S}, X \subset \bigcup_n I_n\}. \tag{13.365}$$

Soit $\epsilon > 0$. Si $A \subset \bigcup_n I_n$, pour chaque $n \geqslant 1$ nous considérons un intervalle ouvert J_n tel que $I_n \subset J_n$ et $\ell(I_n) + \frac{\epsilon}{2^n} \leqslant \ell(J_n)$. Faisant cela pour chacun des découpages de X en intervalles nous trouvons

$$\lambda^*(X) \leqslant \inf\{\sum_n \ell(J_n) \ J_n \text{ est ouvert et } X \subset \bigcup_n J_n\} + \epsilon. \tag{13.366}$$

Étant donné que ϵ est arbitraire nous avons l'égalité. □

Proposition 13.135 ([273])**.**
Si $X \subset \mathbb{R}$ est tel que $\lambda^*(X) < \infty$ alors

(1) Pour tout $\epsilon > 0$ il existe un ouvert Ω_ϵ tel que

$$\begin{cases} X \subset \Omega_\epsilon & (13.367a) \\ \lambda(\Omega_\epsilon) \leqslant \lambda^*(X) + \epsilon. & (13.367b) \end{cases}$$

(2) Il existe une intersection dénombrable d'ouverts G telle que

$$\begin{cases} X \subset G & (13.368a) \\ \lambda(G) = \lambda^*(X). & (13.368b) \end{cases}$$

Démonstration. Pour (1), la proposition 13.134 nous a déjà dit que

$$\lambda^*(X) = \inf\{\sum_n \ell(I_n) \ I_n \text{ est un intervalle ouvert}, X \subset \bigcup_n I_n\}, \tag{13.369}$$

donc si $\epsilon > 0$, il existe des intervalles ouverts I_n tels que

$$\begin{cases} X \subset \bigcup_n I_n & (13.370a) \\ \sum_n \ell(I_n) \leqslant \lambda^*(X) + \epsilon. & (13.370b) \end{cases}$$

Si nous posons $\Omega_\epsilon = \bigcup_n I_n$, alors nous avons bien

$$\begin{cases} X \subset \Omega_\epsilon & (13.371a) \\ \lambda(\Omega_\epsilon) \leqslant \sum_n \ell(I_n) \leqslant \lambda^*(X) + \epsilon. & (13.371b) \end{cases}$$

En ce qui concerne (2), pour chaque $k \geqslant 1$ nous considérons l'ensemble $\Omega_{1/k}$ obtenu comme précédemment avec $\epsilon = 1/k$ et nous posons $G = \bigcap_{k\geqslant 1} \Omega_{1/k}$. Cela est une intersection dénombrable d'ouverts vérifiant $X \subset G$ (parce que $X \subset \Omega_{1/k}$ pour tout k) et donc $\lambda^*(X) \leqslant \lambda^*(G) = \lambda(G)$. De plus pour tout k nous avons

$$\lambda(G) \leqslant (\Omega_{1/k}) \leqslant \lambda^*(X) + \frac{1}{k} \tag{13.372}$$

pour tout k. En faisant $k \to \infty$ nous avons

$$\lambda(G) \leqslant \lambda^*(X). \tag{13.373}$$

Au final

$$\lambda(G) \leqslant \lambda^*(X) \leqslant \lambda(G), \tag{13.374}$$

d'où l'égalité. □

Corolaire 13.136.
Une partir $N \subset \mathbb{R}$ est négligeable[37] *si et seulement si $\lambda^*(N) = 0$.*

Démonstration. Nous savons que si N est négligeable il existe un borélien Y tel que $N \subset Y$ avec $\lambda(Y) = 0$. Par conséquent[38]

$$\lambda^*(N) \leqslant \lambda^*(Y) = \lambda(Y) = 0. \tag{13.375}$$

Pour l'implication inverse nous supposons que $\lambda^*(N) = 0$ et nous prenons l'ensemble G définit par la proposition 13.135(2) : c'est un borélien contenant N et tel que $\lambda(G) = \lambda^*(N) = 0$. L'ensemble N est donc négligeable. □

Théorème 13.137 (Régularité extérieure de la mesure de Lebesgue)**.**
Pour tout mesurable $A \subset \mathbb{R}$ nous avons

$$\lambda(A) = \inf\{\lambda(\Omega); \Omega \text{ ouvert contenant } A\}. \tag{13.376}$$

Démonstration. Nous commençons par le cas où B est un borélien.

Si B borélien, $\lambda(B) < \infty$ Soit $\epsilon > 0$; par la proposition 13.135(1) il existe un ouvert Ω_ϵ contenant B tel que $\lambda(\Omega_\epsilon) \leqslant \lambda^*(B) + \epsilon$. Vu qu'ici B est borélien, $\lambda^*(B) = \lambda(B)$ et nous concluons que pour tout ϵ il existe un ouvert Ω_ϵ tel que

$$\begin{cases} B \subset \Omega_\epsilon & \text{(13.377a)} \\ \lambda(\Omega_\epsilon) \leqslant \lambda(B) + \epsilon, & \text{(13.377b)} \end{cases}$$

et donc

$$\lambda(B) = \inf\{\lambda(\Omega);\ \Omega \text{ ouvert contenant } B\ \}. \tag{13.378}$$

Si B borélien, $\lambda(B) = +\infty$ Dans ce cas l'infimum est pris uniquement sur des ouverts Ω tels que $\lambda(\Omega) = \infty$.

Si A est mesurable non borélien Nous passons maintenant au cas où A est mesurable sans être borélien. Il s'écrit donc $A = B \cup N$ avec B borélien et N négligeable par la proposition 13.57, et par définition $\lambda(A) = \lambda(B)$. Si Y est un borélien tel que $N \subset Y$ et $\lambda(Y) = 0$ alors

$$\begin{aligned}
\lambda(A) = \lambda(B) &= \inf\{\lambda(\Omega) \text{ tel que } \Omega \text{ ouvert}, B \subset \Omega\} && \text{(13.379a)}\\
&\leqslant \inf\{\lambda(\Omega) \text{ tel que } \Omega \text{ ouvert}, B \cup N \subset \Omega\} && \text{(13.379b)}\\
&\leqslant \inf_{\Omega', Y'}\{\lambda(\Omega' \cup Y') \text{ tel que } \Omega', Y' \text{ ouverts}, B \subset \Omega', Y \subset Y'\} && \text{(13.379c)}\\
&\leqslant \inf_{\Omega', Y'}\{\lambda(\Omega') + \lambda(Y') \text{ tel que } \Omega', Y' \text{ ouverts}, B \subset \Omega', Y \subset Y'\} && \text{(13.379d)}\\
&\leqslant \inf_{\Omega'}\{\lambda(\Omega') \text{ tel que } \Omega' \text{ ouvert}, B \subset \Omega\} && \text{(13.379e)}\\
&= \lambda(B). && \text{(13.379f)}
\end{aligned}$$

Justifications :

— (13.379a) Le cas borélien déjà fait.

— (13.379b) Les ouverts Ω tels que $B \cup N \subset \Omega$ vérifient a fortiori $B \subset \Omega$; nous avons donc agrandit l'ensemble sur lequel l'infimum est pris.

— (13.379c) Parmi les ouverts Ω qui recouvrent $B \cup N$, il y a ceux de la forme $\Omega' \cup Y'$ où Ω' recouvre B et Y' est un ouvert contenant Y. Donc nous avons rétréci l'ensemble sur lequel l'infimum est pris et par conséquent agrandit l'infimum.

— (13.379d) Mesure d'une union majorée par la somme des mesures.

37. Définition 13.54.
38. Au péril d'être lourd nous rappelons que λ^* est défini sur toutes les parties de $\mathbb{R}$.

— (13.379e) Vu que Y est borélien, $\lambda(Y) = \inf_{Y' \text{ ouvert}}\{\lambda(Y') \text{ tel que } Y \subset Y'\} = 0$. Donc pour tout Ω' et tout $\epsilon > 0$, nous pouvons trouver un Y' vérifiant les conditions tel que $\lambda(\Omega') + \lambda(Y') \leqslant \lambda(\Omega') + \epsilon$.

Toutes les inégalités sont des égalités en en particulier (13.379b) donne

$$\lambda(A) = \inf\{\lambda(\Omega) \text{ tel que } \Omega \text{ ouvert}, B \cup N \subset \Omega\}, \tag{13.380}$$

ce qu'il fallait.

□

Proposition 13.138 ([273]).
Si A est mesurable dans $\mathbb{R}$ et si $\epsilon > 0$ alors il existe un ouvert Ω_ϵ et un fermé F_ϵ tels que

$$\begin{cases} F_\epsilon \subset A \subset \Omega_\epsilon & (13.381a) \\ \lambda(\Omega_\epsilon \backslash F_\epsilon) \leqslant \epsilon. & (13.381b) \end{cases}$$

Démonstration. Nous commençons par le cas où A est un borélien, que nous noterons B.

Première étape Montrons qu'il existe un ouvert U_ϵ tel que

$$\begin{cases} B \subset U_\epsilon & (13.382a) \\ \lambda(U_\epsilon \backslash B) \leqslant \frac{\epsilon}{2}. & (13.382b) \end{cases}$$

Si $\lambda(B) < \infty$ alors le théorème 13.137 nous donne un ouvert U_ϵ tel que $B \subset U_\epsilon$ et $\lambda(U_\epsilon) \leqslant \lambda(B) + \frac{\epsilon}{2}$. Nous avons alors

$$\lambda(\Omega_\epsilon \backslash B) = \lambda(\Omega_\epsilon) - \lambda(B) \leqslant \frac{\epsilon}{2}. \tag{13.383}$$

Si par contre $\lambda(B) = \infty$, nous posons $B_n = B \cap [-n, n]$ et $\epsilon_n = \epsilon/2^{n+1}$. Pour chaque n nous avons un ouvert Ω_n tel que

$$\begin{cases} B_n \subset \Omega_n & (13.384a) \\ \lambda(\Omega_n \backslash B_n) \leqslant \frac{\epsilon}{2^{n+1}} & (13.384b) \end{cases}$$

Par conséquent en posant $\Omega = \bigcup_{n \geqslant 1} \Omega_n$ nous avons [39]

$$\begin{cases} B \subset \Omega & (13.385a) \\ \lambda(\Omega \backslash B) \leqslant \lambda\big(\bigcup_n (\Omega_n \backslash B_n)\big) \leqslant \sum_{n \geqslant 1} \lambda(\Omega_n \backslash B_n) = \frac{\epsilon}{2}. & (13.385b) \end{cases}$$

La première étape est terminée.

Deuxième étape Nous prouvons à présent qu'il existe un ouvert Ω_ϵ et un fermé F_ϵ tels que

$$\begin{cases} F_\epsilon \subset B \subset \Omega_\epsilon & (13.386a) \\ \lambda(\Omega_\epsilon \backslash B) \leqslant \frac{\epsilon}{2} & (13.386b) \\ \lambda(B \backslash F_\epsilon) \leqslant \frac{\epsilon}{2}. & (13.386c) \end{cases}$$

L'ouvert Ω_ϵ, nous l'avons déjà de l'étape précédente. Pour le fermé, nous appliquons la première étape au borélien B^c ; ce qui nous trouvons est un ouvert G_ϵ tel que

$$\begin{cases} B^c \subset G_\epsilon & (13.387a) \\ \lambda(G_\epsilon \backslash B^c) \leqslant \frac{\epsilon}{2}. & (13.387b) \end{cases}$$

En posant $F_\epsilon = G_\epsilon^c$ nous avons un fermé tel que $F_\epsilon \subset B$ et

$$\lambda(B \backslash F_\epsilon) = \lambda(F_\epsilon^c \backslash B^c) = \lambda(G_\epsilon \backslash B^c) \leqslant \frac{\epsilon}{2}. \tag{13.388}$$

39. Nous utilisons la petite relation ensembliste $\big(\bigcup_n A_n\big) \backslash \big(\bigcup_n B_n\big) \subset \bigcup_n (A_n \backslash B_n)$.

Dernière étape Les ensembles F_ϵ et Ω_ϵ trouvés à la deuxième étape donnent bien les relations (13.381). En effet $\Omega_\epsilon \backslash F_\epsilon = (\Omega_\epsilon \backslash B) \cup (B \backslash F_\epsilon)$, donc

$$\lambda(\Omega_\epsilon \backslash F_\epsilon) \leqslant \lambda(\Omega_\epsilon \backslash B) + \lambda(B \backslash F_\epsilon) = \epsilon. \tag{13.389}$$

Nous passons au cas où $A = B \cup N$ est mesurable. Nous commençons par prendre les Ω_ϵ et F_ϵ qui correspondent à B :

$$\begin{cases} F_\epsilon \subset B \subset \Omega_\epsilon & (13.390a) \\ \lambda(\Omega_\epsilon \backslash F_\epsilon) \leqslant \epsilon. & (13.390b) \end{cases}$$

Soit Y un borélien tel que $N \subset Y$ et $\lambda(Y)$ puis un ouvert Y' tel que $\lambda(Y') \leqslant \epsilon$ et $Y \subset Y'$. L'existence d'un tel Y' est assurée par la proposition 13.137 appliquée à Y. Nous vérifions que les ensembles F_ϵ et $\Omega_\epsilon \cup Y'$ fonctionnent. En effet $\Omega_\epsilon \cup Y' \backslash F_\epsilon \subset (\Omega_\epsilon \backslash F_\epsilon) \cup Y'$, donc

$$\begin{cases} F_\epsilon \subset B \cup N \subset \Omega_\epsilon \cup Y' & (13.391a) \\ \lambda\big((\Omega_\epsilon \backslash F_\epsilon)\big) \leqslant \lambda(\Omega_\epsilon \backslash F_\epsilon) + \lambda(Y') \leqslant 2\epsilon. & (13.391b) \end{cases}$$

Donc en réalité il faut choisir $\Omega_{\epsilon/2}$, $F_{\epsilon/2}$ et $\lambda(Y') \leqslant \epsilon/2$. □

Théorème 13.139 (Régularité intérieure de la mesure de Lebesgue)**.**
Si A est mesurable dans $\mathbb{R}$ alors

$$\lambda(A) = \sup\{\lambda(K); K \text{ compact contenu dans } A\}. \tag{13.392}$$

Démonstration. Par la proposition 13.138 nous avons

$$\lambda(A) = \sup_{F \text{ fermé dans } A} \lambda(F). \tag{13.393}$$

Pour un tel F nous posons $K_n = F \cap [-n, n]$ qui est compact [40] et contenu dans B. De plus le lemme 13.23(2) nous dit que

$$\lambda(F) = \lim_{n\to\infty} \lambda(K_n) \tag{13.394}$$

Donc tous les $\lambda(F)$ peuvent être arbitrairement approchés par un $\lambda(K)$ avec K compact dans A, et le supremum (13.393) n'est pas affecté en nous restreignant à prendre des compacts contenus dans B :

$$\lambda(A) = \sup_{F \text{ fermé dans } A} \lambda(F) = \sup_{K \text{ compact dans } A} \lambda(K). \tag{13.395}$$

□

13.8.3 Fonctions mesurables

Lemme 13.140.
Soit une fonction $f\colon \mathbb{R} \to \mathbb{R}$ mesurable telle que $\lambda(f \neq 0) > 0$. Alors il existe une partie mesurable M et $m > 0$ tels que $\lambda(M) > 0$ et $f(x) > m$ pour tout $x \in M$.

Démonstration. Nous notons

$$D = \{x \in \mathbb{R} \text{ tel que } f(x) > 0\}, \tag{13.396}$$

et nous supposons que $\lambda(D) > 0$ pour fixer les idées (si ce n'est pas le cas, nous prenons pour D la partie où f est strictement négative).

Nous posons

$$A_1 = [1, \infty[\tag{13.397a}$$

$$A_n = [\frac{1}{n}, \frac{1}{n-1}[. \tag{13.397b}$$

40. parce que fermé et borné, théorème de Borel-Lebesgue 7.127.

Ces parties A_n sont disjointes ; donc les parties

$$D_n = \{x \in \mathbb{R} \text{ tel que } f(x) \in A_n\} \tag{13.398}$$

sont également disjointes. Vu que $\bigcup_n D_n =]0, \infty[$, nous avons $D = \bigcup_{n\in\mathbb{N}} D_n$. Vu que

$$\lambda(D) = \sum_{n=1}^{\infty} \lambda(D_n) > 0, \tag{13.399}$$

il existe au moins un N tel que $\lambda(D_N) > 0$. Pour $x \in D_N$ nous avons

$$f(x) \in A_N = [\frac{1}{N}, \frac{1}{N-1}[. \tag{13.400}$$

Donc pour $x \in D_N$ nous avons $f(x) > \frac{1}{N}$. □

13.8.4 Ensemble de Vitali (non mesurable)

Exemple 13.141 (Un ensemble non mesurable au sens de Lebesgue[287])**.**
Nous considérons l'ensemble quotient $\mathbb{R}/\mathbb{Q}$; chaque classe intersecte l'intervalle $[0, 1]$. Grâce à l'axiome du choix (voir 1.6) nous pouvons construire un ensemble V contenant un représentant dans $[0, 1]$ de chaque classe. Un tel ensemble est un **ensemble de Vitali**. Nous allons prouver que V n'est pas mesurable.

Supposons que V soit mesurable. Alors tous les ensembles de la forme $V + q$ ($q \in \mathbb{Q}$) sont mesurables et ont même mesure par la proposition 13.133. Nous posons

$$A = \bigcup_{\substack{q\in\mathbb{Q}\\ -1\leqslant q\leqslant 1}} (V + q) \subset [-1, 2]. \tag{13.401}$$

Cela est une union disjointe d'ensembles mesurables. Donc

$$\lambda(A) = \sum_{\substack{q\in\mathbb{Q}\\ -1\leqslant q\leqslant 1}} \lambda(V + q). \tag{13.402}$$

Vu que $A \subset [-1, 2]$ nous avons $\lambda(A) \leqslant 3$ et donc tous les termes de la somme doivent être nuls. Nous avons donc $\lambda(A) = 0$.

Prouvons toutefois que $[0, 1] \subset A$, ce qui serait une contradiction. Soit $x \in [0, 1]$; il est dans une des classes de $\mathbb{R}/\mathbb{Q}$ et donc il existe $v \in V$ tel que $x - v \in \mathbb{Q}$. De plus $x, v \in [0, 1]$, donc

$$-1 \leqslant x - v \leqslant 1. \tag{13.403}$$

Cela fait que $x \in V + (x - v) \subset A$. Nous avons donc $x \in A$ et donc $[0, 1] \subset A$. En conséquence de quoi nous aurions $\lambda(A) \geqslant 1$. △

13.8.5 Ensemble de Cantor

Nous considérons la fonction donnant l'écriture décimale des nombres définie en (10.149).

Définition 13.142 (Ensemble de Cantor)**.**
Soit $K_0 = [0, 1[$ et les ensembles K_n définis par la récurrence

$$K_{n+1} = \big(\frac{1}{3}K_n\big) \cup \big(\frac{1}{3}(K_n + 2)\big). \tag{13.404}$$

L'ensemble

$$K = \bigcup_{n\geqslant 0} K_n \tag{13.405}$$

*est l'**ensemble triadique de Cantor**.*

Les principales propriétés de l'ensemble de Cantor sont qu'il est non dénombrable (proposition 13.146) et borélien de mesure nulle (proposition 13.147).

13.143.
L'idée de base pour prouver que l'ensemble K est non dénombrable est que ses éléments sont les nombres qui s'écrivent en base 3 sans utiliser le chiffre 1. En prenant un nombre sans 1 écrit en base 3, en changeant tous les 2 en 1 et en lisant le résultat en base 2, nous obtenons tous les nombres possibles en base 2 et donc une quantité non dénombrable. L'idée est donc simple et astucieuse. La mise en musique est un peu plus délicate parce qu'il faut faire attention aux queues de suites ; c'est pour cela que nous avons construit l'ensemble de Cantor en partant de $[0,1[$ et non de $[0,1]$.

Le lemme suivant dit précisément ce que nous entendons en disant que les éléments de l'ensemble de Cantor sont les nombres qui s'écrivent en base 3 sans utiliser le chiffre 1. Nous rappelons que $\mathbb{D}_3$ est l'ensemble des suites constituées de 0, 1 et 2, et qui ne se terminent pas par une suite infinie de 2, voir 10.75 pour une définition précise.

Lemme 13.144 ([1]).
Soit $n \in \mathbb{N}$ et $x \in \mathbb{D}_3$ (définition 10.75) ; nous avons $\varphi_3(x) \in K_n \in$ si et seulement si $x_1, \dots, x_n \in \{0,2\}$.

Démonstration. Nous procédons par récurrence en commençant avec $n = 1$. Si $x_1 = 1$ alors

$$\varphi_3(x) = \frac{1}{3} + \sum_{k=2}^{\infty} \frac{x_k}{3^k} \in [\frac{1}{3}, \frac{2}{3}[. \tag{13.406}$$

Notons que $\varphi_3(x) = \frac{2}{3}$ est impossible parce que ça demanderait une queue de suite de 2. Par conséquent $\varphi_3(x) = [0,1[\setminus[\frac{1}{3}, \frac{2}{3}[= K_1$.

Nous passons à la récurrence.

Sens direct Nous supposons que $x_1, \dots, x_{n+1} \in \{0,2\}$ et nous montrons que $\varphi_3(x) \in K_{n+1}$. La chose surprenante est que nous n'allons pas considérer deux cas suivant que x_{n+1} vaut 0 ou 1 ; nous allons considérer deux cas suivant [41] que x_1 vaut 0 ou 1. Écrivons encore $\varphi_3(x)$:

$$\varphi_3(x) = \sum_{k=1}^{n+1} \frac{x_k}{3^k} + \sum_{k=n+2}^{\infty} \frac{x_k}{3^k}. \tag{13.407}$$

Si $x_1 = 0$ Alors nous avons

$$3\varphi_3(x) = \sum_{k=2}^{\infty} \frac{x_k}{3^{k-1}} = \sum_{k=1}^{\infty} \frac{x_{k+1}}{3^k} = \varphi_3(x_2, \dots, x_n, x_{n+1}, \dots) \tag{13.408}$$

Vu que par hypothèse $x_2, \dots, x_{n+1}$ sont dans $\{0,2\}$ nous avons $3\varphi_3(x) \in K_n$ par hypothèse de récurrence. Cela implique que $\varphi_3(x) \in K_{n+1}$.

Si $x_1 = 2$ Alors

$$\varphi_3(x) = \frac{2}{3} + \sum_{k=2}^{\infty} \frac{x_k}{3^k}, \tag{13.409}$$

et

$$3\varphi_3(x) - 2 = \sum_{k=1}^{\infty} \frac{x_{k+1}}{3^k} = \varphi(x_2, \dots, x_{n+1}, \dots), \tag{13.410}$$

et donc là nous avons $3\varphi_3(x) - 2 \in K_n$, ce qui implique encore $\varphi_3(x) \in K_{n+1}$.

41. Pour comprendre pourquoi, faire un dessin de comment K_n se transforme en K_{n+1} et remarquer dans K_2, les deux premiers segments ne sont pas une division du premier segment de K_1, mais bien une copie des *deux* segments de K_1.

Sens réciproque Nous devons maintenant prouver que $\varphi_3(x) \in K_{n+1}$ implique $x_1, \dots, x_{n+1} \in \{0, 2\}$. Par le même calcul que précédemment nous avons soit

$$3\varphi_3(x) = \varphi_3(x_2, \dots, x_{n+1}, \dots), \tag{13.411}$$

si $x_1 = 0$, soit

$$3\varphi_3(x) - 2 = \varphi_3(x_2, \dots, x_{n+1}, \dots), \tag{13.412}$$

si $x_1 = 2$. Dans les deux cas, si $x_l = 1$ pour un certain $2 \leqslant l \leqslant n+1$, alors l'hypothèse de récurrence donne que ces éléments ne sont pas dans K_n et donc $\varphi_3(x)$ pas dans K_{n+1}.

□

Corolaire 13.145 ([1]).
En posant $\mathbb{E} = \{x \in \mathbb{D}_3$ tel que $x_i \neq 1 \forall i\}$ nous avons $K = \varphi_3(\mathbb{E})$. Et plus précisément, $\varphi_3 \colon \mathbb{E} \to K$ est une bijection.

Démonstration. Nous divisons la preuve en trois étapes.

Image contenue dans K Si $x \in \mathbb{E}$ et $n \in \mathbb{N}$ nous avons $x_1, \dots, x_n \in \{0, 2\}$ et donc $\varphi_3(x) \in K_n$ par la proposition 13.144. Donc

$$\varphi_3(x) \in \bigcup_{n \geqslant 1} K_n = K. \tag{13.413}$$

Injective L'application $\varphi_3 \colon \mathbb{E} \to K$ est injective parce qu'elle est déjà injective depuis $\mathbb{D}_3$.

Surjective Soit $p \in K \subset [0, 1[$. Vu que $\varphi_3 \colon \mathbb{D}_3 \to [0, 1[$ est surjective (théorème 10.78), il existe $x \in \mathbb{D}_3$ tel que $\varphi_3(x) = p$. Pour tout n nous avons $\varphi_3(x) \in K_n$ et donc $x_1, \dots, x_n \in \{0, 2\}$ et donc au final $x \in \mathbb{E}$.

□

Proposition 13.146 ([1]).
L'ensemble de Cantor est non dénombrable.

Démonstration. Nous avons prouvé à la proposition 10.79 que l'ensemble $\mathbb{D}_2$ n'était pas dénombrable. Nous allons à présent prouver que l'application

$$\begin{aligned} \psi \colon \mathbb{D}_2 &\to K \\ c &\mapsto \varphi_3(c \text{ en remplaçant les 1 par des 2}) \end{aligned} \tag{13.414}$$

est une bijection. Le fait que ψ soit injective est une conséquence du fait que ce soit la composition de deux applications injectives (le remplacement et φ_3). Il faut par contre montrer que l'image est égale à K, en notant qu'il n'est pas évident à priori que l'image soit contenue dans K.

L'opération qui consiste à remplacer les 1 par des 2 est une bijection $\mathbb{D}_2 \to \mathbb{E}$. Le corolaire 13.145 nous dit aussi que $\varphi_3 \colon \mathbb{E} \to K$ est une bijection. En tant que composée de bijections, ψ est une bijection.

Étant en bijection avec $\mathbb{D}_2$ qui n'est pas dénombrable par la proposition 10.79, l'ensemble de Cantor n'est pas dénombrable. □

Proposition 13.147 (Ensemble de Cantor).
L'ensemble de Cantor[42] *est borélien, non dénombrable et de mesure nulle.*

Démonstration. Nous reprenons les notations de la définition 13.142. Le fait que l'ensemble de Cantor soit non dénombrable a été prouvé dans la proposition 13.146.

42. Définition 13.142

L'ensemble de Cantor étant une intersection dénombrable de boréliens, il est borélien par le lemme 13.3. Vu que $K_n \subset [0,1[$ nous avons $\frac{1}{3}K_n \leqslant \frac{1}{3}$ et $\frac{1}{3}(K_n+2) \geqslant \frac{2}{3}$, donc K_n est une union disjointe de 2^n intervalles de mesure $2/3^n$. Nous avons donc

$$\lambda(K_n) = \left(\frac{2}{3}\right)^n. \tag{13.415}$$

L'ensemble de Cantor étant contenu dans chacun des K_n, sa mesure est plus petite que la mesure de chacun des K_n (lemme 13.22) et donc $\lambda(K) \leqslant \left(\frac{2}{3}\right)^n$ pour tout n ; ergo $\lambda(K) = 0$. □

13.8.6 Mesure positive sans intervalle

Vu que la mesure de Lebesgue est basée sur la mesure des intervalles et quelques extensions, nous sommes en droit de croire qu'une partie de mesure strictement positive de $\mathbb{R}$ doit toujours contenir un intervalle, éventuellement à partie de mesure nulle près. Eh bien non.

Exemple 13.148 ([288]).
Soient une énumération (q_i) de $\mathbb{Q} \cap]0,1[$ et une suite (r_i) telle que $\sum_{i=0}^{\infty} r_i < \frac{1}{2}$. Quitte à prendre r_i plus petit, supposons de plus que $B(q_i, r_i) \subset [0,1]$.

Nous posons $J_n = B(q_i, r_i)$, $J = \bigcup_{i=0}^{\infty} J_n$ et

$$B = [0,1] \backslash J. \tag{13.416}$$

Les parties J_i ne sont pas disjointes, donc, en notant λ la mesure de Lebesgue,

$$0 < \lambda(J) \leqslant \sum_{i=0}^{\infty} \lambda(J_i) \leqslant \frac{1}{2}. \tag{13.417}$$

Mais, par définition, l'union $[0,1] = B \cup J$ est disjointe, donc

$$1 = \lambda\big([0,1]\big) = \lambda(J) + \lambda(B). \tag{13.418}$$

Nous en déduisons que

$$\frac{1}{2} \leqslant \lambda(B) \leqslant 1. \tag{13.419}$$

Je plaide que cette partie B ne contient non seulement aucun intervalle, mais qu'il est impossible de le compléter par une partie de mesure nulle pour obtenir un intervalle.

Soit un intervalle I dans $[0,1]$. Il existe $q_i \in I$ et donc [43]

$$J_i \subset B \backslash I. \tag{13.420}$$

Donc il n'existe pas de parties de mesure nulle qui, ajoutée à B, contiendrait I. △

Vous voulez un truc dingue à propos de la partie J de l'exemple 13.148 ? Le théorème 13.139 nous dit qu'il existe dans J des compacts de mesure arbitrairement proches de $\lambda(J)$. Il existe donc des compacts non seulement de mesure strictement positive mais même de mesure assez grande, tout en étant infiniment découpés.

13.9 Intégrale par rapport à une mesure

13.149.
Nous n'en avons pas encore terminé avec la théorie de la mesure, mais nous devons quand même définir les intégrales et voir quelques propriétés avant de continuer avec la mesure parce que la définition de la mesure sur un espace mesurable produit [44] passe par une intégrale.

43. C'est ici que nous utilisons le fait que r_i est choisi pour que $B(q_i, r_i)$ ne déborde pas de $[0,1]$. Sinon il aurait fallu chipoter et prendre seulement une partie de la boule.
44. Théorème 13.210.

13.150.
En théorie de l'intégration, la convention est la suivante : pour une fonction $f\colon X \to \mathbb{R}$, nous considérons sur X la tribu des ensembles mesurables au sens de Lebesgue sur X, *tout en gardant celle des boréliens sur l'ensemble d'arrivée.* C'est-à-dire qu'en théorie de l'intégration, c'est

$$f\colon \big(X, \mathcal{L}\mathrm{eb}(X)\big) \to \big(\mathbb{R}, \mathcal{B}\mathrm{or}(\mathbb{R})\big). \tag{13.421}$$

En particulier, $f\colon \mathbb{R}^n \to \mathbb{R}^m$ sera mesurable si pour tout borélien A de $\mathbb{R}^m$ l'ensemble $f^{-1}(A)$ est Lebesgue-mesurable dans $\mathbb{R}^n$.

Étant donné qu'il est franchement difficile de créer des ensembles non mesurables au sens de Lebesgue, il est franchement difficile de créer des fonctions non mesurables à valeurs réelles. L'hypothèse de mesurabilité est donc toujours satisfaite dans les cas pratiques.

Voir aussi le point 13.106, et les résultats qui suivent.

13.9.1 Définition pour les fonctions à valeurs positives

Voir le thème 4.

Une mesure μ sur un espace mesurable $(\Omega, \mathcal{A})$ permet de définir une fonctionnelle linéaire sur l'ensemble des fonctions mesurables $\Omega \to \mathbb{R}$. Cette fonctionnelle linéaire est l'intégrale que nous allons définir à présent.

Définition 13.151.
Soient $(\Omega, \mathcal{A}, \mu)$ un espace mesuré ainsi que $Y \in \mathcal{A}$. Notre but est de définir

$$\int_Y f d\mu \tag{13.422}$$

*que nous nommons **intégrales de** f de f sur Y pour la mesure μ.*

Fonction étagée *Si f est une fonction étagée*[45]*, et si sa forme canonique est $f = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$ alors nous définissons*

$$\int_Y f d\mu = \sum_i \alpha_i \mu(Y \cap A_i). \tag{13.423}$$

Fonction mesurable à valeurs positives *Pour une fonction $\mathcal{A}$-mesurable $f\colon \Omega \to [0, \infty]$ nous définissons l'intégrale de f sur Y par*

$$\int_Y f d\mu = \sup\Big\{ \int_Y \psi d\mu \text{ où } \psi \text{ est une fonction étagée telle que } 0 \leqslant \psi \leqslant f \Big\}. \tag{13.424}$$

Remarque 13.152.
Toute fonction mesurable à valeurs dans $[0, +\infty]$ est intégrable (l'intégrale vaut éventuellement $+\infty$). Au moment où une fonction commence à prendre des valeurs positives et négatives, nous demandons à pouvoir intégrer séparément les parties positive et négative. C'est pour cela que nous disons qu'une fonction f à valeurs dans $\mathbb{R}$ est intégrable si $|f|$ l'est.

13.153.
Le nombre $\int_0^\infty f$ est défini directement par (13.424) complètement indépendamment d'une éventuelle limite $\lim_{x\to 0} \int_0^x f$.

13.154.
Si la fonction n'est pas mesurable ? Alors nous n'avons pas défini son intégrale. Supposons la plus simple des fonctions non mesurables sur Ω : la fonction indicatrice d'une partie non mesurable :

$$f(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon.} \end{cases} \tag{13.425}$$

45. Définition 13.100.

où $A \subset \Omega$ n'est pas mesurable[46].

Nous supposons que l'espace mesuré $(\Omega, \mathcal{F}, \mu)$ est complet (définition 13.56). Vu que A n'est pas mesurable, il n'est pas contenu dans une partie négligeable (parce que l'espace est complet), et nous voulons que l'intégrale ne soit pas nulle ; sinon on se demande bien à quoi sert une intégrale.

Toute fonction étagée minorant f est forcément nulle en dehors de A. Dès que B est une partie mesurable de mesure non nulle dans A, le complémentaire de B dans A est encore non mesurable, et nous voulons encore que l'intégrale de f sur ce complémentaire soit non nul.

Mais comme A n'est pas mesurable et que $\mathbb{1}_A$ n'est le supremum d'aucune suite de fonctions mesurables (lemme 13.89), bien que le supremum qui définirait l'intégrale de f existe (toute partie de $\mathbb{R}$ a un supremum), il est sans espoir que ce supremum ait un sens que l'on puisse interpréter en tant que mesure de f.

13.9.2 Premières propriétés

13.155.
Si $(\Omega^n(\mathfrak{A})\mathcal{A}, \mu)$ est un espace mesurable, et si Y est un élément de $\mathcal{A}$, nous avons l'espace mesurable $(Y, \mathcal{A}_Y, \mu_Y)$ donné par

— $\mathcal{A}_y = \{B \cap Y \text{ tel que } B \in \mathcal{A}\}$,

— $\mu_Y = \mu$.

Et là, nous arrivons à un problème de notations parce que $\int_Y f d\mu$ peut désigner l'intégrale de f sur Y dans $(\Omega, \mathcal{A}, \mu)$ ou l'intégrale de f sur Y dans $(Y, \mathcal{A}_Y, \mu_Y)$.

Heureusement, nous allons tout de suite montrer que ces deux choses sont identiques.

Lemme 13.156.
Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$ ainsi que $Y \in \mathcal{A}$. Nous considérons une fonction $f\colon \Omega \to \mathbb{R}^+$ qui est $\mathcal{A}$-mesurable et intégrable sur Y.

Alors, avec des notations que j'espère être claires,

(1) f est $\mathcal{A}_Y$-mesurable,

(2) f est $(Y, \mathcal{A}_Y, \mu_Y)$-intégrable,

(3) nous avons l'égalité

$$\int_{(Y,\mathcal{A}_Y,\mu_Y)} f|_Y = \int_{(Y\subset\Omega,\mathcal{A},\mu)} f. \tag{13.426}$$

Démonstration. Nous considérons les deux ensembles suivants :

$$S_1 = \{\psi \text{ étagées sur } Y \text{ et majorées par } f|_Y\} \tag{13.427a}$$
$$S_2 = \{\psi \text{ étagées sur } \Omega \text{ et majorées par } f\}. \tag{13.427b}$$

Nous considérons l'application suivante :

$$\begin{aligned} s\colon S_1 &\to S_2 \\ s(\psi)(x) &= \begin{cases} \psi(x) & \text{si } x \in Y \\ 0 & \text{sinon.} \end{cases} \end{aligned} \tag{13.428}$$

L'application s est une bijection.

Pour $\psi \in S_1$ nous avons

$$\psi = \sum_{k=1}^{n} \psi_k \mathbb{1}_{A_k}|_Y \tag{13.429}$$

avec $A_k \in \mathcal{A}_k \subset \mathcal{A}$ et $\mathbb{1}_{A_k}|_Y\colon Y \to \{0,1\}$. Nous avons aussi

$$s(\psi) = \sum_{\psi_k} \mathbb{1}_{A_k} \tag{13.430}$$

46. Ça existe, par exemple 13.141.

avec $\mathbb{1}_{A_k}\colon \Omega \to \{0,1\}$.

En ce qui concerne les intégrales de ces fonctions étagées, nous avons

$$\int_{(Y,\mathcal{A}_Y,\mu_Y)} \psi = \sum_{k=1}^{n} \psi_k \mu_Y(A_k \cap Y) \tag{13.431a}$$

$$= \sum_{k=1}^{n} \psi_k \mu(A_k) \tag{13.431b}$$

$$= \int_{(Y\subset\Omega,\mathcal{A},\mu)} s(\psi). \tag{13.431c}$$

Justifications. Pour passer à (13.431b) nous avons utilisé d'abord que $A_k \subset Y$ et ensuite que $\mu_Y(A_k) = \mu(A_k)$.

Nous sommes maintenant prêts à prouver l'égalité du lemme. Nous avons ceci :

$$\int_{(Y,\mathcal{A}_Y,\mu_Y)} f|_Y = \sup\{\int_{(Y,\mathcal{A}_Y,\mu_Y)} \psi \text{ tel que } \psi \in S_1\} \tag{13.432a}$$

$$= \sup\{\int_{(Y\subset\Omega,\mathcal{A},\mu)} s(\psi) \text{ tel que } \psi \in S_1\} \tag{13.432b}$$

$$= \sup\{\int_{(Y\subset\Omega,\mathcal{A},\mu)} \varphi \text{ tel que } \varphi \in S_2\} \tag{13.432c}$$

$$= \int_{(Y\subset\Omega,\mathcal{A},\mu)} f. \tag{13.432d}$$

□

Lemme 13.157.
Si $(\Omega,\mathcal{A},\mu)$ est un espace mesuré et si $B \in \mathcal{A}$ alors

$$\mu(B) = \int_B 1 d\mu = \int_\Omega \mathbb{1}_B. \tag{13.433}$$

Démonstration. La fonction caractéristique d'une partie mesurable est une fonction étagée dont la forme canonique est $\mathbb{1}_B = 1 \cdot \mathbb{1}_B + 0 \times \mathbb{1}_{B^c}$. Son intégrale est donc

$$\int \mathbb{1}_B d\mu = 1 \times \mu(B) + 0 \times \mu(B^c) = \mu(B) \tag{13.434}$$

parce que $0 \times \mu(B^c) = 0$, même si $\mu(B^c) = \infty$, comme nous l'avons convenu en 13.84. □

Proposition 13.158 ([1]).
Soient une fonction $f\colon (\Omega,\mathcal{A},\mu) \to \mathbb{R}^+$ et une fonction g intégrable sur Ω telle que $f \leqslant g$. Alors f est intégrable.

Démonstration. Une fonction étagée qui minore f minore également g. Donc l'ensemble sur lequel il faut faire le supremum pour définir $\int_\Omega f$ est inclus dans celui pour $\int_\Omega g$. Le second supremum étant fini, le premier l'est également. □

Lemme 13.159.
Soient un espace mesuré $(\Omega,\mathcal{A},\mu)$, une fonction $f\colon \Omega \to \mathbb{R}^+$ et $Y \in \mathcal{A}$. Nous avons :

$$\int_Y f d\mu = \int_\Omega f \mathbb{1}_Y d\mu. \tag{13.435}$$

Démonstration. En plusieurs parties, selon la généralité.

Si f est étagée Nous posons $f = \sum_{k=1}^{n} f_k \mathbb{1}_{A_k}$ avec $f_k \in \mathbb{R}^+$. Dans ce cas,

$$f\mathbb{1}_Y = \sum_k f_k \mathbb{1}_{A_k \cap Y} \tag{13.436}$$

est encore une fonction étagée. Donc nous avons d'une part

$$\int_\Omega f\mathbb{1}_Y = \int_\Omega \sum_k f_k \mathbb{1}_{A_k \cap Y} = \sum_k f_k \mu(A_k \cap Y), \tag{13.437}$$

et d'autre part,

$$\int_Y f d\mu = \sum_k f_k \mu(Y \cap A_k), \tag{13.438}$$

Si f est à valeurs positives Nous posons

$$S_1 = \{\psi \text{ étagées sur } \Omega \text{ tel que } 0 \leqslant \psi \leqslant f\mathbb{1}_Y\} \tag{13.439}$$

et

$$S_2 = \{\psi\mathbb{1}_Y \text{ tel que } \psi \text{ étagée avec } 0 \leqslant \psi \leqslant f\}. \tag{13.440}$$

Nous prouvons que $S_1 = S_2$.
Si $\psi \in S_1$, alors

$$0 \leqslant \psi \leqslant f\mathbb{1}_Y \leqslant f. \tag{13.441}$$

De plus comme $\psi = 0$ hors de Y nous avons $\psi = \psi\mathbb{1}_Y$.
Pour l'autre inclusion, soit $0 \leqslant \psi \leqslant f$ pour une fonction étagée ψ et montrons que $\psi\mathbb{1}_Y \in S_1$. L'application $\psi\mathbb{1}_Y$ est étagée sur Ω et vérifie

$$0 \leqslant \psi\mathbb{1}_Y \leqslant f\mathbb{1}_Y \tag{13.442}$$

parce que $\psi \leqslant f$.

L'égalité à prouver Dans l'égalité 13.435 à prouver, le membre de droite est, d'après la définition 13.424,

$$\int_\Omega f\mathbb{1}_Y = \sup\{\int \psi \text{ tel que } \psi \in S_1\}. \tag{13.443}$$

Il nous reste donc à prouver que $\int_Y f$ se calcule de la même façon avec les éléments de S_2. D'abord nous copions la définition :

$$\int_Y f = \sup\{\int_Y \psi \text{ tel que } 0 \leqslant \psi \leqslant f\}. \tag{13.444}$$

Ensuite nous réfléchissons un peu. Si $0 \leqslant \psi \leqslant f$ avec $\psi = \sum_k \psi_k \mathbb{1}_{A_k}$, alors

$$\int_Y \psi = \sum_k \mu(A_k \cap Y) = \int_Y \psi\mathbb{1}_Y = \int_\Omega \psi\mathbb{1}_Y. \tag{13.445}$$

La dernière égalité est la partie déjà faite, à propos des fonctions étagées. Nous avons donc bien

$$\int_Y f = \sup\{\int_Y s \text{ tel que } s \in S_2\}. \tag{13.446}$$

□

13.9.3 Propriétés plus avancées

13.9.3.1 Convergence monotone

Le théorème suivant est très utile parce que le théorème fondamental d'approximation 13.105 donne les fonctions étagées qu'il faut.

Théorème 13.160 (Théorème de la convergence monotone ou de Beppo-Levi[289])**.**
Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$ *et* (f_n) *une suite croissante de fonctions mesurables à valeurs dans* $[0, \infty]$*. Alors la limite ponctuelle* $\lim_{n\to\infty} f_n$ *existe, est mesurable et*

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega \lim_{n\to\infty} f_n d\mu, \tag{13.447}$$

cette intégrable valant éventuellement ∞*.*

Démonstration. La limite ponctuelle de la suite est la fonction à valeurs dans $[0, \infty]$ donnée par

$$f(x) = \lim_{n\to\infty} f_n(x). \tag{13.448}$$

Ces limites existent parce que pour chaque x la suite $f_n(x)$ est une suite numérique croissante. Nous notons

$$I_0 = \int_\Omega f d\mu. \tag{13.449}$$

Nous posons par ailleurs

$$I_n = \int_\Omega f_n. \tag{13.450}$$

Cela est une suite numérique croissante qui a par conséquent une limite que nous notons $I = \lim_{n\to\infty} I_n$. Notre objectif est de montrer que $I = I_0$. D'abord par croissance de la suite, pour tous n nous avons $I_n \leqslant I_0$, par conséquent $I \leqslant I_0$.

Nous prouvons maintenant l'inégalité dans l'autre sens en nous servant de la définition (13.424). Soit une fonction simple h telle que $h \leqslant f$, et une constante $0 < C < 1$. Nous considérons les ensembles

$$E_n = \{x \in \Omega \text{ tel que } f_n(x) \geqslant Ch(x)\}. \tag{13.451}$$

Ces ensembles vérifient les propriétés $E_n \subset E_{n+1}$ et $\bigcup_{n=1}^\infty E_n = \Omega$. Pour chaque n nous avons les inégalités

$$\int_\Omega f_n \geqslant \int_{E_n} f_n \geqslant C \int_{E_n} h. \tag{13.452}$$

Si nous prenons la limite $n \to \infty$ dans ces inégalités,

$$\lim_{n\to\infty} \int_\Omega f_n \geqslant C \lim_{n\to\infty} \int_{E_n} h = C \int_\Omega h. \tag{13.453}$$

Par conséquent $\lim_{n\to\infty} \int f_n \geqslant C \int_\Omega h$. Mais étant donné que cette inégalité est valable pour tout C entre 0 et 1, nous pouvons l'écrire sans le C :

$$\lim_{n\to\infty} \int_\Omega f_n \geqslant \int_\Omega h. \tag{13.454}$$

Par définition, l'intégrale de f est donné par le supremum des intégrales de h où h est une fonction simple dominée par f. En prenant le supremum sur h dans l'équation (13.454) nous avons

$$\lim_{n\to\infty} \int_\Omega f_n \geqslant \int_\Omega f, \tag{13.455}$$

ce qu'il nous fallait. □

Remarque 13.161.
La proposition 13.105 ainsi que le lemme 13.103 montrent qu'une fonction mesurable peut-être écrite comme limite croissante de fonctions simples. Cela permet de démontrer des théorèmes en commençant par prouver sur les fonctions simples et en utilisant Beppo-Levi pour généraliser.

Remarque 13.162.
Une des raisons de demander la positivité des fonctions f_n est de n'avoir pas d'ambiguïté à parler d'intégrales qui valent ∞. Si par exemple nous prenons $\Omega = [0,1]$ et que nous considérons

$$f_n(x) = \begin{cases} 0 & \text{si } x \leqslant \frac{1}{n} \\ \frac{1}{x} & \text{sinon.} \end{cases} \tag{13.456}$$

Ce sont des fonctions intégrables, mais la limite étant la fonction $1/x$, l'égalité (13.447) est une égalité entre deux intégrales valant ∞.

Corolaire 13.163 (Inversion de somme et intégrales)**.**
Si (u_n) est une suite de fonctions mesurables positives ou nulles, alors

$$\sum_{i=0}^{\infty} \int u_i = \int \sum_{i=0}^{\infty} u_i. \tag{13.457}$$

Démonstration. Nous considérons la suite des sommes partielles de (u_n) : $f_n(x) = \sum_{i=0}^{n} u_n(x)$. Le théorème de la convergence monotone (théorème 13.160) implique que

$$\lim_{n\to\infty} \int f_n = \int \lim_{n\to\infty} f_n. \tag{13.458}$$

Nous remplaçons maintenant f_n par sa valeur en termes des u_i et dans le membre de gauche nous permutons l'intégrale avec la somme finie :

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} \int u_n = \int \sum_{i=0}^{\infty} u_n, \tag{13.459}$$

ce qu'il fallait démontrer. □

13.9.3.2 Lemme de Fatou

Lemme 13.164 (Lemme de Fatou)**.**
Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $f_n \colon \Omega \to [0,\infty]$ une suite de fonctions mesurables. Alors la fonction $f(x) = \liminf f_n(x)$ est mesurable et

$$\int_\Omega \liminf f_n d\mu \leqslant \liminf \int_\Omega f d\mu. \tag{13.460}$$

Démonstration. Nous posons

$$g_n(x) = \inf_{i\geqslant n} f_i(x). \tag{13.461}$$

Cela est une suite croissance de fonctions positives mesurables telles que, par définition,

$$\lim_{n\to\infty} g_n(x) = \liminf f_n(x). \tag{13.462}$$

Nous pouvons y appliquer le théorème de la convergence monotone,

$$\lim_{n\to\infty} \int g_n(x) = \int \liminf f_n(x). \tag{13.463}$$

Par ailleurs, pour chaque $i \geqslant n$ nous avons

$$\int g_n \leqslant \int f_i, \tag{13.464}$$

en passant à l'infimum nous avons

$$\int g_n \leqslant \inf_{i\geqslant n} \int f_i, \tag{13.465}$$

et en passant à la limite nous avons

$$\int \liminf f_n = \lim_{n\to\infty} \int g_n \leqslant \lim_{n\to\infty} \inf_{i\geqslant n} \int f_i = \lim_{i\to\infty} \inf \inf f_i. \tag{13.466}$$

□

L'inégalité donnée dans ce lemme n'est en général pas une égalité, comme le montre l'exemple suivant :

$$f_i = \begin{cases} \mathbb{1}_{[0,1]} & \text{si } i \text{ est pair} \\ \mathbb{1}_{[1,2]} & \text{si } i \text{ est impair.} \end{cases} \tag{13.467}$$

Nous avons évidemment $g_n(x) = 0$ tandis que $\int_{[0,2]} f_i = 1$ pour tout i.

Théorème 13.165 ([273]).
Soient f, g des fonctions étagées positives sur $(\Omega, \mathcal{A}, \mu)$. Alors si $\alpha \in [0, \infty]$ nous avons

(1)

$$\int_\Omega (\alpha f) d\mu = \alpha \int_\Omega f d\mu. \tag{13.468}$$

(2)

$$\int_\Omega (f + g) d\mu = \int_\Omega f d\mu + \int_\Omega g d\mu. \tag{13.469}$$

(3) Si $a_k \in \mathbb{R}^+$ et si les f_k sont étagées positives,

$$\int_\Omega \left(\sum_{k=1}^n a_k f_k \right) = \sum_{k=1}^n a_k \left(\int_\Omega f_k d\mu \right). \tag{13.470}$$

Démonstration. En ce qui concerne le produit par un nombre, tout repose sur le fait que

$$(\alpha f)^{-1}(\alpha a_i) = f^{-1}(a_i), \tag{13.471}$$

ce qui fait que si la représentation canonique de f est $f = \sum_i a_i \mathbb{1}_{A_i}$ alors la représentation canonique de αf est $\alpha f = \sum_i (\alpha a_i) \mathbb{1}_{A_i}$. Donc

$$\int_\Omega \alpha f d\mu = \sum_i \alpha a_i \mu(A_i) = \alpha \sum_i a_i \mu(A_i) = \alpha \int_\Omega f d\mu. \tag{13.472}$$

Pour la somme c'est plus lourd. Soient les formes canoniques

$$f = \sum_i a_i \mathbb{1}_{A_i} \tag{13.473a}$$

$$g = \sum_j b_j \mathbb{1}_{B_i}. \tag{13.473b}$$

Vu que l'union des B_j est Ω nous avons l'union disjointe $A_i = \bigcup_j A_i \cap B_j$ et donc $\mu(A_i) = \sum_j \mu(A_i \cap B_j)$. Nous avons donc pour les intégrales :

$$\int_\Omega f d\mu = \sum_i a_i \sum_j \mu(A_i \cap B_j) \tag{13.474a}$$

$$\int_\Omega g d\mu = \sum_i b_k \sum_l \mu(B_k \cap A_l). \tag{13.474b}$$

Pour la somme :

$$\int_\Omega f d\mu + \int_\Omega g d\mu = \sum_{k,l}(a_k + b_l)\mu(A_k \cap B_l). \tag{13.475}$$

Nous devons maintenant évaluer $\int_\Omega (f+g)d\mu$. Pour cela nous remarquons que si $c \in (f+g)(\Omega)$ (l'ensemble des valeurs atteintes pas $f+g$), alors nous notons

$$I_c = \{(k,l) \text{ tel que } a_k + b_l = c\} \tag{13.476}$$

et nous avons

$$\{f+g=c\} = \bigcup_{(k,l)\in I_c} (A_k \cap B_l), \tag{13.477}$$

et comme cette union est disjointe, nous pouvons faire la somme des mesures :

$$\mu(f+g=c) = \sum_{(k,l)\in I_c} \mu(A_k \cap B_l). \tag{13.478}$$

Cela nous permet de faire le calcul suivant :

$$\int_\Omega (f+g)d\mu = \sum_{c\in(f+g)(\Omega)} c\mu(f+g=c) \tag{13.479a}$$

$$= \sum_{c\in(f+g)(\Omega)} c \sum_{(k,l)\in I_c} \mu(A_k \cap B_l) \tag{13.479b}$$

$$= \sum_{c\in(f+g)(\Omega)} \sum_{(k,l)\in I_c} (a_k + b_l)\mu(A_k \cap B_l) \tag{13.479c}$$

Dans cette double somme, tous les couples (k,l) sont tirés une et une seule fois parce qu'ils sont tous dans un et un seul des I_c, donc

$$\int_\Omega (f+g)d\mu = \sum_{c\in(f+g)(\Omega)} \sum_{(k,l)\in I_c} (a_k + b_l)\mu(A_k \cap B_l) \tag{13.480a}$$

$$= \sum_{(k,l)} (a_k + b_l)\mu(A_k \cap B_l) \tag{13.480b}$$

$$= \int_\Omega f d\mu + \int_\Omega g d\mu. \tag{13.480c}$$

□

Remarque 13.166.
Si $f = \sum_k a_k \mathbb{1}_{A_k}$ n'est pas une décomposition canonique, il n'en reste pas moins que chacun des $\mathbb{1}_{A_k}$ est la forme canonique de lui-même. Donc le théorème 13.165 s'applique et nous avons quand même

$$\int_\Omega f d\mu = \sum_k a_k \mu(A_k). \tag{13.481}$$

Proposition 13.167.
Soient deux fonctions mesurables $f, g\colon \Omega \to [0,+\infty]$. Alors

$$\int_\Omega (f+g) = \int_\Omega f + \int_\Omega g. \tag{13.482}$$

Démonstration. Soient des suites $f_n \to f$ et $g_n \to g$ fournies par le théorème fondamental d'approximation 13.105. Par le théorème de la convergence monotone 13.160 nous avons d'une part

$$\lim_{n\to\infty} \int_\Omega (f_n + g_n) = \int_\Omega \int (f+g), \tag{13.483}$$

et par le théorème 13.165 nous avons d'autre part

$$\lim_{n\to\infty} \int_\Omega (f_n + g_n) = \lim_{n\to\infty} \Big(\int f_n + \int g_n \Big) = \int f + \int g \tag{13.484}$$

où nous avons encore utilisé la convergence monotone.

En égalant les deux, nous avons notre résultat. □

13.9.4 Fonctions à valeurs réelles

L'intégrale d'une fonction à valeurs dans $[0,+\infty]$ étant faite, nous passons aux fonctions à valeurs dans $[-\infty,+\infty]$.

Proposition-définition 13.168 ([1]).
Soit une fonction mesurable $f\colon \Omega \to \bar{\mathbb{R}}$. Nous considérons les deux fonction suivantes à valeurs dans $[0,+\infty]$:

$$f^+(x) = \begin{cases} 0 & \text{si } f(x) < 0 \\ f(x) & \text{si } f(x) \geqslant 0. \end{cases} \tag{13.485a}$$

$$f^-(x) = \begin{cases} 0 & \text{si } f(x) > 0 \\ -f(x) & \text{si } f(x) \leqslant 0. \end{cases} \tag{13.485b}$$

Nous avons $\int_\Omega |f| < \infty$ si et seulement si $\int_\Omega f^+ < \infty$ et $\int_\Omega f^- < \infty$.

*Dans ce cas nous disons que f est **intégrable** au sens de Lebesgue et nous posons*

$$\int_\Omega f = \int_\Omega f^+ - \int_\Omega f^- \tag{13.486}$$

Démonstration. Vu que f est mesurable, les fonctions f^+ et f^- sont également mesurables et nous avons l'égalité

$$|f| = f^+ + f^-. \tag{13.487}$$

La proposition 13.167 nous dit alors que

$$\int_\Omega |f| = \int_\Omega f^+ + \int_\Omega f^-. \tag{13.488}$$

Dans cette égalité, tous les nombres sont dans $[0,\infty]$. Le membre de gauche vaut $+\infty$ si et seulement si au moins un des deux de droite vaut $+\infty$. □

Nous verrons comment donner un sens à $\int_\Omega f$ dans certains cas où f n'est pas intégrable sur Ω dans la section 13.14.7 sur les intégrales impropres.

Nous définissons aussi

$$\mu(f) = \int_\Omega f \tag{13.489}$$

si f est une fonction mesurable sur Ω.

Lemme 13.169.
Pour $f\colon \Omega \to \mathbb{R}$ nous avons $\int_\Omega |f| < \infty$ si et seulement si $\int_\Omega f$ existe et est finie.

Démonstration. Deux sens.

$\Rightarrow$ La proposition 13.168 nous indique que $\int_\Omega f^+$ et $\int_\Omega f^-$ sont finies. Dans ce cas, la partie « définition » de 13.168 donne $\int_\Omega f = \int_\Omega f^+ - \int_\Omega f^- < \infty$.

$\Leftarrow$ Nous n'avons défini $\int_\Omega f$ que dans le cas où les intégrales de f^+ et f^- sont finies.

□

Ce lemme justifie pourquoi nous appelons l'espace L^1 l'espace des « fonctions intégrables ».

Remarque 13.170.
Dans $\mathbb{R}^d$, quasiment toutes les fonctions et ensembles sont mesurables. En effet la construction d'ensembles non mesurables demande obligatoirement l'utilisation de l'axiome du choix ; de tels ensembles doivent être construits « exprès pour ». Il y a très peu de chances pour que vous tombiez sur un ensemble non mesurable de $\mathbb{R}^d$ sans que vous ne vous en rendiez compte.

Il y en a un en l'exemple 13.141.

Remarque 13.171.
« Mesurable » ne signifie pas « intégrable ». Par exemple la fonction

$$\begin{aligned} f\colon \mathbb{R} &\to \bar{\mathbb{R}} \\ \omega &\mapsto \begin{cases} \frac{1}{\omega} & \text{si } \omega \neq 0 \\ \infty & \text{si } \omega = 0. \end{cases} \end{aligned} \tag{13.490}$$

est mesurable, mais non intégrable.

13.9.5 Additivité de l'intégrale

Lemme 13.172.
Soit une fonction $f\colon \Omega \to \mathbb{R}$ telle que $|f(x)| \leqslant g(x)$ pour tout $x \in \Omega$. Si g est intégrable, alors f est intégrable.

Démonstration. La fonction g est manifestement à valeurs réelles positives. La proposition 13.158 nous dit alors que $|f|$ est intégrable. Ensuite c'est au tour de la proposition 13.168 de conclure à l'intégrabilité de f. □

Proposition 13.173.
Soient deux fonctions intégrables sur $(S, \mathcal{F}, \mu)$ et à valeurs dans $\mathbb{C}$. Alors $f + g$ est intégrable et

$$\int_S (f+g) d\mu = \int_S f d\mu + \int_S g d\mu. \tag{13.491}$$

Démonstration. En plusieurs étapes suivant la généralité de f et g.

Si f et g sont étagées et positives C'est le théorème 13.165(2) déjà prouvé.

Si f et g sont à valeurs positives Le théorème fondamental d'approximation 13.105 nous permet de considérer des suites croissantes de fonctions étagées positives (f_k) et (g_k) qui vérifient $f_k \to f$ et $g_k \to g$.
Pour chaque k nous avons

$$\int_S (f_k + g_k) d\mu = \int_S f_k d\mu + \int_S g_k d\mu. \tag{13.492}$$

De plus, la suite $k \mapsto f_k + g_k$ est une suite croissante de fonctions étagées positives convergeant vers $f + g$. Le théorème de la convergence monotone 13.160 nous permet donc de passer à la limité dans (13.492) et de permuter toutes les limites avec toutes les intégrales, des deux côtés.

f et g à valeurs réelles Il faut diviser le domaine en de nombreuses régions suivant les signes de f, g et $f + g$.

□

Nous prouvons à présent l'additivité de l'intégrale pour des unions finie. Une version pour les unions dénombrables sera donnée dans les propositions 13.187 et 13.188.

Proposition 13.174 (σ-additivité finie)**.**
Si $A, B \subset \Omega$ sont des parties disjointes de $(\Omega, \mathcal{A}, \mu)$ et si $f\colon \Omega \to \mathbb{R}$ est intégrable sur $A \cup B$ alors les intégrales $\int_A f$ et $\int_B f$ existent et

$$\int_{A \cup B} f = \int_A f + \int_B f. \tag{13.493}$$

Démonstration. Vu que A et B sont disjoints, $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B$. En utilisant alors le lemme 13.159 et la proposition 13.173 nous avons le calcul

$$\int_{A \cup B} f = \int_\Omega f \mathbb{1}_{A \cup B} = \int_\Omega f \mathbb{1}_A + \int_\Omega f \mathbb{1}_B = \int_A f + \int_B f. \tag{13.494}$$

□

13.9.6 Fonctions à valeurs vectorielles (dimension finie)

Nous voulons intégrer des fonctions du type

$$f\colon \Omega \to V \tag{13.495}$$

où Ω et V sont des espaces vectoriels. Nous expliquons à présent plus précisément le cadre.

13.175.
Nous considérons à présent un espace vectoriel normé $(V, \|.\|)$ de dimension finie, et un espace mesuré $(\Omega, \mathcal{A}, \mu)$.

Attention à ne pas confondre espace de départ et espace d'arrivée. Vu que V est un espace topologique, nous avons bien entendu les boréliens de V, et pour peut que nous ayons une mesure sur V (qui qui n'est pas compliqué à créer à partir de celle canonique de $\mathbb{R}^n$ et un isomorphisme), nous avons déjà une définition de $\int_V f d\mu$ lorsque $f\colon V \to \mathbb{R}$.

Ici nous nous proposons non d'intégrer $f\colon V \to \mathbb{R}$ mais bien $f\colon (\Omega, \mathcal{A}, \mu) \to V$ où V est un espace vectoriel normé.

Le lemme suivant est la point de départ pour définir les intégrales de fonctions à valeurs dans un espace vectoriel de dimension finie. Pour les fonctions à valeurs dans un espace de dimension infine (par exemple de Banach), il existe des choses, mais c'est un peu plus compliqué.

Lemme 13.176 ([1])**.**
Soit un espace vectoriel V réel de dimension finie, muni de la norme N. Soient une base $\{e_i\}$ de V, et une fonction $f\colon (\Omega, \mathcal{A}, \mu) \to V$ telle que la norme $N(f)\colon \Omega \to \mathbb{R}^+$ soit intégrable. Nous notons f_i les composantes de f : $f(x) = \sum_i f_i(x)e_i$.

Alors pour chaque i,

(1) la fonction $|f_i|\colon \Omega \to \mathbb{R}^+$ est intégrable,

(2) la fonction $f_i\colon \Omega \to \mathbb{R}$ est intégrable.

Démonstration. Si V était un espace muni d'un produit scalaire, et si la base $\{e_i\}$ était orthonormée, ce serait facile parce que la norme majore toutes les composantes. Hélas, ce n'est pas spécialement le cas. La base $\{e_i\}$ n'est pas spécialement orthonormée et même la norme N ne dérive pas spécialement d'un produit scalaire.

Nous allons utiliser l'équivalence de toutes les normes en dimension finie (théorème 7.224) pour nous ramener au cas d'une norme euclidienne.

Nous considérons sur V la norme « euclidienne » construite sur la base $\{e_i\}$: $\|\sum_i v_i e_i\| = \sum_i |v_i|^2$. Par équivalence des normes nous avons des nombres non nuls λ_1 et λ_2 tels que

$$N(v) \leqslant \lambda_1 \|v\|, \tag{13.496}$$

et

$$\|v\| \leqslant \lambda_2 N(v) \tag{13.497}$$

pour tout $v \in V$. Pour un i fixé nous avons alors les majorations

$$N\big(f_i(x)e_i\big) \leqslant \lambda_1 \|f_i(x)e_i\| \leqslant \lambda_1 \|f(x)\| \leqslant \lambda_1\lambda_2 N\big(f(x)\big). \tag{13.498}$$

En posant $N_i = N(e_i)$ nous avons la majoration [47]

$$|f_i(x)| \leqslant \frac{\lambda_1\lambda_2}{N(e_i)} N\big(f(x)\big). \tag{13.499}$$

L'application

$$\begin{aligned} |f_i|\colon \Omega &\to \mathbb{R}^+ \\ x &\mapsto |f_i(x)| \end{aligned} \tag{13.500}$$

47. Vous notez l'utilisation de la condition (2) de la définition 7.113 de la norme pour « convertir » la norme N en valeur absolue.

est donc une fonction à valeurs réelles positives, majorée par une fonction intégrable (la fonction $x \mapsto N\big(f(x)\big)$). Elle est donc intégrable par le lemme 13.172.

La fonction f_i elle-même est alors intégrable par la proposition 13.168. □

Notons que ce lemme est en réalité très simple si V est un espace vectoriel normé dont la norme découle d'un produit scalaire, comme c'est le cas pour $\mathbb{C}$. D'ailleurs, il ne faut pas se voiler la face : le cas d'intégrales de fonctions à valeurs dans $\mathbb{C}$ sera dans le Frido le cas de loin le plus courant. À ce propos, nous n'avons pas encore défini ce que nous voulons noter $\int_\Omega f d\mu$ lorsque f est une fonction à valeurs vectorielles. Comblons vite ce manque ...

Proposition-définition 13.177 ([1]).
Soit une fonction $f\colon \Omega \to V$ où V est un espace vectoriel normé de dimension finie. Soit une base $\{e_i\}$ de V. Si la fonction $\|f\|\colon \Omega \to \mathbb{R}^+$ est intégrable, alors

(1) toutes les composantes $f_i\colon \Omega \to \mathbb{R}$ sont intégrables,

(2) le vecteur

$$\sum_i \Big(\int_\Omega f_i\Big) e_i \tag{13.501}$$

ne dépend pas de la base choisie.

*Dans ce cas, la fonction f est dite **intégrable** et nous définissons*

$$\int_\Omega f d\mu = \sum_i \Big(\int_\Omega f_i\Big) e_i. \tag{13.502}$$

Démonstration. Le fait que les composantes soient intégrables est le lemme 13.176. Soient deux bases de V, $\{e_i\}$ et $\{s_\alpha\}$, liées conformément à (4.224) par la relation $s_\alpha = \sum_i Q_{i\alpha} e_i$ pour une certaine matrice inversible Q. Nous avons pour tout $x \in \Omega$:

$$f(x) = \sum_i f_i(x) e_i = \sum_\alpha f_\alpha(x) s_\alpha \tag{13.503}$$

avec $f_\alpha(x) = \sum_i f_i(x) Q^{-1}_{\alpha i}$ par la proposition 4.115.

Notons pour être pointilleux que les ensembles $\{e_i\}$ et $\{s_\alpha\}$ ne sont pas indexés par le même ensemble, de telle sorte que f_i ne peut pas être confondu avec f_α, même lorsqu'on attribue des valeurs à i et à α.

Comme combinaisons linéaires des fonctions f_i qui sont intégrables, les fonctions f_α sont intégrables (proposition 13.173). En écrivant $\int_\Omega f$ par rapport à la base $\{s_\alpha\}$ nous trouvons :

$$\sum_\alpha \Big(\int f_\alpha\Big) s_\alpha = \sum_\alpha \Big(\int \sum_i f_i(x) Q^{-1}_{\alpha i} dx\Big) \sum_j Q_{j\alpha} e_j \tag{13.504a}$$

$$= \sum_j \int \sum_{\alpha i} f_i(x) Q^{-1}_{\alpha i} Q_{j\alpha} dx e_j \tag{13.504b}$$

$$= \sum_j \int f_j(x) dx e_j \tag{13.504c}$$

$$= \sum_j \Big(\int f_i\Big) e_i \tag{13.504d}$$

où nous avons permuté des sommes finies et des intégrales des fonctions f_i, à valeurs dans $\mathbb{R}$ en vertu de la proposition 13.173 □

La proposition suivante est, pour les intégrales à valeurs vectorielles, analogue à la proposition 13.168.

Proposition 13.178.
Soit une fonction mesurable $f\colon \Omega \to (V, \|.\|)$. Soit une base $\{e_i\}$ de V et la décomposition $f = \sum_i f_i e_i$.

Nous avons équivalence entre

(1) $\int_\Omega \|f\| < \infty$
(2) $\int_\Omega |f_i| < \infty$
(3) $\int_\Omega f_i^+ < \infty$ *et* $\int_\Omega f_i^- < \infty$.

Démonstration. L'équivalence entre les points (2) et (3) est la proposition 13.168. Nous démontrons l'équivalence entre (1) et (2).

Vu que toutes les normes sont équivalentes sur V, nous considérons en particulier la norme associée à la base $\{e_i\}$ donnée par

$$N(x) = \sum_i |x_i|. \tag{13.505}$$

Il existe des constantes λ_1 et λ_2 telles que

$$\lambda_1\big(\sum_i |f_i(x)|\big) \leqslant \|f(x)\| \leqslant \lambda_2\big(\sum_i |f_i(x)|\big) \tag{13.506}$$

pour tout $x \in \Omega$.

La première inégalité dit que si $\int_\Omega \|f\| < \infty$, alors $\lambda_1\big(\sum_i \int_\Omega |f - i|\big) < \infty$. Et vu que chacun des termes est positif, ils sont tous finis.

La seconde inégalité donne l'implication dans réciproque. □

13.9.7 Quelques propriétés

Le lemme suivant nous aide à détecter des fonctions presque partout nulles.

Lemme 13.179.
Soit f une fonction mesurable positive ou nulle telle que

$$\int_\Omega f d\mu = 0. \tag{13.507}$$

Alors $f = 0$ μ-presque partout.

Démonstration. L'ensemble des points $x \in \Omega$ tels que $f(x) \neq 0$ peut s'écrire comme une union dénombrable disjointe :

$$\{x \in \Omega \text{ tel que } f(x) \neq 0\} = \bigcup_{i=0}^{\infty} E_i \tag{13.508}$$

avec

$$E_0 = \{x \in \Omega \text{ tel que } f(x) > 1\} \tag{13.509a}$$
$$E_i = \{x \in \Omega \text{ tel que } \frac{1}{i+1} \leqslant f(x) < \frac{1}{i}\}. \tag{13.509b}$$

Si un des ensembles E_i est de mesure non nulle, alors nous pouvons considérer la fonction simple $h(x) = \frac{1}{i+1}\mathbb{1}_{E_i}$ dont l'intégrale sur Ω est strictement positive. Par conséquent le supremum de la définition (13.424) est strictement positif.

Nous savons donc que $\mu(E_i) = 0$ pour tout i. Étant donné que la mesure d'une union disjointe dénombrable est égale à la somme des mesures, nous avons

$$\mu\{x \in \Omega \text{ tel que } f(x) \neq 0\} = 0, \tag{13.510}$$

ce qui signifie que f est nulle μ-presque partout. □

Corolaire 13.180.
Soit f une fonction mesurable sur l'espace mesuré $(\Omega, \mathcal{A}, \mu)$ telle que

$$\int_\Omega f \mathbb{1}_{f>0} d\mu = 0. \tag{13.511}$$

Alors $f \leqslant 0$ presque partout.

Démonstration. Nous avons l'égalité d'ensembles

$$\{f \mathbb{1}_{f>0} \neq 0\} = \{\mathbb{1}_{f>0} \neq 0\}. \tag{13.512}$$

Mais lemme 13.179 implique que $f\mathbb{1}_{f>0}$ est nulle presque partout, c'est-à-dire que la mesure de l'ensemble du membre de gauche est nulle par conséquent

$$\mu\{\mathbb{1}_{f>0} \neq 0\} = 0. \tag{13.513}$$

Cela signifie que la fonction f est presque partout négative ou nulle. □

13.9.8 Permuter limite et intégrale

13.9.8.1 Convergence uniforme

Proposition 13.181 (Permuter limite et intégrale)**.**
Soit $f_n \to f$ uniformément sur un ensemble mesuré A de mesure finie. Alors si les fonctions f_n et f sont intégrables sur A, nous avons

$$\lim_{n\to\infty} \int_A f_n = \int_A \lim_{n\to\infty} f_n. \tag{13.514}$$

Démonstration. Notons f la limite de la suite (f_n). Pour tout n nous avons les majorations

$$\left| \int_A f_n d\mu - \int_A f d\mu \right| \leqslant \int_A |f_n - f| d\mu \tag{13.515a}$$

$$\leqslant \int_A \|f_n - f\|_\infty d\mu \tag{13.515b}$$

$$= \mu(A) \|f_n - f\|_\infty \tag{13.515c}$$

où $\mu(A)$ est la mesure de A. Le résultat découle maintenant du fait que $\|f_n - f\|_\infty \to 0$. □

Il existe un résultat considérablement plus intéressant que cette proposition. En effet, l'intégrabilité de f n'est pas nécessaire. Cette hypothèse peut être remplacée soit par l'uniforme convergence de la suite (théorème 13.182), soit par le fait que les normes des f_n sont uniformément bornées (théorème de la convergence dominée de Lebesgue 13.184).

Théorème 13.182 ([290])**.**
La limite uniforme d'une suite de fonctions intégrables sur un borné est intégrable, et on peut permuter la limite et l'intégrale.

Plus précisément, soit A un ensemble de μ-mesure finie et $f_n \colon A \to \mathbb{R}$ des fonctions intégrables sur A. Si la limite $f_n \to f$ est uniforme, alors f est intégrable sur A et nous pouvons inverser la limite et l'intégrale :

$$\lim_{n\to\infty} \int_A f_n = \int_A \lim_{n\to\infty} f_n. \tag{13.516}$$

Démonstration. Soit $\epsilon > 0$ et n tel que $\|f_n - f\|_\infty \leqslant \epsilon$ (ici la norme uniforme est prise sur A). Étant donné que f_n est intégrable sur A, il existe une fonction simple φ_n qui minore f_n telle que

$$\left| \int_A \varphi_n - \int_A f_n \right| < \epsilon. \tag{13.517}$$

La fonction $\varphi_n + \epsilon$ est une fonction simple qui majore la fonction f. Si ψ est une fonction simple qui minore f, alors

$$\int_A \psi \leqslant \int_A \varphi_n + \epsilon \leqslant \int_A f_n + \epsilon\mu(A). \tag{13.518}$$

Par conséquent le supremum qui définit $\int_A f$ existe, ce qui montre que f est intégrable. Le fait qu'on puisse inverser la limite et l'intégrale est maintenant une conséquence de la proposition 13.181. □

Remarque 13.183.
L'hypothèse sur le fait que A soit de mesure finie est importante. Il n'est pas vrai qu'une suite uniformément convergente de fonctions intégrables est intégrables. En effet nous avons par exemple la suite

$$f_n(x) = \begin{cases} 1/x & \text{si } x < n \\ 0 & \text{sinon} \end{cases} \tag{13.519}$$

qui converge uniformément vers $f(x) = 1/x$ sur $A = [1, \infty[$. Le limite n'est cependant guerre intégrable sur A.

13.9.8.2 Convergence dominée de Lebesgue

Théorème 13.184 (Convergence dominée de Lebesgue)**.**
Soit une suite de fonctions $(f_n)_{n\in\mathbb{N}}$ sur $(\Omega, \mathcal{A}, \mu)$ à valeurs dans $\mathbb{C}$ ou $\mathbb{R}$. Nous supposons que

(1) Pour chaque n nous avons $f_n \in L^1(\Omega, \mathcal{A}, \mu)$,

(2) $f_n \to f$ simplement presque partout sur Ω,

(3) Il existe une fonction $g \in L^1(\Omega)$ telle que

$$|f_n(x)| \leqslant g(x) \tag{13.520}$$

pour presque[48] *tout $x \in \Omega$ et pour tout $n \in \mathbb{N}$.*

Alors

(1) f est intégrable,

(2) $\lim_{n\to\infty} \int_\Omega f_n = \int_\Omega f$,

(3) $\lim_{n\to\infty} \int_\Omega |f_n - f| = 0$.

Démonstration. La fonction limite f est intégrable parce que $|f| \leqslant g$ et g est intégrable[49]. Par hypothèse nous avons

$$-g(x) \leqslant f_n(x) \leqslant g(x). \tag{13.521}$$

En particulier la fonction $g_n = f_n + g$ est positive et mesurable si bien que le lemme de Fatou 13.164 implique

$$\int_\Omega \liminf g_n \leqslant \liminf \int_\Omega g_n. \tag{13.522}$$

Évidemment nous avons $\liminf g_n = f + g$, de telle sorte que

$$\int f + \int g \leqslant \liminf \int g_n = \liminf \int f_n + \int g, \tag{13.523}$$

et le nombre $\int g$ étant fini, nous pouvons le retrancher des deux côtés de l'inégalité :

$$\int f \leqslant \liminf \int f_n. \tag{13.524}$$

48. S'il n'y avait pas le « presque » ici, ce théorème serait à peu près inutilisable en probabilité ou en théorie des espaces L^p, comme dans la démonstration du théorème de Fischer-Riesz 26.41 par exemple.
49. Par le lemme 13.172

Afin d'obtenir une minoration de $\int f$ nous refaisons exactement le même raisonnement en utilisant la suite de fonctions $k_n = -f_n \to k = -f$. Nous obtenons que

$$\int k \geqslant \liminf \int k_n = -\limsup \int f_n, \tag{13.525}$$

et par conséquent

$$\liminf \int f_n \leqslant \int f \leqslant \limsup \int f_n. \tag{13.526}$$

La limite supérieure étant plus grande ou égale à la limite inférieure, les trois quantités dans les inégalités (13.526) sont égales.

Nous prouvons maintenant le troisième point. Soit la suite de fonctions

$$h_n(x) = |f_n(x) - f(x)| \tag{13.527}$$

qui tend ponctuellement vers zéro. De plus

$$h_n(x) \leqslant |f_n(x)| + |f(x)| \leqslant 2g(x), \tag{13.528}$$

ce qui prouve que les h_n majorés par une fonction intégrable. Donc

$$\lim_{n\to\infty} \int_\Omega |f_n - f| = \lim_{n\to\infty} \int_\Omega h_n(x)dx = \int_\Omega \lim_{n\to\infty} |f_n(x) - f(x)| = 0 \tag{13.529}$$

□

Remarque 13.185.
Lorsque nous travaillons sur des problèmes de probabilités, la fonction g peut être une constante parce que les constantes sont intégrables sur un espace de probabilité.

Corolaire 13.186.
Soit $(a_i)_{i\in\mathbb{N}}$ une suite numérique absolument convergente. Alors elle est convergente. Il en est de même pour les séries de fonctions si on considère la convergence ponctuelle.

Démonstration. L'hypothèse est la convergence de l'intégrale $\int_{\mathbb{N}} |a_i| dm(i)$ où dm est la mesure de comptage. Étant donné que $|a_i| \leqslant |a_i|$, la fonction a_i (fonction de i) peut jouer le rôle de g dans le théorème de la convergence dominée de Lebesgue (théorème 13.184). □

13.9.9 Additivité de l'intégrale de Lebesgque

Les propositions 13.187 et 13.188 démontrent la même chose. La différence est la méthode utilisée pour permuter une somme et une intégrale. Dans le premier cas, nous utilisons la convergence monotone (et sommes obligés de séparer le cas où f est positive), alors que dans le second cas, nous utilisons la convergence dominée de Lebesgue, et nous ne devons pas faire de séparation d'après la positivité de f.

Proposition 13.187 (σ-additivité dénombrable[1])**.**
Si $(A_i)_{i\in\mathbb{N}}$ sont des parties mesurables disjointes de $(\Omega, \mathcal{A}, \mu)$ et si $f\colon \Omega \to \mathbb{R}$ est intégrable sur $\bigcup_{i=0}^\infty A_i$ alors les intégrales $\int_{A_i} f d\mu$ existent et

$$\int_{\bigcup_i A_i} f d\mu = \sum_{i=0}^\infty \int_{A_i} f d\mu. \tag{13.530}$$

Démonstration. En deux cas d'après la positivité de f.

Si f est positive Nous posons $f_N = f\mathbb{1}_{\bigcup_{i=0}^N A_i}$. Cette suite de fonctions vérifie la limite

$$\lim_{N\to\infty} f_N = f\mathbb{1}_{\bigcup_{i=0}^\infty}. \tag{13.531}$$

De plus, pour chaque N nous avons

$$\int_\Omega f_N = \int_\Omega f \mathbb{1}_{\bigcup_i A_i} = \int_{\bigcup_i A_i} f = \sum_{i=0}^{N} \int_{A_i} f \tag{13.532}$$

Justifications :

— La proposition 13.159 pour l'introduction de la fonction caractéristique de $\bigcup_i A_i$

— La proposition 13.174 qui traite le cas de la sous-additivité finie pour la dernière égalité.

La suite $(f_N)_{n\in\mathbb{N}}$ est une suite croissante de fonctions mesurables[50] et positives. Donc le théorème de la convergence monotone 13.160 s'applique et

$$\sum_{i=0}^{\infty} \int_{A_i} f = \lim_{N\to\infty} \sum_{i=0}^{N} \int_{A_i} f = \lim_{N\to\infty} \int_{\bigcup_{i=0}^{N} A_i} f = \lim_{N\to\infty} \int_\Omega f_N \tag{13.533a}$$

$$= \int_\Omega \lim_{N\to\infty} f_N = \int_\Omega f \mathbb{1}_{\bigcup_{i=0}^{\infty}} = \int_{\bigcup_i A_i} f d\mu. \tag{13.533b}$$

Si f est à valeurs réelles Si f est à valeurs dans $\mathbb{R}$, alors $f = f_+ - f_-$ où f_+ et f_- sont intégrables. Nous avons alors

$$\int_{\bigcup_{i=0}^{\infty} A_i} f d\mu = \int_{\bigcup_i A_i} f_+ - \int_{\bigcup_i A_i} f_- \tag{13.534a}$$

$$= \sum_{k=0}^{\infty} \int_{A_k} f_+ - \int_{k=0}^{\infty} f_- \tag{13.534b}$$

$$= \sum_{k=0}^{\infty} \Big(\int_{A_k} f_+ - \int_{A_k} f_- \Big) \tag{13.534c}$$

$$= \sum_{k=0}^{\infty} \int_{A_k} f. \tag{13.534d}$$

Justifications :

— Pour (13.534c), c'est l'associativité de la somme, proposition 10.60.

— Pour (13.534d), c'est la proposition 13.173.

□

Proposition 13.188 (σ-additivité[291])**.**
Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$. Nous considérons des parties disjointes $\{A_i\}_{i\in\mathbb{N}}$ de Ω telles que $\bigcup_{k=0}^{\infty} A_k = \Omega$. Si $f \in L^1(\Omega)$, alors

$$\int_\Omega f d\mu = \sum_{k=0}^{\infty} \int_{A_k} f d\mu. \tag{13.535}$$

Démonstration. Nous posons $\Omega_n = \bigcup_{k=0}^{n} A_k$ ainsi que $f_n = f \mathbb{1}_{\Omega_n}$. Pour chaque $N \in \mathbb{N}$ nous avons

$$\sum_{k=0}^{N} \int_{A_k} f d\mu = \int_{\bigcup_{k=0}^{N} A_k} f \tag{13.536a}$$

$$= \int_{\Omega_N} f \tag{13.536b}$$

$$= \int_\Omega f_N. \tag{13.536c}$$

Justifications :

50. La fonction f elle-même est mesurable ; c'est inclus à la définition de « intégrable ».

— Pour (13.536a), c'est la proposition 13.174 qui traite du cas de sommes finies.
— Pour (13.536c) c'est la proposition 13.187.

L'idée est maintenant de passer à la limite des deux côtés de (13.536). Voici le raisonnement :

— Nous montrons qu'à droite, la limite existe et vaut $\int_\Omega f d\mu$.
— Le fait que la limite du membre de droite existe implique l'existence de la limite du membre de gauche.
— La limite du membre de gauche vaut $\sum_{k=0}^\infty \int_{A_k} f d\mu$.

La limite du membre de droite s'établi avec le théorème de la convergence dominée de Lebesgue 13.182.

— Nous avons convergence simple $f_n \to f$ parce que $\bigcup_{n=0}^\infty A_i = \Omega$.
— La fonction $g = |f|$ est intégrable sur Ω parce que $f \in L^1(\Omega)$ par hypothèse.
— Pour tout $n \in \mathbb{N}$ et pour tout $x \in \Omega$ nous avons $|f_n(x)| \leqslant g(x)$ parce que $|f_n(x)|$ est soit égal à $g(x)$ soit égal à zéro suivant que $x \in \Omega_n$ ou non.

Donc le théorème de la convergence dominée est applicable. La limite du membre de droite de (13.536) existe et vaut :

$$\lim_{N\to\infty} \int_\Omega f_N = \int_\Omega f. \tag{13.537}$$

Nous pouvons alors prendre aussi la limite du membre de gauche dans (13.536) et obtenir le résultat attendu. □

13.9.10 Produit d'une mesure par une fonction (mesure à densité)

Proposition-définition 13.189 (Produit d'une mesure par une fonction[1, 292])**.**
Soit un espace mesuré $(\Omega, \mathcal{F}, \mu)$ et une fonction mesurable positive $w\colon \Omega \to \bar{\mathbb{R}}^+$. Alors la formule

$$(w \cdot \mu)(A) = \int_A w d\mu \tag{13.538}$$

*pour tout $A \in \mathcal{F}$ définit une mesure positive sur $(\Omega, \mathcal{F})$ appelée **produit** de la mesure μ par la fonction w. La fonction w est la **densité** de la mesure $w \cdot \mu$ par rapport à la mesure μ.*

Démonstration. D'abord $(w \cdot \mu)(\varnothing) = 0$ parce que le lemme 13.157 donne

$$(w \cdot \mu)(\varnothing) = \int_\Omega w \mathbb{1}_\varnothing d\mu = \int_\Omega 0 d\mu = 0 \times \mu(\Omega) = 0 \tag{13.539}$$

où nous avons (éventuellement) utilisé deux fois la convention $0 \times \infty = 0$.

Ensuite si les ensembles A_i sont des éléments deux à deux disjoints de $\mathcal{F}$ alors nous avons $\mathbb{1}_{\bigcup_{i=1}^\infty A_i} = \sum_{i=1}^\infty \mathbb{1}_{A_i}$, et donc

$$(w \cdot \mu)(\bigcup_{i=0}^\infty A_i) = \int_{\bigcup_{i=0}^\infty A_i} w d\mu = \sum_{i=0}^\infty \int_{A_i} w d\mu = \sum_{i=0}^\infty (w \cdot \mu)(A_i). \tag{13.540}$$

où nous avons utilisé la σ-additivité dénombrable de l'intégrale de la proposition 13.187. □

En particulier nous parlons souvent de mesure à densité par rapport à la mesure de Lebesgue. C'est alors la construction suivante.

Définition 13.190.
*Si μ est une mesure sur $\mathbb{R}^d$, une fonction $f\colon \mathbb{R}^d \to \mathbb{R}$ est une **densité** pour μ si pour tout $A \subset \mathbb{R}^d$ nous avons*

$$\mu(A) = \int_A f(x) dx \tag{13.541}$$

où dx est la mesure de Lebesgue.

*Si la mesure μ admet une densité, nous disons que c'est une **mesure à densité** par rapport à la mesure de Lebesgue.*

Exemple 13.191.
Toutes les mesures n'admettent pas de densité. Par exemple la mesure de Dirac donnée par

$$\nu(A) = \begin{cases} 1 & \text{si } 0 \in A \\ 0 & \text{sinon.} \end{cases} \tag{13.542}$$

n'a pas de densité par rapport à la mesure de Lebesgue. △

La meure ν de l'exemple 13.542 admet, au sens des distributions, la mesure de Dirac δ comme densité, mais c'est une autre histoire qui vous sera contée une autre fois.

Proposition 13.192 ([292])**.**
Soit une fonction mesurable $w\colon (S, \mathcal{F}, \mu) \to \bar{\mathbb{R}}^+$.

(1) Si $f\colon S \to \bar{\mathbb{R}}^+$ est mesurable, alors $f \cdot (w \cdot \mu) = (fg) \cdot \mu$.

(2) Si $f\colon S \to \bar{\mathbb{R}}$ ou $\mathbb{C}$ est mesurable, elle est $w \cdot \mu$-intégrable si et seulement si fw est μ-intégrable. Dans ce cas, nous avons encore $f \cdot (w \cdot \mu) = (fg) \cdot \mu$.

Attention : dans le cas où f est à valeurs dans $\mathbb{C}$, alors il faut que w soit à valeurs finies dans $\mathbb{R}$ parce que nous n'avons pas définit $\infty \times z$ lorsque $z \in \mathbb{C}$.

Démonstration. Nous commençons par prouver le résultat pour la fonction caractéristique de l'ensemble mesurable A. Nous avons : $\mathbb{1}_A \cdot (w \cdot \mu)(B) = \int_B \mathbb{1}_A d(w \cdot \mu)$. Mais par définition, l'intégrale d'une fonction indicatrice est la mesure de l'ensemble indiqué. En passant sur le fait que $\mathbb{1}_A \mathbb{1}_B = \mathbb{1}_{A\cap B}$,

$$\int_B \mathbb{1}_A d(w \cdot \mu) = (w \cdot \mu)(A \cap B) = \int_S \mathbb{1}_{A\cap B} w d\mu = \int_S \mathbb{1}_A \mathbb{1}_B w d\mu = \int_B \mathbb{1}_A w d\mu = (\mathbb{1}_A w) \cdot \mu(B). \tag{13.543}$$

Supposons maintenant que f soit une fonction étagées qui s'écrit $f = \sum_k a_k \mathbb{1}_{A_k}$ où les A_k sont des ensembles mesurables disjoints. Alors le calcul est le suivant, en utilisant le fait que sur A_k, on a $a_k = f(x)$:

$$\begin{aligned} f \cdot (g \cdot \mu)B &= \int_B f d(g \cdot \mu) && \text{(13.544a)} \\ &= \sum_k a_k (g \cdot \mu)(A_k \cap B) && \text{(13.544b)} \\ &= \sum_k a_k \int_{A_k \cap B} g f \mu && \text{(13.544c)} \\ &= \int_{A_k \cap} f(x) g(x) d\mu(x) && \text{(13.544d)} \\ &= \sum_k (fg \cdot \mu)(A_k \cap B) && \text{(13.544e)} \\ &= (fg \cdot \mu)(B) && \text{(13.544f)} \end{aligned}$$

parce que les $A_k \cap B$ forment une partition de l'ensemble B (voir le point (2) de la définition 13.20).

Si $f\colon S \to \bar{\mathbb{R}}^+$ est mesurable, le théorème 13.105 donne une suite croissante f_n de fonctions étagées positives convergeant (ponctuellement) vers f. Vu que la fonction w est positive, nous avons aussi la limite positive et croissante $wf_n \to wf$. Ainsi l'utilisation du théorème de la convergence monotone est justifié dans le calcul suivant :

$$\int_S f d(w \cdot \mu) = \lim_{n\to\infty} \int_S f_n d(w \cdot \mu) = \lim_{n\to\infty} \int_S (wf_n) d\mu = \int_S wf d\mu. \tag{13.545}$$

Nous passons maintenant au cas général où f est une fonction à valeurs dans $\bar{\mathbb{R}}$ ou $\mathbb{C}$ (avec w finie dans ce dernier cas). Nous avons la chaine d'équivalences

$\Leftrightarrow$ f est $(w \cdot \mu)$ intégrable

$\Leftrightarrow$ $|f|$ est $(w \cdot \mu)$-intégrable
$\Leftrightarrow$ $|f|w$ est μ-intégrable
$\Leftrightarrow$ $|fw|$ est μ-intégrable.

Si cela est le cas, la formule se démontre en se ramenant au cas déjà prouvé des fonctions positives en utilisant les $(fw)^+ = f^+w$, $(fw)^- = f^-w$ etc. □

13.9.11 Mesure et topologie

Exemple 13.193 (Un compact n'est pas toujours de mesure finie)**.**
Soit l'espace mesurable $(\mathbb{R}, \mathcal{B}\mathrm{or}(\mathbb{R}))$ réel avec ses boréliens et la fonction

$$\begin{aligned} w\colon \big(\mathbb{R}, \mathcal{B}\mathrm{or}(\mathbb{R})\big) &\to \big(\bar{\mathbb{R}}, \mathcal{B}\mathrm{or}(\bar{\mathbb{R}})\big) \\ x &\mapsto \begin{cases} \frac{1}{|x|} & \text{si } x \neq 0 \\ +\infty & \text{si } x = 0. \end{cases} \end{aligned} \tag{13.546}$$

Essayons d'étudier la mesure de densité w par rapport à la mesure de Lebesgue.

w est mesurable Soit un borélien B de $\bar{\mathbb{R}}$. Si B ne contient pas ∞ alors $w^{-1}(B)$ est un borélien de $\mathbb{R}$ par continuité de l'application restreinte $w\colon \mathbb{R}\backslash\{0\} \to \mathbb{R}$. Ici nous avons par exemple appliqué la proposition 13.110 à chacun des deux intervalles $]-\infty, 0[$ et $]0, \infty[$. Si $+\infty \in B$ alors

$$w^{-1}(B) = w^{-1}\big(B\backslash\{0\}\big) \cup w^{-1}(\{\infty\}) = w^{-1}\big(B\backslash\{0\}\big) \cup \{0\}, \tag{13.547}$$

qui est borélien par union de boréliens.

Mesure produit La proposition 13.189 nous assure alors qu'en posant [51]

$$\mu(B) = \int_B \frac{1}{|x|} d\lambda(x) \tag{13.548}$$

où λ est la mesure de Lebesgue, nous avons une mesure.

Mesure du singleton Pour avoir les idées claires, nous pouvons nous demander la mesure $\mu\big(\{0\}\big)$. Nous cela nous devons calculer

$$\int_{\{0\}} \frac{1}{|x|} d\lambda(x) = \int_{\{0\}} w(x) d\lambda(x) \tag{13.549}$$

où là, l'abus de notation n'est plus possible. Mais quelle que soit la fonction étagée $h = \sum_i \alpha_i \mathbb{1}_{A_i}$ considérée,

$$\int_{\{0\}} h(x) d\lambda(x) = \sum_i \alpha_i \lambda\big(A_i \cap \{0\}\big) = 0. \tag{13.550}$$

Attention : ceci n'a rien de particulier à la fonction $x \mapsto 1/|x|$. Lorsqu'une mesure a une densité par rapport à Lebesgue, la mesure d'un singleton sera toujours nulle.

Mesure de la boule compacte Il n'en reste pas moins que $\mu\big([-1, 1]\big) = \infty$.

△

13.194.
En réalité, il n'y a pas de liens forts entre mesure et topologie. Un espace topologique est une chose, et y mettre une mesure en est une autre. Bien entendu, une topologie étant donnée, nous pouvons considérer la tribu des boréliens et y mettre une mesure un peu quelconque. Il n'y a pas de choix canonique.

Notons que même dans l'exemple de compact de mesure infinie 13.193, la mesure introduite n'est pas sans lien avec la topologie de $\mathbb{R}$. En effet pour avoir une mesure à densité par rapport à Lebesgue, nous avons dû prendre une application mesurable par rapport à la tribu des boréliens, laquelle est éminemment liée à la topologie. Il y a donc parfaitement moyen de construire des espaces mesurés tenant compte de la topologie, et ayant des propriétés qui ne sont pas celle attendues.

51. Avec un mini abus de notation : si $0 \in B$, cette notation n'est pas tout à fait correcte.

Quand les choses sont faciles, ça se passe bien. La proposition suivante dit qu'une fonction continue sur un compact y est intégrable ; sauf que pour dire cela de façon précise, il faut un peu bosser parce qu'il y a de écueils à éviter, tels que l'exemple 13.193.

Proposition 13.195 ([1]).
Soit un espace mesuré $(K, \mathcal{A}, \mu)$ et une fonction $f\colon K \to \mathbb{R}$. Nous supposons pas mal de trucs techniques :

(1) La mesure est finie : $\mu(K) < \infty$.

(2) L'ensemble K est par ailleurs un espace topologique compact[52].

(3) La fonction f est continue pour les topologies de K et de $\mathbb{R}$.

(4) La fonction f est mesurable pour la tribu $\mathcal{A}$ de K et la tribu des boréliens de $\mathbb{R}$.

Alors f est intégrable sur K et $\int_K |f| < \infty$.

L'hypothèse (4) ne se déduit pas nécessairement de l'hypothèse (3). Dans les cas usuels, nous avons bien « continue implique mesurable », mais si $\mathcal{A}$ n'a aucun rapport avec la topologie ... hum ...

Démonstration. Si nous écrivons $f(x) = f^+(x) - f^-(x)$ avec f^+ et f^- prenant des valeurs positives ou nulles[293], en vertu de la proposition 13.168, si nous devons prouver séparément $\int_K f^+ < \infty$ et $\int_K f^- < \infty$. Nous allons donc prouver cette proposition en plusieurs étapes.

Si f est positive La fonction f est continue sur K qui est compact (même en tant qu'espace topologique en soi ; il n'est pas nécessaire d'être compact *dans* quelque chose), donc elle a un maximum par le théorème 7.105 nommons M ce maximum. Donc $f\colon K \to [0, M]$. De plus la mesure μ sur K est finie et vérifie disons $\mu(K) = m$.

Soit une fonction étagée $h\colon K \to \mathbb{R}^+$ majorée par f. Nous notons

$$h(x) = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}(x) \tag{13.551}$$

où les A_i sont des éléments de $\mathcal{A}$. Vu que $0 \leqslant h(x) \leqslant f(x) \leqslant M$, nous avons[53]

$$\int_K h = \sum_{i=1}^{n} \alpha_i \mu(K \cap A_i) \leqslant \sum_{i=1}^{n} M\mu(K \cap A_i) \leqslant M = \mu(K) = Mm \tag{13.552}$$

parce que les A_i sont disjoints et vérifient $\bigcup_i A_i = K$ (lemme 13.102).

Donc tous les éléments de l'ensemble sur lequel nous prenons le supremum dans la définition (13.424) sont contenus dans $[0, Mm]$. Le supremum est donc dans $[0, Mm]$ et est alors strictement plus petit que l'infini.

Si f est positive ou négative Nous appliquons la première partie séparément à f^+ et f^-. Et nous avons alors que f est intégrable et

$$\int_K |f| = \int_K f^+ + \int_K f^- < \infty. \tag{13.553}$$

□

13.10 Propriétés des intégrales

Théorème 13.196 ([292]).
Soient deux espaces mesurables $(S_1, \mathcal{F}_1)$ et $(S_2, \mathcal{F}_2)$ ainsi qu'une application mesurable $\varphi\colon S_1 \to S_2$. Soit encore μ, une mesure positive sur $(S_1, \mathcal{F}_1)$.

Si $f\colon S_2 \to \bar{\mathbb{R}}$ ou $\mathbb{C}$ est mesurable alors,

52. Nous ne prétendons pas que la tribu $\mathcal{A}$ soit liée à la topologie de K.
53. Définition (13.423).

(1) f est $\varphi(\mu)$-intégrale si et seulement si $f\circ\varphi$ est μ-intégrable.

(2) dans le cas où f est $\varphi(\mu)$-intégrable, nous avons

$$\int_{S_2} f d\big(\varphi(\mu)\big) = \int_{S_1} (f\circ\varphi)d\mu. \tag{13.554}$$

Démonstration. L'intégrabilité est la définition 13.168, et demande que $|f|$ soit intégrable. L'égalité (13.554) a un sens si les deux membres sont infinis. Tant que les fonctions considérées sont positives, le point (1) est immédiat. Ce n'est qu'au moment où les fonctions considérées deviennent à valeurs dans $\mathbb{C}$ ou $\mathbb{R}$ que l'intégrabilité de $|f|$ commence à jouer parce qu'il faut que f^+ et f^- soient séparément intégrables.

Nous allons prouver la formule (13.554) pour des fonctions de plus en plis générales. Pour la suite nous notons $\mu' = \varphi(\mu)$.

Pour $f = \mathbb{1}_B$, B mesurable Soit $B \in \mathcal{F}_2$. Nous avons $\mathbb{1}_B\circ\varphi = \mathbb{1}_{\varphi^{-1}(B)}$. Donc en utilisant le lemme 13.157 nous avons

$$\int_{S_2} \mathbb{1}_B d\mu' = \mu'(B) = \mu\big(\varphi^{-1}(B)\big) = \int_{S_1} \mathbb{1}_{\varphi^{-1}(B)} d\mu = \int_{S_1} (\mathbb{1}_B\circ\varphi)d\mu. \tag{13.555}$$

f est étagée positive La fonction f peut être écrite sous la forme

$$f = \sum_{k=1}^{n} a_k \mathbb{1}_{B_k} \tag{13.556}$$

avec $B_k \in \mathcal{F}_2$ et $a_k \in \mathbb{R}^+$. Nous avons alors, en utilisant la sous-additivité de l'intégrale du théorème 13.165(3),

$$\int_{S_2} f d\mu' = \sum_k a_k \int_{S_2} \mathbb{1}_{B_k} d\mu' \tag{13.557a}$$

$$= \sum_k a_k \int_{S_1} (\mathbb{1}_{B_k}\circ\varphi)d\mu \tag{13.557b}$$

$$= \int_{S_1} \Big(\sum_k a_k \mathbb{1}_{B_k}\Big)\circ\varphi d\mu \tag{13.557c}$$

$$= \int_{S_1} (f\circ\varphi)d\mu. \tag{13.557d}$$

f à valeurs dans $\bar{\mathbb{R}}^+$ Vu que f est mesurable, par le théorème 13.105 il existe une suite croissante de fonctions étagées positives convergeant vers f. Soit donc cette suite, $f_n\colon S_2\to\mathbb{R}^+$. Les fonctions $f_n\circ\varphi$ sont étagées et positives et nous avons aussi la limite ponctuelle et croissante $f_n\circ\varphi\to f\circ\varphi$ parce que φ est continue. Le théorème de la convergence monotone (théorème 13.160) permet d'écrire ceci :

$$\int_{S_2} f d\mu' = \lim \int_{S_2} f_n d\mu' = \lim \int_{S_1} (f_n\circ\varphi)d\mu = \int_{S_1} (f\circ\varphi)d\mu. \tag{13.558}$$

Pour $f\colon S_2\to\bar{\mathbb{R}}$ ou $\mathbb{C}$ C'est maintenant que l'intégrabilité va jouer. Nous avons $|f|\circ\varphi = |f\circ\varphi|$, donc

$$\int_{S_2} |f| d\mu' = \int_{S_1} |f|\circ\varphi d\mu = \int_{S_1} |f\circ\varphi| d\mu, \tag{13.559}$$

ce qui montre que f est μ'-intégrable si et seulement si $f\circ\varphi$ est μ-intégrable.

De plus si $f = f^+ - f^-$ alors $f^+\circ\varphi = (f\circ\varphi)^+$, $f^-\circ\varphi = (f\circ\varphi)^-$, et de façon similaire pour les parties imaginaires et réelles.

□

13.11 Mesure à densité

13.11.1 Théorème de Radon-Nikodym

Définition 13.197 ([294]).
Soient μ et ν deux mesures sur l'espace mesurable $(\Omega, \mathcal{A})$. Nous disons que la mesure μ est ***dominée*** *par ν si pour tout ensemble mesurable A, $\nu(A) = 0$ implique $\mu(A) = 0$.*

Si ν est une mesure positive et μ une mesure, nous disons que μ est ***absolument continue*** *par rapport à ν si $\nu(A) = 0$ implique $\mu(A) = 0$. On note aussi $\mu \ll \nu$.*

La mesure μ est **portée** par l'ensemble $E \in \mathcal{A}$ si pour tout $A \in \mathcal{A}$,

$$\mu(A) = \mu(A \cap E). \tag{13.560}$$

Nous écrivons que $\mu \perp \nu$ s'il existe un ensemble $E \in \mathcal{A}$ tel que μ soit porté par E et ν soit porté par $\complement E$.

Théorème 13.198 (Radon-Nikodym[295]).
Soient μ et ν deux mesures σ-finies sur un espace métrisable $(\Omega, \mathcal{A})$.

(1) Il existe un unique couple de mesures μ_1 et μ_2 telles que
   (1a) $\mu = \mu_1 + \mu_2$
   (1b) μ_1 est dominé par ν
   (1c) $\mu_2 \perp \nu$.
   Dans ce cas, les mesures μ_1 et μ_2 sont positives et σ-finies.

(2) À égalité ν-presque partout près, il existe une unique fonction mesurable positive f telle que pour tout mesurable A,

$$\mu_1(A) = \int_A d\mu_1 = \int_\Omega \mathbb{1}_A f d\nu. \tag{13.561}$$

(3) À égalité ν-presque partout près, il existe une unique fonction positive mesurable h telle que $\mu_1 = h\nu$.

Corolaire 13.199.
Si μ es une mesure σ-finie dominée par la mesure σ-finie m, alors μ possède une unique fonction de densité.

Corolaire 13.200.
Soient μ et m, deux mesures positives σ-finies sur $(\Omega, \mathcal{A})$. Alors m domine μ si et seulement si μ possède une densité par rapport à m.

Démonstration. Si μ est dominée par m, alors la décomposition $\mu = \mu + 0$ satisfait le théorème de Radon-Nikodym. Par conséquent il existe une fonction f telle que

$$\mu(A) = \int_A f dm. \tag{13.562}$$

Cette fonction est alors une densité pour μ par rapport à m.

Pour la réciproque, nous supposons que μ a une densité f par rapport à m, et que A est une ensemble de m-mesure nulle :

$$m(A) = \int_\Omega \mathbb{1}_A dm = 0. \tag{13.563}$$

Cela signifie que la fonction $\mathbb{1}_A$ est m-presque partout nulle. La fonction produit $\mathbb{1}_A f$ est également nulle m-presque partout, et par conséquent

$$\mu(A) = \int_\Omega \mathbb{1}_A f dm = 0. \tag{13.564}$$

□

Problèmes et choses à faire

Est-ce que la démonstration de cela ne demande pas la convergence monotone d'une façon ou d'une autre ?

13.11.2 Mesure complexe

Définition 13.201 (Mesure complexe[296])**.**
Si $(\Omega, \mathcal{A})$ est un espace mesurable, une ***mesure complexe*** *est une application $\mu\colon \mathcal{A} \to \mathbb{C}$ telle que*

(1) $\mu(\emptyset) = 0$,

(2) ν *est sous-additive : si les ensembles* $A_i \in \mathcal{A}$, *alors* $\sum_i \mu(A_i) = \mu(\bigcup_i A_i)$.

Notons que la série $\sum_i \mu(A_i)$ est alors nécessairement absolument convergente. En effet changer l'ordre de la somme ne change pas l'union, et donc ne change pas la valeur de la somme. Si $\sigma\colon \mathbb{N} \to \mathbb{N}$ est une permutation,

$$\sum_i \mu(A_{\sigma(i)}) = \mu\big(\bigcup_i A_{\sigma(i)}\big) = \mu\big(\bigcup_i A_i\big) = \sum_i \mu(A_i). \tag{13.565}$$

Le théorème 10.89 dit alors que la somme doit être absolument convergente.

Théorème 13.202 (Radon-NikoDym complexe[54])**.**
Soit μ une mesure positive sur $(\Omega, \mathcal{A})$ et ν une mesure complexe. Alors

(1) Il existe un unique couple de mesures complexes ν_a, ν_s sur $(\Omega, \mathcal{A})$ tel que

(1a) $\nu = \nu_a + \nu_s$

(1b) $\nu_a \ll \mu$

(1c) $\nu_s \perp \mu$.

(2) Ces mesures satisfont alors $\nu_a \perp \nu_s$.

(3) Il existe une fonction intégrable $h\colon \Omega \to \mathbb{C}$ telle que $\nu_a = h\mu$.

(4) La fonction h est unique à μ-équivalence près.

(5) Si de plus $\nu \ll \mu$ alors $\nu = h\mu$.

Démonstration. No proof. □

Remarque 13.203.
Le point (5) est souvent utilisé sous la forme

$$\nu(A) = \int_\Omega \mathbb{1}_A(\omega) h(\omega) d\mu(\omega) = \int_A h(\omega) d\mu(\omega). \tag{13.566}$$

13.11.3 Théorème d'approximation

Lemme 13.204 ([283])**.**
Soit un espace topologique métrique (Ω, d). Nous considérons sa tribu des boréliens[55] $\mathcal{B}or(\Omega)$ *ainsi qu'une mesure finie μ sur $\big(\Omega, \mathcal{B}or(\Omega)\big)$.*

Soit un borélien A de Ω et $\epsilon > 0$.

Il existe un fermé F et un ouvert V de Ω tels que

(1) $F \subset A \subset V$

(2) $\mu(V \backslash F) < \epsilon$.

54. L'histoire du nom de ce théorème est intéressante. Lorsque monsieur et madame Rèmederdonnukodym apprirent que leurs amis, les Rèmedelaboulechevelue avaient appelé leur fils Théo, ils décidèrent d'en faire autant. C'est en souvenir de ces circonstances que monsieur Nikodym (prénommé Radon) décida de faire des math.

55. Définition 13.72.

Démonstration. Soit la famille $\mathcal{D}$ des parties D de Ω qui vérifient la propriété suivante : pour tout $\epsilon > 0$, il existe un fermé F et un ouvert V de Ω tels que $F \subset D \subset V$ et $\mu(V \backslash F) < \epsilon$.

Nous allons prouver que $\mathcal{D}$ est une tribu qui contient tous les ouverts.

$\mathcal{D}$ contient les ouverts Soit un ouvert D. Nous posons

$$F_n = \{x \in \Omega \text{ tel que } d(x, D^c) \geqslant 2^{-n}\}. \tag{13.567}$$

F_n est fermé Le lemme 7.110 montre que le complémentaire F_n^c est ouvert. Donc F_n est fermé.

$D \subset \bigcup_{n\in\mathbb{N}} F_n$ Si $x \in D$, alors il existe $\delta > 0$ tel que $B(x, \delta) \subset D$ (parce que D est ouvert). Donc $d(x, V^c) \geqslant \delta$. Donc $x \in F_n$ pour $2^{-n} < \delta$.

$\bigcup_{n\in\mathbb{N}} F_n \subset D$ Si $x \in F_n$, nous avons $d(x, D^c) > 0$, c'est-à-dire que x n'est pas dans D^c. Autrement dit, $x \in D$.

$\bigcup_{n\in\mathbb{N}} F_n = D$ Nous avons donc l'égalité

$$D = \bigcup_{n\in\mathbb{N}} F_n. \tag{13.568}$$

Vu que $F_n \subset F_{n+1}$, le lemme 13.23(1) nous indique que

$$\lim_{n\to\infty} \mu(F_n) = \mu\Big(\bigcup_{k\in\mathbb{N}} F_k\Big) = \mu(D). \tag{13.569}$$

Étant donné que la mesure est finie, nous pouvons écrire cela sous la forme

$$\mu(D) - \mu(F_n) \to 0. \tag{13.570}$$

Pour chaque n nous avons l'encadrement

$$F_n \subset D \subset D \tag{13.571}$$

où F_n et D sont ouverts. Lorsque ϵ est donné, il suffit de prendre n assez grand pour avoir $\mu(D \backslash F_n) < \epsilon$ pour avoir un encadrement de D par un fermé et un ouvert (D lui-même) dont la différence des mesures est plus petite que e.

Tout cela pour dire que $D \in \mathcal{D}$.

$\mathcal{D}$ est une tribu Il faut vérifier les trois points de la définition 13.1.

$\Omega \in \mathcal{D}$ Nous venons de voir que les ouverts sont dans $\mathcal{D}$. Or Ω est un ouvert.

$D \in \mathcal{D}$ implique $D^c \in \mathcal{D}$ Soit F fermé et V ouvert tels que $F \subset D \subset V$. Nous avons aussi

$$V^c \subset D^c \subset F^c \tag{13.572}$$

où V^c est fermé et F^c est ouvert. De plus $F^c \backslash = V \backslash F$ et donc

$$\mu(F^c \backslash V^c) = \mu(V \backslash F). \tag{13.573}$$

Nous pouvons donc choisir F et V pour avoir $\mu(F^c \backslash V^c) < \epsilon$.

$\bigcup_{i\in\mathbb{N}} D_i \in \mathcal{D}$ Soient $D_i \in t\mathcal{D}$. Pour chaque n nous posons

$$F_n \subset D_n \subset V_n \tag{13.574}$$

en choisissant V_n et F_n de telle sorte que $\mu(V_n \backslash F_n) < 2^{-n}\epsilon$.

Nous posons

$$Y_N = \bigcup_{n=0}^{N} F_n, \tag{13.575}$$

et

$$Y = \bigcup_{n=0}^{\infty} F_n. \tag{13.576}$$

Chacun des Y_N est fermé en tant qu'union finie de fermés (lemme 7.3(2)). Mais Y ne l'est pas spécialement [56]. Le lemme 13.23 nous dit cependant que $\mu(Y) = \lim_{N\to\infty} \mu(Y_N)$.

Nous posons

$$D = \bigcup_{n=0}^{\infty} D_n \tag{13.577}$$

ainsi que

$$V = \bigcup_{n\in\mathbb{N}} V_n. \tag{13.578}$$

La partie V est ouverte dans Ω comme union d'ouverts (c'est dans le définition d'une topologie). Nous avons, pour tout N, l'encadrement

$$Y_N = \bigcup_{n=0}^{N} F_n \subset Y \subset D \subset V. \tag{13.579}$$

Nous prouvons à présent que $\lim_{N\to\infty} \mu(V\backslash Y_N) = 0$, de telle sorte que l'encadrement (13.579) dise que $D \in \mathcal{D}$.

D'abord nous avons

$$V\backslash Y \subset \bigcup_n (V_n\backslash F_n) \tag{13.580}$$

parce que si $x \in V\backslash Y$, alors $x \in V_i$ pour un certain i, mais vu que x n'est pas dans Y, il n'est dans aucun des F_n donc en particulier pas dans F_i et $x \in V_n\backslash F_i$.

Un peu de calcul :

$$\begin{aligned}
\mu(V) - \mu(Y) &= \mu(V\backslash Y) && \text{(13.581a)}\\
&\leqslant \mu\big(\bigcup_n (V_n\backslash F_n)\big) && \text{(13.581b)}\\
&\leqslant \sum_{n=0}^{\infty} \mu(V_n\backslash F_n) && \text{(13.581c)}\\
&= \sum_{n=0}^{\infty} 2^{-n}\epsilon && \text{(13.581d)}\\
&= 2\epsilon. && \text{(13.581e)}
\end{aligned}$$

Justifications :

— Pour (13.581a), c'est le lemme 13.22.

— Pour (13.581b), c'est (13.580).

— Pour (13.581c), c'est le lemme 13.22(3).

— Pour (13.581e), c'est la série géométrique (10.127).

Nous choisissons maintenant N assez grand pour que $\mu(Y) - \mu(Y_N) < \epsilon$. Nous avons alors l'encadrement

$$Y_N \subset Y \subset D \subset V \tag{13.582}$$

avec

$$\mu(V\backslash Y_N) = \mu(V) - \mu(Y_N) = \underbrace{\mu(V) - \mu(y)}_{\leqslant 2\epsilon} + \mu(Y) - \mu(Y_N) \leqslant 2\epsilon + \epsilon = 3\epsilon. \tag{13.583}$$

56. Par exemple $A_n = [1/n, 2]$ sont des fermés dont l'union est $]0, 2]$ qui n'est pas fermé.

Nous avons donc montré que $\mathcal{D}$ était une tribu contenant les ouverts. Donc $\mathcal{D}$ contient tous les boréliens. □

Lemme 13.205 ([283]).
Soit un espace topologique métrique (Ω, d). Nous considérons sa tribu des boréliens $\mathcal{B}or(\Omega)$ ainsi qu'une mesure μ sur $\big(\Omega, \mathcal{B}or(\Omega)\big)$.

Soient un ouvert $W \subset \Omega$ tel que $\mu(W) < \infty$ et un borélien A tel que $A \subset W$. Soit aussi $\epsilon > 0$.

Il existe un fermé F et un ouvert V tels que

(1) $\mu(V) < \infty$,
(2) $\mu(V \backslash F) < \epsilon$,
(3) et $F \subset A \subset V$.

Démonstration. Vu que la mesure de W est finie, nous considérons la mesure finie

$$\begin{aligned} \nu \colon \mathcal{B}\mathrm{or}(\Omega) &\to [0, \mu(W)] \\ B &\mapsto \mu(B \cap W). \end{aligned} \tag{13.584}$$

La partie A étant borélienne ; par le lemme 13.204, nous avons un fermé F et un ouvert V_1 ouvert tels que

$$F \subset A \subset V_1 \tag{13.585}$$

et $\nu(V_1 \backslash F) < \epsilon$. Nous posons $V = V_1 \cap W$; vu que $A \subset W$ et $A \subset V_1$ nous avons aussi $A \subset V_1 \cap W$ et donc l'encadrement

$$F \subset A \subset V \subset W. \tag{13.586}$$

En ce qui concerne la mesure :

$$\mu(V \backslash F) = \mu(V) - \mu(F) = \mu(V \cap W) - \mu(F \cap W) = \nu(B) - \nu(F) < \epsilon. \tag{13.587}$$

□

Théorème 13.206 (Théorème d'approximation, thème 3[283]).
Soit un espace topologique métrique (Ω, d). Nous considérons sa tribu des boréliens $\mathcal{B}or(\Omega)$ ainsi qu'une mesure μ sur $\big(\Omega, \mathcal{B}or(\Omega)\big)$.

Soient un ouvert $W \subset \Omega$ tel que $\mu(W) < \infty$ et un un borélien A tel que $A \subset W$. Soit aussi $\epsilon > 0$.

Il existe un fermé $F \subset W$ et une fonction $f \in C^0(\Omega, \mathbb{R})$ vérifiant

(1) $F \subset A \subset W$,
(2) $f|_F = 1$,
(3) $f|_{W^c} = 0$
(4) $\|f - \mathbb{1}_A\|_{L^1} < \epsilon$

Démonstration. Par le lemme 13.205, il existe un fermé F et un ouvert V tels que

$$F \subset A \subset V \subset W \tag{13.588}$$

et $\mu(V \backslash F) < \epsilon$. Nous posons alors

$$f(x) = \frac{d(x, V^c)}{d(x, V^c) + d(x, F)}. \tag{13.589}$$

Le dénominateur de cette expression ne s'annule jamais parce que si $d(x, V^c) = 0$, c'est que $x \in V^c$. Mais alors x n'est pas dans V et donc pas dans F non plus. La partie F étant fermée, $d(x, F) > 0$ par lemme 7.111. De plus la fonction f est continue par le lemme 7.112.

Pour (2) Si $x \in F$, alors $d(x, F) = 0$, et f devient

$$f(x) = \frac{d(x, V^c)}{d(x, V^c)} = 1 \tag{13.590}$$

Pour (3) Si $x \in W^c$, alors $x \in V^c$ et $d(x, V^c) = 0$ si bien que $f(x) = 0$.

Pour (4) Les premiers points montrent que

$$\mathbb{1}_F \leqslant f \leqslant \mathbb{1}_V. \tag{13.591}$$

Mais nous avons aussi, par ailleurs,

$$\mathbb{1}_F \leqslant \mathbb{1}_A leq \mathbb{1}_V. \tag{13.592}$$

Ces deux encadrement, par le lemme 1.149 donnent l'encadrement

$$|f - \mathbb{1}_A| \leqslant \mathbb{1}_V - \mathbb{1}_F. \tag{13.593}$$

En ce qui concernent les intégrales nous avons alors

$$\int_\Omega |\mathbb{1}_A - f| \leqslant \int_\Omega (\mathbb{1}_V - \mathbb{1}_F) d\mu \tag{13.594a}$$
$$= \mu(V) - \mu(F) \tag{13.594b}$$
$$< \epsilon. \tag{13.594c}$$

Pour (13.594b), c'est le lemme 13.157.

□

13.12 Produit de mesures

Lemme 13.207 (Propriété des sections[280])**.**
Soient $\mathcal{A}_1$ et $\mathcal{A}_2$ des tribus sur les ensembles Ω_1 et Ω_2. Si $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ alors pour tout $x \in \Omega_1$ et $y \in \Omega_2$, les ensembles

$$A_1(y) = \{x \in \Omega_1 \text{ tel que } (x, y) \in A\} \tag{13.595a}$$
$$A_2(x) = \{y \in \Omega_2 \text{ tel que } (x, y) \in A\} \tag{13.595b}$$

sont mesurables.

Démonstration. Soit $y \in \Omega_2$; nous allons prouver le résultat pour $A_1(y)$. Pour cela nous notons

$$S = \{A \in \mathcal{A}_1 \otimes \mathcal{A}_2 \text{ tel que } \forall y \in \Omega_2, A_1(y) \in \mathcal{A}_1\}, \tag{13.596}$$

et nous allons noter que S est une tribu contenant les rectangles. Par conséquent, S sera égal à $\mathcal{A}_1 \otimes \mathcal{A}_2$.

Les rectangles Considérons le rectangle $A = X \times Y$ et si $y \in \Omega_2$ alors

$$A_1(y) = \{x \in \Omega_1 \text{ tel que } (x, y) \in X \times Y\}. \tag{13.597}$$

Donc soit $y \in Y$ alors $A_1(y) = X \in \mathcal{A}_1$, soit $y \notin Y$ et alors $A_1(y) = \varnothing \in \mathcal{A}_1$.

Tribu : ensemble complet Nous avons $\Omega_1 \times \Omega_2 \in S$ parce que c'est un rectangle.

Tribu : complémentaire Soit $A \in S$. Montrons que $A^c \in S$. Nous avons d'abord

$$(A^c)_1(y) = \{x \in \Omega_1 \text{ tel que } (x, y) \in A^c\}. \tag{13.598}$$

D'autre part

$$A_1(y)^c = \{x \in \Omega_1 \text{ tel que } (x, y) \notin A\} = \{x \in \Omega_1 \text{ tel que } (x, y) \in A^c\} = (A^c)_1(y). \tag{13.599}$$

Vu que $\mathcal{A}_1$ est une tribu et que par hypothèse $A_1(y) \in \mathcal{A}_1$, nous avons aussi $A_1(y)^c \in S$, et donc $(A^c)_1(y) \in \mathcal{A}_1$, ce qui prouve que $A^c \in S$.

Tribu : union dénombrable Soit une suite $A_n \in S$. Nous avons

$$(\bigcup_n A_n)_1(y) = \{x \in \Omega_1 \text{ tel que } (x,y) \in \bigcup_n A_n\} \tag{13.600a}$$
$$= \bigcup_n \{x \in \Omega_1 \text{ tel que } (x,y) \in A_n\} \tag{13.600b}$$
$$= \bigcup_n (A_n)_1(y), \tag{13.600c}$$

et ce dernier ensemble est dans $\mathcal{A}_1$ parce que c'est une union dénombrable d'éléments de $\mathcal{A}_1$.

Nous avons donc prouvé que S est une tribu contenant les rectangles, donc S contient au moins $\mathcal{A}_1 \otimes \mathcal{A}_2$. □

Corolaire 13.208.
Si $f\colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ est une fonction mesurable[57] *sur $X \times Y$ alors pour chaque y dans Ω_2, la fonction*

$$\begin{aligned} f_y\colon X &\to \mathbb{R} \\ x &\mapsto f(x,y) \end{aligned} \tag{13.601}$$

est mesurable.

Démonstration. Soit $\mathcal{O}$ un ensemble mesurable de $\mathbb{R}$ (i.e. un borélien), et $y \in \Omega_2$. Nous avons

$$f_y^{-1}(\mathcal{O}) = \{x \in X \text{ tel que } f(x,y) \in \mathcal{O}\} = A_1(y) \tag{13.602}$$

où

$$A = \{(x,y) \in \Omega_1 \times \Omega_2 \text{ tel que } f(x,y) \in \mathcal{O}\} = f^{-1}(\mathcal{O}). \tag{13.603}$$

Ce dernier est mesurable parce que f l'est. □

Théorème 13.209 ([280][58])**.**
Soient $(\Omega_i, \mathcal{A}_i, \mu_i)$ $(i = 1,2)$ deux espaces mesurés σ-finie. Soit $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$. Alors les fonctions[59]

$$x \mapsto \mu_2\big(A_2(x)\big) \tag{13.604a}$$
$$y \mapsto \mu_1\big(A_1(y)\big) \tag{13.604b}$$

sont mesurables et

$$\int_{\Omega_1} \mu_2\big(A_2(x)\big) d\mu_1(x) = \int_{\Omega_2} \mu_2\big(A_1(y)\big) d\mu_2(y). \tag{13.605}$$

Démonstration. Nous supposons d'abord que μ_1 et μ_2 sont finies et nous notons $\mathcal{D}$ le sous-ensemble de $\mathcal{A}_1 \otimes \mathcal{A}_2$ sur lequel le théorème est correct. Nous allons commencer par prouver que $\mathcal{D}$ est un λ-système.

λ-système : différence ensembliste Soient $A, B \in \mathcal{D}$ avec $A \subset B$. Nous avons

$$(B \backslash A)_1(y) = \{x \in \Omega_1 \text{ tel que } (x,y) \in B \backslash A\} \tag{13.606a}$$
$$= \{x \in \Omega_1 \text{ tel que } (x,y) \in B\} \backslash \{x \in \Omega_1 \text{ tel que } (x,y) \in A\} \tag{13.606b}$$
$$= B_1(y) \backslash A_1(y). \tag{13.606c}$$

Vu que $A_1(y) \subset B_1(y)$ et que les mesures sont finies le lemme 13.22 nous donne

$$\mu_1\big((B \backslash A)_1(y)\big) = \mu_1\big(B_1(y)\big) - \mu_1\big(A_1(y)\big), \tag{13.607}$$

57. Définition 13.41.
58. Modèle non contractuel : des notations et la définition de λ-système peuvent varier entre la référence et le présent texte.
59. Voir la notation du lemme 13.595.

et similairement pour $1 \leftrightarrow 2$. Les deux fonctions (de y) à droite étant mesurables, nous avons la mesurabilité de la fonction $y \mapsto \mu_1\big((B\backslash A)_1(y)\big)$.

Prouvons la formule intégrale en nous rappelant que la formule (13.605) est supposée correcte pour A et B séparément :

$$\int_{\Omega_2} \mu_1\big((B\backslash A)_1(y)\big)d\mu_2(y) = \int_{\Omega_2} \mu_1\big(B_1(y)\big)d\mu_2(y) - \int_{\Omega_2} \mu_1\big(A_1(y)\big)d\mu_2(y) \tag{13.608a}$$

$$= \int_{\Omega_1} \mu_2\big(B_2(x)\big)d\mu_1(x) - \int_{\Omega_1} \mu_2\big(A_2(x)\big)d\mu_1(x) \tag{13.608b}$$

$$= \int_{\Omega_1} \mu_2\big((B\backslash A)_2(x)\big)d\mu_1(x). \tag{13.608c}$$

λ-système : limite de suite croissante Soit (A_n) une suite croissante dans $\mathcal{D}$; nous posons $B_n = A_n\backslash A_{n-1}$ et $A_0 = \varnothing$ de telle sorte à travailler avec une suite d'ensembles disjoints qui satisfait $\bigcup_n A_n = \bigcup_n B_n$. Vu que la suite est croissante nous avons $A_{n-1} \subset A_n$ et donc $B_n \in \mathcal{D}$ par le point déjà fait sur la différence ensembliste. Nous avons :

$$\mu_1\big((\bigcup_n B_n)_1(y)\big) = \{x \in \Omega_1 \text{ tel que } (x,y) \in \bigcup_n B_n\} \tag{13.609a}$$

$$= \bigcup_n \{x \in \Omega_1 \text{ tel que } (x,y) \in B_n\} \tag{13.609b}$$

$$= \bigcup_n (B_n)_1(y). \tag{13.609c}$$

Par conséquent, par la propriété (2) d'une mesure nous avons

$$\mu_1\big((\bigcup_n B_n)_1(y)\big) = \sum_n \mu_1\big((B_n)_1(y)\big). \tag{13.610}$$

En tant que somme de fonctions positives et mesurables, la fonction

$$y \mapsto \sum_n \mu_1\big((B_n)_1(y)\big) \tag{13.611}$$

est mesurable par la proposition 13.90. Il faut encore vérifier la formule intégrale. Le gros du boulot est de permuter une somme et une intégrale par le corolaire 13.163 :

$$\int_{\Omega_2} \sum_n \mu_1\big((B_n)_1(y)\big)d\mu_2(y) = \sum_n \int_{\Omega_2} \mu_1\big((B_n)_1(y)\big)d\mu_2(y) \tag{13.612a}$$

$$= \sum_n \int_{\Omega_1} \mu_2\big((B_n)_2(x)\big)d\mu_1(x) \tag{13.612b}$$

$$= \int_{\Omega_1} \sum_n \mu_2\big((B_n)_2(x)\big)d\mu_1(x) \tag{13.612c}$$

$$= \int_{\Omega_1} \mu_2\big((\bigcup_n B_n)_1(y)\big)d\mu_1(x). \tag{13.612d}$$

Maintenant que $\mathcal{D}$ est un λ-système contenant les rectangles, le lemme 13.32 dit que la tribu engendrée par $\mathcal{D}$ (c'est-à-dire $\mathcal{A}_1 \otimes \mathcal{A}_2$) est le λ-système $\mathcal{D}$ lui-même.

La preuve est finie dans le cas de mesures finies. Nous commençons maintenant à prouver dans le cas où les mesures μ_1 et μ_2 sont seulement σ-finies. Nous considérons des suites croissantes $\Omega_{i,n} \to \Omega_i$ d'ensembles mesurables et de mesure finie : $\mu_i(\Omega_{i,n}) < \infty$. D'abord remarquons que

$$\mu_2\Big((A \cap \Omega_{1,j} \times E_{2,j})_2(x)\Big) = \mu_2\Big(A_2(x) \cap \Omega_{2,j}\Big)\mathbb{1}_{\Omega_{1,j}}. \tag{13.613}$$

En effet,

$$\heartsuit = (A \cap \Omega_{1,j} \times E_{2,j})_2(x) \tag{13.614a}$$

$$= \{y \in \Omega_2 \text{ tel que } (x,y) \in A \cap \Omega_{1,j} \times E_{2,j}\} \tag{13.614b}$$

$$= \{y \in \Omega_2 \text{ tel que } (x,y) \in A \times \Omega_{2,j}\} \cap \{y \in \Omega_2 \text{ tel que } (x,y) \in \Omega_{1,j} \times \Omega_{2,j}\}. \tag{13.614c}$$

Si $y \in \Omega_{1,j}$ alors $\{y \in \Omega_2 \text{ tel que } (x,y) \in \Omega_{1,j} \times \Omega_{2,j}\} = \Omega_{2,j}$ et dans ce cas

$$\heartsuit = \{y \in \Omega_2 \text{ tel que } (x,y) \in A \times \Omega_{2,j}\} \cap \Omega_{2,j} = A_2(x) \cap E_{2,j}. \tag{13.615}$$

Et inversement, si $x \notin \Omega_{1,j}$ alors $\heartsuit = \emptyset$. Dans les deux cas nous avons (13.613).

Les ensembles $A \cap \Omega_{1,j} \times \Omega_{2,j}$ étant de mesure finie, nous pouvons leur appliquer la première partie :

$$\int_{\Omega_1} \mu_2\Big((A \cap \Omega_{1,j} \times \Omega_{2,j})_2(x)\Big) d\mu_1(x) = \int_{\Omega_2} \mu_1\Big((A \cap \Omega_{1,j} \times \Omega_{2,j})_1(y)\Big) d\mu_2(u), \tag{13.616}$$

ou encore

$$\int_{\Omega_1} \mu_2\Big(A_2(x) \cap \Omega_{2,j}\Big) \mathbb{1}_{\Omega_{1,j}}(x) d\mu_1(x) = \int_{\Omega_2} \mu_1\Big(A_1(y) \cap \Omega_{1,j}\Big) \mathbb{1}_{\Omega_{2,j}}(y) d\mu_2(y). \tag{13.617}$$

Ce que nous avons dans ces intégrales sont (par rapport à j) des suites croissantes de fonction positives ; nous pouvons donc permuter une limite et une intégrale. En sachant que si $k \to \infty$, alors

$$\mathbb{1}_{1,j}(x) \to 1 \tag{13.618a}$$

$$\mu_2\big(A_2(x) \cap \Omega_2, j\big) \to \mu_2\big(A_2(x)\big), \tag{13.618b}$$

nous trouvons le résultat demandé. □

Théorème-définition 13.210 ([297, 298]).
Soient μ_i des mesures σ-finies sur $(\Omega_i, \mathcal{A}_i)$ $(i = 1, 2)$.

(1) Il existe une et une seule mesure, notée $\mu_1 \otimes \mu_2$, sur $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ telle que

$$(\mu_1 \otimes \mu_2)(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \tag{13.619}$$

pour tout $A_1 \in \mathcal{A}_1$ et $A_2 \in \mathcal{A}_2$.

(2) Cette mesure est donnée par la formule[60]

$$(\mu_1 \otimes \mu_2)(A) = \int_{\Omega_1} \mu_2\big(A_2(x)\big) d\mu_1(x) = \int_{\Omega_2} \mu_1\big(A_1(y)\big) d\mu_2(y). \tag{13.620}$$

*Cette mesure est la **mesure produit** de μ_1 par μ_2.*

(3) La mesure $\mu_1 \otimes \mu_2$ ainsi définie est σ-finie.

Démonstration. La partie « existence » sera divisée en deux parties : l'une pour prouver que les formules (13.620) donnent une mesure et une pour montrer que cette mesure vérifie la condition (13.619).

Unicité L'ensemble des rectangles de $\Omega_1 \times \Omega_2$ engendre la tribu $\mathcal{A}_1 \otimes \mathcal{A}_2$, est fermé par intersection et contient une suite croissante d'ensembles $P_n \times R_n$ de mesure finie ($\mu(P_n \times R_n) < \infty$) telle que $P_n \times R_n \to \Omega_1 \times \Omega_2$. Cette suite est donné par le fait que μ_1 et μ_2 sont σ-finies. En effet si (X_n) et (Y_n) sont des recouvrements dénombrables de Ω_1 et Ω_2 par des ensembles de mesure finie, en posant $P_n = \bigcup_{k=1}^n X_n$ et $R_n = \bigcup_{k=1}^n Y_n$ nous avons bien une suite croissante de rectangles qui tendent vers $\Omega_1 \times \Omega_2$. Avec ces rectangles en main, le théorème 13.33 donne l'unicité.

Les formules définissent une mesure Le théorème 13.209 dit que ces formules ont un sens et que l'égalité entre les deux intégrales est correcte. Nous prouvons à présent qu'elles déterminent effectivement une mesure sur $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$.

Pour tout $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, $\mu(A) \geqslant 0$ parce que μ est donnée par l'intégrale d'une fonction positive.

60. Voir les notations du lemme 13.207.

En ce qui concerne la condition d'unions dénombrable disjointe, soient $A^{(i)}$ des éléments disjoints de $\mathcal{A}_1 \otimes \mathcal{A}_2$; nous commençons par remarquer que

$$\left(\bigcup_{i=1}^{\infty} A^{(i)}\right)_2 (x) = \{y \in \Omega_2 \text{ tel que } (x,y) \in \bigcup_{i=1}^{\infty} A^{(i)}\} \tag{13.621a}$$

$$= \bigcup_{i=1}^{\infty} \{y \in \Omega_2 \text{ tel que } (x,y) \in A^{(i)}\} \tag{13.621b}$$

$$= \bigcup_{i=1}^{\infty} A_2^{(i)}(x). \tag{13.621c}$$

Par conséquent,

$$\mu\left(\bigcup_{i=1}^{\infty} A^{(i)}\right) = \int_{\Omega_1} \mu_2\left(\left(\bigcup_{i=1}^{\infty} A^{(i)}\right)_2 (x)\right) d\mu_1(x) \tag{13.622a}$$

$$= \int_{\Omega_1} \sum_{i=1}^{\infty} \mu_2\big(A_2^{(i)}(x)\big) d\mu_1(x) \tag{13.622b}$$

$$= \int_{\Omega_1} \lim_{n\to\infty} \sum_{i=1}^{n} \mu_2\big(A_2^{(i)}(x)\big) d\mu_1(x). \tag{13.622c}$$

où nous avons utilisé l'additivité de la mesure μ_2. À ce niveau, il serait commode de permuter la somme et l'intégrale. Pour ce faire nous considérons la suite (croissante) de fonctions

$$f_n(x) = \sum_{i=1}^{n} \mu_2\big(A_2^{(i)}(x)\big). \tag{13.623}$$

Nous pouvons permuter la limite et l'intégrale grâce au théorème de la convergence monotone 13.160 ; ensuite la somme se permute avec l'intégrale en tant que somme finie :

$$\mu\left(\bigcup_{i=1}^{\infty} A^{(i)}\right) = \lim_{n\to\infty} \sum_{i=1}^{n} \int_{\Omega_1} \big(A_2^{(i)}(x)\big) d\mu_1(x) \tag{13.624a}$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} \mu(A^{(i)}) \tag{13.624b}$$

$$= \sum_{i=1}^{\infty} \mu(A^{(i)}). \tag{13.624c}$$

Elles vérifient la condition Prouvons que les formules (13.620) se réduisent à (13.619) dans le cas des rectangles. Soit donc $A = X_1 \times X_2$ avec $X_i \in \mathcal{A}_i$. Alors

$$A_1(y) = \{x \in \Omega_1 \text{ tel que } (x,y) \in X_1 \times X_2\} \tag{13.625}$$

et

$$\mu_1\big(A_1(y)\big) = \mathbb{1}_{X_2}(y)\mu_1(X_1), \tag{13.626}$$

donc

$$(\mu_1 \otimes \mu_2)(A) = \int_{\Omega_2} \mu_1\big(A_1(y)\big) d\mu_2(y) \tag{13.627a}$$

$$= \int_{\Omega_2} \mu_1(X_1)\mathbb{1}_{X_2}(y) d\mu_2(y) \tag{13.627b}$$

$$= \mu_1(X_1) \int_{\Omega_2} \mathbb{1}_{X_2}(y) d\mu_2(y) \tag{13.627c}$$

$$= \mu_1(X_1)\mu_2(X_2). \tag{13.627d}$$

Pour cela nous avons utilisé le fait que l'intégrale de la fonction caractéristique d'un ensemble mesurable est la mesure de cet ensemble.

□

Définition 13.211 (Produit d'espaces mesurés)**.**
*Si $(\Omega_i, \mathcal{A}_i, \mu_i)$ sont deux espaces mesurés, l'**espace produit** est l'ensemble $\Omega_1 \times \Omega_2$ muni de la tribu produit $\mathcal{A}_1 \otimes \mathcal{A}_2$ de la définition 13.112 et de la mesure produit $\mu_1 \otimes \mu_2$ définie par le théorème 13.210.*

Remarque 13.212.
Il n'est pas garantit que la tribu $\mathcal{A}_1 \otimes \mathcal{A}_2$ soit la tribu la plus adaptée à l'ensemble $S_1 \times S_2$. Dans le cas de $\mathbb{R}^N$, il se fait que c'est le cas : en prenant des produits des boréliens sur $\mathbb{R}$ on obtient bien les boréliens sur $\mathbb{R}^N$, voir proposition 13.116.

13.13 Tribu et mesure de Lebesgue sur $\mathbb{R}^d$

Définition 13.213 (Mesure de Lebesgue)**.**
En plusieurs étapes.

(1) D'abord nous avons la mesure λ_N sur $\mathbb{R}^n$ définie sur

$$\big(\mathbb{R}^d, \mathcal{B}or(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}or(\mathbb{R})\big) \tag{13.628}$$

comme le produit $\lambda \otimes \ldots \otimes \lambda$ via la définition 13.211.

(2) Ensuite nous nous souvenons du corolaire 13.116 qui donne λ_N comme une mesure sur

$$\big(\mathbb{R}^N, \mathcal{B}or(\mathbb{R}^N)\big). \tag{13.629}$$

(3) Et enfin nous considérons la completion de la mesure λ_N (théorème 13.57), que nous notons encore λ_N.

Proposition 13.214 ([284])**.**
Tout ouvert de $\mathbb{R}^n$ est une union dénombrable et disjointe de cubes semi-ouverts.

Démonstration. Nous allons même montrer que ces cubes peuvent être choisis sur un quadrillage.

Soit G un ouvert de $\mathbb{R}^n$. Soit $\{Q_i^1\}_{i\in\mathbb{N}}$ un découpage de $\mathbb{R}^n$ en cubes semi-ouverts de côté 1 et dont les sommets sont en les coordonnées entières. Ils sont de la forme

$$\prod_{i=1}^{n} [n_i, n_i + 1[\tag{13.630}$$

où les n_i sont des entiers. Ce sont des cubes disjoints. Nous considérons ensuite pour chaque $k > 1$ le découpage $\{Q_i^{(k)}\}_{i\in\mathbb{N}}$ de $\mathbb{R}^n$ en cubes de côtés 2^{-k} qui consiste à découper en 2 les côtés des cubes du découpage $Q^{(k-1)}$. Ces cubes forment encore un découpage dénombrable de $\mathbb{R}^n$ en des cubes disjoints. Ils sont de la forme

$$\prod_{i=1}^{n} [\frac{n_i}{2^k}, \frac{n_i + 1}{2^k}[\tag{13.631}$$

où les n_i sont encore entiers. Ensuite nous considérons $\mathcal{E}$ l'union de tous les $Q_i^{(k)}$ contenus dans G.

Montrons que $\mathcal{E} = G$. D'abord $\mathcal{E} \subset G$ parce que $\mathcal{E}$ est une union d'ensembles contenus dans G. Ensuite si $x \in G$, il existe une boule de rayon r autour de x contenue dans G ; alors un des ensembles $Q_i^{(k)}$ avec $2^{-j} < \frac{r}{2}$ est contenue dans $B(x, r)$ et donc dans $\mathcal{E}$.

Bien entendu l'union qui donne $\mathcal{E}$ n'est pas satisfaisante par ce que les $Q_i^{(k+1)}$ sont contenus dans les $Q_i^{(k)}$; les intersections sont donc loin d'être vides.

Nous faisons ceci :

$$R^{(0)} = \{Q_i^{(1)} \text{contenu dans } G\} \tag{13.632a}$$

$$R^{(k+1)} = \{Q_i^{(k+1)} \text{contenus dans } G \text{ et pas dans } R^{(k)}\}. \tag{13.632b}$$

En fin de compte l'union de tous les ensembles contenus dans les $R^{(k)}$ forment encore $\mathbb{R}^n$, mais sont d'intersection vide. □

Les cubes dont il est question dans cette preuve, de côtés 2^{-k} sont souvent appelés des cubes **dyadiques**.

Corolaire 13.215 ([284]).
Tout ouvert de $\mathbb{R}^n$ est une union dénombrable de cubes presque disjoints[61].

Démonstration. Il suffit de prendre les cubes de la proposition 13.214 et de les fermer. Ce que l'on ajoute est de mesure nulle. □

Remarque 13.216.
La proposition 13.214 est une propriété seulement de la topologie de $\mathbb{R}^n$ alors que le corolaire fait intervenir la mesure de Lebesgue parce qu'il faut bien dire que les intersections sont de mesure (de Lebesgue) nulle.

13.13.1 Ensembles négligeables

Lemme 13.217 ([299]).
L'image d'une partie négligeables de $\mathbb{R}^N$ par une application Lipschitz est négligeable.

Démonstration. Soit N une partie négligeable de $\mathbb{R}^N$ et une application Lipschitz $f\colon N \to \mathbb{R}^N$. Soit $Q \subset \mathbb{R}^N$ un cube borné de côté r. Pour tout $x, x' \in N \cap Q$ nous avons

$$\|f(x) - f(x')\| \leqslant C\|x - x'\| \leqslant Cr. \tag{13.633}$$

Donc $f(N \cap Q)$ est dans une boule de rayon Cr. Mais comme toutes les normes sont équivalentes[62] sur $\mathbb{R}^N$ nous pouvons tout aussi bien prendre la norme $\|.\|_1$ au lieu de la norme $\|.\|_2$ (qui est toujours la norme prise implicitement lorsqu'on parle de $\mathbb{R}^n$), de telle sorte que les boules soient des cubes. Quoi qu'il en soit, $f(N \cap Q)$ est contenu dans un cube de côté $2Cr$ et au niveau de la mesure extérieure,

$$m^*\big(f(N \cap Q)\big) \leqslant (2Cr)^N = (2C)^N r^N, \tag{13.634}$$

ou encore

$$m\big(f(N \cap Q)\big) \leqslant (2C)^N m(Q) \tag{13.635}$$

parce que r^N est la mesure du cube Q.

Soit maintenant $\epsilon > 0$; vu que N est négligeable, il existe un ouvert U contenant N et tel que $m(U) < \epsilon$. Ce U est une union presque disjointe de cubes dyadiques (Q_n) par le corolaire 13.215. Nous avons alors

$$\begin{aligned}
m^*\big(f(N)\big) &= m^*\big(f(\bigcup_n N \cap Q_n)\big) && \text{(13.636a)}\\
&= m^*\big(\bigcup_n f(N \cap Q_n)\big) && \text{(13.636b)}\\
&\leqslant \sum_n m^*(f(N \cap Q_n)) && \text{(13.636c)}\\
&\leqslant \sum_n (2C)^N m(Q_n) && \text{(13.636d)}\\
&= (2C)^N m(U) && \text{(13.636e)}\\
&< (2C)^d \epsilon. && \text{(13.636f)}
\end{aligned}$$

Au final, $m^*\big(f(N)\big) \leqslant (2C)^N \epsilon$. L'ensemble N est donc négligeable parce que le lemme 13.63 le dit : $m^*(N) = 0$. □

Corolaire 13.218.
Un sous-espace vectoriel strict de $\mathbb{R}^N$ est négligeable.

61. « presque » au sens où les intersections éventuelles sont de mesure de Lebesgue nulle.
62. Proposition 7.223

Démonstration. Un sous-espace vectoriel strict de $\mathbb{R}^N$ de dimension $k < N$ est l'image de

$$A = \{t_1 e_1 + \cdots + t_k e_k \text{ tel que } t_i \in \mathbb{R}\} \tag{13.637}$$

par une application linéaire. Ce A est un pavé de mesure de Lebesgue nulle. Donc l'image est négligeable par le lemme 13.217. □

13.13.2 Parties et fonctions mesurables

Pour rappel, la notion d'application de classe C^1 est donnée par la définition 10.128.

Proposition 13.219.
Soient U et V des ouverts de $\mathbb{R}^N$ et $\phi\colon U \to V$ un C^1-difféomorphisme. Si $E \subset U$ est mesurable, alors $\phi(E)$ est mesurable[63].

Démonstration. Si E est mesurable, il existe un borélien B et un ensemble négligeable N tels que $E = B \cup N$. Vu que ϕ est un homéomorphisme, l'application ϕ^{-1} est borélienne parce que continue (théorème 13.49). Nous avons

$$\phi(B) = (\phi^{-1})^{-1}(B), \tag{13.638}$$

c'est-à-dire que $\phi(B)$ est l'image inverse de B par ϕ^{-1}. L'ensemble $\phi(B)$ est donc borélien.

Il reste à voir que $\phi(N)$ est négligeable. Soit $Q \subset U$ une cube compact. L'application $d\phi\colon Q \to \mathcal{L}(\mathbb{R}^N, \mathbb{R}^N)$ est continue et donc bornée (par la remarque 10.129) sur le compact Q. Par les accroissements finis (théorème 10.145), l'application ϕ est donc Lipschitz sur Q. La partie $\phi(N \cap Q)$ est alors négligeable par le lemme 13.217. Pour conclure,

$$\phi(N) = \bigcup_i \phi(N \cap Q_i) \tag{13.639}$$

où les Q_i sont tous des cubes compacts. Donc $\phi(N)$ est une union dénombrable d'ensembles négligeables ; ergo négligeable lui-même par le lemme 13.55. □

Proposition 13.220.
Soient U et V des ouverts de $\mathbb{R}^N$ et $\phi\colon U \to V$ un C^1-difféomorphisme. Si $f\colon V \to \mathbb{C}$ est mesurable, alors $f \circ \phi\colon U \to \mathbb{C}$ l'est.

Démonstration. Soit A une partie mesurable de $\mathbb{C}$. Il nous faut prouver que

$$(f \circ \phi)^{-1}(A) = \phi^{-1}\big(f^{-1}(A)\big) \tag{13.640}$$

soit mesurable. Par hypothèse , $f^{-1}(A)$ est mesurable. Vu que ϕ est un C^1-difféomorphisme, elle et son inverse sont mesurables par la proposition 13.219. Donc l'image du mesurable $f^{-1}(A)$ par ϕ^{-1} est encore mesurable. □

13.13.3 Propriétés d'unicité

Corolaire 13.221.
La mesure λ_N est l'unique mesure sur $(\mathbb{R}^N, \mathcal{B}or(\mathbb{R}^N))$ à satisfaire

$$\mu\Big(\prod_{i=1}^{N}[a_i, b_i]\Big) = \prod_{i=1}^{n} |a_i - b_i| \tag{13.641}$$

Démonstration. Par définition de la mesure produit, λ_N est l'unique mesure sur $(\mathbb{R}^N, \mathcal{B}\mathrm{or}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}\mathrm{or}(\mathbb{R}))$ à satisfaire la condition. La proposition 13.116 conclut. □

Vu que les compacts de $\mathbb{R}^n$ sont les fermés bornés (théorème 7.127), et que tout borné est dans un tel produit d'intervalle, la mesure de Lebesgue est une mesure de Borel (définition 13.80(1)).

63. Ici « mesurable » parle de mesurabilité au sens de la tribu de Lebesgue, c'est-à-dire pas seulement les boréliens.

Théorème 13.222 ([300]).
La mesure de Lebesgue est invariante par translation. Autrement dit si A est mesurable dans $\mathbb{R}^n$ et si $a \in \mathbb{R}^n$ alors $A + a$ est mesurable et

$$\lambda_N(A + a) = \lambda_N(A). \tag{13.642}$$

Démonstration. Nous supposons que A est borélien ; sinon il l'est à ensemble négligeable près. Nous notons t_a la translation et nous nommons μ la mesure donnée par

$$\mu(A) = \lambda_N(A + a). \tag{13.643}$$

Vu que

$$\mu\Big(\prod_{n=1}^{N}[r_n, s_n[\Big) = \lambda_N\Big(\prod_i [r_n + a_n, s_n + a_n[\Big) = \prod_i |s_n - r_n|. \tag{13.644}$$

Vu qu'il y a unicité de la mesure vérifiant cette propriété (corolaire 13.221), nous avons $\mu = \lambda_N$. □

Pour la suite nous notons Q_0 le cube unité de $\mathbb{R}^N$: $Q_0 = \big([0, 1[\big)^N$.

Théorème 13.223 ([300]).
Soit μ une mesure positive sur $\mathbb{R}^N$ telle que

(1) μ soit invariante par translation (des boréliens),

(2) $\mu(Q_0) = 1$.

Alors $\mu = \lambda_N$.

Démonstration. Pour simplifier l'écriture nous faisons $N = 2$. Notre but est de prouver que $\mu([0, r[\times [0, r'[) = rr'$ pour tout $r, r' \in \mathbb{R}$.

Longueur $=1/J$ Soient J, K des entiers. Nous pouvons diviser le cube Q_0 en rectangles de côtés $1/J$ et A/K :

$$Q_0 = \bigcup_{\substack{1\leqslant j\leqslant J\\ 1\leqslant k\leqslant K}} [\frac{j-1}{J}, \frac{j}{J}[\times [\frac{k-1}{K}, \frac{k}{K}[\tag{13.645}$$

où l'union est disjointe. En ce qui concerne la mesure nous commençons par utiliser la sous-additivité :

$$\mu(Q_0) = \sum_{\substack{1\leqslant j\leqslant J\\ 1\leqslant k\leqslant K}} \mu\left([\frac{j-1}{J}, \frac{j}{J}[\times [\frac{k-1}{K}, \frac{k}{K}[\right). \tag{13.646}$$

Nous utilisons ensuite, sur chacun des termes séparément l'invariance par translation selon les vecteurs $(\frac{j-1}{J}, 0)$ et $(0, \frac{k-1}{K})$:

$$1 = \mu(Q_0) = \sum_{\substack{1\leqslant j\leqslant J\\ 1\leqslant k\leqslant K}} \mu\left([0, \frac{1}{J}[\times [0, \frac{1}{K}[\right) = JK\mu\mu\left([0, \frac{1}{J}[\times [0, \frac{1}{K}[\right), \tag{13.647}$$

et donc

$$\mu\left([0, \frac{1}{J}[\times [0, \frac{1}{K}[\right) = \frac{1}{J} \times \frac{1}{K}. \tag{13.648}$$

Longueur L/K Soient L, M des entiers et calculons :

$$\mu\left([\frac{0}{J}, \frac{L}{J}[\times [\frac{0}{K}, \frac{M}{K}[\right) = \sum_{\substack{0\leqslant l\leqslant L-1\\ 0\leqslant m\leqslant M-1}} \mu\left([\frac{l}{J}, \frac{l+1}{J}[\times [\frac{m}{K}, \frac{m+1}{K}[\right) \tag{13.649a}$$

$$= LM\mu\left([\frac{0}{J}, \frac{1}{J}[\times [\frac{0}{K}, \frac{1}{K}[\right) \tag{13.649b}$$

$$= LM \times \frac{1}{J} \times \frac{1}{K}. \tag{13.649c}$$

Nous avons donc, pour tout J, K, L, M :

$$\mu\left([0, \frac{L}{J}[\times [0, \frac{M}{K}[\right) = \frac{L}{J} \times \frac{M}{K}, \tag{13.650}$$

c'est-à-dire que pour tout $r, s \in \mathbb{Q}^+$ nous avons

$$\mu\big([0, r[\times [0, s[\big) = rs. \tag{13.651}$$

Longueur réelle Nous passons au cas de longueur réelle. Soit $a > 0$ et une suite croissante de rationnels $r_n \to a$. Une telle suite existe par la proposition 7.122. L'intervalle $[0, a[$ s'écrit sous la forme d'une union croissante $[0, a[= \bigcup_{n\geqslant 1}[0, r_n[$; le lemme 13.23(1) peut être utilisé et nous avons

$$\mu\big([0, a[\big) = \mu\left(\bigcup_{n\geqslant 1}[0, r_n[\right) = \lim_{n\to\infty} \mu\big([0, r_n[\big) = \lim_{n\to\infty} r_n = a. \tag{13.652}$$

Enfin, si $a, a' \in \mathbb{R}$, l'invariance par translation donne

$$\mu\big([a, a'[\big) = \mu\big([0, a' - a[\big) = a' - a. \tag{13.653}$$

Par unicité de la mesure ayant cette propriété, nous avons $\mu = \lambda_N$. □

Corolaire 13.224.
Si μ est une mesure positive sur $\mathbb{R}^N$ invariante par translation et telle que $\mu(Q_0) = C < \infty$ alors $\mu = C\lambda_N$.

Démonstration. Si $C > 0$ nous considérons la mesure $\frac{1}{C}\mu$ qui vérifie $(\frac{1}{C}\mu)(Q_0) = 1$. En conséquence du théorème 13.223, $\frac{1}{C}\mu = \lambda_N$ et $\mu = C\lambda_N$.

Si au contraire $C = 0$ alors nous pouvons paver $\mathbb{R}^N$ avec des cubes Q_i de côté 1 qui ont tous mesure 0. Par conséquent, $\mathbb{R}^N = \bigcup_{i=1}^{\infty} Q_i$, donc $\mu(\mathbb{R}^N) = \sum_i \mu(Q_i) = 0$. Par conséquent $\mu = 0$ parce que toute partie de $\mathbb{R}^N$ a une mesure au maximum égale à celle de $\mathbb{R}^N$. □

13.13.4 Régularité

Les différentes notions de régularité pour une mesure sont données dans la définition 13.80. Ce sont essentiellement des questions de compatibilité entre la mesure et la topologie.

Proposition 13.225.
La mesure de Lebesgue est une mesure de Radon sur tout ouvert de $\mathbb{R}^N$.

Démonstration. Soit V un ouvert de $\mathbb{R}^N$. C'est localement compact et dénombrable à l'infini. Il suffit de prouver que λ_N est de Borel sur V pour que le théorème 13.81 conclue à la régularité de la mesure de Lebesgue.

Soit K un compact de V. Par la proposition 7.65 c'est également un compact de $\mathbb{R}^N$. Par conséquent K est dans un pavé fermé de $\mathbb{R}^N$ du type

$$K \subset \prod_{n=1}^{N} [a_n, b_n] \tag{13.654}$$

et donc en passant par le corolaire 13.221,

$$\lambda_N(K) \leqslant \prod_{i=1}^{N} (b_n - a_n) < \infty. \tag{13.655}$$

Nous avons démontré que λ_N reste fini sur tout compact de V. □

13.14 Propriétés de l'intégrale de Lebesgue

Un lemme qui a l'air de rien, mais qui au final est souvent utilisé ; tellement qu'on l'oublie un peu.

Lemme 13.226 ([1]).
Soit un compact K de $\mathbb{R}$ et une fonction continue $f\colon \mathbb{R}\to\mathbb{R}$. Alors l'intégrale

$$\int_K f \tag{13.656}$$

existe et est finie.

Démonstration. Vu que f est continue sur le compact K, elle y atteint une borne supérieure [64] que nous nommons M.

Soit R tel que $B(0,R)$ contienne K. La fonction $(M+1)\mathbb{1}_{B(0,R)}$ majore strictement f sur le mesurable $B(0,R)$. L'ensemble sur lequel nous prenons le supremum dans la définition (13.424) de l'intégrale de f contient donc au moins le nombre fini $(M+1)\mu(K)$. Le supremum existe et est fini (proposition 1.157). □

13.14.1 Quelques limites dans les bornes

Dans le cas de l'intégrale de Lebesgue définie par 13.151, si f est une fonction sur $\mathbb{R}$ et si λ est la mesure de Lebesgue, nous avons une définition directe de

$$\int_0^\infty f\lambda. \tag{13.657}$$

Nous sommes cependant en droit de nous demander si nous n'aurions pas également ceci :

$$\lim_{x\to\infty}\int_0^x f\lambda = \int_0^\infty f d\lambda. \tag{13.658}$$

Lorsque l'intégrale considérée est celle de Riemann, l'égalité (13.658) est une définition. Ici, ça va être une propriété.

13.227.
Tant que nous sommes à parler de limites dans les bornes, nous aurions pu vouloir, pour les séries, suivre le chemin suivant :

— Définir l'intégrale de Lebesgue sur un espace mesuré.
— Prendre au passage le cas particulier $\sum_{k=0}^\infty a_k = \int_{\mathbb{N}} a$ où $a\colon \mathbb{N}\to\mathbb{R}$ est une fonction mesurable pour la mesure de comptage.
— Démontrer qu'avec ces définitions, $\sum_{k=0}^\infty = \lim_{N\to\infty}\sum_{k=0}^N a_k$.

Or le dernier point est pris comme définition et son égalité avec l'intégrale pour la mesure de comptage est une propriété [65]. Pourquoi ? Parce que la définition 13.20 de mesure positive demande déjà d'avoir défini les sommes sur $\mathbb{N}$.

Lemme 13.228.
Soit une partie mesurable $A\subset\mathbb{R}^+$ de mesure finie. Alors

$$\lim_{M\to 0}\lambda\big(A\cap[M,\infty[\big) = 0. \tag{13.659}$$

Démonstration. La fonction

$$\begin{aligned} f\colon \mathbb{R}^+ &\to \mathbb{R}^+\\ x &\mapsto \lambda\big(A\cap[x,\infty[\big)\end{aligned} \tag{13.660}$$

64. Nous ne nous lasserons jamais de citer le théorème de Weierstrass 7.105.
65. Proposition 13.231.

est décroissante et bornée vers le bas par 0. Elle possède dont une limite $\ell \geqslant 0$ (corolaire 7.156). Nous allons prouver que $\ell = 0$ en calculant la limite sur les entiers.

Nous posons $J_k = [k, k+1[$. Pour $n \in \mathbb{N}$ nous avons :

$$f(n) = \lambda\big(A \cap [n, \infty[\big) \tag{13.661a}$$

$$= \lambda\big(\bigcup_{k=n}^{\infty} (A \cap J_k)\big) \tag{13.661b}$$

$$= \sum_{k=n}^{\infty} \lambda(A \cap J_k). \tag{13.661c}$$

Mais nous savons par hypothèse sur la mesure de A que

$$\lambda(A) = \sum_{k=0}^{\infty} \lambda(A \cap J_k) < \infty. \tag{13.662}$$

Donc $f(n)$ est une queue de série convergente. Elle tend donc vers zéro par le lemme 10.59. C'est à dire que

$$\lim_{n\to\infty} f(n) = 0. \tag{13.663}$$

Et comme $\lim_{x\to\infty} f(x)$ existe et vaut ℓ, la seule possibilité est $\ell = 0$. □

Lemme 13.229.
Soit une fonction mesurable $f\colon \mathbb{R} \to \mathbb{R}^+$. Alors

$$\lim_{x\to\infty} \int_x^{\infty} f d\lambda = 0. \tag{13.664}$$

Démonstration. Nous posons

$$F(x) = \int_x^{\infty} f d\lambda. \tag{13.665}$$

Nous commençons par prouver que c'est une fonction décroissante. En effet,

$$F(x) - F(x+a) = \int_x^{\infty} f - \int_{x+a}^{\infty} f = \int_x^{x+a} f + \int_{x+a}^{\infty} f - \int_{x+a}^{\infty} = \int_x^{a+a} f \geqslant 0. \tag{13.666}$$

Nous avons utilisé 13.174.

Vu que f prend ses valeurs dans $\mathbb{R}^+$, nous avons $F(x) \geqslant 0$ pour tout x. La fonction F est décroissante et bornée vers le bas. Donc elle a une limite :

$$\lim_{x\to\infty} F(x) = \ell \geqslant 0. \tag{13.667}$$

Supposons $\ell > 0$ et posons $0 < \epsilon < \ell$. Soit M tel que pour tout $x > M$ nous ayons

$$\int_x^{\infty} f > m. \tag{13.668}$$

Soit $a \in \mathbb{R}^+$ tel que

$$|\int_a^{\infty} f(t) - \ell\, dt| < \epsilon. \tag{13.669}$$

En vertu de (13.668) nous considérons $a_0 > a$ tel que

$$\int_{a_0}^{\infty} f = I_0 > m. \tag{13.670}$$

Nous construisons la suite strictement croissante (a_k) de la façon suivante :

$$\int_{a_k}^{\infty} f = I_k > m \tag{13.671}$$

et

$$|\int_{a_k}^{a_{k+1}} f - I_k| < \epsilon. \tag{13.672}$$

Donc pour k nous avons

$$\int_{a_k}^{a_{k+1}} f \geqslant I_k - \epsilon \geqslant m - \epsilon. \tag{13.673}$$

Mais

$$\int_{a_0}^{\infty} = \sum_{k=0}^{\infty} \int_{a_k}^{a_{k+1}} f \geqslant \sum_k (m - \epsilon) = \infty. \tag{13.674}$$

Nous avons une contradiction. □

13.14.2 Mesure de comptage et série

Définition 13.230 (mesure de comptage)**.**
Soit $(S, \mathcal{F})$ un ensemble mesurable. La ***mesure de comptage*** *sur $(S, \mathcal{F})$ est la mesure définie par*

$$m(A) = \begin{cases} \mathrm{Card}(A) & \text{si } A \text{ est fini} \\ +\infty & \text{sinon.} \end{cases} \tag{13.675}$$

Cette mesure est utilisée pour voir des séries comme des intégrales sur $(\mathbb{N}, \mathcal{P}(\mathbb{N}), m)$.

Proposition 13.231.
Soit l'espace mesuré $\big(\mathbb{N}, \mathcal{P}(\mathbb{N}), m\big)$ [66]. Nous considérons une suite $a\colon \mathbb{N} \to \mathbb{R}$.

(1) L'intégrale $\int_{\mathbb{N}} a\, dm$ existe si et seulement si la série $\sum_{n=0}^{\infty} a_n$ existe.
(2) Si $\int_{\mathbb{N}} a\, dm$ existe, alors

$$\int_{\mathbb{N}} a\, dm = \sum_{n=0}^{\infty} a_n. \tag{13.676}$$

Exemple 13.232.
La mesure de comptage m sur $\mathbb{N}$ muni de la tribu de ses parties est σ-finie parce que $E_n = \{0, \dots, n\}$ est de mesure finie et $\bigcup_{n\in\mathbb{N}} E_n = \mathbb{N}$. △

Exemple 13.233.
L'intervalle $I = [0, 1]$ muni de la tribu de toutes ses parties et de la mesure de comptage est un espace mesuré non σ-fini. △

13.14.3 Théorème de la moyenne

Théorème 13.234 ([1])**.**
Soit Q un compact connexe par arcs et une fonction continue $f\colon Q \to \mathbb{R}$. Si λ est la mesure de Lebesgue, alors il existe $a \in Q$ tel que

$$f(a) = \frac{1}{\lambda(Q)} \int_Q f d\lambda \tag{13.677}$$

Démonstration. En posant $I = \int_Q f d\lambda$ nous avons immédiatement

$$\min(f)\lambda(Q) \leqslant I \leqslant \max(f)\lambda(Q) \tag{13.678}$$

où le minimum et le maximum existent parce que f est continue sur un compact. Si une des deux inégalités est une égalité alors la fonction est constante. En effet supposons que la première inégalité soit une égalité ; si la fonction n'était pas constante, il existerait une boule sur laquelle

66. La mesure de comptage sur $\mathbb{N}$ est donnée en la définition 13.230.

f serait strictement supérieure à $\min(f)$. En intégrant d'abord sur cette boule et ensuite sur le complémentaire nous obtenons une intégrale plus grande que $\min(f)\lambda(Q)$.

Soit $\epsilon > 0$. Il existe $\alpha, \beta \in Q$ tels que $f(\alpha) \leqslant \min(f)+\epsilon$ et $f(\beta) \geqslant \max(f)-\epsilon$. Soit $\gamma\colon [0,1] \to Q$ un chemin continu tel que $\gamma(0) = \alpha$ et $\gamma(1) = \beta$. La fonction $f \circ \gamma\colon [0,1] \to \mathbb{R}$ est alors continue et vérifie $(f \circ \gamma)(0) \leqslant \min(f) + \epsilon$ et $(f \circ \gamma)(1) \geqslant \max(f) - \epsilon$.

Si ϵ est assez petit et vu que les inégalités (13.678) sont strictes,

$$\lambda(Q)(f \circ \gamma)(0) \leqslant \min(f)\lambda(Q) + \epsilon\lambda(Q) < I < \max(f)\lambda(Q) - \epsilon\lambda(Q) \leqslant \lambda(Q)(f \circ \gamma)(1). \tag{13.679}$$

Par le théorème des valeurs intermédiaires 11.50, il existe $t_0 \in [0,1]$ tel que $\lambda(Q)(f \circ \gamma)(t_0) = I$. Le point $a = \gamma(t_0)$ vérifie

$$f(a) = \frac{1}{\lambda(Q)} \int_Q f d\lambda. \tag{13.680}$$

□

13.14.4 Primitives et intégrales

En termes de notations, si $a < b$ nous posons

$$\int_a^b f(t)dt = \int_{[a,b]} f. \tag{13.681}$$

Si par contre $a > b$ nous posons $\int_a^b f = -\int_b^a f$.

Proposition 13.235 (Primitive et intégrale[147]).
Soit f une fonction intégrable sur $[a,b]$ et continue sur $]a,b[$. Alors la fonction

$$\begin{aligned} F\colon [a,b] &\to \mathbb{R} \\ x &\mapsto \int_{[a,x]} f(t)dt. \end{aligned} \tag{13.682}$$

est l'unique primitive de f sur $]a,b[$ s'annulant en $x = a$.

Démonstration. Nous devons prouver que F est dérivable et que pour tout $x_0 \in]a,b[$ nous avons $F'(x_0) = f(x_0)$. Soit $\epsilon > 0$. Par continuité de f en x_0, il existe une fonction $\alpha\colon \mathbb{R} \to \mathbb{R}$ telle que

$$f(x_0 + h) = f(x_0) + \alpha(h) \tag{13.683}$$

avec $\lim_{h\to 0} \alpha(h) = 0$. Cette dernière limite signifie qu'il existe un $\delta > 0$ tel que $|\alpha(h)| < \epsilon$ pour tout h tel que $|h| < \delta$, c'est-à-dire pour tout $h \in B(0,\delta)$. À partir de maintenant nous ne considérons plus que de tels h.

Notre travail maintenant est de prouver que F est dérivable en x_0, et de montrer que la dérivée est $f(x_0)$. Pour cela,

$$F(x_0 + h) - F(x_0) = \int_{x_0}^{x_0+h} f(t)dt \tag{13.684a}$$

$$= \int_0^h f(x_0 + t)dt \tag{13.684b}$$

$$= \int_0^h \big[f(x_0) + \alpha(t)\big]dt \tag{13.684c}$$

$$= hf(x_0) + \int_0^h \alpha(t)dt. \tag{13.684d}$$

Nous avons donc montré que pour tout $\epsilon > 0$, il existe un δ (défini via la fonction α) tel que $|h| < \delta$ implique

$$\left| \frac{F(x_0 + h) - F(x_0)}{h} - f(x_0) \right| < \epsilon. \tag{13.685}$$

Cela signifie que

$$\lim_{h\to 0} \frac{F(x_0+h)-F(x_0)}{h} = f(x_0), \tag{13.686}$$

qui n'est rien d'autre que le fait que F est dérivable en x_0 et que sa dérivée est $f(x_0)$.

Le fait que F s'annule en $x = a$ est par sa définition. L'unicité provient du corolaire 11.137. □

Théorème 13.236 (Théorème fondamental du calcul intégral)**.**
Soit f une fonction continue sur un intervalle ouvert I contenant strictement l'intervalle $[a, b] \subset \mathbb{R}$ et F une primitive de f sur I. Alors

$$\int_a^b f(t)dt = F(b) - F(a). \tag{13.687}$$

Démonstration. Nous avons vu par la proposition 13.235 que la fonction

$$\begin{aligned} G\colon [a,b] &\to \mathbb{R} \\ x &\mapsto \int_a^x f(t)dt \end{aligned} \tag{13.688}$$

était l'unique primitive de f sur $]a, b[$ à s'annuler pour $x = a$. Nous avons évidemment

$$\int_a^b f(t)dt = G(b). \tag{13.689}$$

Si F est une primitive quelconque, il suffit de soustraire sa valeur en $x = a$: $G(x) = F(x) - F(a)$ et donc

$$\int_a^b f(t)dt = G(b) = F(b) - F(a), \tag{13.690}$$

comme il fallait le prouver. □

Le théorème fondamental s'écrit souvent sous la forme[67]

$$f(x) = f(a) + \int_a^x f'(t)dt. \tag{13.691}$$

Sous cette forme, il faut penser que nous calculons $f(x)$ en un point pas trop éloigné de a, en sachant $f(a)$ et en intégrant la dérivée entre les deux.

Remarque 13.237.
Le lien entre primitive et intégrale est fondamentalement lié à l'invariance par translation de la mesure de Lebesgue, et non à la construction précise de cette mesure. Mais en même temps, la mesure de Lebesgue est l'unique à être invariante par translation.

Quelque remarques.

(1) Le théorème fondamental du calcul intégral est à utiliser pour calculer des intégrales des fonctions réelles lorsqu'on a des primitives sur un domaine strictement plus large que le domaine sur lequel nous voulons intégrer.

(2) Une version pour les intégrales impropres sera donnée au corolaire 13.250.

(3) Une primitive est forcément une fonction continue parce qu'une primitive est dérivable.

(4) Le théorème fondamental du calcul intégral ne sert pas qu'à calculer des intégrales à partir de primitives. Il sert aussi à démontrer des résultats plus théoriques, comme le théorème 11.311.

(5) En vertu du corolaire 11.137, une fonction ne possède qu'une seule primitive à constante près.

67. Par exemple dans les théorèmes du reste des polynômes de Taylor 14.34 et de Cauchy-Lipschitz 16.40.

13.14.5 Exemples et applications

Si f est une fonction définie sur un intervalle I et y admettant des primitives, nous notons

$$\int f(x)dx \tag{13.692}$$

l'ensemble des primitives de f sur I :

$$\int f(x)dx = \{F(x) + C \text{ tel que } C \in \mathbb{R}\} \tag{13.693}$$

où F est une quelconque primitive de f.

Exemple 13.238.
Une primitive bien connue de $f\colon x \mapsto x^2$ est la fonction $F\colon x \to \frac{x^3}{3}$. Nous écrivons donc

$$\int x^2 dx = \frac{x^3}{3} + C. \tag{13.694}$$

Cela est un abus de notations terrible pour dire en réalité

$$\{x \mapsto \frac{x^3}{3} + C \text{ tel que } C \in \mathbb{R}\}. \tag{13.695}$$

△

En termes de notations, nous posons

$$\int_a^b f(t)dt = \Big[F(t)\Big]_{t=a}^{t=b} = F(b) - F(a). \tag{13.696}$$

Remarque 13.239.
La valeur de l'intégrale ne dépend pas de la primitive qu'on choisi pour le calculer, car si F_1 et F_2 sont deux primitives de f alors $F_1 = F_2 + C$ et $F_1(b) - F_1(a) = (F_2(b) + C) - (F_2(a) + C) = F_2(b) - F_2(a)$.

Remarque 13.240.
Si l'intervalle d'intégration est réduit à un seul point alors la valeur de l'intégrale est zéro. Nous le savions déjà, et cela est cohérent avec le théorème fondamental car $\int_a^a f(t)dt = F(a) - F(a) = 0$.

Remarque 13.241.
Toute intégrale d'une fonction impaire sur un intervalle symétrique par rapport à l'origine est nulle.

Proposition 13.242 ([1])**.**
Soient des espaces vectoriels normés E et F où F est de dimension finie[68]*. Nous considérons une fonction $f\colon E \to F$ de classe C^1 ainsi qu'un chemin $\gamma\colon [0,1] \to E$ de classe C^1 également.*

Alors nous avons l'égalité

$$\int_0^1 (df)_{\gamma(t)}\big(\gamma'(t)\big) = f\big(\gamma(1)\big) - f\big(\gamma(0)\big). \tag{13.697}$$

Démonstration. Nous posons

$$\begin{aligned} f\colon [0,1] &\to F \\ t &\mapsto (f \circ \gamma)(t). \end{aligned} \tag{13.698}$$

Cette fonction vérifie $g'(t) = (df)_{\gamma(t)}\big(\gamma'(t)\big)$ par le lemme 11.223. Le théorème fondamental du calcul intégral[69] nous permet donc d'écrire

$$\int_0^1 (df)_{\gamma(t)}\big(\gamma'(t)\big)dt = \int_0^1 g'(t)df = g(1) - g(0). \tag{13.699}$$

Notons que g est continue grâce aux hypothèses de classe C^1 pour γ et f. □

68. Sinon l'intégrale dont nous allons parler n'est pas définie au sens où nous n'en avons pas donné de définition. Voir 13.175.
69. Théorème 13.236.

13.14.6 Permuter limite et dérivée

13.243.
Voici une preuve alternative du théorème 11.311. Elle utilise des intégrales ; elle demande donc plus de dependences.

Théorème 13.244 ([147]).
Soient une suite de fonctions $f_i\colon \mathbb{R} \to \mathbb{R}$, une fonction $f\colon \mathbb{R} \to \mathbb{R}$ et une fonction $g\colon \mathbb{R} \to \mathbb{R}$ telles que

(1) f_i est de classe C^1 pour tout i,

(2) $f_i \to f$ simplement,

(3) $f_i' \to g$ uniformément sur tout compact.

Alors

(1) f est de classe C^1,

(2) $f' = g$,

(3) $f_i \to f$ uniformément sur tout compact.

Démonstration. Nous commençons par considérer $x_0 \in \mathbb{R}$ et un intervalle compact K contenant x_0. Nous montrons que $f'(x_0) = g(x_0)$ en plusieurs étapes.

Une formule intégrale Par hypothèse, les fonctions f_i sont continues (en particulier sur un ouvert contenant K), et le théorème fondamental de l'analyse 13.236 donne

$$f_i(x) = f_i(x_0) + \int_{x_0}^{x} f_i'(t)dt \tag{13.700}$$

pour tout $x \in K$. Nous avons envie de prendre la limite $i \to \infty$ en permutant la limite avec l'intégrale. Pour cela nous allons utiliser la convergence dominée de Lebesgue.

Convergence dominée La convergence uniforme sur tout compact des fonctions continues f_i' vers g donne la continuité de g, théorème 11.294. En particulier g est bornée et donc intégrable sur le compact $[x_0, x]$. Mais il en faut plus pour le théorème de la convergence dominée de Lebesgue (théorème 13.184). Soit $a > 0$; il existe N tel que pour tout $i > n$ nous ayons $\|f_i' - g\| < a$. Avec cela nous avons

$$|f_i'(x)| < |g(x)| + a \tag{13.701}$$

pour tout $x \in K$. En particulier, la fonction $x \mapsto g(x) + a$ fonctionne pour la convergence dominée et nous pouvons permuter la limite et l'intégrale dans (13.700).

Passage à la limite En passant à la limite $i \to \infty$ dans (13.700) nous trouvons

$$f(x) = f(x_0) + \int_{x_0}^{x} g(t)dt. \tag{13.702}$$

Premières conclusions Il suffit maintenant de prendre la dérivée de (13.702) au point $x = x_0$ grâce à la proposition 13.235 :

$$f'(x_0) = g(x_0). \tag{13.703}$$

Cela nous donne l'égalité $f = g$ parce que x_0 était arbitraire.

De plus g est continue comme limite uniforme des fonctions continues f_i'. Plus précisément, pour voir la continuité de g en x_0, prendre un ouvert borné $B(x_0, r)$ autour de x_0, et ensuite un compact K contenant cet ouvert. La convergence uniforme $f_i' \to g$ sur K implique la convergence uniforme sur $B(x_0, r)$ et donc la continuité sur $B(x_0, r)$ (théorème 11.294).

$f_i \to f$ uniforme sur tout compact Un compact n'étant pas spécialement connexe, nous ne pouvons pas reprendre le travail fait jusqu'ici sans prendre une petite précaution. Soit un compact L. Cette partie de $\mathbb{R}$ étant bornée[70], nous pouvons prendre r assez grand pour que $L \subset \overline{B(0,r)}$. Nous posons $K = \overline{B(0,r)}$ et nous prouvons la convergence uniforme $f_i \to f$ sur K. A fortiori, cela donnera la convergence uniforme sur L.

Prenons la différence entre (13.702) et (13.700) :

$$|f(x) - f_i(x)| = \Big|f(x_0) - f_i(x_0) + \int_{x_0}^{x} g(t) - f_i'(t)dt\Big| \tag{13.704a}$$

$$\leqslant |f(x_0) - f_i(x_0)| + \Big|\int_{x_0}^{x} |g(t) - f_i'(t)|dt\Big| \tag{13.704b}$$

$$\leqslant |(f - f_i)(x_0)| + |x - x_0|\|g - f_i'\|_K. \tag{13.704c}$$

Notez les valeurs absolues autour de l'intégrale dans (13.704b). Elles sont nécessaires parce que x est dans un voisinage de x_0, sans que nous sachions si $x \geqslant x_0$ ou $x \leqslant x_0$ (ça change le signe de l'intégrale).

Nous avons donc

$$\|f - f_i\| \leqslant |(f - f_i)(x_0)| + \operatorname{diam}(K)\|g - f_i'\| \tag{13.705}$$

où $\operatorname{diam}(K)$ est le diamètre de K, c'est-à-dire la plus grande distance entre deux éléments de K c'est un nombre fini parce que K est borné. Il majore évidemment $|x - x_0|$. Le membre de droite tend vers zéro si $i \to \infty$ parce que nous avons convergence simple $f_i \to f$ et donc $(f - f_i)(x_0) \to 0$, et parce que nous avons convergence uniforme sur tout compact, donc $\|g - f_i'\| \to 0$.

Nous avons donc bien $\lim_{i\to\infty} \|f - f_i\| = 0$, c'est-à-dire convergence uniforme de (f_i) vers f sur K.

□

La proposition suivante est la généralisation à $\mathbb{R}$ de la proposition 11.363.

Proposition 13.245.
Pour tout $\alpha \in \mathbb{R}$, si $f_\alpha(x) = x^\alpha$ alors

$$f_\alpha'(x) = \alpha x^{\alpha-1}. \tag{13.706}$$

Au niveau du domaine, c'est $\mathbb{R}$ auquel il faut enlever $\{0\}$ si $\alpha - 1 < 0$.

Démonstration. Soient $\alpha \in \mathbb{R}$ et une suite de rationnels α_i qui converge vers α. Le plus amateurs d'abstraction diront $(\alpha_i) \in \alpha$ en référence à la proposition 1.133.

Nous notons $f_\alpha(x) = x^\alpha$ et $f_i(x) = x^{\alpha_i}$. Par définition nous avon

$$f_i \to f_\alpha \tag{13.707}$$

ponctuellement. De plus en utilisant la proposition 11.363 nous savons que $f_i'(x) = \alpha_i x^{\alpha_i - 1}$. En posant $g(x) = \alpha x^{\alpha-1}$ nous avons donc

$$f_i' \to g. \tag{13.708}$$

ponctuellement. Mais f_i' est continue pour tout i et g également. Donc la convergence $f_i \to f_\alpha$ est uniforme sur tout compact[71]. Le théorème 11.311 nous permet de permuter limite et dérivée pour avoir $g = f_\alpha'$. □

70. Par le théorème de Borel-Lebesgue 7.127
71. Proposition 11.300.

13.14.7 Intégrales impropres

Définition 13.246 ([147]).
*Une fonction $f\colon D \subset \mathbb{R} \to \mathbb{R}$ est **localement intégrable** sur un intervalle I si f est intégrable sur tout intervalle compact contenu dans I.*

Proposition 13.247.
Soit $f\colon [a,b] \to \mathbb{R}$ une fonction intégrable. Alors

$$\int_{[a,b]} f = \lim_{x\to b^-} \int_a^x f. \tag{13.709}$$

Démonstration. Notons que la valeur de f en b n'a strictement aucune importance parce que l'intégrale de Lebesgue ne dépend pas du choix de la valeur de la fonction en un ensemble de mesure nulle ; et en même temps la limite à gauche de (13.709) ne dépend pas non plus de la valeur de f en b. Bref si f n'est pas définie en b, nous pouvons poser $f(b) = 42$.

Notons de plus que du point de vue de l'intégrale de Lebesgue, $\int_{[a,b]}$ et $\int_{[a,b[}$ sont identiques et valent toutes les deux $\int_a^b$ (lorsque ça existe).

Supposons d'abord que f est positive. Alors nous posons $f_n = f\mathbb{1}_{[a,b-\frac{1}{n}]}$. Ponctuellement nous avons la limite croissante $f_n \to f$ et de plus

$$\lim_{x\to b^-} \int_{[a,x]} f = \lim_{n\to\infty} \int_{[a,b]} f_n. \tag{13.710}$$

Chacun des f_n est intégrable sur $[a,b]$. Le théorème de Beppo-Levi 13.160 implique que f est intégrable sur $[a,b]$ et que

$$\lim_{n\to\infty} \int_a^b f_n = \int_a^b f. \tag{13.711}$$

Cela montre que dans le cas d'une fonction f positive nous avons bien (13.709).

Si f n'est pas positif, alors nous la décomposons en partie positive et négative $f = f^+ - f^-$ et par définition de l'intégrale d'une fonction non positive,

$$\lim_{x\to b^-} \int_{[a,x[} f = \lim \int f^+ - \lim \int f^-. \tag{13.712}$$

□

Il peut cependant arriver que la limite $\lim_{x\to b} \int_a^b f$ existe alors que f n'est pas intégrable sur $[a,b]$. C'est l'ennui des fonctions non positives. Un exemple classique est

$$\int_0^\infty \frac{\sin(t)}{t} dt \tag{13.713}$$

Définition 13.248 ([301]).
Si

$$\lim_{x\to b} \int_a^b f \tag{13.714}$$

*existe alors nous disons que l'intégrale est **convergente** en b. Ce procédé de limite est l'intégrale **impropre** de f sur $[a,b]$.*

Exemple 13.249 (Intégale impropre).
Nous considérons la fonction $f\colon [0,\infty[\to \mathbb{R}$ définie par

$$f(x) = \begin{cases} \frac{1}{n} & \text{si } x \in [2n-2, 2n-1[\\ -\frac{1}{n} & \text{si } x \in [2n-1, 2n[. \end{cases} \tag{13.715}$$

Par la divergence de la série harmonique, $\int_0^\infty |f|$ n'existe pas. La fonction f n'est donc pas intégrable au sens de Lebesgue (définition 13.168).

Cependant pour tout n pair nous avons

$$\int_0^n f = 0. \tag{13.716}$$

Du coup pour tout $x \geqslant 0$ nous avons

$$\int_0^x f = \int_{2n}^x f \tag{13.717}$$

où $2n$ est le plus grand nombre pair inférieur à x. Nous avons $|x - 2n| \leqslant 2$ et $|f(x)| \leqslant \frac{1}{n}$ pour $x \in [2n, x]$. Donc

$$\int_{2n}^x f \leqslant \frac{2}{n}. \tag{13.718}$$

Nous avons par conséquent

$$\lim_{x\to\infty} \int_0^x f = 0, \tag{13.719}$$

ce qui signifie que l'intégrale de f sur $[0, \infty[$ converge au sens des intégrales impropres. △

L'intégrale (13.713) est une intégrale convergente mais la fonction n'est pas intégrable (parce que pour être intégrale il faut que $|f|$ soit intégrable). Nous pouvons ainsi dire que cette intégrale converge mais n'existe pas.

Le corolaire suivant nous autorise à utiliser le théorème fondamental du calcul intégral 13.236 même dans les cas limites.

Corolaire 13.250.
Si f est localement intégrable sur $[a, b]$ et si F est une primitive de f sur tout ouvert de $[a, b]$ alors

$$\int_a^b f = \lim_{x\to b^-} F(x) - F(a). \tag{13.720}$$

Démonstration. Pour chaque x dans $[a, b[$ nous avons

$$\int_a^x f = F(x) - F(b). \tag{13.721}$$

La proposition 13.247 nous explique que la limite $x \to b^-$ du membre de gauche existe et vaut $\int_a^b f$. Donc également le membre de droite :

$$\int_a^b f = \lim_{x\to b^-} \int_a^x f = \lim_{x\to b^-} F(x) - F(b). \tag{13.722}$$

□

La convergence des intégrales de fonctions $\frac{1}{x^\alpha}$ en 0 et ∞ est une question classique de l'intégration. De plus ces fonctions servent souvent à utiliser une théorème de comparaison (type intégrale dominée de Lebesgue).

Proposition 13.251.
Deux intégrales remarquables.

(1) Nous avons

$$\int_0^1 \frac{1}{x^\alpha} = \infty \tag{13.723}$$

si et seulement si $\alpha \geqslant 1$.

(2) Nous avons

$$\int_1^\infty \frac{1}{x^\alpha} = \infty \tag{13.724}$$

si et seulement si $\alpha \leqslant 1$.

Démonstration. La fonction $\frac{1}{x^\alpha}$ admet la primitive $F(x) = \frac{1}{1-\alpha}\frac{1}{x^{\alpha-1}}$ sur tout compact de $]0,\infty[$. Le corolaire 13.250 nous permet [72] de dire que $\int_0^1 \frac{1}{x^\alpha}$ vaudra

$$\lim_{x\to 0-+} \frac{1}{1-\alpha}\frac{1}{x^{\alpha-1}}. \tag{13.725}$$

Cela est strictement plus petit que ∞ si et seulement si $\alpha < 1$. □

13.15 Changement de variables dans une intégrale multiple

Dans ce qui suit, U et V sont des ouverts de $\mathbb{R}^N$ et $\phi\colon U \to V$ est un C^1-difféomorphisme. Nous notons $\mathcal{Q}$ l'ensemble des cubes fermés dans U dont les côtés sont parallèles aux axes.

13.15.1 Des lemmes

Lemme 13.252 ([300]).
Soient μ et ν deux mesures de Borel sur l'ouvert U de $\mathbb{R}^N$. Si $\mu(Q) \leqslant \nu(Q)$ pour tout $Q \in \mathcal{Q}$ alors $\mu(B) \leqslant \nu(B)$ pour tout borélien B.

Démonstration. Si Q est un cube semi-ouvert, c'est-à-dire de la forme

$$Q = \prod_{i=1} N[a_n, a_n + h[\subset U \tag{13.726}$$

alors Q est une réunion croissante de cubes fermés du type $[a_n+\epsilon, a_n+h-\epsilon]$, et donc $\mu(Q) \leqslant \nu(Q)$ par le lemme 13.23(1). La propriété est donc vraie pour les cubes semi-ouverts.

Si Ω est un ouvert, alors il est réunion disjointe dénombrable de cubes semi-ouverts par la proposition 13.214. Donc pour tout ouvert $\Omega \subset U$ nous avons $\mu(\Omega) \leqslant \nu(\Omega)$. En vertu de la proposition 13.81 et de la remarque 13.82, les mesures μ et ν sont régulières, et l'inégalité au niveau des ouverts se répercute en inégalité pour tout boréliens de U :

$$\mu(B) \leqslant \nu(B) \tag{13.727}$$

pour tout $B \in \mathcal{B}\mathrm{or}(U)$. Notons que U étant ouvert dans $\mathbb{R}^N$, les boréliens de U sont exactement les boréliens de $\mathbb{R}^N$ inclus dans U par le corolaire 13.51. □

Lemme 13.253 ([300]).
Soit une application $\theta\colon U \to \mathbb{R}^N$ de classe C^1 où U est ouvert dans $\mathbb{R}^N$. Pour tout $Q \in \mathcal{Q}$ nous avons

$$\lambda_N\big(\theta(Q)\big) \leqslant \sup_{s\in Q} \|d\theta_s\|^N \lambda_N(Q). \tag{13.728}$$

Démonstration. Nous notons h la longueur du côté du cube. Le théorème des accroissements finis 11.257, pour la composante θ_i donne, pour $u, v \in Q$:

$$\big|\theta_i(u) - \theta_i(v)\big| \leqslant \sup_{s\in Q} \|(d\theta_i)_s\| \|u-v\| \leqslant \sum_{s\in Q} \|(d\theta_i)_s\| h. \tag{13.729}$$

D'autre part nous avons (nous écrivons pour $N = 2$ pour être plus court) :

$$d\theta_s(u) = \frac{d}{dt}\Big[\theta_1(s+tu)e_1 + \theta_2(s+tu)e_2\Big]_{t=0} = (d\theta_1)_s(u)e_1 + (d\theta_2)_s(u)e_2. \tag{13.730}$$

72. Tout ce que nous avons fait avec la borne b de l'intégrale $\int_a^b$ reste valable avec la borne a.

Donc pour chaque i : $\|d\theta_s\| \geqslant \|(d\theta_i)_s\|$, et nous continuons la majoration (13.729) :

$$\big|\theta_i(u) - \theta_i(v)\big| \leqslant\leqslant \sum_{s\in Q} \|(d\theta_i)_s\| h \leqslant \sup_{s\in Q} \|d\theta_s\| h. \tag{13.731}$$

Les points $\theta(u)$ et $\theta(v)$ sont donc dans un cube de côté $\sup_{s\in Q} \|d\theta_s\| h$, ce qui permet de majorer $\lambda_N\big(\theta(Q)\big)$ par

$$\lambda_N\big(\theta(Q)\big) \leqslant \Big(\sup_{s\in Q} \|d\theta_s\| h\Big)^N = \Big(\sup_{s\in Q} \|d\theta_s\|\Big)^N \lambda_N(Q) \tag{13.732}$$

où le dernier facteur provient de l'égalité $h^N = \lambda_N(Q)$. □

13.15.2 Déterminant et mesure de Lebesgue

Dans la suite, Q_0 désigne le cube unité : $Q_0 = \big([0,1[\big)^N$.

Théorème 13.254 (Interprétation géométrique du déterminant[300])**.**
Soit une application linéaire $T\colon \mathbb{R}^N \to \mathbb{R}^N$. Alors pour tout borélien B de $\mathbb{R}^N$,

$$\lambda_N\big(T(B)\big) = |\det(T)|\lambda_N(B). \tag{13.733}$$

Démonstration. Nous considérons la mesure positive μ donnée par $\mu(B) = \lambda_N\big(T(B)\big)$, qui est bien une mesure par la proposition 13.77. Cette mesure est invariante par translation parce que λ_N l'est :

$$\mu(B+a) = \lambda_N\big(T(B)+a\big) = \lambda_N\big(T(B)\big) = \mu(B). \tag{13.734}$$

De plus, $T(Q_0)$ est borné et nous notons $\mu(Q_0) = C$. Nous avons $\mu = C\lambda_N$ par le corolaire 13.224.

<u>$C(T_1T_2) = C(T_1)C(T_2)$</u> Par définition,

$$\begin{aligned} C(T_1T_2)\lambda_N(B) &= \lambda_N\big((T_1T_2)(B)\big) && (13.735a)\\ &= \lambda_N\big(T_1(T_2B)\big) = C(T_1)\lambda_N\big(T_2(B)\big) = C(T_1)C(T_2)\lambda_N(B). && (13.735b)\end{aligned}$$

Par conséquent la fonction C est multiplicative :

$$C(T_1T_2) = C(T_1)C(T_2). \tag{13.736}$$

Et en plus, $C(\mathrm{Id}) = 1$.

<u>**Matrice diagonale**</u> En guise de T, nous considérons l'application linéaire diagonale donnée par $De_i = d_ie_i$, ou, sous forme matricielle, $D = \mathrm{diag}(d_1,\dots,d_N)$ qui fait

$$T(Q_0) = [0,d_1[\times \dots \times [0,d_N[\tag{13.737}$$

La mesure de cela est $|d_1\cdots d_N|$, ce qui nous donne

$$C(D) = |d_1\dots d_N| = |\det(D)|. \tag{13.738}$$

<u>**Matrice orthogonale**</u> Nous considérons maintenant $T = U$ où U est une matrice orthogonale ($UU^t = 1$). Une matrice orthogonale est une isométrie[73] qui conserve donc la boule unité : $UB(0,1) = B(0,1)$. Nous avons

$$\lambda_N\big(B(0,1)\big) = \lambda_N\big(UB(0,1)\big) = C(U)\lambda_N\big(B(0,1)\big) \tag{13.739}$$

par conséquent $C(U) = 1$, et 1 est justement le déterminant de U.

73. Proposition 9.42.

Matrice quelconque Nous savons par le corolaire 12.32 de la décomposition polaire que toute matrice peut être écrite sous la forme $T = U_1 D U_2$ où U_i sont orthogonales et D est diagonale. Donc $C(T) = C(U_1)C(D)C(U_2) = \det(U_1)\det(D)\det(U_2) = \det(U_2 D U_2) = \det(T)$ parce que le déterminant est multiplicatif (proposition 9.10(1)).

□

Ce théorème donne une interprétation géométrique du déterminant en tant que facteur de dilatation des volumes lors de l'utilisation d'une application linéaire. Si T est une application linéaire quelconque,

$$\lambda_N\big(T(Q_0)\big) = |\det(T)|\lambda_N(Q_0) = |\det(T)|. \tag{13.740}$$

Le déterminant de T est le volume de l'image du cube unité par l'application T.

De la même façon, en utilisant l'application linéaire $T(x) = ax$ nous avons pour tout borélien B :

$$\lambda_N(aB) = a^N \lambda_N(B). \tag{13.741}$$

Une dilatation d'un facteur a des longueurs provoque une multiplication par a^N des volumes.

13.15.3 Le théorème et sa démonstration

Théorème 13.255 (Changement de variable[299, 300]).
Soient U et V des ouverts de $\mathbb{R}^N$ ainsi qu'un C^1-difféomorphisme $\phi\colon U \to V$. Nous notons J_ϕ la fonction

$$\begin{aligned} J_\phi\colon \mathbb{R}^N &\to \mathbb{R} \\ a &\mapsto \det(d\phi_a). \end{aligned} \tag{13.742}$$

Alors :

(1) Si $E \subset U$ est borélien, alors $\phi(E)$ est borélien et

$$\lambda_N\big(\phi(E)\big) = \int_E |J_\phi| d\lambda_N, \tag{13.743}$$

c'est-à-dire $\phi^{-1}(\lambda_N) = |J_\phi| \cdot \lambda_N$.

(2) Si $f\colon V \to [0, +\infty]$ est mesurable alors la fonction

$$(f \circ \phi) \times |J_\phi|\colon U \to [0, \infty] \tag{13.744}$$

l'est également et[74]

$$\int_V f d\lambda_N = \int_U (f \circ \phi)(x)|J_\phi(x)| d\lambda_N(x). \tag{13.745}$$

(3) Si $f\colon V \to \mathbb{C}$ est mesurable alors elle est intégrable si et seulement si $(f \circ \phi) \times |J_\phi|\colon U \to \mathbb{C}$ est intégrable. Si c'est le cas, alors nous avons encore la formule de changement de variables :

$$\int_V f d\lambda_N = \int_{\phi^{-1}(V)} (f \circ \phi)|J_\phi| d\lambda_N. \tag{13.746}$$

Démonstration. Attention : la preuve va être longue.

(1) Le fait que $\phi(E)$ soit borélien lorsque E l'est est la proposition 13.219. En ce qui concerne la formule annoncée, il faut travailler.

74. L'intégrabilité d'une fonction est la définition 13.168 qui stipule que l'intégrale de $|f(x)|$ est finie. L'égalité proposée a un sens si les deux membres sont infinis. Il n'y a donc pas d'hypothèses d'intégrabilité obligatoire pour écrire une intégrale lorsque la fonction a des valeurs positives.

Inégalité dans un sens (cubes) Nous commençons par prouver l'inégalité

$$\lambda_N\big(\phi(Q)\big) \leqslant \int_Q |J_\phi(x)|dx \tag{13.747}$$

pour tout $Q \in \mathcal{Q}$. On peut diviser le côté du cube Q en k éléments de longueurs égales. Le cube est alors divisé en k^N petits cubes d'intérieurs disjoints. Nous les nommons Q_i $(i = 1, \dots, k^N)$ Nous avons alors

$$\sum_i \lambda_N(Q_i) = \sum_i \lambda_N\big(\operatorname{Int}(Q_i)\big) = \lambda_N\big(\bigcup_i \operatorname{Int}(Q_i)\big) \leqslant \lambda_N(Q) \leqslant \sum_i \lambda_N(Q_i). \tag{13.748}$$

La dernière inégalité est le fait que les intersections ne sont pas disjointes. Toutes ces inégalités sont en réalité des égalités et en particulier : $\lambda_N(Q) = \sum_i \lambda_N(Q_i)$.
Soit $a \in Q_i$. Posons

$$\begin{aligned} \theta\colon U \quad & \to U \\ \theta = (d\phi_a)^{-1} \circ \phi \quad & \end{aligned} \tag{13.749}$$

Cela appelle deux commentaires. D'abord l'application $d\phi_a\colon U \to V$ est inversible parce que ϕ est un difféomorphisme (lemme 10.137). Ensuite, l'application θ est la composée de $(d\phi_a)$ (qui est linéaire) et de ϕ qui est de classe C^1 ; donc θ est de classe C^1. Donc le lemme 13.253 s'applique. La différentielle de θ n'est pas trop compliquée à écrire parce que nous avons la formule de différentielle d'une composée (théorème 10.135) et le fait que $(d\phi_a)^{-1}$ qui est linéaire et donc sa propre différentielle (lemme 10.131). Nous avons donc $d\theta = (d\phi_a)^{-1} \circ d\phi$, et le lemme donne

$$\lambda_N\left((d\phi_a)^{-1}\phi(a)\right) \leqslant \sup_{s\in Q_i} \|(d\phi_a)^{-1} \circ d\phi_s\|^N \lambda_N(Q_i) \tag{13.750}$$

Étant donné que $(d\phi_a)^{-1}$ est une application linéaire, la proposition 13.254 s'applique, et donc

$$\lambda_N\left((d\phi_a)^{-1}\phi(a)\right) = |\det(d\phi_a)^{-1}|\lambda_N\big(\phi(a)\big). \tag{13.751}$$

Le déterminant d'une application réciproque est donné par la proposition 9.10(4) :

$$\det\left((d\phi_a)^{-1}\right) = \frac{1}{\det\left(d\phi_a\right)} = \frac{1}{J_\phi(a)}. \tag{13.752}$$

Recollant les morceaux,

$$\lambda_N\big(\phi(Q_i)\big)\frac{1}{J_\phi(a)} \leqslant \sup_{s\in Q_i} \|(d\phi_a)^{-1} \circ d\phi_s\|^N \lambda_N(Q_), \tag{13.753}$$

ou encore :

$$\lambda_N\big(\phi(Q_i)\big) \leqslant |J_\phi(a)| \sup_{s\in Q_i} \|(d\phi_a)^{-1} \circ d\phi_s\|^N \lambda_N(Q_i). \tag{13.754}$$

Vu que a et s sont proches l'un de l'autre (on peut choisir encore la taille du cube), nous pouvons espérer que $(d\phi_a)^{-1}$ ne soit pas loin d'être l'inverse de $d\phi_s$. Et c'est en effet le cas. Pour s'en assurer, remarquons que l'application

$$d\phi\colon Q_i \to \mathcal{L}(\mathbb{R}^N, \mathbb{R}^N) \tag{13.755}$$

est continue et même uniformément continue parce que Q_i est compact. De plus la composition de différentielles étant un produit de matrices nous pouvons permuter la limite dans le calcul suivant :

$$\lim_{s\to a}(d\phi_a)^{-1} \circ d\phi_s = (d\phi_a)^{-1} \circ \lim_{s\to a} d\phi_s = \mathbb{1}. \tag{13.756}$$

Donc si $\epsilon > 0$ est donné, il existe δ tel que pour tout $s \in B(a, \delta)$, $\|(d\phi_a)^{-1} \circ d\phi_s - \mathbb{1}\| \leqslant \epsilon$. En ce qui concerne les normes, si $\|A - \mathbb{1}\| \leqslant \epsilon$ alors $\|A\| \leqslant \|A - \mathbb{1}\| + \|\mathbb{1}\| \leqslant \epsilon + 1$.

Cela étant dit, nous nous souvenons que nous avions découpé U en un nombre fini de cubes Q_i d'égales dimensions ; il suffit de prendre k suffisamment grand pour que la diagonale des cubes sot plus petite que le minimum des δ_i. Avec un tel découpage,

$$\sup_{s\in Q_i} \|(d\phi_a)^{-1}\circ d\phi_s\| \leqslant 1+\epsilon \tag{13.757}$$

et par conséquent

$$\lambda_N\big(\phi(Q_i)\big) \leqslant (1+\epsilon)^N |J_\phi(a_i)|\lambda_N(Q_i) \tag{13.758}$$

où nous avons ajouté un indice i au point a pour nous rappeler que nous avons choisi $a\in Q_i$.

Le théorème de la moyenne 13.234 appliqué à l'intégrale $\int_{Q_i}|J_\phi(t)|d\lambda_N(t)$ donne l'existence d'un $a_i\in Q_i$ tel que

$$|J_\phi(a_i)| = \frac{1}{\lambda_N(Q_i)}\int_{Q_i}|J_\phi|d\lambda_N. \tag{13.759}$$

Ce point a_i vérifie l'inégalité (13.758) comme tout point de Q_i. Nous sommons ces inégalités sur tous les i :

$$\lambda_N\big(\phi(Q)\big) \leqslant \sum_i \lambda_N\big(\phi(Q_i)\big) \tag{13.760a}$$

$$\leqslant (1+\epsilon^N \sum_i \left(\frac{1}{\lambda_N(Q_i)\int_{Q_i}|J_\phi|d\lambda_N}\right)\lambda_N(Q_i) \tag{13.760b}$$

$$= (1+\epsilon)^N\sum_i\int_{Q_i}|J_\phi|d\lambda_N \tag{13.760c}$$

$$= (1+\epsilon)^N\int_Q|J_\phi|d\lambda_N \tag{13.760d}$$

où nous avons utilisé le fait que $\mathbb{1}_Q=\sum_i\mathbb{1}_{Q_i}$ presque partout. En prenant le limite $\epsilon\to 0$ nous trouvons

$$\lambda_N\big(\phi(Q)\big) \leqslant \int_Q|J_\phi|d\lambda_N. \tag{13.761}$$

L'inégalité (13.747) est prouvée.

Inégalité pour les boréliens Soit B un borélien de U. Vu que U et V sont des ouverts de $\mathbb{R}^N$, les mesures de Lebesgue sur U et sur V sont les mêmes que celles sur $\mathbb{R}^n$ par le corolaire 13.51.

Par les définitions 13.189 et 13.77, les applications μ et n définies par $\mu=\phi^{-1}(\lambda_N)$ et $\nu=|J_\phi|\lambda_N$ sont des mesures positives sur U (de Borel, qui plus est). L'inégalité (13.747) à peine prouvée s'écrit $\mu(Q)\leqslant\nu(Q)$ pour tout cube Q. Le lemme 13.252 nous dit alors que l'inégalité tient pour tout borélien.

Inégalité dans l'autre sens En utilisant la notation de la mesure image et du produit d'une mesure par une fonction [75], nous pouvons écrire l'inégalité prouvée sous la forme $\phi^{-1}(\lambda_N)\leqslant|J_\phi|\lambda_N$. En inversant les rôles de U et V (et donc de ϕ et ϕ^{-1}) nous avons aussi

$$\phi(\lambda_N)\leqslant|J_{\phi^{-1}}|\lambda_N. \tag{13.762}$$

En y appliquant ϕ^{-1} et le lemme 13.77,

$$\lambda_N\leqslant\phi^{-1}\big(|J_{\phi^{-1}}|\lambda_N\big). \tag{13.763}$$

75. Définition 13.77 et 13.189

Nous prouvons à présent que $\phi^{-1}\big(|J_{\phi^{-1}}|\cdot\lambda_N\big)=\Big(|J_{\phi^{-1}}|\circ\phi\Big)\cdot\phi^{-1}(\lambda_N)$ en appliquant à un borélien B de U. D'une part

$$\phi^{-1}\big(|J_{\phi^{-1}}|\cdot\lambda_N\big)(B)=\big(|J_{\phi^{-1}}|\cdot\lambda_N\big)\phi(B) \tag{13.764a}$$
$$=\int_{\phi(B)}|J_{\phi^{-1}}|d\lambda_N, \tag{13.764b}$$

et d'autre part,

$$\big(|J_{\phi^{-1}}|\circ\phi\big)\cdot\phi^{-1}(\lambda_N)B=\int_{\mathbb{R}^N}\mathbb{1}_B(x)\big(|J_{\phi^{-1}}|\circ\phi\big)(x)d\big(\phi^{-1}(\lambda_N)\big)(x) \tag{13.765a}$$
$$=\int_{\mathbb{R}^N}\mathbb{1}_B\big(\phi^{-1}(x)\big)\big(|J_{\phi^{-1}}|\circ\phi\big)\big(\phi^{-1}(x)\big)d\lambda_N(x) \tag{13.765b}$$
$$=\int_{\mathbb{R}^N}\mathbb{1}_{\phi(B)}|J_{\phi^{-1}}| \tag{13.765c}$$
$$=\int_B|J_{\phi^{-1}}|d\lambda_N. \tag{13.765d}$$

Justification :

— Pour (13.765b), le théorème 13.196(2).

L'équation (13.763) devient alors

$$\lambda_N\leqslant\big(|J_{\phi^{-1}}|\circ\phi\big)\cdot\phi^{-1}(\lambda_N). \tag{13.766}$$

Nous allons faire le produit de cette mesure par $|J_\phi|$ en nous souvenant que $J_\phi(x)=\det\big(d\phi_x\big)$. Par le lemme 10.137 nous avons aussi $(d\phi_x)^{-1}=d\phi^{-1}_{\phi(x)}$ et donc, par la propriété 9.10(3) du déterminant,

$$J_\phi(x)=\frac{1}{\det\big(d\phi^{-1}_{\phi(x)}\big)}=\frac{1}{J_{\phi^{-1}}\big(\phi(x)\big)}. \tag{13.767}$$

Nous avons

$$|J_\phi|\cdot\lambda_N\leqslant|J_\phi|\cdot\big(|J_{\phi^{-1}}|\circ\phi\big)\cdot\phi^{-1}(\lambda_N). \tag{13.768}$$

En utilisant la proposition 13.192, il s'agit de multiplier la mesure $\phi^{-1}(\lambda_N)$ par la fonction

$$x\mapsto|J_\phi(x)J_{\phi^{-1}}\big(\phi(x)\big)|=1. \tag{13.769}$$

Nous avons donc bien

$$|J_\phi|\cdot\lambda_N\leqslant\phi^{-1}(\lambda_N), \tag{13.770}$$

et donc l'égalité

$$|J_\phi|\cdot\lambda_N=\phi^{-1}(\lambda_N), \tag{13.771}$$

c'est-à-dire le point (1).

(2) Le fait que la fonction proposée soit mesurable est le fait que la mesurabilité n'est pas affectée par produit et composition (propositions 13.93 et 13.43), et le fait que pour les mêmes raisons, l'application $J_\phi\colon U\to\mathbb{R}$ est également mesurable. En ce qui concerne la formule nous allons la démontrer dans le cas de fonctions de plus en plus générales.

Pour les fonctions indicatrices Soit B un borélien de U. Considérons la fonction $f=\mathbb{1}_{\phi(B)}$. Alors

$$\int_V fd\lambda_N=\int_{\mathbb{R}^N}\mathbb{1}_{\phi(B)}(y)\mathbb{1}_V(y)d\lambda_N(y)=\int_{\mathbb{R}^N}\mathbb{1}_{\phi(B)}d\lambda_N=\lambda_N\big(\phi(B)\big). \tag{13.772}$$

parce que $V = \phi(U)$ et $B \subset U$, donc $\mathbb{1}_{\phi(B)}\mathbb{1}_{\phi(U)} = \mathbb{1}_{\phi(B)}$. D'autre part, pour calculer l'autre membre de (13.745) nous remarquons que $f = \mathbb{1}_{\phi(B)} = \mathbb{1}_B \circ \phi^{-1}$, ce qui donne

$$\int_U f\big(\phi(x)\big)|J_\phi(x)|d\lambda_N(x) = \int_U \mathbb{1}_B|J_\phi|d\lambda_N = \int_B |J_\phi|d\lambda_N. \tag{13.773}$$

L'ensemble B étant borélien, il est extrêmement mesurable, ce qui fait que le point (1) s'applique : les expressions (13.772) et (13.773) sont égales.

Pour les fonctions étagées Soit $f\colon V \to \mathbb{R}^+$ une fonction étagée :

$$f(x) = \sum_{i=1}^n a_i \mathbb{1}_{A_i}(x) \tag{13.774}$$

Nous pouvons faire le calcul suivant :

$$\begin{aligned}
\int_V f d\lambda_N &= \int_V \sum_i a_i \mathbb{1}_{A_i} d\lambda_N && (13.775a)\\
&= \sum_i a_i \int_V \mathbb{1}_{A_i} d\lambda_N && (13.775b)\\
&= \sum_i \int_U (\mathbb{1}_{a_i} \circ \phi)(x)|J_\phi(x)|d\lambda_N(x) && (13.775c)\\
&= \sum_i a_i \int_U \mathbb{1}_{\phi^{-1}(A_i)}|J_\phi(x)|d\lambda_N(x) && (13.775d)\\
&= \int_V \underbrace{\sum_i a_i \mathbb{1}_{\phi^{-1}(A_i)}(x)}_{=(f\circ\phi)(x)} |J_\phi(x)|d\lambda_N(x) && (13.775e)\\
&= \int_V (f \circ \phi)|J_\phi|d\lambda_N. && (13.775f)
\end{aligned}$$

Justifications :

— Pour (13.775b) : linéarité de l'intégrale, théorème 13.165(2) [76]

— Pour (13.775c) : le cas des fonctions indicatrices est utilisé pour chaque i entre 1 et n.

Fonction mesurable positive Soit $f\colon V \to [0,\infty]$. Par le théorème fondamental d'approximation 13.105, il existe une suite croissante de fonctions étagées et mesurables $\varphi_n\colon V \to [0,\infty[$ dont la limite ponctuelle est f. Nous avons alors le calcul suivant :

$$\begin{aligned}
\int_V f d\lambda_N &= \lim_{n\to\infty} \int_V \varphi_n d\lambda_N && (13.776a)\\
&= \lim_{n\to\infty} \int_U (\varphi_n \circ \phi)|J_\phi|d\lambda_N && (13.776b)\\
&= \int_U \lim_{n\to\infty} (\varphi_n \circ \phi)|J_\phi|d\lambda_N && (13.776c)\\
&= \int_U (f \circ \phi)|J_\phi|d\lambda_N. && (13.776d)
\end{aligned}$$

Justifications :

— Pour (13.776a), c'est le théorème de la convergence monotone 13.160.

— Pour (13.776b), c'est le présent théorème pour la fonction étagée φ_n.

— Pour (13.776c), c'est encore la convergence dominée, justifiée par le fait que $\varphi_n \circ \phi$ est également une suite croissante : si $x \in U$ alors $\varphi_{n+1}\big(\phi(x)\big) \geqslant \varphi_n\big(\phi(x)\big)$.

76. Il est remarquable que nous n'utilisons cette linéarité que pour les fonctions étagées.

— Pour (13.776d), c'est la limite ponctuelle $\varphi_n\big(\phi(x)\big) \to f\big(\phi(x)\big)$.

(3) La partie sur l'intégrabilité repose sur le fait que $|f| \circ \phi = |f \circ \phi|$. Ici $|.|$ est le module et non une valeur absolue. Les faits suivants sont équivalents :

— la fonction $f\colon V \to \mathbb{C}$ est intégrable

— la fonction $|f|\colon V \to \mathbb{R}$ est intégtrable

— la fonction $(|f| \circ \phi)|J_\phi|\colon U \to \mathbb{R}$ est intégrable (par le point (2)).

— la fonction $(f \circ \phi)|J_\phi|\colon U \to \mathbb{R}$ est intégrable.

En ce qui concerne la formule, il s'agit seulement d'appliquer le point (2) aux parties positives, négatives, imaginaires et réelles de f.

□

Notons que la formule peut être écrite sous la forme

$$\langle f, g\rangle_V = \langle f \circ \phi, (g \circ \phi)|J|\rangle_U, \tag{13.777}$$

qui est plus pratique lorsqu'on parle de produits scalaires. Pour rappel, $\phi\colon U \to C$ est un C^1-difféomorphisme.

13.256.
La formule de changement de variables peut être comprise de la façon suivante. Si ϕ est linéaire alors le facteur $|J_\phi|$ est la mesure de l'image par ϕ d'une portion de $\mathbb{R}^p$ de mesure 1, sinon $|J_\phi|$ est le rapport entre la mesure de l'image d'un élément infinitésimale de volume de $\mathbb{R}^p$ et sa mesure originale.

Soit $\phi(u, v) = g(u, v)e_1 + h(u, v)e_2$ un difféomorphisme dans $\mathbb{R}^2$. Soit (x_0, y_0) l'image par ϕ de (u_0, v_0). On considère le petit rectangle R de sommets (u_0, v_0), $(u_0 + \Delta u, v_0)$, $(u_0 + \Delta u, v_0 + \Delta v)$ et $(u_0, v_0 + \Delta v)$. L'image de R n'est pas un rectangle en général, mais peut être bien approximée par le rectangle de sommets (x_0, y_0), $(x_0, y_0) + \phi_u\Delta u$, $(x_0, y_0) + \phi_u\Delta u + \phi_v\Delta v$ et $(x_0, y_0) + \phi_v\Delta v$ et son aire est $\|\phi_u \times \phi_v\|\Delta u\Delta v$. La valeur $|\phi_u \times \phi_v|$ est exactement $|J_\phi|$

13.15.4 Exemples

Un exemple avec une exponentielle est donnée dans l'exemple 17.62.

Énormément d'exemples sont disponibles avec les coordonnées polaires et toutes leurs variations. Cependant les fonctions trigonométriques ne seront vues que plus tard ; les coordonnées polaires, cylindrique et sphériques seront vues en section 17.12 et les exemples d'utilisation pour les intégrales seront dans la section 17.14.

13.16 Changement d'espace mesuré

Proposition 13.257 ([1]).
Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$. Soit un ensemble Ω' et une bijection $\varphi\colon \Omega \to \Omega'$. Nous posons

(1) $\mathcal{A}' = \varphi(\mathcal{A})$,

(2) $\mu'(B) = \mu\big(\varphi^{-1}(B)\big)$ pour tout $B \in \mathcal{A}'$.

Soit enfin une fonction mesurable $f\colon \Omega \to X$.

Alors

(1) Le triple $(\Omega', \mathcal{A}', \mu')$ est un espace mesuré.

(2) L'application $f \circ \varphi^{-1}\colon \Omega' \to X$ est mesurable.

(3) Nous avons l'égalité

$$\int_\Omega f d\mu = \int_{\Omega'} (f \circ \varphi^{-1}) d\mu'. \tag{13.778}$$

Démonstration. La proposition 13.52 montre déjà que $(\Omega', \mathcal{A}', \mu')$ est un espace mesuré.

Soit une partie S mesurable dans X. Alors $f^{-1}(S)$ est mesurable dans Ω par hypothèse sur f, c'est-à-dire que $f^{-1}(S) \in \mathcal{A}$. Ensuite $(\varphi \circ f^{-1})(S)$ est mesurable dans Ω' par hypothèse sur φ. Cela prouve que $f \circ \varphi^{-1}$ est une application mesurable.

Nous avons encore à prouver l'égalité d'intégrale. Par la définition 13.151 nous avons

$$\int_\Omega f d\mu = \sup\{\sum_{i=1}^n a_i \mu(A_i)\} \tag{13.779}$$

où le supremum est sur tous les n et tous les choix de $A_i \in \mathcal{A}$, $a_i \in \mathbb{R}^+$ tels que $f|_{A_i} > a_i$. Vu que $\mathcal{A}' = \varphi(\mathcal{A})$, si $A_i \in \mathcal{A}$ et a_i sont choisis, nous avons aussi

$$f \circ^{-1} |_{\varphi(A_i)} \geqslant a_i \tag{13.780}$$

avec $\varphi(A_i) \in \mathcal{A}'$. Donc pour un choix de $\{(A_i, a_i)\}$ donné,

$$\sum_{i=1}^n a_i \mu(A_i) = \sum_{i=1}^n a_i \mu'\big(\varphi(A_i)\big). \tag{13.781}$$

Au final,

$$\int_\Omega f d\mu = \sup\{\sum_{i=1}^n a_i \mu(A_i)\} = \sup\{\sum_i a_i \mu'\big(\varphi(A_i)\big)\} = \int_{\varphi(\Omega)} f \circ \varphi^{-1} d\mu'. \tag{13.782}$$

□

Remarque 13.258 (Ce n'est pas la mesure que nous voulons)**.**
La mesure donnée par la proposition 13.257 n'est pas celle que nous voulons d'habitude sur Ω'. Anticipons un peu pour comprendre. Prenons l'exemple de la partie C de $\mathbb{R}$ donnée par

$$C = \{(x, y) \in \mathbb{R}^2 \text{ tel que } y = x^2, x \in]0, 3[\}. \tag{13.783}$$

(1) La façon correcte de définir la longueur de C est de prendre une limite d'approximations par des morceaux de droites, comme fait à la définition 20.4.

(2) Cette définition de la longueur peut être exprimée sous forme intégrale par le théorème 20.10 qui nous assure que

$$l(C) = \int_0^3 \|\varphi'(t)\| dt = \int_0^3 \sqrt{1 + 4t^2} dt \neq \mu'(C). \tag{13.784}$$

En effet, $\mu'(C) = \mu\big(\varphi^{-1}(C)\big) = \mu\big(]0, 3[\big) = 3$, alors que pour tout t nous avons $\sqrt{1 + 4t^2} > 1$ et donc $l(C) > 3$.

(3) Donc μ' n'est pas exactement ce que nous aurions pu vouloir appeler la « mesure » de C.

(4) La mesure à considérer sur C doit donc plutôt être quelque chose comme le produit de la mesure μ' par la fonction $\|\varphi'\|$. Mais cela est une autre histoire qui vous sera contée une autre fois.

(5) Dans le cas de S^1, nous avons $\varphi(x) = e^{ix}$, et $\|\varphi'(x)\| = 1$. Donc la mesure donné ici est probablement bien celle que nous voulons. Peut-être à coefficient $\frac{1}{2\pi}$ près pour avoir une normalisation $\mu'(S^1) = 1$. Cela est également une autre histoire qui vous sera contée une autre fois ; par exemple dans la proposition 17.52.

13.17 Théorème de Fubini-Tonelli et de Fubini

Nous rappelons que $\mathbb{R}^n$ muni de la mesure de Lebesgue est un espace mesuré σ-fini, conformément à la définition 13.20.

Le théorème de Fubini-Tonelli parle de fonctions réelles et non complexes, et même positives. Le truc est que ce théorème va servir de base pour construire les autres. Si nous avons une fonction à valeurs complexes, elle se décompose en parties réelles et imaginaires qui elles-mêmes se décomposent en parties positives et négatives. Au final, les preuves pour $f\colon \Omega \to \mathbb{C}$ se ramènent à appliquer quatre fois le théorème pour $f\colon \Omega \to \bar{\mathbb{R}}^+$.

Théorème 13.259 (Fubini-Tonelli[280])**.**
Soient $(\Omega_i, \mathcal{A}_i, \mu_i)$ deux espaces mesurés σ-finis, et $(\Omega, \mathcal{A}, \mu)$ l'espace produit. Soit une fonction $f\colon \Omega_1 \times \Omega_2 \to \mathbb{R}$ une fonction mesurable et positive (valant éventuellement ∞ à certains endroits) Alors :

(1) Les fonction

$$F_1\colon x \mapsto \int_{\Omega_2} f(x,y)d\mu_2(y) \tag{13.785}$$

et

$$F_2\colon y \mapsto \int_{\Omega_1} f(x,y)d\mu_1(x) \tag{13.786}$$

sont mesurables.

(2) Toutes les intégrales imaginables existent et sont égales :

$$\int_{\Omega_1\times\Omega_2} f(x,y)d(\mu_1\otimes\mu_2)(x,y) = \int_{\Omega_1}\left[\int_{\Omega_2} f(x,y)d\mu_2(y)\right]d\mu_1(x) \tag{13.787a}$$

$$= \int_{\Omega_2}\left[\int_{\Omega_1} f(x,y)d\mu_1(x)\right]d\mu_2(y) \tag{13.787b}$$

où tous les membres de l'égalité valent éventuellement $+\infty$.

Démonstration. Commençons par prouver le théorème dans le cas d'une fonction caractéristique d'un ensemble mesurable : $f(x,y) = \mathbb{1}_A(x,y)$ pour un certain ensemble $A \subset \Omega_1 \times \Omega_2$. Dans ce cas,

$$F_1(x) = \int_{\Omega_2} \mathbb{1}_A(x,y)d\mu_2(y) = \int_{\Omega_2} \mathbb{1}_{A_1(y)}(x)d\mu_2(y) = \mu_2\big(A_1(x)\big), \tag{13.788}$$

et nous avons déjà vu au théorème 13.209 que cette fonction F_1 était alors mesurable. En utilisant maintenant les égalités (13.620) ainsi que le fait que $\mathbb{1}_A(x,y) = \mathbb{1}_{A_2(x)}(y)$ nous avons

$$\int_{\Omega_1\times\Omega_2} \mathbb{1}_A(x,y)d(\mu_1\otimes\mu_2)(x,y) = (\mu_1\otimes\mu_2)(A) \tag{13.789a}$$

$$= \int_{\Omega_1} \mu_2\big(A_2(x)\big)d\mu_1(x) \tag{13.789b}$$

$$= \int_{\Omega_1}\left[\int_{\Omega_2} \mathbb{1}_{A_2(x)}(y)d\mu_2(y)\right]d\mu_1(x) \tag{13.789c}$$

$$= \int_{\Omega_1}\left[\int_{\Omega_2} \mathbb{1}_{A}(x,y)d\mu_2(y)\right]d\mu_1(x). \tag{13.789d}$$

Le théorème étant valable pour les fonctions caractéristiques, il est valable pour les fonctions simples (définition 13.100) par linéarité de l'intégrale.

Si f n'est pas une fonction simple, alors la proposition 13.105 nous donne une suite croissante de fonctions simples et positives convergeant ponctuellement vers f. La partie du théorème sur les fonctions simples dit que pour chaque n l'intégrale

$$\int_{\Omega_1\times\Omega_2} f_n(x,y)d(\mu_1\otimes\mu_2)(x,y) \tag{13.790}$$

peut être décomposée comme il faut en suivant la formule (13.787). Il faut pouvoir permuter la limite et l'intégrale dans chacun de cas. D'abord le théorème de la convergence monotone 13.160 appliqué à l'espace $\Omega_1 \times \Omega_2$ dit que

$$\lim_{n\to\infty} \int_{\Omega_1\times\Omega_2} f_n(x,y)d(\mu_1 \otimes \mu_2)(x,y) = \int_{\Omega_1\times\Omega_2} f(x,y)d(\mu_1 \otimes \mu_2)(x,y). \tag{13.791}$$

Ensuite, pour chaque $x \in \Omega_1$, les fonctions

$$\sigma_n(y) = \int_{\Omega_1} f_n(x,y)d\mu_1(x) \tag{13.792}$$

forment une suite croissante de fonctions mesurables ; nous leur appliquons encore le théorème de la convergence monotone :

$$\lim_{n\to\infty} \int_{\Omega_2} \left[\int_{\Omega_1} f_n(x,y)d\mu_1(x)\right] d\mu_2(y) = \lim_{n\to\infty} \int_{\Omega_2} \sigma_n(y)d\mu_2(y) \tag{13.793a}$$

$$= \int_{\Omega_2} \left[\lim_{n\to\infty} \int_{\Omega_1} f_n(x,y)d\mu_1(x)\right] d\mu_2(y) \tag{13.793b}$$

$$= \int_{\Omega_2} \left[\int_{\Omega_1} f(x,y)d\mu_1(x)\right] d\mu_2(y) \tag{13.793c}$$

où nous avons utilisé une seconde fois Beppo-Levi. □

Remarque 13.260.
Les formules (13.787) sont bien, mais ne garantissent en aucun cas que $f \in L^1(\Omega_1 \times \Omega_2)$: il faut encore que ces intégrales soient finies.

Corolaire 13.261 ([298])**.**
Soient $(\Omega_i, \mathcal{A}_i, \mu_i)$ deux espaces mesurés σ-finis, et $(\Omega, \mathcal{A}, \mu)$ l'espace produit[77]*. Soit une fonction mesurable $f\colon \Omega \to \mathbb{R}$ ou $\mathbb{C}$. Alors les conditions suivantes sont équivalentes*

(1) $f \in L^1(\Omega_1 \times \Omega_2)$,

(2)

$$\int_{\Omega_1} \left[\int_{\Omega_2} |f| d\mu_2\right] d\mu_1 < \infty, \tag{13.794}$$

(3)

$$\int_{\Omega_2} \left[\int_{\Omega_1} |f| d\mu_1\right] d\mu_2 < \infty. \tag{13.795}$$

Démonstration. Nous commençons par supposer que f est à valeurs dans $\mathbb{R}$. La notation $|f|$, pour l'instant, dénote donc bien la valeur absolue et non le module.

La fonction $|f|$ est mesurable et positive par hypothèse et par le fait que si f est mesurable, alors $|f|$ l'est également par le corolaire 13.96. Le théorème 13.259(2) nous dit alors que les intégrales suivantes existent et sont égales :

$$\int_{\Omega_1\times\Omega_2} |f| d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \left[\int_{\Omega_2} |f(x,y)| d\mu_2(y)\right] d\mu_1(x) = \int_{\Omega_2} \left[\int_{\Omega_1} |f(x,y)| d\mu_1(x)\right] d\mu_2(y). \tag{13.796}$$

Attention : rien ne dit encore que ces intégrales sont finies.

(1) implique (2) et (3) Si $f \in L^1(\Omega_1 \times \Omega_2)$ alors $|f|$ y est également. Cela implique que le membre de droite de (13.796) est fini. Les deux autres sont alors également finis.

(2) ou (3) implique (1) Les expressions à droite de (13.796) sont finies. Donc celle de gauche également. Cele signifie que $|f| \in L^1(\Omega_1 \times \Omega_2)$. Par conséquent f est également dans $L^1(\Omega_2 \times \Omega_2)$.

77. Définition 13.211.

Nous passons maintenant au cas où f est à valeurs dans $\mathbb{C}$. Nous décomposons

$$f = f_R + if_I \tag{13.797}$$

où f_R et f_I sont des fonctions réelles. Nous avons

$$\int_\Omega |f| \leqslant \int_\Omega |f_R| + \int_\Omega |f_I|. \tag{13.798}$$

Donc si f_R et f_I sont dans $L^1(\Omega)$, la fonction f le sera aussi. De même,

$$\int_\Omega |f_R| \leqslant \int_\Omega |f|, \tag{13.799}$$

qui donne l'inverse : si $f \in L^1(\Omega)$ alors $f_R, f_I \in L^1(\Omega)$. Bref, f est intégrable sur Ω si et seulement si f_R et f_I le sont.

Supposons que $f \in L^1(\Omega_1 \times \Omega_2)$. Alors

$$\int_{\Omega_1} \left[\int_{\Omega_2} |f| d\mu_2 \right] d\mu_1 \leqslant \int_{\Omega_1} \left[\int_{\Omega_2} |f_R| \right] + \int_{\Omega_1} \left[\int_{\Omega_2} |f_I| \right] < \infty \tag{13.800a}$$

où nous avons appliqué (1) implique (2) aux fonctions f_R et f_I qui sont dans $L^1(\Omega_1 \times \Omega_2)$ parce que f y est.

Dans l'autre sens, si

$$\int_{\Omega_1} \left[\int_{\Omega_2} |f| \right] < \infty, \tag{13.801}$$

alors en remplaçant $|f|$ par $|f_R|$ ou par $|f_I|$ nous restons fini. En appliquant alors « (2) implique (1) » nous trouvons que f_R et f_I sont dans $L^1(\Omega_1 \times \Omega_2)$. Et cela implique que $f \in L^1(\Omega_1 \times \Omega_2)$. □

Théorème 13.262 (Fubini[298]).
Soient $(\Omega_i, \mathcal{A}_i, \mu_i)$ deux espaces mesurés σ-finis, et $(\Omega, \mathcal{A}, \mu)$ l'espace produit. Soit

$$f \in L^1\big((\Omega, \mathcal{A}), \mathbb{C}\big), \tag{13.802}$$

c'est-à-dire une fonction à valeurs mesurable et intégrable sur Ω. Alors :

(1) Pour presque tout $x \in \Omega_1$, la fonction $y \mapsto f(x, y)$ est $L^1(\Omega_2)$.

(2) Si nous posons

$$\varphi_f(x) = \int_{\Omega_2} f(x, y) d\mu_2(y); \tag{13.803}$$

alors $\varphi_f \in L^1(\Omega_1)$.

(3) Nous avons la formule d'inversion d'intégrale

$$\int_\Omega f d(\mu_1 \otimes \mu_2) = \int_{\Omega_1} \varphi_f d\mu_1 \tag{13.804a}$$

$$= \int_{\Omega_1} \left[\int_{\Omega_2} f(x, y) d\mu_2(y) \right] d\mu_1(x) \tag{13.804b}$$

$$= \int_{\Omega_2} \left[\int_{\Omega_1} f(x, y) d\mu_1(x) \right] d\mu_2(y). \tag{13.804c}$$

Démonstration. Nous commençons par supposer que f est à valeurs réelles : $f \in L^1\big((\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2), \mathbb{R}\big)$. Nous décomposons la fonction f en parties positives et négatives : $f = f^+ - f^-$ avec f^+ et f^- positives ou nulles. Nous avons évidemment

$$\int_{\Omega_1 \times \Omega_2} |f^+| \leqslant \int_{\Omega_1 \times \Omega_2} |f| < \infty. \tag{13.805}$$

Donc f^+ et f^- sont des éléments de $L^1(\Omega_1 \times \Omega_2)$.

Pour (1) Nous posons

$$\varphi_{f^+}(x) = \int_{\Omega_2} f^+(x,y)d\mu_2(y) \tag{13.806}$$

pour tous les $x \in \Omega_1$ pour lesquels cette intégrale est bien définie. Vu que f^+ est positive et mesurable, le théorème de Fubini-Tonelli 13.259(1) s'applique donc pour nous dire que φ_{f^+} est mesurable.

De plus le résultat (13.787) appliqué à f^+ donne

$$\int_{\Omega_1} \varphi_{f^+} d\mu_1 = \int_{\Omega_1 \times \Omega_2} f^+ d(\mu_1 \otimes \mu_2) < \infty. \tag{13.807}$$

Le fait que le tout soit fini est une conséquence du fait déjà mentionné que $f^+ \in L^1(\Omega_1 \times \Omega_2$. Vu que φ_{f^+} est une fonction positive, l'inégalité (13.807) signifie que $\varphi_{f^+} \in L^1(\Omega_1, \mu_1)$.

En particulier, $\varphi_{f^+}(x) < \infty$ pour presque tout $x \in \Omega_1$. C'est-à-dire pour presque tout $x \in \Omega_1$:

$$\int_{\Omega_2} f^+(x,y)d\mu_2(y) < \infty, \tag{13.808}$$

et sachant que $f^+ \geqslant 0$ nous avons $f^+(x, \ \cdot \ ) \in L^1(\Omega_2)$ pour presque tout x.

Pour (2) Partout où φ_{f^+} et φ_{f^-} sont finies nous avons

$$\varphi_f = \varphi_{f^+} - \varphi_{f^-}, \tag{13.809}$$

et comme cela a lieu presque partout, nous pouvons considérer une partie mesurable $A \subset \Omega_1$ telle que $\mu_1(A) = 0$ et $\varphi_f(x) = \varphi_{f^+}(x) - \varphi_{f^-}(x)$ pour tout x hors de A. Bref, nous posons

$$g(x) = \begin{cases} \varphi_{f^+} - \varphi_{f^-}(x) & \text{si } x \in A^c \\ 0 & \text{si } x \in A. \end{cases} \tag{13.810}$$

Cette fonction g est mesurable et $g = \varphi_f$ presque partout. De plus

$$\int_{\Omega_1} |g| d\mu_1 = \int_{A^c} |g| \leqslant \int_{A^c} \varphi_{f^+} + \int_{A^c} \varphi_{f^-} < \infty. \tag{13.811}$$

La dernière inégalité est le fait que $\varphi_{f^\pm}$ sont dans $L^1(\Omega_1)$. Et notons au passage que nous aurions pu laisser toutes les intégrales sur Ω_1 sans faire de précisions sur la distinction entre Ω_1 et A^c parce que la partie de Ω_1 sur laquelle $\varphi_{f^\pm}$ sont infinies est trop petite pour changer la valeur de l'intégrale.

Nous avons donc $g \in L^1(\Omega_1)$, et par conséquent également $\varphi_f \in L^1(\Omega_1)$ parce que ces deux fonctions sont égales presque partout (les classes sont égales).

Pour (3) En utilisant l'équation (13.807) nous avons

$$\int_{\Omega_1} \varphi_f d\mu_1 = \int g d\mu_1 = \int_{\Omega_1} \varphi_{f^+} - \int_{\Omega_1} \varphi_{f^-} \tag{13.812a}$$

$$= \int_{\Omega_1 \times} f^+ d\mu - \int_{\Omega_1 \times \Omega_2} f^- d\mu \tag{13.812b}$$

$$= \int_{\Omega_1 \times \Omega 2} f d\mu. \tag{13.812c}$$

Et toutes ces intégrales sont finies.

Et c'est maintenant que nous considérons le cas complexe. Nous décomposons $f = f_R + if_I$ avec des fonctions réelles f_R et f_I. Comme déjà mentionné autour de (13.798), les fonctions f_R et f_I sont intégrables. Nous leur appliquons le théorème.

Les valeurs de x pour lesquelles $f_R(x, \cdot)$ et $f_I(x, \cdot)$ ne sont pas dans $L^1(\Omega_2)$ forment un ensemble de mesure nulle, nommons le A. En posant

$$g(x,y) = \begin{cases} f_R(x,y) + if_I(x,y) & \text{si } x \in A^c \\ 0 & \text{si } x \in A, \end{cases} \tag{13.813}$$

nous avons que $g(x, \cdot)$ est intégrable pour tout $x \in A^c$. Vu que pour ces valeurs de x nous avons $g(x,y) = f(x,y)$ nous en déduisons que pour $x \in A^c$ nous avons aussi $f(x, \cdot) \in L^1(\Omega_2)$.

Les autres points se traitent de la même façon[78]. □

13.263.
En pratique, il n'est pas toujours évident qu'une fonction soit intégrable sur $\Omega_1 \times \Omega_2$. Pour permuter des intégrales sur une fonction à deux paramètres nous faisons comme suit.

(1) Nous testons l'intégrabilité en chaine de $|f|$, et si c'est bon, le corolaire 13.261 nous donne $f \in L^1(\Omega_1 \times \Omega_2)$.

(2) Nous utilisons le théorème de Fubini 13.262 pour séparer et permuter les intégrales comme des ingénieurs.

Si la fonction $(x,y) \mapsto f(x)g(y)$ satisfait aux hypothèse du théorème de Fubini alors

$$\int_{\Omega_1 \times \Omega_2} f(x)g(y)dx \otimes dy = \left(\int_{\Omega_1} f(x)dx\right)\left(\int_{\Omega_2} g(y)dy\right). \tag{13.814}$$

Le théorème de Fubini est souvent utilisé sous cette forme.

Exemple 13.264 (Nécessité d'avoir des mesures σ-finies)**.**
Nous montrons que le théorème ne tient pas si une des deux mesures n'est pas σ-finie. Soit $I = [0,1]$. Nous considérons l'espace mesuré

$$(I, \mathcal{B}\text{or}(I), \lambda) \tag{13.815}$$

où $\mathcal{B}\text{or}(I)$ est la tribu des boréliens sur I et λ est la mesure de Lebesgue (qui est σ-finie). D'autre part nous considérons l'espace mesuré

$$(I, \mathcal{P}(I), m) \tag{13.816}$$

où $\mathcal{P}(I)$ est l'ensemble des parties de I et m est la mesure de comptage. Cette dernière n'est pas σ-finie parce que les seuls ensembles de mesure finie pour la mesure de comptage sont des ensembles finis, or une union dénombrable d'ensemble finis ne peut pas recouvrir l'intervalle I.

Nous allons montrer que dans ce cadre, l'intégrale de la fonction indicatrice de la diagonale sur I^2 ne vérifie pas le théorème de Fubini. Étant donné que $\mathcal{B}\text{or}(I) \subset \mathcal{P}(I)$ nous avons

$$\mathcal{B}\text{or}(I^2) \subset \mathcal{B}\text{or}(I) \otimes \mathcal{P}(I). \tag{13.817}$$

Soit $\Delta = \{(x,x) \text{ tel que } x \in I\}$. La fonction

$$\begin{aligned} g\colon I^2 &\to \mathbb{R} \\ (x,y) &\mapsto x - y \end{aligned} \tag{13.818}$$

est continue et $\Delta = g^{-1}(\{0\})$ est donc fermé dans I^2. L'ensemble Δ est donc un borélien de I^2 et par conséquent un élément de la tribu $\mathcal{B}\text{or}(I) \otimes \mathcal{P}(I)$. La fonction indicatrice $\mathbb{1}_\Delta$ est alors mesurable pour l'espace mesuré

$$(I \times I, \mathcal{B}\text{or}(I) \otimes \mathcal{P}(I), \lambda \otimes m). \tag{13.819}$$

Pour x fixé nous avons

$$\mathbb{1}_\Delta(x,y) = \begin{cases} 1 & \text{si } y = x \\ 1 & \text{si } y \neq x \end{cases} = \mathbb{1}_{\{x\}}(y), \tag{13.820}$$

78. Attention : je n'ai pas vérifié explicitement. C'est juste une intuition. Vérifiez et écrivez-moi pour dire si c'est bon ou non.

et donc

$$A_1 = \int_I \left(\int_I \mathbb{1}_\Delta(x,y) dm(y) \right) d\lambda(x) \tag{13.821a}$$
$$= \int_I \left(\int_I \mathbb{1}_{\{x\}}(y) dm(y) \right) d\lambda(x) \tag{13.821b}$$
$$= \int_I \Big(m(\{x\}) \Big) d\lambda(x) \tag{13.821c}$$
$$= \int_I 1 d\lambda(x) \tag{13.821d}$$
$$= 1. \tag{13.821e}$$

Par contre le support de $\mathbb{1}_\Delta$ étant de mesure nulle pour la mesure de Lebesgue, nous avons

$$\int_I \mathbb{1}_\Delta(x,y) d\lambda(x) = 0 \tag{13.822}$$

et par conséquent

$$A_2 = \int_I \left(\int_I \mathbb{1}_\Delta(x,y) d\lambda(x) \right) dm(y) = 0. \tag{13.823}$$

Nous voyons donc que le théorème de Fubini ne s'applique pas. △

Exemple 13.265.
Nous nous proposons de calculer l'intégrale suivante en utilisant le théorème de Fubini :

$$G = \int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi} \tag{13.824}$$

alors que la fonction $x \mapsto e^{-x^2}$ n'a pas de primitives parmi les fonctions élémentaires.

Nous allons le faire de deux façons. Une première directe en utilisant le théorème de Fubini sur un domaine non borné, et une seconde en utilisant Fubini sur un domaine borné, et en passant à la limite ensuite.

Fubini, domaine non borné Par symétrie nous pouvons nous contenter de calculer

$$G_+ = \int_0^\infty e^{-x^2} dx. \tag{13.825}$$

L'astuce est de passer par l'intermédiaire

$$H = \int_{\mathbb{R}^+ \times \mathbb{R}^+} e^{-(x^2+y^2)} dx dy \tag{13.826a}$$
$$= \int_{\mathbb{R}^+} \left(\int_{\mathbb{R}^+} e^{-x^2} e^{-y^2} dx \right) dy \tag{13.826b}$$
$$= \left(\int_{\mathbb{R}^+} e^{-x^2} dx \right)^2 \tag{13.826c}$$
$$= G_+^2 \tag{13.826d}$$

L'intégrale (13.826a) se calcule en passant aux coordonnées polaires et le résultat est $H = \frac{\pi}{4}$. Nous avons alors $G = \frac{\sqrt{\pi}}{2}$ et

$$\int_{\mathbb{R}} e^{-x^2} = \sqrt{\pi}. \tag{13.827}$$

Fubini, domaine borné, puis limite Une variante, qui n'applique pas Fubini sur un domaine non borné. Nous commençons par écrire

$$I = \int_{-\infty}^{+\infty} e^{-x^2} dx := \lim_{R \to +\infty} \int_{-R}^{+R} e^{-x^2} dx \tag{13.828}$$

et puis nous faisons le calcul

$$\begin{aligned} I^2 &= \lim_{R\to+\infty}\left((\int_{-R}^{+R} e^{-x^2}dx)(\int_{-R}^{+R} e^{-y^2}dy)\right) \\ &= \lim_{R\to+\infty}\left(\iint_{K_R} e^{-(x^2+y^2)}dxdy\right) \\ &= \lim_{R\to+\infty}\left(\iint_{C_R} e^{-(x^2+y^2)}dxdy\right) \end{aligned} \tag{13.829}$$

où K est le carré de demi-côté R centré à l'origine et de côtés parallèles aux axes et C_R est le cercle de rayon R centré à l'origine.

La première étape à justifier est simplement l'application de Fubini. Pour le passage de l'intégrale du carré vers le cercle, définissons

$$I_K(r) = \int_{K_r} f, \quad I_C(r) = \int_{C_r} f \tag{13.830}$$

où K_r est la carré de demi-côté r et C_r est le cercle de rayon r. Le demi-côté du carré inscrit à C_r est $\sqrt{2}$, donc pour tout r nous avons

$$I_K(\sqrt{2}r) \leqslant I_C(r) < I_K(r), \tag{13.831}$$

et en prenant la limite, nous avons évidement

$$\lim_{r\to\infty} I_K(\sqrt{2}r) = \lim_{r\to\infty} I_K(r), \tag{13.832}$$

et donc cette limite est également égale à $\lim_{r\to\infty} I_C(t)$.

Il ne reste qu'à calculer la dernière intégrale sur le cercle en passant aux coordonnées polaires :

$$\int_{C_R} e^{-(x^2+y^2)}dxdy = \int_0^{2\pi} d\theta \int_0^R re^{-r^2}dr = \pi(1-e^{-R^2}). \tag{13.833}$$

La limite donne π, nous en déduisons que

$$\int_{-\infty}^{\infty} e^{-x^2}dx = \sqrt{\pi}. \tag{13.834}$$

△

Le théorème de Fubini-Tonelli nous permet également d'inverser des sommes et des séries. En effet une somme n'est rien d'autre qu'une intégrale pour la mesure de comptage :

$$\sum_{n=0}^{\infty} a_n = \int_{\mathbb{N}} a_n dm(n). \tag{13.835}$$

La proposition suivante montre comment il faut faire.

Proposition 13.266.
Soient les espaces mesurés $(\mathbb{N}, \mathcal{P}(\mathbb{N}), m)$, $(\mathbb{R}^n, \mathcal{B}or(\mathbb{R}^n), \lambda)$ *où* λ *est la mesure de Lebesgue ainsi qu'une suite de fonctions positives* $f_n\colon \mathbb{R}^d \to \mathbb{R}$. *Nous supposons de plus que la fonction* f_n *soit intégrable pour tout* n *et que les résultats forment une suite sommable. Alors*

$$\sum_{n=0}^{\infty}\int_{\mathbb{R}^n} f_n(x)dx = \int_{\mathbb{R}^d}\sum_{n\in\mathbb{N}} f_n(x)dx. \tag{13.836}$$

Démonstration. Nous pouvons la récrire le membre de gauche sous la forme

$$\int_{\mathbb{N}} \left(\int_{\mathbb{R}^n} f(n,x)dx \right) dm(n) \tag{13.837}$$

avec la notation évidente $f(n,x) = f_n(x)$. Prouvons que la fonction $f\colon \mathbb{N} \times \mathbb{R}^d \to \mathbb{R}$ ainsi définie est une fonction mesurable pour l'espace mesuré

$$\big(\mathbb{N} \times \mathbb{R}^d, \mathcal{P}(\mathbb{N}) \otimes \mathcal{B}\mathrm{or}(\mathbb{R}^d), m \otimes \lambda\big). \tag{13.838}$$

Si $A \subset \mathbb{R}$, nous avons

$$f^{-1}(A) = \bigcup_{n \in \mathbb{N}} \{n\} \times f_n^{-1}(A). \tag{13.839}$$

Chacun des ensembles dans l'union appartient à la tribu $\mathcal{P}(\mathbb{N}) \times \mathcal{B}\mathrm{or}(\mathbb{R}^d)$ tandis que les tribus sont stables sous les unions dénombrables. La fonction f est donc mesurable. Comme nous avons supposé que f était positive, le théorème de Fubini-Tonelli s'applique et nous avons

$$\int_{\mathbb{R}^d} \left(\int_{\mathbb{N}} f(n,x)dm(n) \right) dx = \int_{\mathbb{R}^d} \sum_{n \in \mathbb{N}} f_n(x)dx. \tag{13.840}$$

□

Théorème 13.267 (Fubini)**.**
Soit $(x,t) \mapsto f(x,y) \in \bar{\mathbb{R}}$ une fonction intégrable sur $B_n \times B_m \subset \mathbb{R}^{n+m}$ où B_n et B_m sont des ensembles mesurables de $\mathbb{R}^n$ et $\mathbb{R}^m$. Alors :

(1) pour tout $x \in B_n$, sauf éventuellement en les points d'un ensemble $G \subset B_n$ de mesure nulle, la fonction $y \in B_m \mapsto f(x,y) \in \bar{\mathbb{R}}$ est intégrable sur B_m

(2) la fonction

$$\begin{aligned} B_n \backslash G &\to \mathbb{R} \\ x &\mapsto \int_{B_n} f(x,y)dy \end{aligned} \tag{13.841}$$

est intégrable sur $B_n \backslash G$.

(3) On a

$$\int_{B_n \times B_m} f(x,y)dxdy = \int_{B_n} \left(\int_{B_m} f(x,y)dy \right) dx. \tag{13.842}$$

Notons en particulier que si $f(x,y) = \varphi(x)\phi(y)$, alors $\int_{B_m} \varphi(y)dy$ est une constante qui peut sortir de l'intégrale sur B_n, et donc

$$\int_{B_n \times B_m} \varphi(x)\phi(y)dxdy = \int_{B_n} \varphi(x)dx \int_{B_m} \phi(y)dy. \tag{13.843}$$

Chapitre 14

Suites et séries de fonctions

Les généralités sur les suites et séries de fonctions, c'est dans la section 11.30.

14.1 Séries de fonctions

14.1.1 Intégration de séries de fonctions

Théorème 14.1.
La somme uniforme de fonctions intégrables sur un ensemble de mesure fini est intégrable et on peut permuter la somme et l'intégrale.

En d'autres termes, supposons que $\sum_{n=0}^{\infty} f_n$ converge uniformément vers F sur A avec $\mu(A) < \infty$. Si F et f_n sont des fonctions intégrables sur A alors

$$\int_A F(x)d\mu(x) = \sum_{n=0}^{\infty} \int_A f_n(x)d\mu(x). \tag{14.1}$$

Démonstration. Ce théorème est une conséquence du théorème 13.182. En effet nous définissons la suite des sommes partielles

$$F_N = \sum_{n=0}^{N} f_n. \tag{14.2}$$

La limite $\lim_{N\to\infty} F_N = F$ est uniforme. Par conséquent la fonction F est intégrable et

$$\int_A F = \lim_{N\to\infty} \int_A F_N = \lim_{N\to\infty} \int_A \sum_{n=0}^{N} f_n = \lim_{N\to\infty} \sum_{n=0}^{N} \int_A f_n = \sum_{n=0}^{\infty} \int_A f_n. \tag{14.3}$$

La première égalité est le théorème 13.182, les autres sont de simples manipulations rhétoriques. □

Le théorème suivant est une paraphrase du théorème de la convergence dominée de Lebesgue (13.184).

Théorème 14.2.
Soient des fonctions $(f_n)_{n\in\mathbb{N}}$ telles que $\sum_{n=0}^{N} f_n$ soit intégrable sur $(\Omega, \mathcal{A}, \mu)$ pour chaque N. Nous supposons que la somme converge simplement vers

$$f(x) = \sum_{n=0}^{\infty} f_n(x) \tag{14.4}$$

et qu'il existe une fonction g telle que

$$\left| \sum_{n=0}^{N} f_n \right| < g \tag{14.5}$$

pour tout $N \in \mathbb{N}$. Alors

(1) $\sum_{n=0}^{\infty} f_n$ *est intégrable,*

(2) on peut permuter somme et intégrale :

$$\lim_{N\to\infty}\int_\Omega \sum_{n=0}^{N} f_n d\mu = \int_\Omega \sum_{n=0}^{\infty} f_n, \tag{14.6}$$

(3)

$$\lim_{N\to\infty}\int_\Omega \left|\sum_{n=0}^{N} f_n - \sum_{n=0}^{\infty} f_n\right| = \lim_{N\to\infty}\int_\Omega \left|\sum_{n=N}^{\infty} f_n\right| = 0. \tag{14.7}$$

Théorème 14.3.
Soit f_n *des fonctions* $C^1[a,b]$ *telles que*

(1) la série $\sum_n f_n(x_0)$ *converge pour un certain* $x_0 \in [a,b]$*,*

(2) la série des dérivées $\sum_n f_n'$ *converge uniformément sur* $[a,b]$*.*

Alors la série $\sum_n f_n$ *converge vers une fonction* F *et*

(1) La convergence est uniforme sur $[a,b]$*.*

(2) La fonction F *est dérivable*

(3) $F'(x) = \sum_n f_n'(x)$*.*

14.1.2 Différentiabilité

Lemme 14.4.
Soient E *et* F *deux espaces vectoriels normés. Si la suite* (T_n) *converge vers* T *dans* $\mathcal{L}(E,F)$*, alors pour tout* $v \in E$ *nous avons*

$$\left(\sum_{n=0}^{\infty} T_n\right)(v) = \sum_{n=0}^{\infty} T_n(v). \tag{14.8}$$

Problèmes et choses à faire

À mon avis si on a un ouvert connexe par arcs dans un espace vectoriel normé, alors il est connexe par arcs de classe C^1, c'est-à-dire que deux points peuvent être liés par un chemin de classe C^1.

Je n'en suis pas certain.

Si vous êtes sûr de vous, vous pouvez affaiblir les hypothèses du théorème 14.6 et supprimer la définition 14.5 qui ne sert à rien d'autre.

Définition 14.5.
Soit un espace vectoriel normé E*. Un ouvert* Ω *est dit connexe par arcs de classe* C^1 *si pour tout choix de* $a, b \in \Omega$*, il existe une application* $\gamma\colon [0,1] \to \Omega$ *de classe* C^1 *telle que* $\gamma(0) = a$ *et* $\gamma(1) = b$*.*

Théorème 14.6 ([302])**.**
Soient E *et* F *deux espaces vectoriels normés,* Ω *un ouvert connexe par arcs de classe* C^1 *de* E*. Soit* (u_n) *une suite de fonctions* $u_n\colon \Omega \to F$ *telle que*

(1) pour tout n*, la fonction* u_n *est de classe* C^1 *sur* Ω*,*

(2) la série $\sum_n u_n$ *converge simplement sur* Ω*,*

(3) la série des différentielles $\sum_n (du_n)$ *converge normalement sur tout compact de* Ω*.*

Alors la somme $u = \sum_n u_n$ *est de classe* C^1 *sur* Ω *et sa différentielle est donnée par*

$$du = \sum_{n=0}^{\infty} du_n. \tag{14.9}$$

Démonstration. Pour chaque n, la fonction $du_n\colon \Omega \to \mathcal{L}(E,F)$ est une fonction continue parce que u_n est de classe C^1. La série convergeant normalement, la fonction $\sum_{n=0}^{\infty} du_n$ est également continue par la proposition 11.306. La difficulté de ce théorème est donc de prouver que cela est bien la différentielle de la fonction $\sum_n u_n$, c'est-à-dire que

$$d\left(\sum_{n=0}^{\infty} u_n\right) = \sum_{n=0}^{\infty} du_n. \tag{14.10}$$

Soient $a, x \in \Omega$. Nous considérerions bien le segment $[a, x]$, mais vu que Ω n'est supposé que connexe par arcs de classe C^1 (définition 14.5), nous ne pouvons pas faire mieux pour joindre a à x que choisir un chemin de classe C^1

$$\gamma\colon [0,1] \to \Omega \tag{14.11}$$

tel que $\gamma(0) = a$ et $\gamma(1) = b$.

L'astuce est de poser

$$\begin{aligned} f_n\colon [0,1] &\to \mathbb{R} \\ t &\mapsto (du_n)_{\gamma(t)}\big(\gamma'(t)\big), \end{aligned} \tag{14.12}$$

et d'en étudier l'intégrale [1].

Permuter somme et intégrale Nous voudrions permuter la somme et l'intégrale dans l'expression $\int_0^1 \sum_i f_i(t)dt$. Pour cela nous commençons par regarder quelques majorations de normes. D'abord γ est de classe C^1, ce qui fait que γ' est continue. Vu que la norme est une application continue, la fonction $t \mapsto \|\gamma'(t)\|$ est également continue sur le compact $[0,1]$. Elle est donc majorée par une constante que nous nommons M. C'est le théorème de Weierstrass 7.105.

Ensuite nous avons le calcul

$$\|f_i(t)\| = \|(du_i)_{\gamma(t)}\big(\gamma'(t)\big)\| \leqslant \|(du_i)_{\gamma(t)}\|\|\gamma'(t)\| \leqslant M\|du_i\|_\infty < \infty. \tag{14.13}$$

Justifications :

— Pour la première inégalité. C'est le lemme 10.10.

— Pour la seconde inégalité. Il s'agit de l'inégalité évidente

$$\|du_i\|_\infty = \sup_{x\in\gamma\big([0,1]\big)} \|(du_i)_x\| \tag{14.14}$$

Notons que la norme $\|.\|_\infty$ ne réfère pas à un supremum sur E, mais seulement sur l'image de γ. Nous aurions pu faire preuve d'un peu de créativité dans les notations.

— L'application du_i est continue sur le compact $\gamma\big([0,1]\big)$. Donc le supremum est fini et atteint.

Maintenant nous posons

$$g_n(t) = \sum_{i=0}^n f_i(t). \tag{14.15}$$

Nous avons la majoration

$$\|g_n(t)\| \leqslant \sum_{i=0}^n \|f_i(t)\| \leqslant M \sum_{i=0}^n \|du_i\|_\infty < \infty. \tag{14.16}$$

Le fait que le tout soit fini est l'hypothèse de convergence normale sur tout compact. Le compact en question est $\gamma\big([0,1]\big)$.

C'est le moment d'utiliser le théorème de la convergence dominée de Lebesgue 13.184. Attention aux notations un peu décalées. Nous avons $g_n \to \sum_{i=1}^\infty f_i$ (convergence simple) et $\|g_n(t)\| \leqslant A$ où A est une constante que nous voyons comme une fonction constante inté-

1. Cela revient à étudier l'intégrale de la forme différentielle du_n sur le chemin γ. Voir la définition 19.47 et tout ce qui s'en suit.

grable sur le compact $[0,1]$. Nous permutons la limite et l'intégrale :

$$\begin{aligned}
\int_0^1 \sum_{i=0}^{\infty} f_i(t)dt &= \int_0^1 (\lim_{n\to\infty} g_n)(t)dt && (14.17a)\\
&= \lim_{n\to\infty} \int_0^1 g_n(t)dt && (14.17b)\\
&= \lim_{n\to\infty} \int_0^1 \sum_{i=0}^{n} f_i(t)dt && (14.17c)\\
&= \lim_{n\to\infty} \sum_{i=0}^{n} \int_0^1 f_i(t)dt && (14.17d)\\
&= \sum_{i=0}^{\infty} \int_0^1 f_i(t)dt. && (14.17e)
\end{aligned}$$

Accroissements Nous pouvons maintenant faire le petit calcul suivant :

$$\sum_n \int_0^1 (du_n)_{\gamma(t)}\big(\gamma'(t)\big)dt = \sum_n \Big(u_n\big(\gamma(1)\big) - u_n\big(\gamma(0)\big)\Big) = \sum_n \big(u_n(x) - u_n(a)\big) = u(x) - u(a) \tag{14.18}$$

où nous avons utilisé le théorème fondamental du calcul intégral sous la forme de la proposition 13.242.

Nous retenons l'égalité

$$u(x) = u(a) + \int_0^1 \sum_n (du_n)_{\gamma(t)}\big(\gamma'(t)\big)dt. \tag{14.19}$$

Remarque La formule (14.19) n'est pas une forme de formule des accroissements finis qui parlerait d'évaluer une fonction u en x en partant de a et en intégrant du le long d'un chemin joignant a et x.

Ce serait le cas si nous pouvions permuter la somme et la différentielle qui se trouvent dans l'intégrale. Or permuter somme et différentielle est précisément l'objet du théorème que nous sommes en train de prouver.

Différentielle Forts de la formule (14.19), nous calculons $du_a(v)$, c'est-à-dire la différentielle de u au point a appliquée au vecteur $v \in F$. Pour cela, nous savons que Ω est ouvert, donc Ω contient une boule de rayon r autour de a, ce qui nous permet de dire que pour un donné, le point $a + sv$ est dans Ω pour tout $s \in B(0,\epsilon)$ lorsque ϵ n'est pas trop grand. Pour chacun de ces s, nous considérons un chemin de classe C^1 joignant a à $a + sv$. Ce chemin sera noté

$$\gamma_s \colon [0,1] \to \Omega \tag{14.20}$$

et $\gamma(0) = a$, $\gamma(1) = a + sv$. Nous avons le calcul

$$\begin{aligned}
du_a(v) &= \frac{d}{ds}\Big[u(a+sv)\Big]_{s=0} && (14.21a)\\
&= \frac{d}{ds}\Big[\int_0^1 \sum_{n=0}^{\infty} (du_n)_{\gamma_s(t)}\big(\gamma_s'(t)\big)dt\Big]_{s=0} && (14.21b)\\
&= \frac{d}{ds}\Big[\sum_{n=0}^{\infty} \int_0^1 (du_n)_{\gamma_s(t)}\big(\gamma_s'(t)\big)dt\Big]_{s=0} && (14.21c)\\
&= \frac{d}{ds}\Big[\sum_n \big(u_n(a+sv) - u_n(a)\big)\Big]_{s=0} && (14.21d)\\
&= \frac{d}{ds}\Big[(\sum_n u_n)(a+sv)\Big]_{s=0} && (14.21e)\\
&= \sum_n \frac{d}{ds}\Big[u_n(a+sv)\Big]_{s=0} && (14.21f)\\
&= \sum_n (du_n)_a(v) && (14.21g)\\
&= \Big(\sum_n (du_n)_a\Big)(v) && (14.21h)\\
&= \Big(\sum_n du_n\Big)_a(v). && (14.21i)
\end{aligned}$$

Justifications :

— Pour 14.21c. Permuter la somme et l'intégrale comme plus haut.

— Pour 14.21f. Permuter une somme et une dérivée classique des fonctions $\mathbb{R} \to F$ données par $s \mapsto u_n(a + sv)$. Il s'agit d'utiliser le théorème 11.311 sur chaque composantes dans F.

— Pour 14.21i. Chaque du_n est une application $du_n \colon E \to \mathcal{L}(E, F)$. Au fait près que la notation est plus lourde, il s'agit simplement d'une définition de la somme pontuelle d'une suite de fonctions : $\sum_n f_n(a) = (\sum_n f_n)(a)$. Dans ce cas-ci, le tout est encore un élément de $\mathcal{L}(E, F)$ que nous appliquons à v.

□

14.2 Séries entières

Dans cette section nous allons parler de séries complexes autant que de séries réelles. L'étude des propriétés à proprement parler complexes des séries entières (holomorphie) sera effectuée dans le chapitre dédié, voir le théorème 25.14 et ses conséquences.

14.2.1 Disque de convergence

Une **série de puissance** est une série de la forme

$$\sum_{k=0}^{\infty} c_k (z - z_0)^k \tag{14.22}$$

où $z_0 \in \mathbb{C}$ est fixé, (c_k) est une suite complexe fixée, et z est un paramètre complexe. Nous disons que cette série est *centrée* en z_0.

Définition 14.7.
*Une **série entière** est une somme de la forme*

$$\sum_{n=0}^{\infty} a_n z^n \tag{14.23}$$

avec $a_n, z \in \mathbb{C}$.

Une série entière peut définir une fonction

$$f(z) = \sum_n a_n z^n. \tag{14.24}$$

Le but de cette section est d'étudier des conditions sur la suite (a_n) qui assurent la continuité de f ou la possibilité de dériver ou intégrer la série terme à terme.

Définition 14.8.
Soit $\sum_{n\in\mathbb{N}} a_n z^n$ *une série entière. Le* ***rayon de convergence*** *de cette série est le nombre*

$$R = \sup\{r \in \mathbb{R}^+ \text{ tel que la suite } (a_n r^n) \text{ est bornée}\} \in [0, \infty]. \tag{14.25}$$

La boule $B(0, R)$ *est le* ***disque de convergence*** *de la série.*

Le rayon de convergence d'une série ne dépend que des réels $|a_n|$, même si à la base $a_n \in \mathbb{C}$.

Remarque 14.9.
Si pour tout n nous avons $|b_n| \geqslant |a_n|$ alors le rayon de convergence de la série $\sum_n a_n z^n$ est au moins aussi grand que celui de la série $\sum_n b_n z^n$. Cela y compris lorsque l'un ou l'autre des rayons de convergences est infini.

Lemme 14.10 (Critère d'Abel)**.**
Soit $R > 0$ *le rayon de convergence de la somme* $\sum_n a_n z^n$ *et* $z \in \mathbb{C}$.

(1) Si $|z| < R$ *alors la série converge absolument.*

(2) Si $R < \infty$ *et si* $|z| > R$ *alors la série diverge.*

Démonstration. Démonstration en deux parties.

(1) Si $|z| < R$ alors la suite $(a_n z^n)$ est bornée et il existe un nombre $M \in \mathbb{R}$ tel que $|a_n| r^n \leqslant M$ pour tout n. Nous considérons alors un r tel que $|z| < r < R$ et nous pouvons calculer :

$$|a_n z^n| = |a_n| r^n \left(\frac{|z|}{r}\right)^n \leqslant M \left(\frac{|z|}{r}\right)^n \tag{14.26}$$

Vu que $|z| < r$ nous tombons sur la série géométrique (10.124) qui converge. Par le critère de comparaison [2] la série $\sum_{n=0}^{\infty} |a_n z^n|$ converge.

(2) Par définition du rayon de convergence, la suite $(a_n z^n)$ n'est donc pas bornée et la série ne peut pas converger à cause de la proposition 10.58.

□

Le critère d'Abel parle bien de convergence absolue, et non de convergence normale. Pour chaque t, la série $\sum_k |a_n t^k|$ converge. Si par contre nous posons $u_k(t) = a_k t^k$, nous n'avons à priori pas la convergence normale $\sum_k \|u_k\|_\infty$, même pas si la norme est la norme supremum sur $B(0, R)$ [3]. Prenons comme exemple simplement $a_k = 1$ pour tout k. Pour tout $|t| < 1$, la série $\sum_k t^k$ converge absolument (série géométrique), mais nous aurions $\|u_k\|_\infty = 1$ et donc divergence évidente de $\sum_k \|u_k\|_\infty$.

La proposition suivante sera surtout utile lorsqu'on parlera de dérivée.

Proposition 14.11 ([303])**.**
Quel que soit le nombre $\alpha \in \mathbb{R}$, *les séries* $\sum_n a_n z^n$ *et* $\sum_n n^\alpha a_n z^n$ *ont même rayon de convergence.*

2. Lemme 10.61.
3. Il y aurait par contre bien convergence sur tout compact ? Cher lecteur, dites moi ce que vous en pensez

Démonstration. Nous posons

$$E = \{r \in \mathbb{R}^+ \text{ tel que } (a_n r^n) \text{ est borné }\} E' = \{r \in \mathbb{R}^+ \text{ tel que } (n^\alpha a_n r^n) \text{ est borné }\} \tag{14.27a}$$

Et aussi $R = \sup(E)$, $R' = \sup(E')$. Le fait que $E' \geqslant E$ est facile. Nous supposons $R > 0$ et nous considérons $r < R$ (c'est-à-dire $r \in E$). Nous allons montrer que $r \in E'$. Pour cela nous prenons un nombre s tel que $r < s < R$. Nous avons

$$n^\alpha a_n r^n = n^\alpha a_n \left(\frac{r}{s}\right)^n s^n = n^\alpha \left(\frac{r}{s}\right)^n a_n s^n. \tag{14.28}$$

Mais $r/s < 1$, donc le lemme 11.352 dit que $n^\alpha (r/s)^n \to 0$. Cela est donc borné par une constante M. Donc

$$n^\alpha a_n r^n \leqslant M a_n s^n. \tag{14.29}$$

Mais la suite $(a_n s^n)$ est bornée. Donc la suite $n^\alpha a_n r^n$ est également bornée, ce qui prouve que $r \in E'$. □

Remarque 14.12.
Au fond, cette proposition n'est rien d'autre que dire que dans $n^\alpha r^n$, l'effet « convergent » est r^n qui est une décroissance exponentielle tandis que l'effet « divergent » est n^α qui a une croissance seulement polynomiale.

Théorème 14.13 (Formule de Hadamard)**.**
Le rayon de convergence[4] *de la série entière* $\sum_n c_n z^n$ *est donné par une des deux formules*

$$\frac{1}{R} = \limsup \sqrt[k]{|a_k|} \tag{14.30}$$

ou

$$\frac{1}{R} = \lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| \tag{14.31}$$

lorsque a_k *est non nul à partir d'un certain* k*.*

Si une de ces formules donne $1/R = 0$*, alors le rayon de convergence est infini.*

Le disque $|z - z_0| \leqslant R$ est le **disque de convergence** de la série $\sum_n a_n (z - z_0)^n$. Notons que le critère d'Abel ne dit rien pour les points tels que $|z - z_0| = R$. Il faut traiter ces points au cas par cas. Et le pire, c'est qu'une série donnée peut converger pour certain des points sur le bord du disque, et diverger en d'autres. Le théorème d'Abel radial (théorème 14.21) nous donnera quelques informations sur le sujet.

Il y a un dessin à la figure 14.1.

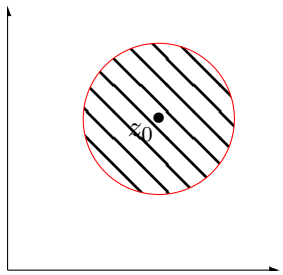


FIGURE 14.1 – À l'intérieur du disque de convergence, la convergence est absolue. En dehors, la série diverge. Sur le cercle proprement dit, tout peut arriver.

Si les suites a_n et b_n sont équivalentes, alors les séries correspondantes auront le même rayon de convergence. Cela ne signifie pas que sur le bord du disque de convergence, elles aient même comportement. Par exemple nous avons

$$\frac{1}{\sqrt{n}} \sim \frac{1}{\sqrt{n}} + \frac{(-1)^n}{n}. \tag{14.32}$$

4. Définition 14.8.

En même temps, en $z = -1$ la série

$$\sum_{n\geqslant 1} \frac{z^n}{\sqrt{n}} \tag{14.33}$$

converge par le critère des séries alternées[5]. Par contre la série

$$\sum_{n\geqslant 1} \left(\frac{1}{\sqrt{n}} + \frac{(-1)^n}{n} \right) z^n \tag{14.34}$$

ne converge pas pour $z = -1$.

Exemple 14.14.
Soit $\alpha \in \mathbb{R}$ et considérons la série $\sum_{n\geqslant 1} a_n z^n$ où a_n est la n-ième décimale de α. Si α est un nombre décimal limité, la suite (a_n) est finie et le rayon de convergence est infini. Sinon, pour tout N il existe un $n > N$ tel que $a_n \neq 0$ et la suite (a_n) ne tend pas vers zéro. Par conséquent la série

$$\sum_n a_n z^n \tag{14.35}$$

diverge pour $z = 1$ et le rayon de convergence satisfait $R \leqslant 1$. Nous avons aussi $|a_n| \leqslant 9$, de telle manière à ce que la série soit bornée et par conséquent majorée en module par $9z^n$, ce qui signifie que $R \geqslant 1$.

Nous déduisons alors $R = 1$. △

14.2.2 Propriétés de la somme

Le théorème suivant donne une formule (dit « produit de Cauchy ») pour le produit de deux séries entières. Nous en donnons une adaptation dans le cas de séries de puissances dans une algèbre normée dans la proposition 14.45.

Théorème 14.15.
Soient $\sum_n a_n z^n$ et $\sum b_n z^n$ deux séries de rayon de convergences respectivement R_a et R_b.

(1) Si R_s est le rayon de convergence de $\sum_n (a_n + b_n) z^n$, nous avons

$$R_s \geqslant \min\{R_a, R_b\} \tag{14.36}$$

et nous avons l'égalité si pour tout $|z| \leqslant \min\{R_a, R_b\}$, $\sum (a_n + b_n) z^n = \sum_n a_n z^n + \sum_n b_n z^n$.

(2) Si $\lambda \neq 0$ la série $\sum_n (\lambda a_n) z^n$ a le même rayon de convergence que la série $\sum_n a_n z^n$ et si $|z| < R_a$ nous avons

$$\sum_{n=0}^{\infty} (\lambda a_n) z^n = \lambda \sum_{n=0}^{\infty} a_n z^n. \tag{14.37}$$

*(3) Le **produit de Cauchy** des deux séries est donné par*

$$\left(\sum_n a_n z^n \right) \left(\sum_k b_k z^k \right) = \sum_{n=0}^{\infty} \left(\sum_{i+j=n} a_i b_j \right) z^n = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} a_k b_{n-k} \right) z^n. \tag{14.38}$$

Si R_p est le rayon de convergence de ce produit nous avons

$$R_p \geqslant \min\{R_a, R_b\} \tag{14.39}$$

et si $|z| < \min\{R_a, R_b\}$ alors

$$\sum_{n=0}^{\infty} \left(\sum_{i+j=n} a_i b_j \right) z^n = \left(\sum_{n=0}^{\infty} a_n z^n \right) \left(\sum_{n=0}^{\infty} b_n z^n \right). \tag{14.40}$$

5. Théorème 10.72.

Démonstration. Nous prouvons la partie sur le produit de Cauchy. En utilisant la propriété du produit de la somme par un scalaire nous avons

$$\left(\sum_{n=0}^{\infty} a_n z^n\right)\left(\sum_{m=0}^{\infty} b_m z^m\right) = \sum_{n=0}^{\infty}\left(\sum_{m=0}^{\infty} b_m a_n z^{m+n}\right) \tag{14.41a}$$

$$= \lim_{N\to\infty}\lim_{M\to\infty}\sum_{n=0}^{N}\sum_{m=0}^{M} b_m a_n z^{m+n} \tag{14.41b}$$

$$= \lim_{N\to\infty}\lim_{M\to\infty}\sum_{k=0}^{N+M}\sum_{i+k=k} b_i a_j z^k \tag{14.41c}$$

$$= \lim_{N\to\infty}\sum_{k=0}^{\infty}\sum_{i+k=k} b_i a_j z^k \tag{14.41d}$$

$$= \sum_{k=0}^{\infty}\sum_{i+j=k} b_i a_j z^k. \tag{14.41e}$$

□

Exemple 14.16.
Montrons un produit de Cauchy dont le rayon de convergence est strictement plus grand que le minimum. D'abord nous considérons

$$A = 1 - z, \tag{14.42}$$

c'est-à-dire $a_0 = 1$, $a_1 = -1$, $a_{n\geqslant 2} = 0$ avec $R_a = \infty$. Ensuite nous considérons

$$B = \sum_n z^n, \tag{14.43}$$

c'est-à-dire $B = (1-z)^{-1}$ et $R_b = 1$. Le produit de Cauchy de ces deux séries valant 1, le rayon de convergence est infini. △

Exemple 14.17.
Nous montrons que

$$\sum_{n=0}^{\infty}(n+1)x^n = \frac{1}{(1-x)^2} \tag{14.44}$$

pour $x \in]-1,1[$.

Étant donné que pour tout r dans $]-1,1[$ la suite $(n+1)r^n$ est bornée, le rayon de convergence est correct. Pour les x dans ce domaine nous avons

$$\frac{1}{(1-x)^2} = \frac{1}{(1-x)}\frac{1}{(1-x)} = \left(\sum_{n=0}^{\infty} x^n\right)\left(\sum_{m=0}^{\infty} z^m\right). \tag{14.45}$$

Nous devons expliciter ce produit de Cauchy en utilisant le théorème 14.15. Pour tout i nous avons $a_i = b_i = 1$. Par conséquent le produit (14.45) devient

$$\sum_{n=0}^{\infty}\sum_{i+j=n} x^n = \sum_{n=0}^{\infty}(n+1)x^n. \tag{14.46}$$

△

Théorème 14.18.
Une série entière converge normalement sur tout disque fermé inclus au disque de convergence.

Démonstration. Toute boule fermée inclue à $B(0,R)$ est inclue à la boule $\overline{B(0,r)}$ pour un certain $r < R$. Nous nous concentrons donc sur une telle boule fermée.

Pour chaque n nous posons $u_n(z) = a_n z^n$ que nous voyons comme une fonction sur $\overline{B(0,r)}$. Pour tout $n \in \mathbb{N}$ et tout $z \in \overline{B(0,r)}$ nous avons

$$\|u_n\|_\infty \leqslant |a_n z^n| \leqslant |a_n| r^n. \tag{14.47}$$

Étant donné que $r < R$ la série $\sum_n |a_n| r^n$ converge et la série $\sum_n \|u_n\|$ est convergente. La série $\sum_n a_n z^n$ est alors normalement convergente. □

Exemple 14.19.
Encore une fois nous n'avons pas d'informations sur le comportement au bord. Par exemple la série $\sum_n z^n$ a pour rayon de convergence $R = 1$, mais $\sup_{z \in B(0,1)} |z^n| = 1$ et nous n'avons pas de convergence normale sur la boule fermée. △

La convergence normale n'est donc pas de mise sur tout l'intérieur du disque de convergence. La continuité, par contre est effective sur la boule. En effet si $z_0 \in B(0,R)$ alors il existe un rayon $0 < r < R$ tel que $B(z_0, r) \subset B(0,R)$. Sur $B(z_0,r)$ nous avons convergence normale et donc continuité en z_0.

La différence est que la continuité est une propriété locale tandis que la convergence normale est une propriété globale.

Proposition 14.20.
Soit $f(z) = \sum_n a_n z^n$ avec un rayon de convergence R. Si $\sum |a_n| R^n$ converge alors

(1) la série $\sum_n a_n z^n$ converge normalement sur $\overline{B(0,R)}$,

(2) f est continue sur $\overline{B(0,R)}$.

Démonstration. La conclusion est claire dans l'intérieur du disque de convergence. En ce qui concerne le bord, chacune des sommes partielles est une fonction continue. De plus nous avons $\|u_n\| \leqslant |a_n| R^n$, dont la série converge. Par conséquent nous avons convergence normale sur le disque fermé. □

Le théorème suivant permet de donner, dans le cas de fonctions réelle, des informations sur la convergence en une des deux extrémités de l'intervalle de convergence.

Théorème 14.21 (Convergence radiale de Abel)**.**
Soit $f(x) = \sum_n a_n x^n$ une série réelle de rayon de convergence $0 < R < \infty$.

(1) Si $\sum a_n R^n$ converge, alors f est continue sur $[0,R]$.

(2) Si $\sum_n a_n (-R)^n$ converge, alors f est continue sur $[-R,0]$.

La proposition 14.74 donnera un exemple d'utilisation pour la série de $\ln(1-x)$ (qui n'est pas encore définie à ce moment).

Le résultat suivant permet d'identifier deux séries complexes lorsque leurs valeurs sur $\mathbb{R}$ sont identiques.

Proposition 14.22.
Soient les séries $f(z) = \sum a_n z^n$ et $g(z) = \sum b_n z^n$ convergentes dans $B(0,R)$. Si $f(x) = g(x)$ pour $x \in [0,R[$ alors $a_n = b_n$.

Démonstration. Soit n_0 le plus petit entier tel que $a_{n_0} \neq b_{n_0}$. Pour tout $z \in B(0,R)$ nous avons

$$f(z) - g(z) = \sum_{n=n_0}^{\infty} (a_n - b_n) z^n = z^{n_0} \varphi(z) \tag{14.48}$$

où

$$\varphi(z) = \sum_{n \geqslant 0} (a_{n+n_0} - b_{n+n_0}) z^n. \tag{14.49}$$

Par le théorème 14.15(1) le rayon de convergence de φ est plus grand que R et la fonction φ est continue en 0. Étant donné que $\varphi(0) = a_{n_0} - b_{n_0} \neq 0$ et que φ est continue nous avons un ρ tel

que $\varphi \neq 0$ sur $B(0,\rho)$. Or cela n'est pas possible parce que au moins sur la partie réelle de cette dernière boule, φ doit être nulle. □

Proposition 14.23 ([304, 1]).
Si la série entière $\sum_{n\geqslant 0} a_n z^n$ a un rayon de convergence R alors

(1) La somme est une fonction holomorphe dans le disque de convergence.

(2) La somme est différentiable et

$$du_{z_0}(z) = \sum_{n=1}^{\infty} n a_n z_0^{n-1} z. \tag{14.50}$$

(3) De plus pour tout $z_0 \in B(0,R)$, on pose[6]

$$S(z) = \sum_{n\geqslant 0} a_n z^n \tag{14.51a}$$

$$T(z) = \sum_{n\geqslant 1} n a_n z^{n-1} = \sum_{n=0}^{\infty} (n+1) a_{n+1} z^n. \tag{14.51b}$$

Alors nous avons

$$\lim_{z\to z_0} \frac{S(z)-S(z_0)}{z-z_0} = T(z_0). \tag{14.52}$$

Démonstration. Nous allons prouver, en utilisant le théorème 14.6, que la somme est une fonction différentiable et que la différentielle est $\mathbb{C}$-linéaire. La proposition 11.254 nous dira alors que la somme est $\mathbb{C}$-dérivable.

Nous posons $u_n(z) = a_n z^n$, qui est une fonction de classe C^1. En ce qui concerne sa différentielle nous considérons $z_0 \in B(0,R)$ et nous avons (si $n=0$ alors la différentielle est nulle)

$$\begin{aligned}
(du_n)_{z_0}(z) &= \frac{d}{dt}\Big[u_n(z_0+tz)\Big]_{t=0} && (14.53a)\\
&= \frac{d}{dt}\Big[a_n(z_0+tz)^n\Big]_{t=0} && (14.53b)\\
&= \frac{d}{dt}\Big[n a_n (z_0^{n-1} tz)\Big]_{t=0} && (14.53c)\\
&= n a_n z_0^{n-1} z. && (14.53d)
\end{aligned}$$

En cours de calcul nous avons développé $(z_0+tz)^n$ et gardé seulement les termes de degré 1 en t. Il y en a n et ils sont tous égaux à $z_0^{n-1}tz$.

La convergence simple $\sum_n u_n$ est dans les hypothèses. Il reste à prouver que la somme des différentielles converge uniformément sur tout compact autour de z_0 ne débordant pas du disque ouvert de convergence. Soit K un compact autour de z_0. Dans le calcul suivant nous utilisons une première fois la norme uniforme de du_n vu comme fonction de K vers $\mathcal{L}(\mathbb{C},\mathbb{C})$ et une fois la norme opérateur[7] de $(du_n)_{z_0}$ comme application linéaire $\mathbb{C}\to\mathbb{C}$:

$$\begin{aligned}
\|du_n\|_k &= \sup_{z_0\in K} \|(du_n)_{z_0}\| && (14.54a)\\
&= \sup_{z_0\in K} \sup_{|z|=1} |(du_n)_{z_0}(z)| && (14.54b)\\
&= \sup_{z_0\in K} \sup_{|z|=1} |n a_n z_0^{n-1} z| && (14.54c)\\
&= \sup_{z_0\in K} n|a_n||z_0|^{n-1}. && (14.54d)
\end{aligned}$$

6. Pour rappel, dans tout ce texte, $B(a,r)$ est une boule *ouverte*.
7. Définition 10.3.

Vu que $z \mapsto |z|^{n-1}$ est une application continue sur le compact K, elle atteint son maximum (théorème 7.105). Nous considérons z_K, un point qui réalise le supremum. Ce nombre est dans le disque de convergence parce que K est un compact autour de z_0.

Nous devons prouver que $\sum_n n|a_n||z_K|^{n-1}$ converge. Vu que $|z_K|$ est une constante (par rapport à n) nous pouvons étudier la convergence en écrivant $|z_K|^n$ au lieu de $|z_K|^{n-1}$.

La suite $(a_n|z_K|^n)$ est une suite bornée. Soit M tel que $|a_n||z_K|^n < M$ pour tout n. Nous considérons de plus r de telle sorte que $K \subset B(0,r) \subset B(0,R)$. En particulier $|z_K| < r$ et nous avons

$$n|a_n||z_K|^n \leqslant n|a_n|r^n \left(\frac{|z_K|}{r}\right)^n \leqslant nM \left(\frac{|z_K|}{r}\right)^n. \tag{14.55}$$

Nous savons que ce qui est dans la parenthèse est plus petit que 1, mais que $\sum_n nx^n$ converge dès que $|x| < 1$. Par conséquent

$$\sum_n \|du_n\|_K \tag{14.56}$$

converge et le théorème 14.6 fonctionne : $du = \sum_{n=1}^{\infty} du_n$ et la somme $\sum_n u_n$ est de classe C^1.

La différentielle de $\sum_n u_n$ s'exprime explicitement par

$$du_{z_0}(z) = \sum_{n=1}^{\infty} na_n z_0^{n-1} z. \tag{14.57}$$

Cette forme montre que du_{z_0} est une application $\mathbb{C}$-linéaire et donc la somme est $\mathbb{C}$-dérivable par la proposition 11.254. Ergo holomorphe sur le disque de convergence par définition 11.248.

En ce qui concerne la formule (14.52), elle provient de la formule (11.637) : $f'(z_0)$ est donné par la facteur multiplicatif de du_{z_0}. En l'occurrence la formule (14.57) nous donne

$$f'(z_0) = \sum_{n\geqslant 1} na_n z_0^{n-1}. \tag{14.58}$$

□

14.2.3 Dérivation

Lemme 14.24.
Soit une série entière $\sum a_n z^n$ de rayon de convergence R. Les séries

$$\sum \frac{a_n}{n+1} z^{n+1} \tag{14.59}$$

et

$$\sum_{n\geqslant 1} na_n z^{n-1} \tag{14.60}$$

ont même rayon de convergence R.

Notons toutefois que nonobstant ce lemme, les séries dont il est question peuvent se comporter différemment sur le bord du disque de convergence. En effet la série

$$\sum \frac{1}{n} z^n \tag{14.61}$$

diverge pour $z = 1$ alors que

$$\sum \frac{1}{n(n+1)} z^{n+1} \tag{14.62}$$

converge pour $z = 1$.

Les théorèmes de dérivation et d'intégration de séries de fonctions (théorèmes 14.1 et 14.3) fonctionnent bien dans le cas des séries entières. Ils donnent la proposition 14.25 pour la dérivation et 14.30 pour l'intégration.

Proposition 14.25.
Soit la série entière

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \tag{14.63}$$

de rayon de convergence R. Alors la fonction f est C^1 sur $]-R, R[$ et se dérive terme à terme :

$$f'(x) = \sum_{n=1}^{\infty} n a_n x^{n-1} \tag{14.64}$$

pour tout $x \in]-R, R[$.

Démonstration. Nous savons que la série $\sum_{n=1}^{\infty} n a_n x^{n-1}$ a le même rayon de convergente que celui de la série f. En particulier cette série des dérivées converge normalement sur tout compact dans $]-R, R[$ et la somme est continue. Le théorème 14.3 conclu. □

Remarque 14.26.
À part lorsqu'on parle de fonction $\mathbb{R} \to \mathbb{R}$, la notion de classe C^k s'entend au sens de la différentielle, et non de la dérivée, voir les définitions 11.267. C'est cela qui explique la structure de la démonstration de la proposition 14.23.

Corolaire 14.27 ([304, 1])**.**
La somme d'une série entière est de classe C^∞ sur le disque ouvert de convergence.

Démonstration. La proposition 14.23 a démontré en réalité nettement plus : sur le disque ouvert de convergence, la somme est une fonction holomorphe. Il n'est cependant pas possible de conclure ainsi parce que le fait qu'une fonction holomorphe est C^∞ ne sera démontré qu'au coût de nombreux efforts dans le théorème 25.14(3).

Cas réel Nous considérons la série entière $\sum_n a_n x^n$ pour $x \in \mathbb{R}$ de rayon de convergence R. Une simple récurrence sur la proposition 14.25 donne le résultat.

Cas complexe Attention : le fait d'être de classe C^k est le fait d'être k fois *différentiable*. Rien à voir avec la $\mathbb{C}$-dérivabilité.

En ce qui concerne la différentiabilité nous avons la proposition 14.23 qui dit que dans le disque de convergence, la fonction $u(z) = \sum_n a_n z^n$ a pour différentielle l'application $du\colon \mathbb{C} \to \mathcal{L}_{\mathbb{C}}(\mathbb{C}, \mathbb{C})$,

$$\begin{aligned} du\colon \mathbb{C} &\to \mathcal{L}_{\mathbb{C}}(\mathbb{C}, \mathbb{C}) \\ du_{z_0}(z) &= \Big(\sum_{n=0}^{\infty} (n+1) a_{n+1} z_0^n \Big) z. \end{aligned} \tag{14.65}$$

Nous allons éviter de considérer la différentielle seconde comme une application

$$d^2u\colon \mathbb{C} \to \mathcal{L}\big(\mathbb{C}, \mathcal{L}(\mathbb{C}, \mathbb{C})\big) \tag{14.66}$$

parce que ça nous mènerait trop loin pour parler de la différentielle k^{e}. Au lieu de cela nous allons considérer l'isomorphisme d'espace vectoriel

$$\begin{aligned} \psi\colon \mathbb{C} &\to \mathcal{L}_{\mathbb{C}}(\mathbb{C}, \mathbb{C}) \\ z_0 &\mapsto \psi(z_0) z = z_0 z. \end{aligned} \tag{14.67}$$

Dans cette optique nous écrivons :

$$du_{z_0} = \psi\Big(\sum_{n=0}^{\infty} (n+1) a_{n+1} z_0^n \Big) \tag{14.68}$$

ou encore :

$$(\psi^{-1} \circ d) u(z_0) = \sum_{n \geqslant 0} (n+1) a_{n+1} z_0^n. \tag{14.69}$$

Nous allons prouver par récurrence que l'égalité suivante est vraie (y compris le fait que la somme converge) :

$$(\psi^{-1}\circ d)^k u(z_0)=\sum_{n=0}^{\infty}\frac{(n+k)!}{n!}a_{n+k}z_0^n. \tag{14.70}$$

Prouvons d'abord que cette somme converge pour tout k. Nous avons $(n+k)!/n!<(n+k)^k$ et donc il suffit de prouver que la série de coefficients $n^k a_n$ converge. C'est le cas par la proposition 14.11.

Nous pouvons calculer la différentielle de $(\psi^{-1}\circ d)^k u$ en dérivant terme à terme en utilisant (encore) la proposition 14.23(2) :

$$d\big((\psi^{-1}\circ d)^k u\big)_{z_0}(z)=\sum_{n=1}^{\infty}\frac{(n+k)!}{n!}a_{n+k}na_0^{n-1}z \tag{14.71a}$$

$$=\sum_{n=0}^{\infty}\frac{(n+k+1)!}{n!}a_{n+k+1}z_0^n z. \tag{14.71b}$$

Nous appliquons ψ^{-1} à cela :

$$(\psi^{-1}\circ d)^{k+1}u(z_0)=\sum_{k=0}^{\infty}\frac{(n+k+1)!}{n!}a_{n+k+1}z_0^n. \tag{14.72}$$

Dérouler à l'envers Nous allons maintenant utiliser la proposition 11.218 pour montrer que u est de classe C^k pour tout k. Nous avons démontré que $(\psi^{-1}\circ d)^k u$ était différentiable. Par conséquent, $d\big((\psi^{-1}\circ d)^{k-1}u\big)$ est différentiable et donc $(\psi^{-1}\circ d)^{k-1}$ est de classe C^1. En continuant ainsi, $(\psi^{-1}\circ d)^{k-l}u$ est de classe C^l et u est de classe C^k.

□

Le lemme suivant est encore essentiellement valable dans un espace de Banach (proposition 10.152).

Lemme 14.28.
La série entière $\sum_{n\geqslant 0}z^{nk}$ a un rayon de convergence 1 et converge vers la fonction

$$\sum_{n\geqslant 0}z^{nk}=\frac{1}{1-z^k}. \tag{14.73}$$

Lorsque $|\omega|=1$ nous avons aussi un rayon de convergence 1 pour la série

$$\frac{1}{\omega-z}=\sum_{k\geqslant 0}\omega^{-k-1}z^k. \tag{14.74}$$

Sous les mêmes hypothèses sur ω nous avons encore la série

$$\frac{1}{(\omega-z)^k}=\frac{1}{(k-1)!}\sum_{s=0}^{\infty}\omega^{-s-1-k}\frac{(s+k-1)!}{s!}z^s \tag{14.75}$$

Démonstration. Les coefficients de la série sont $a_n=1$ lorsque n est multiple de k et $a_n=0$ autrement. Donc pour $r=1$ la suite $r^n a_n$ reste bornée[8]. Cela prouve que le rayon de convergence est au moins 1. Par ailleurs si $r>1$ alors clairement la suite $(a_n r^n)$ n'est pas bornée. Cela prouve le rayon de convergence égal à 1.

Soit donc $z\in B(0,1)$. Nous avons

$$\left(\sum_{n\geqslant 0}z^{nk}\right)(1-z^k)=\sum_{n\geqslant 0}z^{nk}-\sum_{n\geqslant 0}z^{(n+1)k}. \tag{14.76}$$

8. Utilisation directe de la définition 14.8.

Le premier terme de la première somme vaut 1 tandis que tous les autres termes s'annulent deux à deux.

En ce qui concerne la série (14.74), elle s'obtient facilement :

$$\frac{1}{\omega - z} = \frac{1}{\omega}\frac{1}{1-\frac{z}{\omega}} = \frac{1}{\omega}\sum_{s=0}^{\infty}\left(\frac{z}{\omega}\right)^s = \sum_s \omega^{-s-1}z^s. \tag{14.77}$$

La troisième série s'obtient en dérivant la seconde, ce qui est permis dans le disque de convergence par la proposition 14.25. □

Remarque 14.29.
Sur le bord du disque de convergence, la série $\sum_n z^{nk}$ ne converge pas. En effet le rayon étant 1, sur le bord nous avons la série $\sum_n e^{ink\theta}$ dont la norme du terme général ne tend pas vers zéro.

14.2.4 Intégration

Proposition 14.30.
Soit la série entière $\sum a_n x^n$ de rayon de convergence R.

(1) Pour tout segment $[a,b] \subset]-R,R[$ nous pouvons intégrer terme à terme :

$$\int_a^b \left(\sum_{n=0}^{\infty} a_n x^n\right) dx = \sum_{n=0}^{\infty} a_n \int_a^b x^n dx. \tag{14.78}$$

(2) La série entière obtenue en intégrant terme à terme a le même rayon de convergence que celui de la série de départ.

Démonstration. La première assertion est un cas particulier du théorème général 14.1. Pour le rayon de convergence, le lemme 14.24 fait le travail. □

Vu que le rayon de convergence ne varie pas par la dérivation ou par l'intégration et qu'une série entière est de classe C^∞ sur son disque de convergence, nous pouvons dériver terme à terme autant de fois que nous le voulons sans faire de fautes dans le disque de convergence.

14.3 Séries de Taylor

14.31.
Avant de commencer, une petite formule de dérivation toute simple que nous allons utiliser souvent :

$$(z^k)^{(l)} = \begin{cases} 0 & \text{si } l > k \\ \frac{k!}{(k-l)!}z^{k-l} & \text{sinon.} \end{cases} \tag{14.79}$$

Dans les cas où il est permis de dériver terme à terme, nous avons la formule

$$f^{(p)}(x) = \sum_k a_k (x^k)^{(p)} = \sum_{k=p}^{\infty} a_k \frac{k!}{(k-p)!}x^{k-p} \tag{14.80}$$

14.3.1 Polynôme de Taylor d'une série entière

Le polynôme de Taylor d'une fonction définie par une série entière s'obtient en tronquant la série. Cela est une assez bonne nouvelle que nous allons démontrer maintenant.

Proposition 14.32 ([1]).
Soit une série entière

$$f(x) = \sum_k a_k x^k \tag{14.81}$$

de rayon de convergence $R > 0$.

Pour tout $n \in \mathbb{N}$, il existe une fonction α telle que

$$f(x) = \sum_{k=0}^{n} a_k x^k + \alpha(x) x^n \tag{14.82}$$

et

$$\lim_{t \to 0} \alpha(t) = 0. \tag{14.83}$$

Tout ceci étant convenu que

— *l'égalité (14.82) est uniquement valable sur le disque de convergence,*

— *La fonction α dépend de n.*

Démonstration. Le corolaire 14.27 nous indique que f est de classe C^∞ sur $]-R, R[$ et que nous pouvons dériver terme à terme.

En utilisant la formule (14.80) et en l'évaluant en $x = x_0$, tous les termes s'annulent sauf $k = p$:

$$f^{(p)}(0) = p! a_p. \tag{14.84}$$

Le théorème de Taylor 11.375 nous indique alors qu'il existe $\alpha \colon \mathbb{R} \to \mathbb{R}$ telle que $\lim_{t \to 0} \alpha(t) = 0$ et

$$f(x) = \sum_{k=0}^{n} a_k x^k + \alpha(x) x^n. \tag{14.85}$$

□

14.3.2 Une majoration pour le reste

Lemme 14.33.
Soit une fonction $f \colon \mathbb{R} \to \mathbb{R}$ dérivable $n+1$ fois sur $B(a, R)$. Alors pour tout $x \in B(a, r)$,

$$f(x) = f(a) + \sum_{k=1}^{n-1} \frac{(x-a)^k}{k!} f^{(k)}(a) + \int_a^x \dots \int_a^{u_{n-1}} f^{(n)}(u_n) du_n \dots du_1. \tag{14.86}$$

Démonstration. Nous allons intensivement utiliser le théorème fondamental du calcul intégral 13.236 sous la forme de la formule (13.691). Nous avons d'abord

$$f(x) = f(a) + \int_a^x f'(u_1) du_1 = \int_a^x \big[f'(a) + \int_a^{u_1} f''(u_2) du_2 \big] du_1. \tag{14.87}$$

Toute l'astuce de ce théorème est de continuer à substituer $f^{(k)}(t)$ par $f^{(k)}(a)$ plus une intégrale de a à t de $f^{(k+1)}(u)$. Nous démontrons ainsi par récurrence que

$$f(x) = f(a) + \sum_{k=1}^{n-1} \frac{(x-a)^k}{k!} f^{(k)}(a) + \int_a^x \cdots \int_a^{u_{n-1}} f^{(n)}(u_n) du_n \dots du_1. \tag{14.88}$$

La preuve de cela se fait en substituant

$$f^{(n)}(u_n) = f^{(n)}(a) + \int_a^{u_n} f^{(n+1)}(u_{n+1}) du_{n+1} \tag{14.89}$$

et en remarquant (encore par récurrence par exemple) que

$$\int_a^x \dots \int_a^{u_{n-1}} du_n \dots du_1 = \frac{(x-a)^n}{n!}. \tag{14.90}$$

□

Le théorème suivant donne majoration du reste du polynôme de Taylor. Il est un premier pas dans la démonstration de formules comme

$$\lim_{n\to\infty} P_n(x) = f(x) \tag{14.91}$$

lorsque P_n est un polynôme de Taylor autour d'un point $a \neq x$. Nous ne saurions trop insister sur le fait que de telles formules ne seraient valables que pour une classe relativement restreintes de fonctions.

Théorème 14.34 (Inégalité de Taylor[305])**.**
Soit une fonction $f\colon \mathbb{R} \to \mathbb{R}$ dérivable $n+1$ fois et telle que $|f^{(n+1)}(x)| \leqslant M_N$ sur $B(a,d)$. Alors

$$|R_n(x)| \leqslant \frac{M_n}{(n+1)!}|x-a|^{n+1} \tag{14.92}$$

où $R(x) = f(x) - P_n(x)$ et où P_n sont les polynômes de Taylor autour de $a \in \mathbb{R}$.

Démonstration. Nous pouvons écrire la formule du lemme 14.33 pour $n+1$ au lieu de n ; cela donne

$$f(x) = P_n(x) + \int \cdots , \tag{14.93}$$

et donc

$$|R_n(x)| = |P_n(x) - f(x)| = \int_a^x \cdots \int_a^{u_n} f^{(n+1)}(x) du_n \cdots du_1 \tag{14.94}$$

En effectuant toutes les intégrales nous trouvons[9]

$$|R_n(x)| \leqslant \frac{M_n}{(n+1)!}|x-a|^{n+1}. \tag{14.95}$$

□

Cette formule pour le reste est très bien, mais pour l'exploiter au maximum de ses possibilités, il faudra la notion de convergence de suite de fonctions, et en particulier la notion de série de fonctions, pour pouvoir écrire

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} x^k \tag{14.96}$$

lorsque cela est possible. Nous renvoyons donc aux séries de Taylor, section 14.3, et en particulier aux fonctions analytiques de la sous-section 14.3.3.

14.3.3 Fonctions analytiques

Nous avons vu les polynômes de Taylor et déjà noté qu'il n'est pas en général vrai que $\lim_{n\to\infty} P_n(x) = f(x)$ pour des x même proches du point autour duquel les polynômes de Taylor P_n sont calculés.

Nous allons maintenant étudier la classe des fonctions pour lesquelles la série de Taylor est égale à la fonction de départ. D'abord une proposition montrant que les coefficients de Taylor sont les seuls pour lesquels il est possible d'espérer avoir une telle propriété.

Proposition 14.35 ([306])**.**
Soit une fonction donnée par la série entière

$$f(x) = \sum_{k=0}^{\infty} c_n (x-a)^n \tag{14.97}$$

sur la boule de convergence $B(a,R)$ avec $R > 0$ (hypothèse : le rayon de convergence est strictement positif). Alors

$$c_n = \frac{f^{(n)}(a)}{n!}. \tag{14.98}$$

9. Je me demande si je n'ai pas une faute entre n et $n+1$ quelque part. Relisez attentivement et écrivez-moi si vous trouvez une faute.

Démonstration. Par hypothèse, nous avons un rayon de convergence $R > 0$, et le corolaire 14.27 nous indique que f y est de classe C^∞. Et nous pouvons dériver terme à terme par la proposition 14.25. Cela pour dire qu'il nous est autorisé d'utiliser la formule (14.80) pour calculer les dérivées de f au point a. Nous avons d'abord

$$f^{(p)}(x) = \sum_{n=p}^{\infty} c_n \frac{n!}{(n-p)!}(x-a)^{n-p}, \tag{14.99}$$

et donc

$$f^{(p)}(a) = c_p p! \tag{14.100}$$

qui donne immédiatement le résultat. □

Proposition 14.36.
Soit l'intervalle $I = B(a, r)$. Si il existe M tel que

$$|f^{(n)}(x)| \leqslant \frac{M}{r^n} n! \tag{14.101}$$

pour tout $x \in B(a, r)$. Alors nous avons la convergence simple

$$P_n \to f \tag{14.102}$$

sur $B(a, r)$. Ici, P_n est le polynôme de Taylor d'ordre n pour la fonction f autour du point a[10].

Démonstration. Vu que nous avons $|f^{(n)}(x)| \leqslant \frac{M}{r^n} n!$ pour tout x, nous pouvons poser

$$M_n = \frac{M}{r^n} n! \tag{14.103}$$

dans le théorème 14.34 pour le faire fonctionner. Nous avons alors

$$|R_n(x)| \leqslant \frac{M}{r^n} n! \frac{1}{(n+1)!}|x-a|^{n+1} = \frac{M}{n+1}|x-a| \left|\frac{x-a}{r}\right|^n. \tag{14.104}$$

Vu que $x \in B(a, r)$ nous avons $|x - a| < r$ et donc $|(x - a)/r|^n < 1$. Nous pouvons aussi majorer $|x - a|$ par r et écrire

$$|R_n(x)| \leqslant \frac{rM}{n+1}. \tag{14.105}$$

Nous avons donc bien $\lim_{n\to\infty} R_n(x) \to 0$. □

14.4 Exponentielle sur une algèbre normée

14.4.1 Définition

Dans ce qui suit, nous considérons une algèbre commutative.

Proposition-définition 14.37 (Exponentielle[1])**.**
Soit $(A, \|.\|)$ une algèbre[11] *commutative de dimension finie sur $\mathbb{C}$ munie d'une norme d'algèbre. Pour $x \in A$ nous définissons*

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}. \tag{14.106}$$

Cette définition a les propriétés suivantes :

10. Pour être complet, il faut préciser que P_n est calculé dans ZFC. C'est pour cela que nous n'écrivons pas des lourdeurs comme $P_{n,a}(f)(x)$; si il fallait donner tout le contexte dans la notation, on n'en sortirait pas.

Ah, et tant que j'y suis si vous ne savez pas ce qu'est ZFC, je vous déconseille fortement de répéter cela à un jury d'agrég, entre autres parce que vous allez attirer la question « vraiment ? Vous utilisez C ? Où ? Pourquoi ? ». Et là, bonne chance.

11. Définition 3.83.

(1) C'est bien défini pour tout $x \in A$. C'est-à-dire que pour chaque x, la série (14.106) converge.

(2) Cela donne une application continue $\exp\colon A \to A$.

(3) La fonction exp *est différentiable et*

$$(d\exp)_x(y) = \exp(x)y, \tag{14.107}$$

le dernier produit étant la structure d'algèbre sur A.

Démonstration. Pour la différentiabilité de exp, nous voulons utiliser le théorème 14.6. Pour cela nous posons

$$u_k(x) = \frac{x^k}{k!} \tag{14.108}$$

Convergence simple Nous prouvons la convergence simple, c'est-à-dire pour chaque x séparément, de la série (14.106) dans deux buts. D'abord de nous assurer que la définition posée de exp a un sens, et ensuite pour commencer à vérifier les hypothèses du théorème 14.6.

Nous montrons que les sommes partielles forment une suite de Cauchy. Nous fixons $x \in A$ et nous posons

$$s_n = \sum_{k=0}^{\infty} \frac{x^k}{k!}. \tag{14.109}$$

Soient $p > q$, deux entiers. Nous avons :

$$\|s_p - s_q\| = \| \sum_{k=q+1}^{p} \frac{x^k}{k!} \| \leqslant \sum_{k=q+1}^{p} \frac{\|x^k\|}{k!} \leqslant \sum_{k=q+1}^{p} \frac{\|x\|^k}{k!} \tag{14.110}$$

où nous avons utilisé le fait que la norme sur A soit une norme d'algèbre.

C'est le moment d'utiliser la série exponentielle donnée dans l'exemple 10.69 que nous appliquons avec $t = \|x\|$. La série donnée par les coefficients $a_k = \|x\|^k/k!$ converge et ses sommes partielles forment en particulier une suite de Cauchy. Donc ce que nous avons à droite dans (14.110) peut être rendu arbitrairement petit lorsque p et q sont grands.

u_k est continue Il s'agit de remarquer que $(x+h)^k = x^k + hC(x,h)$ où C est une fonction bornée de h (lorsque h est dans un voisinage de $0 \in A$). Donc

$$\|(x+h)^k - x^k\| \leqslant \|h\|\|C(x,h)\| \to 0. \tag{14.111}$$

Candidat différentielle de u_k Nous trouvons à présent un candidat à être différentielle de u_k. Pour cela nous faisons le calcul suivant, sans trop nous soucier de la rigueur :

$$(du_k)_x(y) = \frac{d}{dt}\Big[u_k(x+ty)\Big]_{t=0} = k\frac{1}{k!}x^{k-1}y = u_{k-1}(x)y. \tag{14.112}$$

u_k est différentiable Nous fixons $x \in A$ et nous posons $T(y) = u_{k-1}(x)y$. Ensuite nous vérifions que cela vérifie la définition de la différentielle : nous devons calculer

$$\lim_{h\to 0} \frac{u_k(x+h) - u_k(x) - T(h)}{\|h\|} = \lim_{h\to 0} \frac{(x+h)^k - x^k - kx^{k-1}h}{k!\|h\|} = \clubsuit. \tag{14.113}$$

Vous vous souvenez de la formule pour $(x+h)^k$? Essayez de vous en souvenir. Le premier terme est x^k, et le second est $kx^{k-1}h$. Pour le reste c'est un polynôme dont tous les termes contiennent au moins h^2. Nous avons donc

$$\clubsuit = \lim_{h\to 0} \frac{h^2 P(x,h)}{k!\|h\|} = 0. \tag{14.114}$$

Nous en concluons que u_k est différentiable et que

$$(du_k)_x(y) = u_{k-1}(x)y. \tag{14.115}$$

u_k est de classe C^1 Nous devons démontrer que la différentielle est continue ; cela est la continuité de l'application

$$\begin{aligned} du_k \colon A &\to \mathcal{L}(A,A) \\ x &\mapsto (du_k)_x. \end{aligned} \tag{14.116}$$

La topologie sur A est celle de la norme, et celle sur $\mathcal{L}(A,A)$ est celle de la norme opérateur associée à la norme sur A. Nous avons[12] :

$$\lim_{h\to 0} \|(du_k)_{x+h} - (du_k)_x\| = \lim_{h\to 0} \sup_{\|y\|=1} \|u_{k-1}(x+h)y - u_{k-1}(x)y\| \tag{14.117a}$$

$$\leqslant \lim_{h\to 0} \sup_{\|y\|=1} \|u_{k+1}(x+h) - u_{k-1}(x)\|\|y\| \tag{14.117b}$$

$$= \lim_{h\to 0} \|u_{k+1}(x+h) - u_{k-1}(x)\|. \tag{14.117c}$$

Le fait que cette limite valle zéro est maintenant la continuité de u_{k-1}.

Convergence normale sur tout compact Soit un compact K de A. Par le théorème de Borel-Lebesgue 7.127, K est fermé et borné. C'est pour ceci que nous avons supposé que A était de dimension finie sur $\mathbb{R}$. Soit donc $R > 0$ tel que $\|y\| < R$ pour tout $y \in K$. Nous avons

$$\|du_k\|_K = \sup_{x\in K} \|(du_k)_x\| = \sup_{x\in K} \frac{\|x^{k-1}\|}{(k-1)!} \leqslant \sup_{x\in K} \frac{\|x\|^{k-1}}{(k-1)!} \leqslant \frac{R^{k-1}}{(k-1)!}. \tag{14.118}$$

Mais la série $\sum_{k=0}^{\infty} \frac{R^k}{k!}$ converge. Nous avons donc la convergence normale demandée.

Conclusion Le théorème 14.6 conclu que l'exponentielle est de classe C^1 et que sa différentielle est donnée par la formule

$$(d\exp)_x(y) = \sum_{k=0}^{\infty} (du_k)_x(y) = \sum_{k=1}^{\infty} (du_k)_x(y) = \sum_{k=0}^{\infty} u_k(x)y = \exp(x)y. \tag{14.119}$$

Notez le jeu d'indices : $du_k = 0$ lorsque $k = 0$ (ce qui permet de faire commencer la somme à 1) et ensuite du_k fait intervenir u_{k-1} (ce qui fait revenir le départ de la somme à $k = 0$).

□

14.38.
Lorsque nous disons que la différentielle de l'exponentielle est l'exponentielle elle-même, nous référons au point 14.37(3) : la différentielle de exp en x est l'opérateur de multiplication par $\exp(x)$.

Nous pouvons comprendre maintenant que exp est même de classe C^∞ parce qu'à chaque différentiation nous tombons sur la même fonction, laquelle est de classe au moins C^1.

Cependant, pour formaliser ça, il faut un peut travailler. Le cauchemar des différentielles successives d'une application $A \to A$ est que les espaces en jeu sont des emboîtements terribles de $\mathcal{L}(A, \mathcal{L}(A, \mathcal{L}(A,A)))$.

Ce qui nous sauve est que l'espace $\mathcal{L}(A,V)$ est un A-module, quel que soit V. En particulier lorsque V est lui-même déjà un emboîtement. Faisons un lemme pour voir comment ça fonctionne.

14.4.2 Différentielles

Lemme 14.39 ([1]).
Soient deux espaces vectoriels normés E et V tels que V soit un E-module[13]*. Nous supposons les normes soient telles que $\|xv\|_V \leqslant \|x\|_E\|v\|_V$.*

Soit une fonction différentiable $f \colon E \to V$ telle que la différentielle $df \colon E \to \mathcal{L}(E,V)$ soit de la forme

$$df_x(y) = yg(x) \tag{14.120}$$

12. N'oubliez pas de faire à part le cas $k = 0$ parce que ce qui suit n'est correct que pour $k \geqslant 1$.
13. Définition 3.67.

pour une certains fonction différentiable $g\colon E \to V$.

Alors f est C^1, et deux fois différentiable telle que

$$\begin{aligned} d^2 f\colon E &\to \mathcal{L}\big(E, \mathcal{L}(E,V)\big) \\ (d^2 f)_x(y)z &= z(dg_x)(y) \end{aligned} \tag{14.121}$$

pour tout $x, y, z \in E$.

Démonstration. En plusieurs étapes.

f est C^1 Nous savons, par hypothèse, que f est différentiable. Il faut montrer que sa différentielle est continue, en remarquant déjà que g est continue parce que différentiable.

Soit $x_k \xrightarrow{E} x$, et calculons $\|df_{x_k} - df_x\|$:

$$\begin{aligned} \|df_{x_k} - df_x\| &= \sup_{\|y\|=1} \|df_{x_k}(y) - df_x(y)\| \\ &= \sup_{\|y\|=1} \|\big(g(x_k) - g(x)\big)y\| \\ &\leqslant \sup_{\|y\|=1} \|g(x_k) - g(x)\|\|y\| \\ &= \|g(x_k) - g(x)\|. \end{aligned} \tag{14.122}$$

Donc nous avons bien $df_{x_k} \xrightarrow{\mathcal{L}(E,V)} df_x$, ce qui signifie la continuité de df. Donc f est de classe C^1.

f est deux fois différentiable Pour montrer que df est différentiable, nous mettons directement dans la définition (10.127) le candidat

$$\begin{aligned} T_x(h)\colon R &\to V \\ T_x(h)z &= z dg_x(y). \end{aligned} \tag{14.123}$$

Nous devons vérifier la limite suivante :

$$\lim_{h \xrightarrow{E} 0} \frac{df_{x+h} - df_x - T_x(h)}{\|h\|} = 0. \tag{14.124}$$

Étudions la norme du numérateur :

$$\|df_{x+h} - df_x - T_x(h)\| = \sup_{\|y\|=1} \|df_{x+h}(y) - df_x(y) - T_x(h)y\| \tag{14.125a}$$

$$= \sup_{\|y\|=1} \|yg(x+h) - yg(x) - ydg_x(h)\| \tag{14.125b}$$

$$\leqslant \sup_{\|y\|=1} \|y\|\|g(x+h) - g(x) - dg_x(h)\|. \tag{14.125c}$$

La limite (14.124) se déduit donc de la différentiabilité de g.

Note : la partie démontrant que f est C^1 n'est pas strictement obligatoire parce qu'en vérifiant que f est deux fois différentiable, nous vérifions de facto que df est en particulier continue. □

Lemme 14.40 ([1]).
Soient des algèbres normées A et V telles que V soit un A-module vérifiant $\|xv\| \leqslant \|x\|\|v\|$ pour tout $x \in A$ et $v \in V$. Alors $\mathcal{L}(A,V)$ est un A-module vérifiant $\|x\alpha\| \leqslant \|x\|\|\alpha\|$ pour tout $x \in A$ et $\alpha \in \mathcal{L}(A,V)$.

Démonstration. C'est un simple calcul utilisant la norme opérateur :

$$\|x\alpha\| = \sup_{\|y\|=1} \|(x\alpha)y\| = \sup_{\|y\|=1} \|x\alpha(y)\| \leqslant \sup_{\|y\|=1} \|x\|\|\alpha(y)\| = \|x\| \sup_{\|y\|=1} \|\alpha(y)\| = \|x\|\|\alpha\|. \tag{14.126}$$

□

Proposition 14.41 ([1]).
La fonction $\exp\colon A \to A$ *est de classe* C^∞ *et vérifie, pour tout* $k \geqslant 1$ *la récurrence*

$$(d^k \exp)_x(y) = y(d^{k-1}\exp)_x. \tag{14.127}$$

Démonstration. La formule proposée fonctionne avec $k = 1$:

$$(d\exp)_x(y) = y\exp(x). \tag{14.128}$$

C'est la relation 14.107.

Nous considérons $k > 1$, nous supposons que exp est de classe C^{k-1} et k fois différentiable. Nous allons prouver que exp est alors de classe C^k et $k+1$ fois différentiable, et que la différentielle de $d^k\exp$ est donné par la formule

$$(d^{k+1}\exp)_x(y) = y(d^k\exp)_x. \tag{14.129}$$

Pour nous mettre au clair avec les espaces en présence, nous supposons que

$$d^{k-1}\exp\colon A \to \mathcal{L}(A, V) \tag{14.130a}$$
$$d^k\exp\colon A \to \mathcal{L}\big(A, \mathcal{L}(A, V)\big) \tag{14.130b}$$

pour un certain espace vectoriel normé V, lequel est un de ces terrifiants emboîtement de type $\mathcal{L}\Big(A, \mathcal{L}\big(A, \mathcal{L}(A, A)\big)\Big)$. Il est bien un espace vectoriel normé, et également un A-module parce qu'on peut toujours définir la multiplication d'un élément $v \in V$ par un élément $x \in A$ comme étant la multiplication par x du résultat final de l'évaluation emboîtée, laquelle se termine par un élément de A. Donc tout se met bien.

Quoi qu'il en soit, nous posons

$$T_x(y) = y(d^k\exp)_x \tag{14.131}$$

et nous vérifions ce que cela donne dans la définition de la différentielle. Si nous avons

$$\lim_{h\to 0} \frac{(d^k\exp)_{x+h} - (d^k\exp)_x - T_x(h)}{\|h\|} = 0 \tag{14.132}$$

alors nous aurons prouvé tout ce qu'il nous faut.

Le numérateur est une application $A \to \mathcal{L}(A, V)$; nous en écrivons la norme comme il se doit :

$$\|(d^k\exp)_{x+h} - (d^k\exp)_x - T_x(h)\| = \sup_{\|y\|=1} \|(d^k\exp)_{x+h}(y) - (d^k\exp)_x(y) - h(d^k\exp)_x y\| \tag{14.133a}$$
$$= \sup_{\|y\|=1} \|y(d^{k-1}\exp)_{x+h} - y(d^{k-1}\exp)_x - h(d^k\exp)_x y\| \tag{14.133b}$$
$$= \sup_{\|y\|=1} \|y(d^{k-1}\exp)_{x+h} - y(d^{k-1}\exp)_x - hy(d^{k-1}\exp)_x\| \tag{14.133c}$$
$$\leqslant \|(d^{k-1}\exp)_{x+h} - (d^{k-1}\exp)_x - h(d^{k-1}\exp)_x\| \tag{14.133d}$$
$$= \|(d^{k-1}\exp)_{x+h} - (d^{k-1}\exp)_x - (d^k\exp)_x(h)\|. \tag{14.133e}$$

Dans ce calcul nous avons utilisé le lemme 14.40 et $T_x(h)y = h(d^k\exp)_x y$. Maintenant, la limite

$$\lim_{h\to 0} \frac{\|(d^{k-1}\exp)_{x+h} - (d^{k-1}\exp)_x - (d^k\exp)_x(h)\|.}{\|h\|} \tag{14.134}$$

n'est rien d'autre que la limite arrivant dans la définition du fait que $d^k\exp$ est la différentielle de $d^{k-1}\exp$. Cette limite est donc zéro comme nous voulions le prouver. □

Le théorème suivant est très important parce qu'il permet de définir l'exponentielle d'une matrice. Et les exponentielles de matrices sont utiles, entre très nombreuses autres choses pour résoudre certaines équations différentielles.

Théorème-définition 14.42 ([1]).
Soit une algèbre normée A (pas spécialement commutative). La formule

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{14.135}$$

définit une fonction différentiable dont la différentielle est donnée par[14]

$$(d\exp)_x(y) = \sum_{j,j\in\mathbb{N}} \frac{x^i y x^j}{(i+j+1)!} \tag{14.136}$$

14.43.
Nous ne démontrons pas cela ici.

Il s'agit d'une adaptation de la proposition 14.37. Là où il faut faire attention, c'est dans l'équation (14.113) : il n'y a pas k termes $x^{k-1}h$ dans $(x+h)^k$, mais k termes de la forme $x^i h x$. C'est pour cela que la différentielle n'est pas donnée par $T(y) = u_{k-1}(x)y$, mais bien par la somme (14.136).

M'est avis en réalité que toute la démonstration du théorème 14.116 passe facilement au cas présent.

14.4.3 Séries dans une algèbre normée

Nous allons parler d'exponentielle de matrice. Avant cela, quelques propriétés qui sont valables sur des algèbres normées. Le principal exemple que nous avons en tête est $\mathbb{A} = \mathbb{M}(n, \mathbb{C})$.

Proposition 14.44 ([1]).
Soit une algèbre normée $\mathbb{A}$. Soient une suite d'éléments $A_k \in \mathbb{A}$ et un élément B. Nous supposons que la somme $\sum_{k=0}^{\infty} A_k$ converge. Alors

$$B\sum_k A_k = \sum_k (BA_k). \tag{14.137}$$

Démonstration. Soit $N \in \mathbb{N}$. Nous avons :

$$\| \sum_{k=0}^{N} BA_k - B\sum_{k=0}^{\infty} A_k \| = \| B \sum_{k=N+1}^{\infty} A_k \| \tag{14.138a}$$

$$\leqslant \|B\| \| \sum_{k=N+1}^{\infty} A_k \| \tag{14.138b}$$

Justifications :

— Pour (14.138a), c'est la linéarité du produit matriciel.

— Pour (14.138b), c'est que la norme est une norme d'algèbre[15].

À droite, la limite $N \to \infty$ donne zéro parce que $\|B\|$ est un simple nombre, et $\| \sum_{k=N+1}^{\infty} A_k \|$ est une queue de suite convergente par hypothèse.

Nous avons donc bien convergence

$$\lim_{N\to\infty} \sum_{k=0}^{N} BA_k = B\sum_{k=0}^{\infty} A_k. \tag{14.139}$$

□

14. La fonction exponentielle est, j'en suis quasiment certain, de classe C^∞. Si vous connaissez un moyen pas trop douloureux de prouver cela, faites-le moi savoir.

15. Définition 10.7. Pour rappel, la norme opérateur en est une par le lemme 10.13.

Nous adaptons le produit de Cauchy du théorème 14.15 au cas d'une algèbre normée.

Proposition 14.45 ([1]).
Soient une algèbre normée $\mathbb{A}$, un élément $A \in \mathbb{A}$, ainsi que des séries convergentes $\sum_{k=0}^{\infty} a_k A^k$ et $\sum_{l=0}^{\infty} b_l A^l$. Alors

$$\left(\sum_k a_k A^k\right)\left(\sum_l b_l A^l\right) = \sum_{n=0}^{\infty}\big(\sum_{m=0}^{n} a_m b_{n-m}\big)A^n. \tag{14.140}$$

Démonstration. Un calcul :

$$\left(\sum_k a_k A^k\right)\left(\sum_l b_l A^l\right) = \sum_k \big(\sum_l b_l A^l\big) a_k A^k \tag{14.141a}$$

$$= \sum_k \big(\sum_l b_l a_k A^{l+k}\big) \tag{14.141b}$$

$$= \lim_{K\to\infty} \sum_{k=0}^{K} \big(\lim_{L\to\infty} \sum_{l=0}^{L} b_l a_k A^{k+l}\big) \tag{14.141c}$$

$$= \lim_{K\to\infty} \lim_{L\to\infty} \sum_{k=0}^{K} \sum_{l=0}^{L} b_l a_k A^{k+l} \tag{14.141d}$$

$$= \lim_{K\to\infty} \lim_{L\to\infty} \sum_{n=0}^{K+L} \sum_{m=0}^{n} a_m b_{n-m} A^n \tag{14.141e}$$

$$= \lim_{K\to\infty} \sum_{n=0}^{\infty} \sum_{m=0}^{n} a_m b_{n-m} A^m \tag{14.141f}$$

$$= \sum_{n=0}^{\infty} \sum_{m=0}^{n} a_m b_{n-m} A^m \tag{14.141g}$$

Justifications :

— Pour (14.141a), la proposition 14.44 nous permet d'entrer l'élément $\sum_l b_l A^l \in \mathbb{A}$ dans la somme sur k.

— Pour (14.141b), c'est la même chose.

— Pour (14.141d), la somme sur k étant finie (pour chaque K), elle commute avec la limite sur L.

— Pour (14.141e), c'est une manipulation de sommes finies. On regroupe les termes selon les puissances de A.

— Pour (14.141f), c'est effectuer la limite sur L pour K fixé.

— Pour (14.141g), l'expression dans la limite sur K ne dépend pas de K. Donc nous pouvons simplement supprimer la limite.

□

14.4.4 Exponentielle de matrice

Proposition 14.46 ([307]).
Soient des matrices $A, P \in \mathbb{M}(n, \mathbb{C})$ telle que P soit inversible. Alors[16]

$$e^{P^{-1}AP} = P^{-1}e^A P. \tag{14.142}$$

Démonstration. Pour chaque $m \in \mathbb{N}$ nous avons $(P^{-1}AP)^m = P^{-1}A^m P$. Ensuite,

$$e^{P^{-1}AP} = \sum_k \frac{(P^{-1}AP)^k}{k!} = \sum_k \frac{P^{-1}A^k P}{k!} = P^{-1}\sum_k \frac{A^k}{k!}P. \tag{14.143}$$

16. La définition de l'exponentielle de matrice est 14.37 où la convergence de la somme est celle de la norme opérateur 10.3.

Nous avons utilisé la proposition 14.44 pour sortir P^{-1} à gauche et P à droite de la somme. □

Proposition 14.47 ([307]).
L'exponentielle de matrice vérifie

(1) $e^0 = \mathrm{Id}$

(2) $A^m e^A = e^A A^m$

(3) $(e^A)^t = e^{(A^t)}$

(4) Si $AB = BA$ *alors* $Ae^B = e^B A$ *et* $e^A e^B = e^B e^A$.

Démonstration. Point par point.

Pour (1) Juste substituer $A = 0$ dans la définition. Tous les termes tombent sauf le premier. Il faut utiliser le fait que $A^0 = \mathrm{Id}$.

Pour (2) Il faut utiliser la proposition 14.44 pour écrire

$$A^m \sum_k \frac{A^k}{k!} = \sum_k \frac{A^m A^k}{k!} = \sum_k \frac{A^k A^m}{k!} = \sum_k \frac{A^k}{k!} A^m. \tag{14.144}$$

Pour (3) Pour chaque k nous avons l'égalité $(A^k)^t = (A^t)^k$. En utilisant encore le coup de la queue de suite qui converge vers zéro,

$$\| \sum_{k=0}^{N} \frac{(A^t)^k}{k!} - (e^A)^t \| = \| \sum_{k=N+1}^{\infty} \frac{(A^t)^k}{k!} \| \to 0. \tag{14.145}$$

Pour (4) Pour prouver $Ae^B = e^B A$, c'est le même genre de manipulations que (14.144).

Maintenant, vu que A et e^B commutent, l'égalité à peine prouvée montre que e^A et e^B commutent.

□

Proposition 14.48 ([307]).
Soient $A \in \mathbb{M}(n, \mathbb{C})$ *ainsi que* $s, t \in \mathbb{C}$. *Alors*

$$e^{sA} e^{tA} = e^{(s+t)A}. \tag{14.146}$$

Démonstration. Nous calculons le produit $e^{sA} e^{tA}$ par le produit de Cauchy de la proposition 14.45 :

$$\clubsuit = \left(\sum_k \frac{t^k}{k!} A^k \right) \left(\sum_l \frac{s^l}{l!} A^l \right) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{t^m}{m!} \frac{s^{n-m}}{(n-m)!} A^n. \tag{14.147}$$

À ce point, nous multiplions et divisons par $n!$ et nous réarrangons la somme de la façon suivante :

$$\clubsuit = \sum_{n=0}^{\infty} \frac{A^n}{n!} \sum_{m=0}^{n} \frac{n!}{m!(n-m)!} t^m s^{n-m}. \tag{14.148}$$

Nous reconnaissons la somme sur m comme étant un binôme de Newton [17] pour $(t+s)^n$. Nous avons donc finalement

$$\clubsuit = \sum_{n=0}^{\infty} \frac{\big((t+s)A\big)^n}{n!} = e^{(t+s)A}. \tag{14.149}$$

□

La proposition suivante dit que les exponentielles de matrices sont inversibles. Elle ne dit pas que toutes les matrices inversibles sont des exponentielles. Ce sera la proposition 14.96.

17. Proposition 3.47.

Proposition 14.49 ([307]).
Si $A \in \mathbb{M}(n, \mathbb{C})$, alors e^A est inversible et

$$(e^A)^{-1} = e^{-A}. \tag{14.150}$$

Démonstration. Il suffit de prendre $s = 1$ et $t = -1$ dans la proposition 14.48 et nous avons

$$e^A e^{-A} = e^0 = \mathbb{1}. \tag{14.151}$$

Cela prouve que e^A est inversible et que son inverse est e^{-A}. □

Proposition 14.50 ([307]).
Soit $A \in \mathbb{M}(n, \mathbb{C})$. Nous considérons l'application

$$\begin{aligned} f\colon \mathbb{R} &\to \mathrm{GL}(n, \mathbb{C}) \\ t &\mapsto e^{tA}. \end{aligned} \tag{14.152}$$

Nous avons la formule de dérivation

$$f'(t) = Ae^{tA}. \tag{14.153}$$

Démonstration. Il s'agit de calculer la limite en utilisant la proposition 14.48 :

$$\begin{aligned} f'(t) &= \lim_{\epsilon \to 0} \frac{e^{(t+\epsilon)A} - e^{tA}}{\epsilon} && (14.154a) \\ &= e^{tA} \lim_{\epsilon \to 0} \frac{e^{\epsilon A} - \mathbb{1}}{\epsilon} && (14.154b) \\ &= e^{tA} \lim_{\epsilon \to 0} \frac{1}{\epsilon} \big(\epsilon A + \sum_{k=2}^{\infty} \frac{(\epsilon A)^k}{k!}\big) && (14.154c) \\ &= e^{tA} \lim_{\epsilon \to 0} \big(A + \sum_{k=2}^{\infty} \frac{\epsilon^{k-1} A}{k!}\big) && (14.154d) \\ &= e^{tA} A && (14.154e) \\ &= Ae^{tA} && (14.154f) \end{aligned}$$

où à la toute fin nous avons aussi utilisé la commutation de la proposition 14.47(2). □

Le théorème suivant montre que le produit d'exponentielle de matrices suit la règle usuelle tant que les matrices commutent. Cela est cependant plutôt l'exception que la règle. À priori nous avons $e^A e^B \neq e^{A+B}$.

Théorème 14.51 ([307]).
Soient $A, B \in \mathbb{M}(n, \mathbb{C})$ telles que $AB = BA$. Alors

$$e^{A+B} = e^A e^B. \tag{14.155}$$

Démonstration. Vu que A et B commutent nous avons $Ae^{tB} = e^{tB}A$ (proposition 14.47(4)). Ensuite nous posons

$$g(t) = e^{t(A+B)} e^{-tB} e^{-tA}. \tag{14.156}$$

Nous calculons la dérivée de g en utilisant la règle de Leibnitz et la proposition 14.50 :

$$\begin{aligned} g'(t) = &(A + B)e^{t(A+B)} e^{-tB} e^{-tA} \\ &+ e^{t(A+B)}(-B)e^{-tB} e^{-tA} \\ &+ e^{t(A+B)} e^{-tB}(-A)e^{-tA}. \end{aligned} \tag{14.157}$$

Vu que A, B et $A+B$ commutent, nous pouvons réarranger les facteurs en

$$\begin{aligned} g'(t) &= (A+B)e^{t(A+B)}e^{-tB}e^{-tA} \\ &- Be^{t(A+B)}e^{-tB}e^{-tA} \\ &- Ae^{t(A+B)}e^{-tB}e^{-tA}. \end{aligned} \tag{14.158}$$

Enfin, cela fait

$$g'(t) = (A+B-B-A)e^{t(A+B)}e^{-tB}e^{-tA} = 0. \tag{14.159}$$

Donc g est constante et nous avons

$$e^{t(A+B)}e^{-tB}e^{-tA} = g(0) = \mathbb{1}. \tag{14.160}$$

En multipliant à droite par $e^{tA}e^{tB}$ nous trouvons

$$e^{t(A+B)} = e^{tA}e^{tB} \tag{14.161}$$

comme annoncé. □

14.5 Exponentielle et logarithme dans les réels

Pour avoir une vue synthétique du plan, voir le thème 58.

14.5.1 L'équation différentielle

Pour la suite nous notons y une solution de l'équation $y' = y$, $y(0) = 1$, et nous allons en donner des propriétés indépendamment de l'existence, donnée par le théorème 14.54.

Proposition 14.52.
Quelques propriétés de y (si elle existe) :

(1) Pour tout $x \in \mathbb{R}$ nous avons $y(x)y(-x) = 1$.

(2) $y(x) > 0$ pour tout x.

(3) y est strictement croissante.

Démonstration. Nous posons $\varphi(x) = y(x)y(-x)$ et nous dérivons :

$$\varphi'(x) = y'(x)y(-x) - y(x)y'(-x) = 0. \tag{14.162}$$

Donc φ est constante[18]. Vu que $\varphi(0) = 1$ nous avons automatiquement $y(x)y(-x) = 1$ pour tout x.

Les deux autres allégations sont simples : si $y(x_0) < 0$ alors il existe $t \in]x_0, 1[$ tel que $y(t) = 0$, ce qui est impossible parce que $y(t)y(-t) = 1$. La stricte croissance de y s'ensuit. □

Proposition 14.53.
Quelques formules pour tout $a, b \in \mathbb{R}$ et $n \in \mathbb{Z}$:

(1) $y(a+b) = y(a)y(b)$

(2) $y(na) = y(a)^n$

(3) $y\left(\frac{a}{n}\right) = \sqrt[n]{y(a)}$.

Démonstration. Nous posons $h(x) = y(a+b-x)y(x)$ et nous avons encore $h'(x) = 0$ dont nous déduisons que h est constante. De plus

$$h(0) = y(a+b)y(0) = y(a+b) \tag{14.163}$$

18. Proposition 11.125.

et

$$h(b) = y(a)y(b). \tag{14.164}$$

Vu que h est constante, ces deux expressions sont égales : $y(a+b) = y(a)y(b)$.

Forts de cette relation, une récurrence donne $y(na) = y(a)^n$ pour tout $n \in \mathbb{N}$. De plus

$$y(a) = y\left(\frac{a}{n} \times n\right) = y\left(\frac{a}{n}\right)^n, \tag{14.165}$$

ce qui donne $y(a) = y(a/n)^n$ ou encore $y(a/n) = \sqrt[n]{y(a)}$.

Enfin pour les négatifs, si $n \in \mathbb{N}$,

$$y(-na) = \frac{1}{y(na)} = \frac{1}{y(a)^n} = y(a)^{-n}. \tag{14.166}$$

Et de la même façon,

$$y\left(-\frac{a}{n}\right) = \frac{1}{y\left(\frac{a}{n}\right)} = \sqrt[n]{\frac{1}{y(a)}} = \sqrt[-n]{y(a)}. \tag{14.167}$$

□

14.5.2 Existence

Jusqu'ici nous avons donné des propriétés d'une éventuelle fonction y qui vérifierait l'équation différentielle. Il est temps de montrer qu'une telle fonction existe.

Théorème 14.54.
La série entière

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{14.168}$$

définit une fonction dérivable solution de

$$\begin{cases} y' = y & \text{(14.169a)} \\ y(0) = 1. & \text{(14.169b)} \end{cases}$$

Démonstration. La formule de Hadamard (théorème 14.13) donne le rayon de convergence de la série (14.168) par

$$\frac{1}{R} = \lim_{k\to\infty} \frac{\frac{1}{(k+1)!}}{\frac{1}{k!}} = \lim_{k\to\infty} \frac{1}{k+1} = 0. \tag{14.170}$$

Donc nous avons un rayon de convergence infini. La fonction y est définie sur $\mathbb{R}$ et la proposition 14.25 nous dit que y est dérivable. Nous pouvons aussi dériver terme à terme :

$$y'(x) = \sum_{k=0}^{\infty} \frac{kx^{k-1}}{k!} = \sum_{k=1}^{\infty} \frac{kx^{k-1}}{k!} = \sum_{k=1}^{\infty} \frac{x^{k-1}}{(k-1)!} = \sum_{k=0}^{\infty} \frac{x^k}{k!} = y(x). \tag{14.171}$$

Notez le petit jeu d'indice de départ de k. Dans un premier temps, nous remarquons que $k = 0$ donne un terme nul et nous le supprimons, et dans un second temps nous effectuons la simplification des factorielles (qui ne fonctionne pas avec $k = 0$). □

14.55.
Nous savons que la fonction y existe parce qu'une solution de l'équation différentielle $y' = y$, $y(0) = 1$ est donnée par la fameuse série (théorème 14.54). À part cela, ce qui a été fait avec cette équation différentielle ne permet pas de prouver l'existence de y. Donc, du point de vue de « définir l'exponentielle par son équation différentielle », c'est pas encore gagné. Notons au passage que le nombre e n'est pas encore bien défini via l'équation différentielle.

14.5.3 Le nombre de Neper e

Nous savons par le théorème 14.54 que $x \mapsto \exp(x)$ est une solution de l'équation différentielle exponentielle (avec la bonne condition initiale). Or une telle solution est unique par la proposition 11.358.

Définition 14.56 (Le nombre de Neper)**.**
Nous notons e le nombre $\exp(1)$.

Proposition 14.57.
Pour tout $x \in \mathbb{R}$, nous avons

$$\exp(x) = e^x. \tag{14.172}$$

Démonstration. Soit y vérifiant la fameuse équation différentielle. Nous savons que $y = \exp$ parce que c'est l'unique solution (proposition 11.358). Nous avons :

$$y(x) = y(1)^x. \tag{14.173}$$

Si $q \in \mathbb{Q}$ alors $q = a/b$ et

$$y(q) = y\left(\frac{a}{b}\right) = y\left(a \times \frac{1}{b}\right) = y\left(\frac{1}{b}\right)^a = \left(\sqrt[b]{y(1)}\right)^a = y(1)^{a/b} = y(1)^q. \tag{14.174}$$

Le résultat est prouvé pour les rationnels.

En ce qui concerne un élément général $x \in \mathbb{R}$, la fonction $x \mapsto y(x)$ est continue sur $\mathbb{R}$, et la fonction $x \mapsto e^x$ également (proposition 11.340). Ces deux fonctions étant égales sur $\mathbb{Q}$, elles sont égales sur $\mathbb{R}$ par la proposition 7.269). □

Une conséquence des propositions 14.57 et 11.335 est que

$$\lim_{x \to -\infty} e^x = 0 \tag{14.175a}$$

$$\lim_{x \to +\infty} e^x = +\infty, \tag{14.175b}$$

et en particulier,

$$\begin{aligned} \exp\colon \mathbb{R} &\to]0, \infty[\\ x &\mapsto e^x \end{aligned} \tag{14.176}$$

est une bijection.

14.5.4 Application réciproque : logarithme

Proposition-définition 14.58.
L'application $\exp\colon \mathbb{R} \to]0, \infty[$ *est une bijection. L'application réciproque*

$$\ln\colon]0, \infty[\to \mathbb{R} \tag{14.177}$$

est le ***logarithme****.*

Démonstration. Le fonction exponentielle est dérivable, toujours strictement positive, donc strictement croissante. Les limites en $\pm\infty$ sont 0 et $+\infty$. Le théorème des valeurs intermédiaires 11.50 nous dit que c'est une bijection. En effet, l'injectivité est la stricte croissance. En ce qui concerne la surjection, soit $y \in]0, \infty[$. Vu que la limite en $-\infty$ est zéro, il existe $A \in \mathbb{R}$ tel que $\exp(x) < y$ pour tout $x < A$, et de la même façon, il existe $B \in \mathbb{R}$ tel que $\exp(x) > y$ pour tout $x > B$. Si $a < A$ et $b > B$ alors $\exp(a) < y$ et $\exp(b) > y$, donc y est dans l'image de $[a, b]$ par l'exponentielle. □

Lemme 14.59 ([1])**.**
Le logarithme est une fonction continue.

Démonstration. C'est une conséquence du théorème de la bijection 11.61(4), et de la continuité de l'exponentielle sur $\mathbb{R}$, qui est une partie du théorème 14.54. □

Proposition 14.60 ([1]).
Pour tout $x, y \in \mathbb{R}$ et pour $a > 0$ nous avons

$$\ln(\frac{1}{x}) = -\ln(x), \tag{14.178}$$

et

$$\ln(xy) = \ln(x) + \ln(y), \tag{14.179}$$

et

$$\ln(a^x) = x\ln(a) \tag{14.180}$$

et

$$a^x = e^{x\ln(a)}. \tag{14.181}$$

Démonstration. Nous avons, par la proposition 11.341,

$$e^{-\ln(x)} = \frac{1}{e^{\ln(x)}} = \frac{1}{x}. \tag{14.182}$$

En prenant le logarithme des deux côtés nous trouvons

$$-\ln(x) = \ln\left(\frac{1}{x}\right). \tag{14.183}$$

Nous pouvons continuer avec la suivante.

Par définition, $\ln(xy)$ est donné par $\exp\big(\ln(xy)\big) = xy$. Mais nous avons aussi, par la proposition 11.341 :

$$e^{\ln(x)+\ln(y)} = e^{\ln(x)}e^{\ln(y)} = xy. \tag{14.184}$$

Nous avons donc démontré (14.179).

La relation (14.180) de démontre d'abord pour $x \in \mathbb{N}$, puis pour $x \in \mathbb{Q}$ et enfin pour $x \in \mathbb{R}$. Si $n \in \mathbb{N}$ alors la relation (14.179) donne immédiatement

$$\ln(a^n) = n\ln(a). \tag{14.185}$$

pour tout $x \in \mathbb{R}$.

Si $m, n \in \mathbb{N}$, le nombre $a^{n/m}$ est par définition le $x > 0$ tel que

$$x^m = a^n. \tag{14.186}$$

En prenant le logarithme des deux côtés : $\ln(x^m) = \ln(a^n)$ et en utilisant la relation déjà démontrée pour $\mathbb{N}$ nous trouvons $m\ln(x) = n\ln(a)$ et donc

$$\ln(a^{m/n}) = \ln(x) = \frac{m}{n}\ln(a). \tag{14.187}$$

La relation est donc démontré pour $\ln(a^q)$ avec $q \in \mathbb{Q}^+$.

Nous passons à $q = -m/n \in \mathbb{Q}^-$, c'est-à-dire toujours $m, n \in \mathbb{N}$. Nous avons, en utilisant la proposition 14.60,

$$\ln(a^{-q}) = \ln(\frac{1}{a^q}) = -\ln(a^q) = -q\ln(a). \tag{14.188}$$

Enfin si $x \in \mathbb{R}$ nous considérons une suite de rationnels $x_k \to x$. Pour chaque k nous avons

$$\ln(a^{x_k}) = x_k\ln(a). \tag{14.189}$$

Nous prenons la limite deux deux côtés. À droite nous avons tout de suite $x\ln(a)$, et à gauche, par continuité de la fonction ln (lemme 14.59) et de la fonction puissance (définition 11.340) nous trouvons $\ln(a^x)$. □

La formule (14.180) en particulier est pratique pour réexprimer des fonctions puissances compliquées en écrivant

$$a^x = e^{\ln(a^x)} = e^{x\ln(a)}. \tag{14.190}$$

Cela aide à calculer la dérivée de $x \mapsto a^x$.

Notons que certains prennent (14.190) comme définition de la fonction puissance.

14.5.5 Approximations numériques de e

Nous donnons maintenant quelques approximations numériques de e, particulièrement inefficaces.

Lemme 14.61.
Nous avons

$$2 < e < 3. \tag{14.191}$$

Démonstration. Nous savons que $y(0) = 1$ et $y'(0) = 1$. La fonction y est strictement croissante (et donc sa dérivée aussi). Nous avons donc $y'(x) > 1$ pour tout $x \in]0, 1]$, et donc

$$y(1) > 1 + 1 \times 1 = 2. \tag{14.192}$$

Sachant que $2 > y'(x)$ pour tout $x \in]0, 1[$ nous pouvons refaire le coup de l'approximation affine, cette fois en majorant :

$$y(1) < 1 + 2 \times 1 = 3. \tag{14.193}$$

□

De la même façon nous savons que

$$y(\frac{1}{n}) > 1 + \frac{1}{n} \tag{14.194}$$

parce que y' est minoré par 1 sur $]0, \frac{1}{n}[$. Avec cela nous avons aussi la majoration

$$y(\frac{1}{n}) < 1 + \frac{1}{n} \times \left(1 + \frac{1}{n}\right) = 1 + \frac{1}{n} + \frac{1}{n^2}. \tag{14.195}$$

Et enfin nous pouvons donner l'encadrement, valable pour tout n :

$$\left(1 + \frac{1}{n}\right)^n < y(1) < \left(1 + \frac{1}{n} + \frac{1}{n^2}\right)^n. \tag{14.196}$$

Pour $n = 10$ nous trouvons

$$2.50 < e < 2.83. \tag{14.197}$$

Bien que ce soit à mon avis humainement pas possible à faire à la main nous avons, pour $n = 100$:

$$2.70 < e < 2.7317 \tag{14.198}$$

Cela reste un encadrement très modeste.

Une méthode plus efficace consiste à calculer directement le développement de définition

$$e = \exp(1) = \sum_{k=0}^{\infty} \frac{1}{n!}. \tag{14.199}$$

```
def u(k):
    """
    return the kth term in the expansion of 'e'
    """
    return 1/factorial(k)

def sum_u(n):
    """
    return the sum of the 'n' first terms, that is with
    k froim 0 to n-1.
    """
```

```
    L=[  u(k) for k in range(0,n)  ]
    return sum(L)

s = sum_u(5)          # This is a fraction
print(s)
print( numerical_approx(s)  )
```

tex/sage/sageSnip013.sage

Problèmes et choses à faire

Comment trouver, avec cette méthode, un *encadrement pour e ?*

Ce petit programme, avec 5 termes donne $e \simeq 65/24 \simeq 2.708$. Avouez que c'est déjà bien mieux.

14.5.6 Résumé des propriétés de l'exponentielle

Théorème 14.62.
Les choses que nous savons sur l'exponentielle :

(1) Il y a unicité de la solution à l'équation différentielle

$$\begin{cases} y' = y & (14.200a) \\ y(0) = 1. & (14.200b) \end{cases}$$

(2) L'équation différentielle (14.200) possède une solution donnée par la série entière

$$\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{14.201}$$

(3) Cette solution est une bijection $y\colon \mathbb{R} \to]0,\infty[$.

(4) La fonction y ainsi définie est de classe C^∞.

(5) Elle est également donnée par la formule

$$\exp(x) = e^x \tag{14.202}$$

où e est définit par $e = \exp(1)$.

(6) Elle vérifie

$$e^{a+b} = e^a e^b \tag{14.203}$$

Nous nommons **exponentielle** cette fonction.

Démonstration. Point par point.

(1) C'est la proposition 11.358.

(2) C'est le théorème 14.54.

(3) Le rayon de convergence de la série (14.201) est infini (théorème 14.54) ; elle est donc définie sur $\mathbb{R}$. Le fait que ce soit une bijection est dû au fait qu'elle est strictement croissante (proposition 14.52) ainsi qu'aux limites (14.175).

(4) Vu que $y = y'$, y est dérivable. Mais comme y' est alors égale à une fonction dérivable, y' est dérivable. En dérivant l'égalité $y' = y$ nous obtenons $y'' = y'$ et le jeu continue.

(5) C'est la proposition 14.57.

(6) C'est la proposition 14.53(1).

□

Exemple 14.63 (Un endomorphisme sans polynôme annulateur[176]).
l'exponentielle permet de donner un exemple d'un endomorphisme n'ayant pas de polynôme annulateur [19] : l'endomorphisme de dérivation

$$\begin{aligned} D\colon C^\infty(\mathbb{R},\mathbb{R}) &\to C^\infty(\mathbb{R},\mathbb{R}) \\ f &\mapsto f' \end{aligned} \tag{14.204}$$

n'a pas de polynôme annulateur. En effet supposons que $P = \sum_{k=0}^{p} a_k X^k$ en soit un, et considérons les fonctions $f_\lambda\colon t \mapsto e^{\lambda t}$. Nous avons

$$0 = P(D)f_\lambda = \sum_k a_k D^k(f_\lambda) = \sum_k a_k \lambda^k f_\lambda = P(\lambda) f_\lambda. \tag{14.205}$$

Par conséquent λ est une racine de P pour tout $\lambda \in \mathbb{R}$. Cela implique que $P = 0$.

D'ailleurs si on y pense bien, cet exemple n'est qu'un habillage de l'exemple 9.95. △

Proposition 14.64.
Quelques propriétés du logarithme.

(1) Le logarithme est une application dérivable et strictement croissante.

(2) Le logarithme est la primitive de $x \mapsto \frac{1}{x}$ qui s'annule en $x = 1$.

Démonstration. Elle est donc bijective, d'inverse continue et dérivable par le théorème 11.61 et la proposition 11.120.

La dérivée de la fonction logarithme peut être calculée en utilisant la formule (11.278), mais aussi de façon plus piettone en écrivant l'expression suivante, valable pour tout $x \in \mathbb{R}$:

$$\ln\big(\exp(x)\big) = x, \tag{14.206}$$

que nous pouvons dériver en utilisant le théorème de dérivation des fonctions composées :

$$\ln'\big(\exp(x)\big)\exp'(x) = 1. \tag{14.207}$$

Mais $\exp'(x) = \exp(x)$, donc

$$\ln'(y) = \frac{1}{y} \tag{14.208}$$

pour tout y dans l'image de exp, c'est-à-dire pour tout y dans l'ensemble de définition de ln.

Par ailleurs, $\exp(0) = 1$ donc

$$\ln(1) = \ln\big(\exp(0)\big) = 0. \tag{14.209}$$

En ce qui concerne l'unicité d'une primitive s'annulant en $x = 1$, c'est le corolaire 11.137. □

14.5.7 Dérivée de la fonction puissance

Exemple 14.65.
Soit la fonction $f(x,y) = x^y$, définie en 11.340. Nous allons en calculer les dérivées partielles au point $(1,2)$. Notons que f n'est pas définie pour $x < 0$, mais que cela n'a pas d'importance parce que nous pouvons nous restreindre à un voisinage du point $(1,2)$. La première dérivée partielle est facile :

$$\partial_x f(1,2) = (yx^{y-1})_{(x,y)=(1,2)} = 2.$$

Pour la seconde, il faut utiliser les propriétés de l'exponentielle et du logarithme. D'abord le logarithme est par définition l'application réciproque de l'exponentielle (définition 14.58), donc

$$x^y = \exp\big(\ln(x^y)\big). \tag{14.210}$$

Ensuite nous calculons en utilisant la proposition 14.60 :

$$\partial_y f(1,2) = \partial_y \left(e^{y\ln x}\right)_{(x,y)=(1,2)} = \left(\ln x e^{y\ln x}\right)_{(x,y)=(1,2)} = \ln\left(1 - e^{2\ln(1)}\right) = 0.$$

△

19. Voir la définition 9.91 et ce qui suit.

Cet exemple est facilement généralisable aux fonctions de la forme $x \mapsto u(x)^{v(x)}$. Voici une proposition qui dit comment faire.

Proposition 14.66 ([1])**.**
Soit une fonction dérivable $u\colon \mathbb{R} \to \mathbb{R}$ et $a > 0$. Nous avons

$$(a^u)' = u' \ln(a) a^u. \tag{14.211}$$

Si de plus $u(x) > 0$ pour tout x, nous avons

$$(u^a)' = a u' u^{a-1}. \tag{14.212}$$

Démonstration. Nous considérons la fonction $f(x) = a^{u(x)}$. Vu que $f(x) > 0$ pour tout x, nous pouvons en prendre le logarithme et écrire l'égalité, valable pour tout x :

$$f(x) = e^{\ln(a^{u(x)})} = \exp\big(u(x)\ln(a)\big). \tag{14.213}$$

Sachant la dérivée de l'exponentielle, cela n'est rien d'autre que la dérivée d'une fonction composée :

$$f'(x) = \ln(a) u'(x) e^{u(x)\ln(a)}. \tag{14.214}$$

Pour l'autre, nous posons

$$g(x) = u(x)^a, \tag{14.215}$$

qui peut encore s'écrire sous la forme

$$g(x) = e^{a \ln\big(u(x)\big)}. \tag{14.216}$$

Ici encore, c'est la dérivée de fonctions composées qui donne le résultat. □

14.5.8 Dérivée du logarithme

Lemme 14.67.
Si $u\colon \mathbb{R} \to]0,\infty[$ est dérivable alors $\ln(u)' = \dfrac{u'}{u}$.

Démonstration. Cela est une conséquence du théorème de dérivation des fonctions composées : si $g(x) = \ln(u(x))$ alors

$$g'(x) = \ln'\big(u(x)\big)u'(x) = \frac{1}{u(x)}u'(x). \tag{14.217}$$

□

14.5.9 Taylor pour l'exponentielle

Proposition 14.68 (Développement de l'exponentielle)**.**
Pour tout entier n, il existe une fonction $\alpha\colon \mathbb{R} \to \mathbb{R}$ telle que $\lim_{t\to 0}\alpha(t) = 0$ et

$$e^x = \sum_{k=0}^{n} \frac{x^k}{k!} + x^n \alpha(x). \tag{14.218}$$

Démonstration. Il s'agit de la proposition 14.32 appliquée à la série entière (14.37). □

14.5.10 Analycité

Vu que $\exp(x)$ est défini par une série entière (définition 14.37) et vu la proposition 14.35, il n'est pas étonnant que exp soit analytique. Traitons ce cas.

Exemple 14.69 (Analycité de l'exponentielle)**.**
Soient $a \in \mathbb{R}$ et $R > 0$. Nous démontrons que exp est analytique sur $B(a, R)$. Si $f(x) = e^x$, alors $f^{(n)}(x) = e^x$ pour tout n (équation (14.169a)). Nous avons donc

$$|f^{(n)}(x)| < e^{a+R} \tag{14.219}$$

pour tout $x \in B(a, R)$. Nous partons de l'expression (14.34) du reste :

$$|R_n(x)| \leqslant \frac{M_n}{(n+1)!}|x-a|^{n+1} \leqslant \frac{e^{a+R}}{(n+1)!}R^{n+1}. \tag{14.220}$$

Mais nous avons la limite

$$\lim_{n\to\infty} \frac{R^{n+1}}{(n+1)!} = 0 \tag{14.221}$$

pour tout R.

Donc avec les polynômes de Taylor P_n calculés en a, nous avons $P_n \to \exp$ simplement sur $\mathbb{R}$.

Nous pouvons donc développer la fonction exponentielle autour de n'importe quel point, et avoir convergence des polynômes vers l'exponentielle sur tout $\mathbb{R}$. Vous accepterez cependant que si a et x sont éloignés, la convergence $P_n(x) \to \exp(x)$ peut être extrêmement lente. △

14.5.11 Autres propriétés et petits calculs

Lemme 14.70.
Si $a, b \in]0, \infty[$ alors

$$\ln(ab) = \ln(a) + \ln(b) \tag{14.222}$$

et

$$\ln\left(\frac{1}{b}\right) = -\ln(b). \tag{14.223}$$

Démonstration. Nous posons $f(x) = \ln(ax)$ qui est une fonction dérivable. Alors $f'(x) = \frac{a}{ax} = \frac{1}{x}$. Cette fonction f est donc une primitive de $\frac{1}{x}$ et il existe une constante K telle que

$$f(x) = \ln(x) + K. \tag{14.224}$$

Vu que $\ln(1) = 0$ nous avons $K = f(1) = \ln(a)$. Donc

$$\ln(ax) = \ln(x) + \ln(a). \tag{14.225}$$

En ce qui concerne la seconde formule à démontrer, nous avons

$$\ln(1) = \ln\left(\frac{1}{b}b\right) = \ln\left(\frac{1}{b}\right) + \ln(b). \tag{14.226}$$

Étant donné que $\ln(1) = 0$ nous en déduisons la formule (14.223). □

Lemme 14.71.
Si les suites (u_n) et (v_n) sont équivalentes[20] *et si (v_n) admet une limite l différente de 1, alors les suites $(\ln u_n)$ et $(\ln v_n)$ sont équivalentes.*

Démonstration. En effet si $u_n = v_n\alpha(n)$ alors en utilisant la formule du lemme 14.70,

$$\ln(u_n) = \ln(v_n) + \ln\big(\alpha(n)\big) = \ln(v_n)\left(1 + \frac{\ln\big(\alpha(n)\big)}{\ln(v_n)}\right), \tag{14.227}$$

et comme $\alpha(n) \to 1$, la parenthèse tend vers 1. □

20. Définition 7.136.

14.5.12 Taylor pour le logarithme

Vu que $\ln(0)$ n'existe pas, il n'est pas question de développer ln autour de $x = 0$. À la place, nous allons le développer autour de $x = 1$ et plus précisément nous allons étudier Taylor pour la fonction $f(x) = \ln(1 + x)$. Les résultats seront résumés dans la proposition 14.74.

Proposition 14.72 ([1]).
Soit la fonction

$$\begin{aligned} f\colon]-1,\infty[&\to \mathbb{R} \\ x &\mapsto \ln(1+x). \end{aligned} \tag{14.228}$$

Pour tout n, il existe une fonction $\alpha\colon \mathbb{R} \to \mathbb{R}$ telle que $\lim_{t\to 0} \alpha(t) = 0$ et

$$f(x) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} x^k + \alpha(x)x^n \tag{14.229}$$

pour tout x dans le domaine de f.

Notez la somme qui part de $k = 1$ et non $k = 0$.

Démonstration. Nous utilisons la formule de Taylor-Young (proposition 11.391). La première dérivée de f se calcule en utilisant le lemme 14.67 :

$$f'(x) = \frac{1}{1+x}. \tag{14.230}$$

Pour les dérivées suivantes, c'est juste du calcul et nous pouvons prouver par récurrence que

$$f^{(k)}(x) = \frac{(k-1)!(-1)^{k+1}}{(1+x)^k}. \tag{14.231}$$

En ce qui concerne l'évaluation en zéro :

$$f^{(k)}(0) = \begin{cases} 0 & \text{si } k = 0 \\ (k-1)!(-1)^{k+1} & \text{sinon.} \end{cases} \tag{14.232}$$

Du fait que $f^{(0)}(0) = \ln(1) = 0$, la somme commence à $k = 1$ et non $k = 0$. Nous avons

$$f(x) = \sum_{k=1}^{n} \frac{f^{(k)}(0)}{k!} x^k + \alpha(x)x^n = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} x^k + \alpha(x)x^n. \tag{14.233}$$

□

Nous étudions les polynômes de la série de Taylor pour

$$\begin{aligned} f\colon]-1,\infty[&\to \mathbb{R} \\ x &\mapsto \ln(1+x). \end{aligned} \tag{14.234}$$

Les dérivées successives de f ont déjà été calculées en (14.231). Nous développons autour de $x = 0$. Donc $f(0) = \ln(1) = 0$ et pour les autres,

$$f^{(k)}(0) = (-1)^{k+1}(k-1)!. \tag{14.235}$$

Pour les polynômes de Taylor, nous avons

$$P_n(x) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} x^k \tag{14.236}$$

où vous noterez la somme qui part de $k = 1$ et non de $k = 0$. Nous avons aussi la série de Taylor de f donnée par

$$T(x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k. \tag{14.237}$$

La somme est une limite ponctuelle, là où elle existe.

Jusqu'à présent, la seule certitude à props de T est que $T(0) = f(0) = 0$. Pour le reste :

— Rien ne dit que $T(x)$ existe pour d'autres x que $x = 1$.

— Et même si $T(x)$ existait pour d'autres x (c'est-à-dire si le rayon de convergence de (14.237) était strictement plus grand que zéro), rien n'assurerait que la valeur serait celle de f.

— Et même si $T(x)$ convergeait vers f sur son disque de convergence, ce ne serait pas encore assez pour dire que f est analytique, parce que l'analycité demande que les séries de Taylor autour de *chaque* point converge vers f. Or ici nous ne parlons encore que de T qui est la série autour de $x = 0$.

Lemme 14.73.
La série de Taylor de $x \mapsto \ln(1 + x)$ autour de $x = 0$ converge sur $]-1, 1]$. Elle ne converge pas pour $x = -1$.

Démonstration. En ce qui concerne le rayon de convergence de T, nous utilisons la formule de Hadamard [21] avec

$$a_k = \frac{(-1)^{k+1}}{k}. \tag{14.238}$$

Ce que nous trouvons est

$$\frac{1}{R} = \lim_{k\to\infty} \left|\frac{a_{k+1}}{a_k}\right| = \lim_{k\to\infty} \frac{k}{k+1} = 1. \tag{14.239}$$

Le rayon de convergence de T est donc 1. Nous avons donc que $P_n \to T$ sur $]-1, 1[$, et peut-être que $P_n \to T$ en $x = \pm 1$.

Pour $x = -1$. L'intuition nous dit que ce serait $\ln(0)$ qui n'est pas défini. C'est le cas parce que

$$P_n(-1) = \sum_{k=1}^{n} \frac{(-1)^{k+1}(-1)^k}{k} = -\sum_{k=1}^{n} \frac{1}{k}. \tag{14.240}$$

La limite $n \to \infty$ diverge. Donc T n'est pas définie en $x = -1$.

Pour $x = 1$ par contre,

$$P_n(1) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k}. \tag{14.241}$$

Le critère des séries alternées [22] nous donne la convergence de cette série. □

Nous savons maintenant que la série de Taylor T converge sur $]-1, 1]$, et que $T(0) = f(0) = \ln(1) = 0$. Le premier gros morceau intéressant vient maintenant : nous allons prouver que $T(x)$ converge vers ce que nous croyons, c'est-à-dire $\ln(1 + x)$ en personne.

Proposition 14.74.
Pour tout $x \in]-1, 1]$ nous avons

$$\ln(1 + x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k \tag{14.242}$$

De plus nous avons

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = \ln(2). \tag{14.243}$$

Démonstration. Il s'agit d'utiliser l'expression du reste fourni par le théorème 11.378. Pour tout $x \in]-1, \infty[$, il existe un $c \in]0, x[$ (le c dépend de x) tel que

$$P_n(x) - f(x) = \frac{f^{(n+1)}(c)}{(n+1)!} x^{n+1}. \tag{14.244}$$

21. Théorème 14.13.
22. Théorème 10.72.

Cela est parce que f est de classe C^∞. Calculons un peu :

$$P_n(x) - f(x) = \frac{f^{(n+1)}(c)}{(n+1)!}x^{n+1} \tag{14.245a}$$

$$= \frac{(-1)^n n!}{(1+c)^{n+1}}\frac{1}{(n+1)!}x^{n+1} \tag{14.245b}$$

$$= \frac{(-1)^n}{n+1}\left(\frac{x}{1+c}\right)^{n+1}. \tag{14.245c}$$

Lorsque $x > 1$, il n'y a aucune garantie sur la convergence de cela pour $n \to \infty$. Pour rappel, $c \in]0, x[$. Si par contre $x \in]-1, 1[$, alors nous savons que

$$\left|\frac{x}{1+c}\right| < 1, \tag{14.246}$$

et donc convergence $P_n(x) - f(x) \to 0$.

Jusqu'ici nous avons prouvé que pour la série de Taylor converge vers $\ln(1+x)$ pour $x \in]-1, 1[$. Nous avons également vu que la série converge pour $x = 1$. Donc la fonction

$$g(x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}x^k \tag{14.247}$$

est de continue sur $]-1, 1]$ et égale à $\ln(x+1)$ sur $]-1, 1[$. Vu que $f\colon x \mapsto \ln(x+1)$ est continue sur $]-1, \infty[$, nous avons également $g(1) = f(1) = \ln(2)$.

Ceci nous mène au dernier point de notre proposition : $g(1) = \ln(2)$ s'écrit précisément

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} = \ln(2). \tag{14.248}$$

□

14.75.
La formule (14.243) peut sembler très chouette pour trouver des approximations de $\ln(2)$. Le problème est qu'elle ne donne aucune idée de l'erreur commise en tronquant la série.

Vous pouvez, certre écrite

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} = \frac{47}{60} \simeq 0.78. \tag{14.249}$$

Ce calcul n'a aucune valeur pour affirmer que $\ln(2)$ doit être proche de 0.78. Ni même pour affirmer que $\ln(2) < 1$.

Avoir des valeurs numériques de $\ln(2)$ (c'est-à-dire que « chiffres corrects devant ou derrière la virgule ») demande d'avoir un encadrement. Cela doit donc se faire avec des formules de séries avec reste ; les formules exactes qui somment jusqu'à l'inifini sont inutiles pour avoir des approximations numériques.

Dans le cas de $\ln(2)$, une approximation numérique sera donnée à l'aide de Taylor avec reste intégrale dans la proposition 19.145.

Lemme 14.76.
Soit la fonction[23]

$$f(x) = \frac{\ln(1+x)}{x} \tag{14.250}$$

(1) Elle admet un prolongement de classe C^∞ sur $]-1, \infty[$.

(2) $f(0) = 1$.

23. Pour la définition du logarithme, c'est la définition 14.58.

La seconde condition étant évidemment avec un abus de notation entre f et son prolongement, parce que f n'est pas définie en zéro.

Démonstration. La difficulté étant de voir que f a un prolongement en zéro et qu'elle y est de classe C^∞.

La 14.74 nous donne l'égalité

$$\ln(1+x) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} x^k \tag{14.251}$$

pour tout $x \in]-1,0]$; en particulier pour $x = 0$. Nous faisons le petit calcul suivant :

$$\frac{1}{x}\ln(1+x) = \frac{1}{x}\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^n \tag{14.252a}$$

$$= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} x^{n-1} \tag{14.252b}$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^{k}}{k+1} x^{k}. \tag{14.252c}$$

Ce calcul n'est pas valable pour $x = 0$, mais ça ne nous empèche pas de poser

$$T(x) = \sum_{n=0}^{\infty} \frac{(-1)^{k}}{k+1} x^{k}, \tag{14.253}$$

qui, lui, est bien définie en zéro. Le rayon de convergence de la série T est égal à 1, de telle sorte que

$$T\colon]-1,1[\to \mathbb{R} \tag{14.254}$$

de classe C^∞, et est égale à f sur $]-1,1[\backslash\{0\}$.

La série T est donc le prolongement demandé. En ce qui concerne $f(0)$, c'est un abus pour écrire $T(0)$ qui vaut immédiatement 1. □

Notons qu'un calcul de limite

$$\lim_{x\to 0} \frac{\ln(1+x)}{x} \tag{14.255}$$

donnait la valeur $f(0) = 1$. Donc prolonger avec $f(0) = 1$ était la seule possibilité pour avoir une fonction continue. De là à dire que le prolongement ainsi créé est de classe C^∞, c'est une autre histoire, qui est résoue par les séries entières.

14.5.13 Développements et calcul de limites

Lors d'un calcul de limite, développer une partie d'une expression peut être utile.

Exemple 14.77.
À calculer :

$$\lim_{x\to 0} \frac{\ln(1+x)}{x}. \tag{14.256}$$

Cela est une indétermination de type $\frac{0}{0}$. Le développement limité du numérateur[24] nous donne une fonction $\alpha(x)$ telle que $\lim_{x\to 0}\alpha(x) = 0$ et

$$\frac{\ln(1+x)}{x} = \frac{x - \frac{x^2}{2} + x^2\alpha(x)}{x} = 1 - \frac{x}{2} + x\alpha(x). \tag{14.257}$$

Sur le membre de droite la limite est facile à calculer :

$$\lim_{x\to 0} \frac{\ln(1+x)}{x} = \lim_{x\to 0}\Big(1 - \frac{x}{2} + x\alpha(x)\Big) = 1. \tag{14.258}$$

△

24. Proposition 14.72.

14.6 Vitesses des puissances, de l'exponentielle et du logarithme

14.6.1 Un peu de théorie

Voici une série de résultats qui lient les vitesses des polynômes, du logarithme et de l'exponentielle.

Proposition 14.78.
Nous avons :

(1) *Pour tout $\alpha > 0$, il existe N tel que $\ln(n) \leqslant n^\alpha$ pour tout $n \geqslant N$.*

(2) *Pour tout $p > 0$ et tout $\alpha > 0$, il existe N tel que*

$$\ln(n)^p < n^\alpha \tag{14.259}$$

pour tout $n \geqslant N$.

(3) *Pour tout $n \geqslant 1$ nous avons la limite*

$$\lim_{x\to 0^+} x^n \ln(x) = 0. \tag{14.260}$$

(4) *Nous avons*

$$\lim_{x\to 0^+} \frac{\ln(1-x)}{x} = -1. \tag{14.261}$$

(5) *L'exponentielle croit plus vite que tout polynôme, et plus vite que que logarithme :*

$$\lim_{t\to\infty} e^{-t}(\ln t)^n t^\alpha = 0 \tag{14.262}$$

pour tout n et pour tout α.

(6) *Pour tout $n > 0$, nous avons la limite*

$$\lim_{x\to 0^+} x^n e^{1/x} = \infty. \tag{14.263}$$

Le point (1) et sa généralisation (2) nous font dire que le logarithme croît moins vite que n'importe quel polynôme.

Démonstration. En plusieurs parties.

Pour (1) En effet, nous avons, par la règle de l'Hospital (proposition 11.132),

$$\lim_{x\to\infty} \frac{x^\alpha}{\ln(x)} = \lim_{x\to\infty} \frac{\alpha x^{\alpha-1}}{1/x} = \lim_{x\to\infty} \alpha x^\alpha = \infty \tag{14.264}$$

quand $\alpha > 0$. La dérivée du logarithme est dans la proposition 14.64.

Pour (2)

Pour (3) Lorsque $x \neq 0$ nous avons

$$x^n \ln(x) = \frac{\ln(x)}{1/x^n}, \tag{14.265}$$

qui est un cas $\frac{\infty}{\infty}$. Nous nous en remettons à la règle de l'Hospital 11.133. D'abord nous nous assurons de la limite des dérivées :

$$\lim_{x\to 0^+} \frac{1/x}{-nx^{-n-1}} = \lim_{x\to 0^+} -\frac{1}{n}\frac{x^{n+1}}{x} = 0. \tag{14.266}$$

La règle de l'Hospital conclu à l'existence de la limite demandée et à son égalité à 0.

Pour (4) En effet, par la règle de l'Hospital 11.132,

$$\lim_{x\to 0} -\frac{\ln(1-x)}{x} = \lim_{x\to 0} -\frac{\frac{-1}{1-x}}{1} = \lim_{x\to 0} \frac{1}{1-x} = 1 \tag{14.267}$$

Pour (5)

Pour (6) Nous passons au logarithme :

$$\ln(x^n e^{1/x}) = \ln(x^n) + \ln(e^{1/x}) = n\ln(x) + \frac{1}{x} = \frac{nx\ln(x)+1}{x}. \tag{14.268}$$

Grâce à la limite déjà prouvée en (3), le numérateur tend vers 1 lorsque $x \to 0^+$. Donc le tout tend vers $+\infty$. Au final,

$$\lim_{x\to 0^+} x^n e^{1/x} = \lim_{x\to 0^+} e^{\ln(x^n e^{1/x})} = \infty. \tag{14.269}$$

□

Lemme 14.79.
Si P est un polynôme et si $a > 0$, alors

$$\lim_{x\to\infty} e^{-ax}P(x) = 0 \tag{14.270}$$

Démonstration. Nous prouvons par récurrence que pour tout n, nous avons $e^{-ax}x^n \to 0$. D'abord nous écrivons [25]

$$f(x) = e^{-ax}x = \frac{x}{e^{ax}}, \tag{14.271}$$

et ensuite la règle de l'Hospital 11.133 nous donne

$$\lim_{x\to\infty} \frac{x}{e^{ax}} = \lim_{x\to\infty} \frac{1}{ae^{ax}} = 0. \tag{14.272}$$

En ce qui concerne la récurrence, c'est encore la règle de l'Hospital :

$$\lim_{x\to\infty} \frac{x^n}{e^{ax}} = \frac{n}{a}\lim_{x\to\infty} \frac{x^{n-1}}{e^{ax}} = 0. \tag{14.273}$$

□

Exemple 14.80.
Le lemme 11.352 a déjà prouvé la limite

$$\lim_{n\to\infty} n^\alpha a^n \tag{14.274}$$

pour tout $\alpha > 0$ et $a < 1$.

L'utilisation de propriétés de l'exponentielle nous permet de donner une nouvelle preuve, plus courte [26].

Le théorème 14.62 et la proposition 14.60 nous permettent de passer à l'exponentielle. Pour chaque n nous avons :

$$n^\alpha a^n = e^{\alpha\ln(n)+n\ln(a)}. \tag{14.275}$$

Ce qui est dans l'exponentielle est

$$\alpha\ln(n) + n\ln(a) = n\big(\alpha\frac{\ln(n)}{n} + \ln(a)\big). \tag{14.276}$$

Dans la parenthèse, $\ln(a) < 0$ et $\frac{\ln(n)}{n} \to 0$. Donc ce qui est dans l'exponentielle (14.275) tend vers $-\infty$ et au final l'expression demandée tend vers zéro. △

25. En utilisant 11.341(3).
26. C'est toujours facile de prétendre qu'une preuve est plus courte qu'une autre lorsqu'on utilise en une ligne des très gros théorèmes qui ont mis dix pages à être démontrés.

Remarque 14.81.
Vous ne pouvez pas à priori considérer l'exemple 14.80 comme une preuve alternative au lemme 11.352, parce que vous n'êtes pas sûr que dans toute la théorie permettant de définir l'exponentielle (en particulier la convergence de $\sum_k x^k/k!$), le lemme n'est pas utilisé [27].

Proposition 14.82.
Pour tout polynôme P et pour tout $a > 0$ la fonction $f(x) = P(x)e^{-ax}$ est intégrable [28] *sur* $[0, \infty[$.

Démonstration. Nous avons $f(x) = P(x)e^{-ax/2}e^{-ax/2}$, et par la vitesse comparée des exponentielles et polynômes, pour un certain $M > 0$ nous pouvons affirmer que $P(x)e^{-ax/2} < 1$ sur $[M, 0[$. Dès lors

$$|f(x)| < e^{-ax/2}, \tag{14.277}$$

qui est intégrable. □

Exemple 14.83.
La fonction logarithme (définition 14.58) n'est pas définie pour $x \leqslant 0$. Par conséquent la fonction $f(x) = x\ln(|x|)$ n'est pas définie en $x = 0$. Elle est bien définie pour $x < 0$ et vérifie

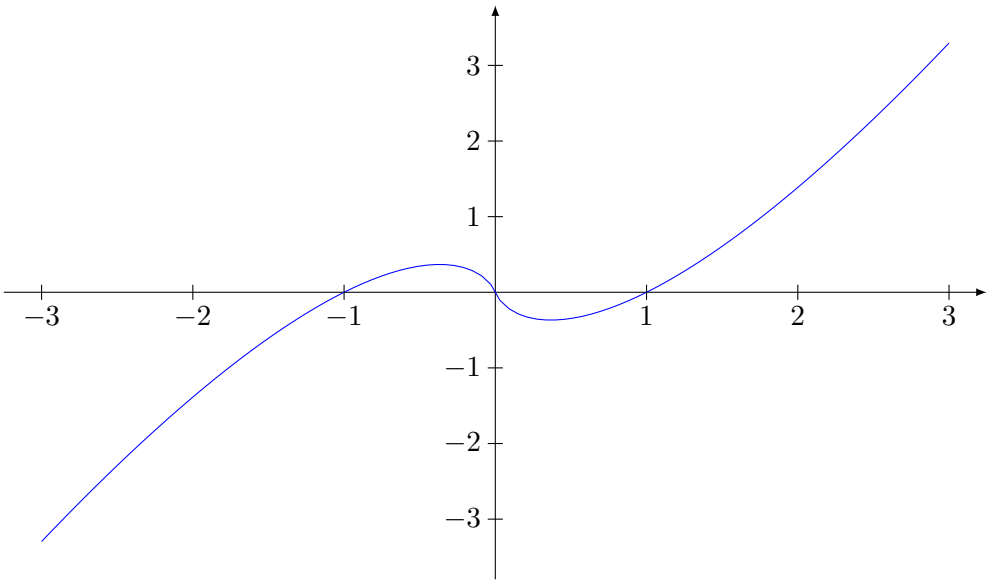
$$\lim_{x\to 0} x\ln(|x|) = 0. \tag{14.278}$$

Nous pouvons donc définir la fonction

$$\begin{aligned} \tilde{f}\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} x\ln(|x|) & \text{si } x \neq 0 \\ 0 & \text{si } x = 0. \end{cases} \end{aligned} \tag{14.279}$$

Contrairement à la fonction initiale f, cette fonction $\tilde{f}$ est définie et continue en 0.

Notez que sur le graphe de la fonction $\tilde{f}$, la courbe est bien régulière en $x = 0$.



△

Exemple 14.84.
Prenons deux suites $\{a_n\}$ et $\{b_n\}$ qui tendent toutes les deux vers l'infini (resp. 0). On dira que la suite $\{a_n\}$ converge plus vite (resp. plus lentement) que la suite $\{b_n\}$ si $\lim_{n\to\infty} \frac{a_n}{b_n} = \infty$, aussi vite si $\lim_{n\to\infty} \frac{a_n}{b_n}$ existe et est finie, et plus lentement (resp. plus vite) si $\lim_{n\to\infty} \frac{a_n}{b_n} = 0$.

27. Faites la vérification et dites moi si c'est bon.
28. Définition 13.168.

(1) Montrer qu'il existe deux suites qui tendent vers ∞ (ou 0) mais qui n'ont pas la même vitesse d'approche.
(2) Montrer que pour toute suite qui tend vers ∞ (ou 0), il existe une suite qui tend vers ∞ (ou 0) plus vite.
(3) Donner une suite non exponentielle qui tend vers l'infini plus vite que la suite $x_k = e^k$.

Voici quelque éléments de réponse.

(1) $x_n = n$ et $y_n = n^2$, et les inverses pour des suites qui tendent vers zéro.
(2) Si $x_n \to \infty$, la suite x_n^2 tend plus vite.
(3) La suite $x_n = n!$ tend vers ∞ vite que l'exponentielle. En effet, le nombre e^k n'est rien d'autre que le produit itéré $e \cdot e \cdot \ldots \cdot e$. Comparez

$$e \cdot e \cdot e \cdot \ldots \cdot e \tag{14.280}$$

avec

$$1 \cdot 2 \cdot 3 \cdot \ldots \cdot 10. \tag{14.281}$$

Étant donné que $e < 3$, nous avons

$$\frac{e^k}{k!} < \frac{e^2}{2} \cdot \left(\frac{e}{3}\right)^{k-2} \to 0. \tag{14.282}$$

△

<++>

14.6.2 Nombres premiers

Théorème 14.85.
Soit P, l'ensemble des nombres premiers. Alors la somme $\sum_{p\in P} \frac{1}{p}$ diverge et plus précisément,

$$\sum_{\substack{p\leqslant x\\ p\in P}} \frac{1}{p} \geqslant \ln(\ln(x)) - \ln(2). \tag{14.283}$$

Démonstration. Nous posons

$$S_x = \{q \leqslant x \text{ avec } q \text{ sans facteurs carrés}\} \tag{14.284}$$

et

$$P_x = \{p \in P \text{ tel que } p \leqslant x\}. \tag{14.285}$$

Si

$$K_x = \{(q,m) \text{ tels que } q \text{ n'a pas de facteurs carrés et } qm^2 \leqslant x\}, \tag{14.286}$$

alors nous avons

$$K_x = \bigcup_{q\in S_x} \bigcup_{m\leqslant\sqrt{x/q}} (q,m). \tag{14.287}$$

Par définition et par le lemme 3.33 nous avons aussi

$$\{n \leqslant x\} = \{qm^2 \text{ tel que } (q,m) \in K_x\}. \tag{14.288}$$

Tout cela pour décomposer la somme

$$\sum_{n\leqslant x} \frac{1}{n} = \sum_{q\in S_x} \sum_{m\leqslant\sqrt{x/q}} \frac{1}{m^2} \leqslant \sum_{q\in S_x} \frac{1}{q} \underbrace{\sum_{m\geqslant 1} \frac{1}{m^2}}_{=C}. \tag{14.289}$$

Nous avons aussi

$$\prod_{p\in P_x}\left(1+\frac{1}{p}\right)=1+\sum_{p\in P_x}\frac{1}{p}+\sum_{\substack{p,q\in P_x\\ p<q}}\frac{1}{pq}+\sum_{\substack{p,q,r\in P_x\\ p<q<r}}\frac{1}{pqr}+\dots \tag{14.290a}$$

$$\geqslant 1+\sum_{p\in P_x}\frac{1}{p}+\sum_{\substack{p,q\in P_x\\ pq\leqslant x}}\frac{1}{pq}+\sum_{\substack{p,q,r\in P_x\\ pqr\leqslant x}}\frac{1}{pqr}+\dots \tag{14.290b}$$

Les sommes sont finies. Les sommes s'étendent sur toutes les façons de prendre des produits de nombres premiers distincts de telle sorte de conserver un produit plus petit que x ; c'est-à-dire que les sommes se résument en une somme sur les éléments de S_x :

$$\exp\left(\sum_{p\in P_x}\frac{1}{p}\right)\geqslant\prod_{p\in P_x}\left(1+\frac{1}{p}\right)\geqslant\sum_{q\in S_x}\frac{1}{q}. \tag{14.291}$$

La première inégalité est simplement le fait que $1+u\leqslant e^u$ si $u\geqslant 0$ (directe de la définition 14.62). Les inégalités suivantes proviennent du fait que le logarithme est une primitive de la fonction inverse (proposition 14.64) :

$$\ln(x)\leqslant\sum_{n\geqslant x}\int_n^{n+1}\frac{dt}{t}\leqslant\sum_{n\geqslant x}\frac{1}{n}. \tag{14.292}$$

Nous prolongeons ces inégalités avec les inégalités (14.289) et (14.291) :

$$\ln(x)\leqslant\sum_{n\geqslant x}\frac{1}{n}\leqslant C\sum_{q\in S_x}\frac{1}{q}\leqslant C\leqslant\exp\left(\sum_{p\in P_x}\frac{1}{p}\right). \tag{14.293}$$

En passant au logarithme,

$$\ln\big(\ln(x)\big)\leqslant\ln(C)+\sum_{p\in P_x}\frac{1}{p}. \tag{14.294}$$

Ceci montre la divergence de la série de droite. Nous cherchons maintenant une borne pour C. Pour cela nous écrivons

$$\sum_{n=1}^{N}\frac{1}{n^2}\leqslant 1+\sum_{n=2}\frac{1}{n(n-1)} \tag{14.295a}$$

$$=1+\sum_{n=2}^{N}\left(\frac{1}{n-1}-\frac{1}{n}\right) \tag{14.295b}$$

$$=1+1-\frac{1}{N} \tag{14.295c}$$

$$\leqslant 2. \tag{14.295d}$$

Donc $C\leqslant 2$. □

Ce théorème prend une nouvelle force en considérant le théorème de Müntz 16.7 qui dit qu'alors l'ensemble $\mathrm{Span}\{x^p$ tel que p est premier$\}$ est dense dans les fonctions continues sur $[0,1]$ muni de la norme uniforme ou $\|.\|_2$.

14.6.3 Quelques limites

Nous voyons à présent quelques calculs de limite et de développements mettant en scène des logarithmes et exponentielles.

Exemple 14.86.
Pour trouver le développement de la fonction $f(x)=e^{-2x}$, il suffit d'écrire celui de e^t et de remplacer ensuite t par $-2x$. Le développement à l'ordre 3 de la fonction exponentielle est :

$$e^t=1+t+\frac{t^2}{2}+\frac{t^3}{6}+t^3\alpha(t). \tag{14.296}$$

Le développement de $f(x) = e^{-2x}$ sera donc

$$f(x) = 1 - 2x + \frac{4x^2}{2} - \frac{8x^3}{6} - 8x^3\alpha(-2x). \tag{14.297}$$

Donc le polynôme de degré 3 partie régulière de g est :

$$1 - 2x + 2x^2 - \frac{4}{3}x^3, \tag{14.298}$$

et la fonction reste correspondante est :

$$\alpha_g(x) = -8\alpha(-2x). \tag{14.299}$$

△

Exemple 14.87.
Nous savons les développements

$$f(x) = \ln(1+x) \sim x - \frac{x^2}{2} + \frac{x^3}{3} \tag{14.300}$$

et

$$\sin(x) \sim x - \frac{x^3}{6}. \tag{14.301}$$

Nous obtenons le développement d'ordre 3 de la fonction $x \mapsto \ln\big(1+\sin(x)\big)$ en écrivant

$$\ln\big(1+\sin(x)\big) \sim \big(x - \frac{x^3}{6}\big) - \frac{1}{2}\left(x - \frac{x^3}{6}\right)^2 + \frac{1}{3}\left(x - \frac{x^3}{6}\right)^3. \tag{14.302}$$

Il s'agit maintenant de trouver les termes qui sont de degré inférieur ou égale à 3.

D'abord

$$\left(x - \frac{x^3}{6}\right)^2 = x^2 - \frac{x^4}{3} + \frac{x^6}{36} \sim x^2 \tag{14.303}$$

Nous avons alors aussi

$$\left(x - \frac{x^3}{6}\right)^6 \sim x^2\left(x - \frac{x^3}{6}\right) \sim x^3. \tag{14.304}$$

En replaçant tout ça dans (14.302) nous trouvons

$$\ln\big(1+\sin(x)\big) \sim x - \frac{x^2}{2} + \frac{x^3}{6}. \tag{14.305}$$

△

Exemple 14.88.
Calculer

$$\lim_{x\to\infty} e^{1/x}\sqrt{1+4x^2} - 2x. \tag{14.306}$$

Nous allons effectuer un développement asymptotique de la partie « difficile » de l'expression posant d'abord $x = 1/h$. Si $f(x) = e^{1/x}\sqrt{1-4x^2}$ alors

$$g(h) = \frac{1}{|h|}e^h\sqrt{h^2+4} = \frac{1}{h}\big(1+h+h\alpha(h)\big)\big(2+h\beta(h)\big). \tag{14.307}$$

La première parenthèse est le développement de e^h et la seconde celui de $\sqrt{h^2+4}$. Nous nous apprêtons à faire la limite $x \to \infty$ qui correspond à $h \to 0^+$, nous pouvons donc supposer que $h > 0$ et omettre la valeur absolue. En effectuant le produit et en regroupant tous les termes contenant h^2, $\alpha(h)$ ou $\beta(h)$ dans un seul terme $h\gamma(h)$,

$$f(h) = \frac{1}{h}\big(2+2h+h\gamma(h)\big) = \frac{2}{h} + 2 + \gamma(h) = 2x + 2 + \gamma(1/x) \tag{14.308}$$

où γ est une fonction vérifiant $\lim_{t\to 0}\gamma(t) = 0$.

Nous sommes maintenant en mesure de calculer la limite (14.306) :

$$\lim_{x\to\infty} e^{1/x}\sqrt{1+x^2} - 2x = \lim_{x\to\infty}\big(2x + 2 + \gamma(1/x) - 2x\big) = 2. \tag{14.309}$$

△

14.7 Trigonométrie hyperbolique

Définition 14.89.
Les fonctions ***cosinus hyperbolique*** *et* ***sinus hyperbolique*** *sont les fonctions définies sur* $\mathbb{R}$ *par les formules suivantes :*

$$\cosh(x) = \frac{e^x + e^{-x}}{2} \tag{14.310a}$$

$$\sinh(x) = \frac{e^x - e^{-x}}{2}. \tag{14.310b}$$

Si vous ne vous rappelez plus la définition de e^x*, c'est 14.37.*

Proposition 14.90.
Quelques propriétés algébriques des fonctions trigonométriques hyperboliques.

(1) $\cosh(-x) = \cosh(x)$
(2) $\sinh(-x) = -\sinh(x)$
(3) $\cosh^2(x) - \sinh^2(x) = 1$
(4) $\cosh(x)\cosh(y) + \sinh(x)\sinh(y) = \cosh(x+y)$
(5) $\cosh(x)\cosh(y) - \sinh(x)\sinh(y) = \cosh(x-y)$
(6) $\cosh(x)\sinh(y) + \sinh(x)\cosh(y) = \sinh(x+y)$
(7) $\cosh(x)\sinh(y) - \sinh(x)\cosh(y) = -\sinh(x-y)$.

Démonstration. Si s'agit simplement de remplacer les définitions et d'utiliser les formules concernant les puissances, dont la formule (11.898). □

Proposition 14.91.
Quelques propriétés analytiques des fonctions trigonométriques hyperboliques.

(1) $\cosh'(x) = \sinh(x)$
(2) $\sinh'(x) = \cosh(x)$.
(3) $\cosh(x) \geqslant 1$.

Démonstration. Pour les dérivées, il s'agit d'utiliser la dérivation de l'exponentielle, laquelle est facile par le théorème 14.62(1).

Pour (3), nous commençons par les $x \geqslant 0$. D'abord $\cosh(0) = 1$. Ensuite $\cosh'(x) = \sinh(x) = \frac{e^x - e^{-x}}{2}$. Vu que $x > 0$ nous avons $e^x > e^{-x} > 0$. Donc la dérivée de cosh est strictement positive sur $]0, \infty[$. La fonction y est donc partout plus grande que $\cosh(0) = 1$.

Pour les $x < 0$, nous avons la fait que cosh est paire. □

Proposition 14.92.
La fonction $\sinh\colon \mathbb{R} \to \mathbb{R}$ *est bijective.*

Démonstration. En deux parties.

Injective Si $\sinh(a) = \sinh(b)$, alors le théorème de Rolle 11.127 affirme qu'il existe $c \in]a, b[$ tel que $\sinh'(c) = 0$. Mais la proposition 14.91 nous dit que $\sinh'(x) = \cosh(c) \geqslant 1$. Donc impossible.

Surjective Nous avons

$$\lim_{x \to -\infty} \sinh(x) = -\infty \tag{14.311}$$

et

$$\lim_{x \to \infty} \sinh(x) = \infty. \tag{14.312}$$

Soit $y \in \mathbb{R}$. Il existe $m < 0$ tel que $\sinh(m) < y$ et $M > 0$ tel que $\sinh(M) > y$. Le théorème des valeurs intermédiaires 11.50 nous enseigne qu'il existe $x \in [m, M]$ tel que $\sinh(x) = y$.

□

Proposition 14.93 ([1]).
Soient $a, b \in \mathbb{R}$ tels que $a - b^2 = 1$. Il existe un unique $(x, \sigma) \in \mathbb{R} \times \{\pm 1\}$ tel que

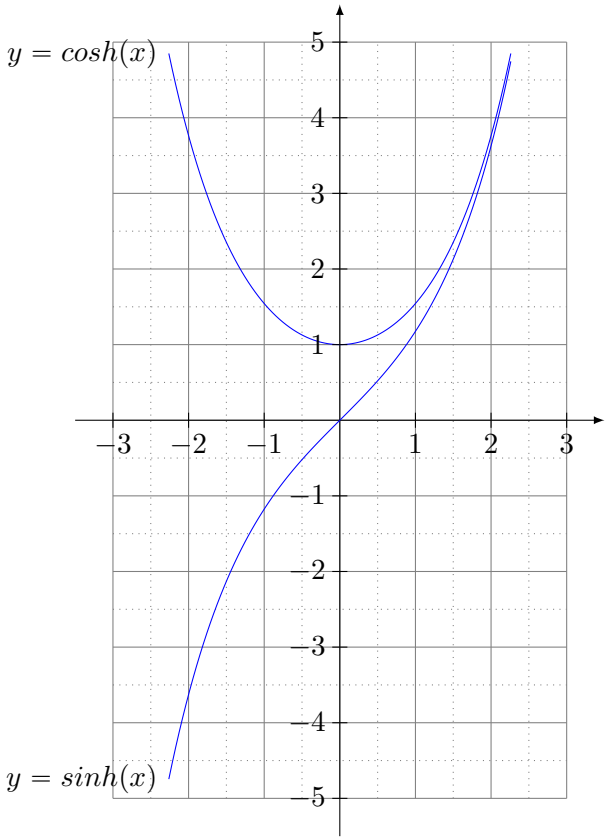
$$\begin{cases} a = \sigma \cosh(x) & (14.313a) \\ b = \sinh(x). & (14.313b) \end{cases}$$

Démonstration. Vu que le sinus hyperbolique est une bijection [29], il existe un unique $x \in \mathbb{R}$ tel que $\sinh(x) = b$. Maintenant un petit calcul :

$$a^2 = 1 + \sinh(x)^2 = 1 + \frac{e^{2x} + e^{-2x} - 2}{4} = \frac{e^{2x} + e^{-2x} + 2}{4} = \cosh(x)^2. \tag{14.314}$$

Vu que $\cosh(x)^2 = a^2$, il existe un unique $\sigma \in \{\pm 1\}$ tel que $\sigma \cosh(x) = a$. □

Les représentations graphiques sont ceci :



La **tangente hyperbolique** est donnée par le quotient

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}. \tag{14.315}$$

14.8 Séries entières de matrices

14.8.1 Différentiabilité

Proposition 14.94.
Soit (a_n) une suite dans $\mathbb{C}$ de rayon de convergence R et la fonction

$$\begin{aligned} f \colon \mathbb{M}(n, \mathbb{R}) &\to \mathbb{M}(n, \mathbb{R}) \\ A &\mapsto \sum_{k=0}^{\infty} a_k A^k \end{aligned} \tag{14.316}$$

29. Proposition 14.92.

Alors

(1) La différentielle de f sur $B(0,R)$ est

$$df_A(U) = \sum_{k=0}^{\infty} a_k \sum_{l=0}^{k-1} A^l U A^{k-1-l}, \tag{14.317}$$

c'est-à-dire que l'on peut différentier terme à terme. (Ici c'est A qui est dans $B(0,R)$)

(2) La convergence de la somme 14.317 est absolue.

(3) La convergence de la somme 14.317 est normale sur tout compact.

(4) La fonction f est de classe C^1 sur $B(0,R)$, c'est-à-dire que la fonction $A \mapsto df_A$ est continue.

Notons que df_A n'est pas tout à fait une série entière. Cependant, en ce qui concerne les normes, c'est tout comme si ça l'était.

Démonstration. Nous posons $u_k(A) = a_k A^k$, qui est une fonction de classe C^∞ et dont la différentielle est donnée par

$$(du_k)_A(U) = \frac{d}{dt}\Big[u_k(A+tU)\Big]_{t=0} = a_k \frac{d}{dt}\Big[(A+tU)^k\Big]_{t=0}; \tag{14.318}$$

en distribuant le produit nous trouvons tout un tas de termes dont seuls ceux contenant exactement une fois tU ne vont pas s'annuler. Étant donné que U et A ne commutent pas nous avons l'expression un peu moche

$$(du_k)_A(U) = \sum_{l=0}^{k-1} a_k A^l U A^{k-1-l}. \tag{14.319}$$

En ce qui concerne la norme, nous regardons celle de $(du_k)_A$ pour un A fixé ; c'est-à-dire que nous en regardons la norme opérateur :

$$\|(du_k)_A\| = \sup_{\|U\|=1} \|\sum_{l=0}^{k-1} a_k A^l U A^{k-1-l}\| \leqslant \sum_{l=0}^{k-1} |a_k| \|A\|^l \|A\|^{k-1-l} \leqslant k|a_k|\|A\|^{k-1}. \tag{14.320}$$

Pour donner la convergence nous considérons un nombre r tel que $\|A\| < r < R$, de telle sorte que la suite $(a_n r^n)$ soit bornée par un nombre M et que nous puissions écrire

$$\|(du_k)_A\| \leqslant k|a_k|\|A\|^{k-1} = \frac{k|a_k|\|A\|^k}{\|A\|} = \frac{k|a_k|}{\|A\|} r^k \left(\frac{\|A\|}{r}\right)^k \leqslant \frac{M}{\|A\|} k \left(\frac{\|A\|}{r}\right)^k, \tag{14.321}$$

dont la série converge. Nous avons donc convergence absolue de la série

$$\sum_{k=0}^{\infty} (du_k)_A. \tag{14.322}$$

Passons à la convergence normale sur tout compact. Nous nous fixons $r < R$ et nous nous intéressons à la norme de du_k sur $\overline{B(0,r)}$, c'est-à-dire

$$\|du_k\|_\infty = \sum_{x \in \overline{B(0,r)}} \|(du_k)_A\|. \tag{14.323}$$

Vu que $\overline{B(0,r)}$ est compact, ce supremum est un maximum et nous pouvons noter A_k la matrice qui le réalise. Nous réalisons alors les mêmes manipulations que pour (14.321) :

$$\|du_k\|_\infty = \|(du_k)_{A_k}\| \leqslant k|a_k|\|A_k\|^{k-1} \leqslant k|a_k|r^{k-1} = \frac{1}{r} k|a_k| r^k. \tag{14.324}$$

Nous prenons maintenant $r < r_0 < R$ et M, un majorant de $(a_n r_0^n)$, de telle sorte qu'en multipliant et divisant par r_0^k,

$$\|du_k\|_\infty \leqslant \frac{k|a_k| r_0^k}{r} \frac{r^k}{r_0^k} \leqslant \frac{kM}{r} \left(\frac{r}{r_0}\right)^k, \tag{14.325}$$

dont la série converge. Nous avons donc convergence normale sur tout compact. Par voie de ~~fait~~ conséquences nous avons continuité de la série

$$\sum_{k=0}^{\infty}(du_k)_A \tag{14.326}$$

et convergence vers df_A par le théorème 14.6. □

Proposition 14.95.
Si le rayon de convergence de la série $u(A) = \sum_{k=0}^{\infty} a_k A^k$ est R, alors

(1) elle converge normalement sur tout compact de $B(0,R)$;

(2) la fonction u y est de classe C^∞.

Démonstration. Nous posons

$$\begin{aligned} u_k\colon \mathbb{M}(n,\mathbb{R}) &\to \mathbb{M}(n,\mathbb{R}) \\ A &\mapsto a_k A^k \end{aligned} \tag{14.327}$$

qui est évidemment une fonction de classe C^∞. Nous étudions la j^{e} différentielle en m, pour $k > j$ (dans une série, nous ne nous intéressons pas aux premiers termes). La j^{e} différentielle appliquée à v_1 appliquée à v_2, etc s'exprime de la façon suivante :

$$(d^j u_k)_m(v_1,\dots,v_j) = \frac{d}{dt_1}\dots\frac{d}{dt_j}\Big(u_k(m + t_1v_1 + \dots + t_jv_j)\Big)_{t_i=0}. \tag{14.328}$$

Dans le produit $(m+t_1v_1+\dots+t_jv_j)^k$, seuls les termes contenant exactement une fois chacun des t_i ne s'annulera pas après avoir fait la dérivée et évalué en $t_i = 0$. Combien de termes cela fait ? Parmi les k facteurs, il faut en placer j qui ne sont pas m (cela fait $\binom{k}{j}$ possibilités), et puis il faut ordonner ces j termes, cela fait encore $j!$ possibilités. Au final,

$$\|(d^j u_k)_m\| \leqslant |a_k|\binom{k}{j}j!\|m\|^{k-j} = |a_k|P(k)\|m\|^{k-j} \tag{14.329}$$

où $P(k) = \frac{k!}{(k-j)!}$ est un polynôme de degré j.

Afin d'étudier la convergence normale sur tout compact de la série des $d^j u_k$, nous considérons $r < r_0 < R$ et nous allons prouver la convergence normale sur $\overline{B(0,r)}$. Vu que c'est un compact, il existe une matrice $m_k \in \overline{B(0,r)}$ telle que

$$\begin{aligned} \|d^j u_k\|_\infty &= \|(d^j u_k)_{m_k}\| && (14.330a)\\ &\leqslant |a_k|P(k)\|m_k\|^{k-j} && (14.330b)\\ &\leqslant |a_k|P(k)r^{k-j} && (14.330c)\\ &= \frac{|a_k|P(k)}{r^j}r^k && (14.330d)\\ &= \frac{|a_k|r_0^kP(k)}{r^j}\left(\frac{r}{r_0}\right)^k && (14.330e)\\ &\leqslant \frac{M}{r^j}P(k)\left(\frac{r}{r_0}\right)^k && (14.330f) \end{aligned}$$

où M est un majorant de $a_n r^n$. Vu que $r_0/r < 1$, la somme sur k converge et nous avons convergence normale sur tout compact de

$$d^j\sum_{k=0}^{\infty}a_kA^k = \sum_{k=0}^{\infty}d^j(a_kA^k) \tag{14.331}$$

avec un peu d'abus de notation. □

14.9 Exponentielle de matrices

Proposition 14.96.
Une matrice complexe est inversible si et seulement si elle est une exponentielle.

Autrement dit :

$$\mathrm{GL}(n,\mathbb{C}) = e^{\mathbb{M}(n,\mathbb{C})}. \tag{14.332}$$

Démonstration. Nous avons déjà prouvé dans la proposition 14.49 que toutes les exponentielles étaient inversibles. Ici nous nous concentrons sur la réciproque.

Soit $A \in \mathrm{GL}(n,\mathbb{C})$; nous allons donner une matrice $B \in \mathbb{M}(n,\mathbb{C})$ telle que $A = \exp(B)$. D'abord remarquons qu'il suffit de prouver le résultat pour une matrice par classe de similitude. En effet si $A = \exp(B)$ et si M est inversible alors

$$\exp(MBM^{-1}) = \sum_k \frac{1}{k!}(MBM^{-1})^k \tag{14.333a}$$

$$= \sum_k \frac{1}{k!} MB^k M^{-1} \tag{14.333b}$$

$$= M\exp(B)M^{-1}. \tag{14.333c}$$

Donc $MAM^{-1} = \exp(MBM^{-1})$. Nous pouvons donc nous contenter de trouver un logarithme pour les blocs de Jordan. Nous supposons donc que $A = (\mathbb{1} + N)$ avec $N^m = 0$. En nous inspirant de (14.242), nous posons [30]

$$D(t) = tN - \frac{t^2}{2}N^2 + \cdots + (-1)^m \frac{t^{m-1}}{m-1}N^{m-1} \tag{14.334}$$

et nous allons prouver que $e^{D(1)} = \mathbb{1} + N$. Notons que N étant nilpotente, cette somme ainsi que toutes celles qui viennent sont finies. Il n'y a donc pas de problèmes de convergences dans cette preuve (si ce n'est les passages des équations (14.333)).

Nous posons $S(t) = e^{D(t)}$ (la somme est finie), et nous avons

$$S'(t) = D'(t)e^{D(t)} \tag{14.335}$$

Afin d'obtenir une expression qui donne S' en termes de S, nous multiplions par $(\mathbb{1} + tN)$ en remarquant que $(\mathbb{1} + tN)D'(t) = N$ nous avons

$$(\mathbb{1} + tN)S'(t) = NS(t). \tag{14.336}$$

En dérivant à nouveau,

$$(\mathbb{1} + tN)S''(t) = 0. \tag{14.337}$$

La matrice $(\mathbb{1} + tN)$ est inversible parce que son noyau est réduit à $\{0\}$. En effet si $(\mathbb{1} + tN)x = 0$, alors $Nx = -\frac{1}{t}x$, ce qui est impossible parce que N est nilpotente. Ce que dit l'équation (14.337) est alors que $S''(t) = 0$. Si nous développons $S(t)$ en puissances de t nous nous arrêtons au terme d'ordre 1 et nous avons

$$S(t) = S(0) + tS'(0) = \mathbb{1} + tD'(0) = 1 + tN. \tag{14.338}$$

En $t = 1$ nous trouvons $S(1) = \mathbb{1} + N$. La matrice $D(1)$ donnée est donc bien un logarithme de $\mathbb{1} + N$. □

30. Le logarithme d'un nombre n'est pas encore définit à ce moment, mais cela ne nous empêche pas de poser une définition ici pour une application des réels vers les matrices.

14.9.1 Diagonalisabilité d'exponentielle

Proposition 14.97 ([71]).
Si $A \in \mathbb{M}(n, \mathbb{R})$ a un polynôme caractéristique scindé, alors A est diagonalisable si et seulement si e^A est diagonalisable.

Démonstration. Si A est diagonalisable, alors il existe une matrice inversible M telle que $D = M^{-1}AM$ soit diagonale (c'est la définition 9.125). Dans ce cas nous avons aussi $(M^{-1}AM)^k = M^{-1}A^kM$ et donc $M^{-1}e^AM = e^{M^{-1}AM} = e^D$ qui est diagonale.

La partie difficile est donc le contraire.

Qui est diagonalisable et comment ? Nous supposons que e^A est diagonalisable et nous écrivons la décomposition de Dunford (théorème 9.196) :

$$A = S + N \tag{14.339}$$

où S est diagonalisable, N est nilpotente, $[S, N] = 0$. Nous avons besoin de prouver que $N = 0$.

Les matrices A est S commutent ; en passant au développement nous en déduisons que A et e^S commutent, puis encore en passant au développement que e^A et e^S commutent. Vu que S est diagonalisable, e^S l'est et par hypothèse e^A est également diagonalisable. Donc e^A et e^{-S} sont simultanément diagonalisables par la proposition 9.131.

Étant donné que A et S commutent, nous avons $e^N = e^{A-S} = e^Ae^{-S}$, et nous en déduisons que e^N est diagonalisable vu que les deux facteurs e^A et e^{-S} sont simultanément diagonalisables.

Unipotence Si r est le degré de nilpotence de N, nous avons

$$e^N - \mathbb{1} = N + \frac{N^2}{2} + \cdots + \frac{N^{r-1}}{(r-1)!}. \tag{14.340}$$

Donc

$$(e^N - \mathbb{1})^k = \left(N + \frac{N^2}{2} + \cdots + \frac{N^{r-1}}{(r-1)!}\right)^k \tag{14.341}$$

où le membre de droite est un polynôme en N dont le terme de plus bas degré est de degré k. Donc $(e^N - \mathbb{1})$ est nilpotente et e^N est unipotente.

Si M est la matrice qui diagonalise e^N, alors la matrice diagonale $M^{-1}e^NM$ est tout autant unipotente que e^N elle-même. En effet,

$$(M^{-1}e^NM - \mathbb{1})^r = \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} M^{-1}(e^N)^kM \tag{14.342a}$$

$$= M^{-1}\left(\sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k}(e^N)^k\right) M \tag{14.342b}$$

$$= M^{-1}(e^N - \mathbb{1})^rM \tag{14.342c}$$

$$= 0. \tag{14.342d}$$

La matrice $M^{-1}e^NM$ est donc une matrice diagonale et unipotente ; donc $M^{-1}e^NM = \mathbb{1}$, ce qui donne immédiatement que $e^N = \mathbb{1}$.

Polynômes annulateurs En reprenant le développement (14.340) sachant que $e^N = \mathbb{1}$, nous savons que

$$N + \frac{N^2}{2} + \cdots + \frac{N^{r-1}}{(r-1)!} = 0. \tag{14.343}$$

Dit en termes pompeux (mais non moins porteurs de sens), le polynôme

$$Q(X) = X + \frac{X^2}{2} + \cdots + \frac{X^{r-1}}{(r-1)!} \tag{14.344}$$

est un polynôme annulateur de N.

La proposition 9.96 stipule que le polynôme minimal d'un endomorphisme divise tous les polynômes annulateurs. Dans notre cas, X^r est un polynôme annulateur et donc le polynôme minimal de N est de la forme X^k. Donc il est X^r lui-même.

Nous avons donc $X^r \mid Q$. Mais Q est un polynôme contenant le monôme X donc X^r ne peut diviser Q que si $r = 1$. Nous en concluons que X est un polynôme annulateur de N. C'est-à-dire que $N = 0$.

Conclusion Vu que Dunford[31] dit que $A = S + N$ et que nous venons de prouver que $N = 0$, nous concluons que $A = S$ avec S diagonalisable.

□

14.10 Étude d'asymptote

Lorsqu'une fonction tend vers l'infini pour $x \to \infty$, une question qui peut venir est : à quelle vitesse tend-t-elle vers l'infini ?

Il est « visible » que la fonction logarithme ne tend pas très vite vers l'infini : certes

$$\lim_{x\to\infty} \ln(x) = +\infty, \tag{14.345}$$

mais par exemple $\ln(100000) \simeq 11.5$ tandis que $e^{100000} \simeq 10^{43429}$. Sans contestations possibles, l'exponentielle croit plus vite que le logarithme.

Soient f et g deux fonctions dont la limite $x \to \infty$ est ∞. Si

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0 \tag{14.346}$$

nous disons que g tend vers ∞ plus vite que f ; si

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = \infty \tag{14.347}$$

nous disons que f tend vers ∞ plus vite que g, et si

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = a \in \mathbb{R} \tag{14.348}$$

avec $a \neq 0$ alors nous disons que f tend vers l'infini à la même vitesse que $ag(x)$.

Exemple 14.98.
La fonction $x \mapsto x^2$ tend vers l'infini plus vite que la fonction $x \mapsto \sqrt{x}$. △

Dans cette section nous allons nous contenter de déterminer les fonctions qui tendent vers l'infini aussi vite qu'une droite oblique, que nous appellons asymptote et que nous voulons déterminer.

Exemple 14.99.
Déterminer les asymptotes obliques (s'ils existent) de la fonction

$$f(x) = e^{1/x}\sqrt{1 + 4x^2}. \tag{14.349}$$

Tout d'abord nous remarquons que $\lim_{x\to\infty} f(x) = \infty$. Nous sommes donc en présence d'une branche du graphe qui tend vers l'infini. Ensuite,

$$\lim_{x\to\infty} \frac{f(x)}{x} = \lim_{x\to\infty} e^{1/x}\sqrt{\frac{1}{x^2} + 4} = 2. \tag{14.350}$$

31. Théorème 9.196.

Donc le graphe de f tend vers l'infini à la même vitesse que le graphe de la fonction $y = 2x$. Nous aurons donc une asymptote oblique de coefficient directeur 2. De façon imagée, nous pouvons penser que le graphe de f et celui de $y = 2x$ sont presque parallèles si x est assez grand. Afin de déterminer l'ordonnée à l'origine de l'asymptote, il nous reste à voir quelle est la « distance » entre le graphe de f et celui de $y = 2x$:

$$\lim_{x\to\infty} f(x) - 2x = \lim_{x\to\infty} e^{1/x}\sqrt{1+4x^2} - 2x. \tag{14.351}$$

Cette limite a été calculée dans l'exemple 14.88 et vaut 2.

Nous concluons que le graphe de la fonction f admet l'asymptote

$$y = 2x + 2. \tag{14.352}$$

△

14.11 Développement en série

14.11.1 Série génératrice d'une suite

Soit u_n une suite telle que le rayon de convergence de

$$f(z) = \sum_{n=0}^{\infty} u_n z^n \tag{14.353}$$

soit strictement positif. Alors la série f est la **série génératrice** de la suite (u_n).

Grâce au théorème 14.25 nous pouvons la dériver terme à terme autour de $z = 0$. En utilisant la petite formule (14.79) nous trouvons

$$f^{(l)}(z) = \sum_{n=l}^{\infty} u_n \frac{n!}{(n-l)!} z^{n-l}, \tag{14.354}$$

et donc

$$u_l = \frac{f^{(l)}(0)}{l!}. \tag{14.355}$$

D'où le nom de série génératrice. Cela est évidemment intéressant seulement si nous connaissons une autre forme pour f par ailleurs.

Nous en utiliserons une pour déterminer les partitions d'un nombre en parts fixes, proposition 25.41.

14.11.2 Développement en série et Taylor

Définition 14.100.
*Soit une fonction $f\colon \mathbb{C} \to \mathbb{C}$ et $z_0 \in \mathbb{C}$. Nous disons que f est **développable en série entière** dans un voisinage de z_0 s'il existe une série $\sum_n a_n z^n$ de rayon de convergence $R > 0$ et $r \leqslant R$ tel que*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n \tag{14.356}$$

pour tout $z \in B(z_0, r)$.

Proposition 14.101.
Si V est un ouvert dans $\mathbb{C}$ alors l'ensemble des fonctions $V \to \mathbb{C}$ développables en série entière forme une $\mathbb{C}$-algèbre.

Démonstration. Les séries entières passent aux sommes et aux produits en gardant des rayons de convergence non nuls. □

Proposition 14.102.
Si f est développable en série entière à l'origine alors elle est C^∞ sur un voisinage de l'origine et le développement est celui de ***Taylor*** *:*

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n \tag{14.357}$$

pour tout x dans un voisinage de 0.

Démonstration. Si $f(x) = \sum a_n x^n$, nous savons que f est C^1 et que nous pouvons dériver terme à terme (au moins dans un voisinage). De plus le fait de dériver ne change pas le domaine. Par récurrence, la fonction est C^∞ sur le voisinage. En dérivant k fois la série $\sum a_n x^n$ nous trouvons

$$f^{(k)}(x) = \sum_{n=k}^{\infty} n(n-1)\dots(n-k+1)a_n x^{n-k}. \tag{14.358}$$

En calculant en $x = 0$ nous trouvons

$$f^{(k)}(0) = k! a_k, \tag{14.359}$$

d'où le terme général

$$a_k = \frac{f^{(k)}(0)}{k!}. \tag{14.360}$$

□

Si f est une fonction et si la série

$$T_f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n \tag{14.361}$$

converge, alors cette série est la **série de Taylor** de f.

Remarque 14.103.
La série de Taylor d'une fonction n'est pas liée à sa fonction de façon aussi raide qu'on pourrait le croire. Même dans le cas d'une fonction C^∞ il peut arriver que $T_f(x) \neq f(x)$.

Il peut aussi arriver que f ne soit pas développable en série entières.

Exemple 14.104.
Nous considérons la fonction

$$f(x) = \begin{cases} e^{-1/x^2} & \text{si } x \neq 0 \\ 0 & \text{si } x = 0. \end{cases} \tag{14.362}$$

Nous avons

$$f'(x) = \begin{cases} \frac{2}{x^3} e^{-1/x^2} & \text{si } x \neq 0 \\ 0 & \text{si } x = 0. \end{cases} \tag{14.363}$$

Note : pour la seconde ligne nous devons faire explicitement le calcul

$$f'(0) = \lim_{t \to 0} \frac{f(t) - f(0)}{t} = \lim_{t \to 0} \frac{1}{t} e^{-1/t^2} = 0. \tag{14.364}$$

Plus généralement nous avons $f^{(k)}(0) = 0$, et par conséquent la série de Taylor converge (trivialement) vers la fonction identiquement nulle.

Cette fonction n'est donc pas développable en série entière vu qu'il n'existe aucun voisinage de zéro sur lequel la série de f coïncide avec f. △

Exemple 14.105.
Développement de $f(x) = \arctan(x)$. Nous savons que

$$f'(x) = \frac{1}{1+x^2}, \tag{14.365}$$

alors que nous connaissons le développement

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n \tag{14.366}$$

pour tout $x \in B(0,1)$. Nous avons donc successivement

$$\frac{1}{1+x} = \sum_{n=0} (-x)^n \tag{14.367a}$$

$$\frac{1}{1+x^2} = \sum_{n=0} (-1)^n x^{2n} \tag{14.367b}$$

$$\arctan(x) = \sum_{n=1}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} + C. \tag{14.367c}$$

Notons que dans la dernière nous avons évité d'écrire la somme depuis $n = 0$ (qui serait un terme constant) et nous avons écris explicitement « $+C$ ». Étant donné que $\arctan(0) = 0$, nous devons poser $C = 0$ et donc

$$\arctan(x) = \sum_{n=1}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1}. \tag{14.368}$$

△

14.11.3 Resommer une série

Nous avons vu comment trouver la série correspondant à une fonction donnée. Un exercice difficile consiste à trouver la fonction qui correspond à une somme donnée.

14.11.3.1 Les sommes du type $\sum_n P(n)x^n$

Pour calculer

$$\sum_{n=0}^{\infty} P(n)x^n \tag{14.369}$$

où P est un polynôme de degré m nous commençons par écrire

$$P(n) = \alpha_0 + \alpha_1(n+1) + \alpha_2(n+1)(n+2) + \cdots + \alpha_m(n+1)\dots(n+m). \tag{14.370}$$

Nous décomposons alors la somme en m sommes de la forme

$$\sum_{n=0}^{\infty} \alpha_k \frac{(n+k)!}{n!} x^n = \alpha_k \left(\sum_{n=0}^{\infty} x^{n+k} \right)^{(k)}. \tag{14.371}$$

Effectuons par exemple

$$\sum_{n=0}^{\infty} x^{n+3} = \frac{1}{1-x} - 1 - x - x^2 \tag{14.372}$$

Notons que dans un usage pratique, ce terme devra être ensuite dérivé trois fois, de telle manière que les termes « correctifs » n'interviennent pas. Cette méthode ne demande donc que de calculer les dérivées successives de $1/(1-x)$.

Exemple 14.106.
Calculons la fonction

$$f(x) = \sum_{n=0}^{\infty} n^3 x^n. \tag{14.373}$$

D'abord nous écrivons

$$n^3 = -1 + 7(n+1) - 6(n+1)(n+2) + (n+1)(n+2)(n+3). \tag{14.374}$$

Nous avons

$$\sum_{n=0}^{\infty}(n+1)x^n = \left(\sum_{n=0}^{\infty} x^{n+1}\right)' = \left(\frac{1}{1-x} - 1\right)' = \frac{1}{(x-1)^2}. \tag{14.375}$$

De la même façon,

$$\sum_n (n+1)(n+2)x^n = \left(\sum x^{n+2}\right)'' = \frac{-2}{(x-1)^3} \tag{14.376a}$$

$$\sum_n (n+1)(n+2)(n+3)x^n = \frac{6}{(x-1)^4}. \tag{14.376b}$$

En remettant tout ensemble nous obtenons

$$\sum_{n=0}^{\infty} n^3 x^n = -\frac{1}{1-x} + \frac{7}{(x-1)^2} + \frac{12}{(x-1)^3} + \frac{6}{(x-1)^4}. \tag{14.377}$$

Nous pouvons vérifier ce résultat en traçant les deux courbes et en remarquant qu'elles coïncident.

```
----------------------------------------------------------------------
| Sage Version 4.7.1, Release Date: 2011-08-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: n=var('n')
sage: S(x)=sum(  [ n**3*x**n for n in range(0,30)  ]   )
sage: f(x)=-1/(1-x)+7/((x-1)**2)+12/((x-1)**3)+6/( (x-1)**4  )
sage: S(0.1)
0.214906264288980
sage: f(0.1)
0.214906264288981
sage: f.plot(-0.5,0.5)+S.plot(-0.5,0.5)
```

△

14.11.3.2 Les sommes du type $\sum_n x^n/P(n)$

Si $P(n)$ a des racines entières, nous pouvons le décomposer en fractions simples et utiliser la somme

$$\sum_{n=1}^{\infty} \frac{x^n}{n} = -\ln(1-x). \tag{14.378}$$

Nous avons par exemple

$$\sum_{n=0}^{\infty} \frac{x^n}{n+1} = \frac{1}{x}\sum_{n=0} \frac{x^{n+1}}{n+1} \tag{14.379a}$$

$$= \frac{1}{x}\sum_{n=1}^{\infty} \frac{x^n}{n} = -\frac{\ln(1-x)}{x}. \tag{14.379b}$$

Notez le changement de point de départ de la somme au passage.

Autre exemple :

$$\sum_{n=0}^{\infty}\frac{x^n}{n+3}=\frac{1}{x^3}\left(\sum_{n=1}^{\infty}\frac{x^n}{n}-x-\frac{x^2}{2}\right) \tag{14.380a}$$

$$=-\frac{\ln(x-1)}{x^3}-\frac{1}{x^2}-\frac{1}{2x}. \tag{14.380b}$$

Si le polynôme possède des racines non entières, les choses se compliquent.

Exemple 14.107.
Calculons

$$\sum_{n=0}^{\infty}\frac{x^n}{2n+1}. \tag{14.381}$$

Si $x\geqslant$, en posant $t=\sqrt{x}$ nous trouvons

$$\sum_{n=0}^{\infty}\frac{x^n}{2n+1}=\frac{1}{t}\sum_{n=0}^{\infty}\frac{t^{2n+1}}{2n+1}. \tag{14.382}$$

Étudions

$$H(t)=\sum_{n=0}^{\infty}\frac{t^{2n+1}}{2n+1}. \tag{14.383}$$

Nous avons

$$H'(t)=\sum_{n=0}^{\infty}t^{2n}=\sum_{n=0}(t^2)^n=\frac{1}{1-t^2}. \tag{14.384}$$

Une primitive de cette fonction est

$$\frac{1}{2}\ln\left|\frac{t+1}{t-1}\right|. \tag{14.385}$$

En $t=0$, cette fonction vaut 0 qui est la bonne valeur. Donc nous avons bien

$$H(t)=\frac{1}{2}\ln\left|\frac{t+1}{t-1}\right|. \tag{14.386}$$

Notons que ce que l'équation (14.384) nous dit est que $H(t)$ est une primitive de $1/(1-t^2)$. Il faut choisir la bonne primitive en fixant une valeur.

Nous avons donc

$$\sum_{n=0}^{\infty}\frac{x^n}{2n+1}=\frac{1}{2\sqrt{x}}\ln\left|\frac{\sqrt{x}+1}{\sqrt{x}-1}\right| \tag{14.387}$$

pour $x>0$. Nous devons encore trouver ce que cela vaut pour $x<0$.

Nous posons successivement $X=-x$ puis $g(X)=f(-X)$. Ce que nous devons calculer est

$$g(t)=\frac{1}{t}\sum_{n=0}^{\infty}\frac{(-1)^nt^{2n+1}}{2n+1}. \tag{14.388}$$

Si nous posons

$$h(t)=\sum\frac{(-1)^nt^{2n+1}}{2n+1}, \tag{14.389}$$

alors

$$h'(t)=\sum(-1)^nt^{2n}=\sum(-t^2)^n=\frac{1}{1+t^2}, \tag{14.390}$$

par conséquent $h(t)=\arctan(t)$ (cela avait déjà été déduit à l'envers dans l'exemple 14.105).

Au final

$$f(x) = \sum_{n=0}^{\infty} \frac{x^n}{2n+1} = \begin{cases} \frac{1}{2\sqrt{x}} \ln \left| \frac{\sqrt{x}+1}{\sqrt{x}-1} \right| & \text{si } x > 0 \\ \frac{\arctan(\sqrt{-x})}{\sqrt{-x}} & \text{si } x < 0 \\ 1 & \text{si } x = 0. \end{cases} \tag{14.391}$$

Notons qu'elle est continue en zéro à gauche et à droite.

△

Exemple 14.108.
Nous considérons l'exemple suivant :

$$f(x) = \sum_{n=0}^{\infty} \frac{x^n}{3n+2}. \tag{14.392}$$

Nous posons $t = \sqrt[3]{x}$, et nous substituons :

$$\frac{x^n}{3n+2} = \frac{t^{3n}}{3n+2} = \frac{1}{t^2} \frac{t^{3n+2}}{3n+2}. \tag{14.393}$$

Nous devons étudier la fonction

$$g(t) = \sum_{n=0}^{\infty} \frac{t^{3n+2}}{3n+2} \tag{14.394}$$

Nous avons

$$g'(t) = \sum_{n=0} t^{3n+1} = t \sum_{n=0} t^{3n} = \frac{t}{1-t^3}. \tag{14.395}$$

Notons que $g(0) = 0$. △

Exemple 14.109.
Calculer le nombre

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1}. \tag{14.396}$$

Nous aurions envie de dire que cela est $f(-1)$ pour la fonction f donnée en (14.391). Le problème est que le rayon de convergence de f étant 1, rien n'est garantit quand au fait que la fonction y soit continue en $x = -1$. En particulier nous devons justifier le fait que

$$\lim_{x \to -1} \sum_n \frac{x^n}{2n+1} = \lim_{x \to -1} \frac{1}{\sqrt{-x}} \arctan(\sqrt{-x}). \tag{14.397}$$

Ce qui nous sauve est le critère d'Abel radial (théorème 14.21). En effet la série

$$\sum \frac{r^n}{2n+1} \tag{14.398}$$

étant convergente avec $r = -1$, la série correspondante est continue sur $[-1, 0]$. Nous pouvons donc calculer la série (14.396) en posant $x = -1$ dans (14.391) :

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} = \frac{\pi}{4}. \tag{14.399}$$

Note : la série (14.398) ne converge pas avec $r = 1$. La fonction f n'est pas continue en $x = 1$. △

Exemple 14.110.
Nous avons

$$\sum_{n=1}^{\infty} n x^{n-1} = \frac{1}{(1-x)^2}. \tag{14.400}$$

En effet si nous désignons par f la somme à gauche, nous trouvons que $f = g'$ avec

$$g(x) = \sum_{n=1}^{\infty} x^n. \tag{14.401}$$

Nous savons par ailleurs que $g(x) = 1/(1-x)$. Par conséquent

$$f(x) = \left(\frac{1}{1-x}\right)' = \frac{1}{(1-x)^2}. \tag{14.402}$$

△

14.11.3.3 Sage, primitives et logarithme complexe

14.111.
Attention : Sage pourrait nous induire en erreur si nous n'y prenions pas garde. En effet ce que vous ne savez pas mais que Sage sait, c'est que

$$\ln(-1) = i\pi. \tag{14.403}$$

Par conséquent Sage se permet de donner des primitives sans valeurs absolues dans le logarithme :

```
sage: f(x)=1/x
sage: f.integrate(x)
x |--> log(x)
```

La primitive à laquelle on s'attend d'habitude est $\ln(|x|)$. Ici la réponse est correcte parce que si x est négatif nous avons

$$\ln(x) = \ln\big((-1)|x|\big) = \ln(-1) + \ln(|x|). \tag{14.404}$$

Cette fonction est donc décalée de la primitive usuelle seulement de la constante $\ln(-1)$.

Un exemple plus élaboré :

```
sage: h(x)=1/(1-x**2)
sage: H=h.integrate(x)
sage: H
x |--> -1/2*log(x - 1) + 1/2*log(x + 1)
sage: H(0)
-1/2*I*pi
```

Exemple 14.112.
Encore une fois il faut faire attention en demandant la primitive à Sage :

```
----------------------------------------------------------------------
| Sage Version 4.7.1, Release Date: 2011-08-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: f(x)=x/(1-x**3)
sage: F=f.integrate(x)
sage: F(0)
-1/3*I*pi - 1/3*sqrt(3)*arctan(1/3*sqrt(3))
```

Cette fois la primitive proposée diffère de celle qu'on cherche de la constante complexe

$$-\frac{\pi}{3}i. \tag{14.405}$$

Mais il y a pire si nous voulons tracer. Nous voudrions définir la fonction $F_2(x) = F(x) - F(0)$. Mathématiquement c'est bien de cette fonction que nous parlons, mais :

```
sage: F2(x)=F(x)-F(0)
sage: F2(x)
1/3*I*pi - 1/3*sqrt(3)*arctan(1/3*(2*x + 1)*sqrt(3)) +
    +1/3*sqrt(3)*arctan(1/3*sqrt(3)) - 1/3*log(x - 1) + 1/6*log(x^2 + x + 1)
sage: F2.plot(x,-0.1,0.1)
verbose 0 (4101: plot.py, generate_plot_points)
        WARNING: When plotting, failed to evaluate function at 200 points.
verbose 0 (4101: plot.py, generate_plot_points)
        Last error message: 'unable to simplify to float approximation'
```

Il refuse de tracer. Pourquoi ? La partie complexe de l'expression de F_2 est mathématiquement nulle, mais elle est en deux parties :

$$\frac{\pi}{3} + \text{la partie imaginaire de} - \frac{1}{3}\ln(x-1). \tag{14.406}$$

Lorsque Sage tente de tracer, il donne à x un certain nombre de valeurs et calcule une *valeur approchée* de $\ln(x-1)$. Cette dernière ne se simplifie pas avec le nombre *exact* $\pi/3$. Sage reste donc avec une partie imaginaire qu'il ne peut pas tracer.

Notez la nuance :

```
sage: ln(-0.1)
-2.30258509299405 + 3.14159265358979*I
sage: ln(-1/10)
I*pi + log(1/10)
```

Du coup nous avons aussi

```
sage: F2(-0.1)
1/3*I*pi - 1/3*sqrt(3)*arctan(0.266666666666667*sqrt(3))
    + 1/3*sqrt(3)*arctan(1/3*sqrt(3)) - 0.0474885065133152 - 1.04719755119660*I
```

△

14.11.3.4 Nombres de Bell

Ici nous montrerions bien le théorème 14.126 sur les nombres de Bell parce que c'est essentiellement un résultat sur les séries entières et leurs manipulations. Hélas, il demande un tout petit peu d'équation différentielle (presque rien). Donc il est postposé jusqu'en page 1100.

14.12 Séries entières de matrices

Nous nous proposons d'étudier des séries de la forme

$$\sum_{k=0}^{\infty} a_k A^k \tag{14.407}$$

où A est une matrice. L'essentiel de la théorie va rester. Nous considérons une norme algébrique (définition 10.7), c'est-à-dire $\|AB\| \leqslant \|A\|\|B\|$.

14.12.1 Rayon de convergence

La notion de rayon de convergence de cette série reste la même : c'est la définition 14.8 qui ne dépend que des coefficients a_k et pas du tout de ce qu'on met à côté dans la somme. Évidemment il faudra montrer que dans le cas des matrices, le nom « rayon de convergence » n'est pas usurpé.

Proposition 14.113.
Soit (a_n) une suite dans $\mathbb{C}$ de rayon de convergence R et $A \in \mathbb{M}(n, \mathbb{R})$ une matrice vérifiant $\|A\| < R$. Alors la série

$$\sum_{k=0}^{\infty} a_k A^k \tag{14.408}$$

converge absolument, c'est-à-dire que $\sum_k \|a_k A^k\| < \infty$.

Démonstration. Nous avons les majorations

$$\|a_n A^n\| \leqslant |a_n| \|A^n\| \leqslant |a_n| \|A\|^n. \tag{14.409}$$

Par hypothèse $\|A\| < R$ et R est un supremum, donc il existe r tel que $\|A\| < r < R$ avec $(a_n r^n)$ borné. Nommons M un majorant de la suite $(a_n r^n)$. Alors nous avons

$$\|A_n A^n\| \leqslant |a_n| r^n \frac{\|A\|^n}{r^n} \leqslant M \left(\frac{\|A\|}{r} \right)^n. \tag{14.410}$$

La série du membre de droite converge parce que c'est une série géométrique de raison plus petite que 1 ; voir l'exemple 10.67. □

14.12.2 Convergence et rayon spectral

Le concept de rayon spectral permet aussi de donner des informations sur la convergence de séries de matrices. Pour rappel le rayon spectral d'une matrice est le maximum du module de ses valeurs propres (définition 10.8). Le rayon spectral de la matrice A est noté $\rho(A)$.

La proposition suivante sera redémontrée indépendamment dans le théorème 14.115.

Proposition 14.114 ([71])**.**
Si $A \in \mathbb{M}(n, \mathbb{C})$ est telle que $\rho(A) < 1$, alors $A^n \to 0$.

Démonstration. Nous nous plaçons dans une base des espaces caractéristiques [32] de A, c'est-à-dire que nous supposons que la matrice A a la forme

$$A = \begin{pmatrix} \lambda_1 \mathbb{1} + N_1 & & \\ & \ddots & \\ & & \lambda_s \mathbb{1} + N_s \end{pmatrix} \tag{14.411}$$

où les λ_i sont les valeurs propres de A et les N_i sont nilpotentes. En effet nous savons que l'espace caractéristique F_{λ_i} est l'espace de nilpolence de $A - \lambda_i \mathbb{1}$. Si nous notons A_i la restriction de A à cet espace, la matrice $N_i = A_i - \lambda_i \mathbb{1}$ est nilpotente. Du coup $A_i = \lambda_I \mathbb{1} + N_i$ et nous avons bien la décomposition (14.411).

Nous avons donc $A^n \to 0$ si et seulement si $(N_i + \lambda_i \mathbb{1})^n \to 0$ pour tout i. Soit donc N nilpotente et $\lambda < 1$ (parce que nous savons que toutes les valeurs propres de A sont inférieures à un). Nous avons

$$(\lambda \mathbb{1} + N)^n = \sum_{k=0}^{n} \binom{n}{k} \lambda^{n-k} N^k = \sum_{k=0}^{r-1} \binom{n}{k} \lambda^{n-k} N^k. \tag{14.412}$$

Nous voyons que le nombre de termes dans la somme ne dépend pas de n. De plus pour chacun de termes, la puissance de N ne dépend pas non plus de n. Le terme

$$\binom{n}{k} \lambda^{n-k} \leqslant P(n) \lambda^{n-k} \tag{14.413}$$

où P est un polynôme tend vers zéro lorsque n devient grand parce que c'est un cas polynôme fois exponentielle. □

32. Voir le théorème 9.195

Théorème 14.115 (Thème 44[209]).
Soit $A \in \mathbb{M}(n, \mathbb{K})$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$). Les affirmations suivantes sont équivalentes.

(1) $\lim_{k\to\infty} A^k = 0$
(2) $\rho(A) < 1$
(3) $\sum_{k=0}^{\infty} A^k$ *converge.*

Dans le cas où ces conditions sont vérifiées, nous avons aussi

— $\mathbb{1} - A$ *est inversible,*
— $\sum_{k=0}^{\infty} A^k = (\mathbb{1} - A)^{-1}$

Démonstration. Nous supposons qu'une norme est donnée sur $\mathbb{K}^n$ et nous considérons sur $\mathbb{M}(n, \mathbb{K})$ la topologie associée à la norme subordonnée[33]. Nous subdivisons la preuves en différentes implications.

(1) implique (2) Si $\rho(A) \leqslant 1$, en combinant la proposition 10.15 avec la proposition 10.17, nous avons

$$\|A^m\| \geqslant \big(\rho(A)\big)^m \geqslant 1 \tag{14.414}$$

Mais la limite $A^k \xrightarrow{\mathbb{M}(n,\mathbb{K})} 0$ signifie la limite $\|A^k\| \xrightarrow{\mathbb{R}} 0$. Le fait que tous les éléments de la suite soient plus grand que 1 empêche cette limite.

(2) implique (1) Vu que $\rho(A) < 1$, il existe $\epsilon > 0$ tel que $\rho(A) + \epsilon < 1$. Par le lemme 10.16 il existe une norme N sur $\mathbb{M}(n, \mathbb{K})$ telle que $N(A) \leqslant \rho(A) + \epsilon < 1$. Notons que cette norme N dépend de A et de ϵ.

Avec cette norme nous avons

$$N(A^k) \leqslant N(A)^k \xrightarrow{\mathbb{R}} 0. \tag{14.415}$$

Cela signifie que $A^k \xrightarrow{N} 0$. L'équivalence entre toutes les normes sur $\mathbb{M}(n, \mathbb{K})$ donne alors la convergence $A^k \xrightarrow{\|.\|} 0$.

Pour une preuve alternative de cette implication, voir la proposition 14.114.

(3) implique (1) La convergence d'une série implique que la norme du terme général converge vers zéro par la proposition 10.58. Nous avons donc $\|A^k\| \to 0$, ce qui signifie $A^k \to 0$, et donc $\rho(A) < 1$ parce que (1) implique (2).

$\rho(A) < 1$ implique $\mathbb{1} - A$ est inversible Si μ est une valeur propre de $\mathbb{1} - A$ alors

$$\det\big((\mathbb{1} - A) - \mu\mathbb{1}\big) = \det\big(A - (1 - \mu)\mathbb{1}\big), \tag{14.416}$$

donc $1 - \mu$ est une valeur propre de A. Donc les valeurs propres de $\mathbb{1} - A$ sont les nombres $1 - \lambda_i$ où les λß sont les valeurs propres de A. Par hypothèse, nous avons $\lambda_i < 1$ pour tout i, donc les valeurs propres de $\mathbb{1} - A$ sont toutes non nulles. Donc $\mathbb{1} - A$ est inversible (pas de noyau).

Le reste Nous montrons à présent que si $\rho(A) < 1$ alors $\sum_{k=0}^{\infty} A^k$ converge vers $\mathbb{1} - A$. Pour cela nous savons déjà que $\mathbb{1} - A$ est inversible. Nous posons

$$B_m = \mathbb{1} + A + \ldots + A^m, \tag{14.417}$$

ce qui donne immédiatement $AB_m = A + A^2 + \ldots + A^{m+1}$. Nous avons donc

$$(\mathbb{1} - A)B_m = \mathbb{1} - A^{m+1}. \tag{14.418}$$

Nous savons que $\lim_{m\to 0} A^m = 0$, donc

$$(\mathbb{1} - A)\sum_{k=0}^{\infty} A^k = \lim_{k\to\infty}(\mathbb{1} - A)B_k = \lim_{k\to\infty}(\mathbb{1} - A^{k+1}) = \mathbb{1}. \tag{14.419}$$

Notez au passage que nous avons permuté la somme avec le produit matriciel (voir 10.1.7).

□

33. Si on parle de convergence d'une suite, c'est qu'il y a une topologie quelque part.

14.12.3 Exponentielle et logarithme de matrice

La définition de l'exponentielle dans le cas des matrices est celle sur les algèbres normées non commutatives, 14.42.

Proposition 14.116.
L'application

$$\begin{aligned} \exp\colon \mathbb{M}(n,\mathbb{R}) &\to \mathbb{M}(n,\mathbb{R}) \\ A &\mapsto \sum_{k=0}^{\infty} \frac{A^k}{k!} \end{aligned} \tag{14.420}$$

est une application de classe C^∞. Sa différentielle en zéro est l'identité : $(d\exp)_0 = \mathrm{Id}$.

Démonstration. En ce qui concerne la continuité, nous savons que le rayon de convergence de la suite $\frac{1}{k!}$ est infini ; la proposition 14.95 conclu.

Pour la différentielle, c'est la proposition 14.94 qui nous permet d'écrire

$$d\exp_0(U) = \frac{d}{dt}\Big[\exp(tU)\Big]_{t=0} = \frac{d}{dt}\Big[\sum_{k=0}^{\infty}\frac{t^kU^k}{k!}\Big]_{t=0} = \sum_{k=0}^{\infty}\frac{kt^{k-1}U^k}{k!}\Bigg|_{t=0} = U \tag{14.421}$$

parce que seul le terme $k=1$ n'est pas nul. □

Nous avons vu par la proposition 14.96 que toute matrice complexe inversible a un logarithme. Nous allons maintenant parler de logarithme de matrices réelles avec une condition sur la norme. La formule ci-dessous montre explicitement que le logarithme est réel.

$$\begin{aligned} \ln\colon \{A\in\mathbb{M}(n,\mathbb{R}) \text{ tel que } \|A-\mathbb{1}\|<1\} &\to \mathbb{M}(n,\mathbb{R}) \\ A &\mapsto \sum_{k=0}^{\infty}(-1)^k\frac{(A-\mathbb{1})^{k+1}}{k+1}. \end{aligned} \tag{14.422}$$

Lemme 14.117.
Si $\|m\|<1$ dans $\mathbb{M}(n,\mathbb{R})$, alors nous posons

$$\ln(\mathbb{1}+m) = \sum_{k=0}^{\infty}(-1)^k\frac{m^{k+1}}{k+1}. \tag{14.423}$$

Cette fonction a les propriétés suivantes.

(1) Elle est de classe C^∞.

(2) Elle est un bon logarithme au sens où

$$e^{\ln(\mathbb{1}+m)} = \mathbb{1}+m. \tag{14.424}$$

(3) Elle vérifie l'approximation

$$\ln(1+m) = m+\sigma(m) \tag{14.425}$$

où σ a la propriété que

$$\lim_{k\to\infty} k\sigma\Big(\frac{m}{k}\Big) = 0. \tag{14.426}$$

Démonstration. Le rayon de convergence de la suite $a_k = \frac{(-1)^k}{k+1}$ est 1. Donc l'application donnée est C^∞ sur $B(0,1)$ par le théorème 14.95.

D'après la formule (14.423) nous avons

$$\sigma(m) = \sum_{l=1}^{\infty}(-1)^l\frac{m^{l+1}}{l+1}. \tag{14.427}$$

Nous avons alors

$$k\sigma(\frac{m}{k}) = \sum_{l=1}^{\infty}(-1)^l \frac{m^{l+1}}{k^l(l+1)}, \tag{14.428}$$

et donc

$$\|k\sigma(\frac{m}{k})\| \leqslant \sum_{l=1}^{\infty} \frac{\|m\|^{l+1}}{k^l(l+1)} \leqslant \frac{1}{k}\sum_{l=1}^{\infty} \frac{\|m\|^{l+1}}{l+1} \overset{k\to\infty}{\to} 0 \tag{14.429}$$

Cela prouve la dernière assertion. □

Proposition 14.118.
Soit V un espace vectoriel de dimension finie et $A \in \mathrm{End}(V)$. Nous considérons la fonction

$$\begin{aligned} f\colon \mathbb{R} &\to \mathrm{End}(V) \\ t &\mapsto e^{tA}. \end{aligned} \tag{14.430}$$

Cette fonction vérifie

$$f'(t) = \big(e^{tA}\big)' = Ae^{tA}. \tag{14.431}$$

Démonstration. Si nous posons $f_k(t) = \frac{t^k A^k}{k!}$ alors la fonction f est la somme : $f = \sum_{k=0}^{\infty} f_k$. Nous allons permuter la somme et la dérivation à l'aide du théorème 14.6. Vu que

$$f_k'(t) = \frac{kt^{k-1}A^k}{k!}, \tag{14.432}$$

la suite suite des dérivées converge normalement sur $\mathbb{R}$, nous pouvons dériver terme à terme pour obtenir

$$\Big(\sum_{k=0}^{\infty} t^k \frac{A^k}{k!}\Big) = \sum_{k=0}^{\infty} kt^{k-1}\frac{A^k}{k!} = \sum_{k=1}^{\infty} kt^{k-1}\frac{A^k}{k!} = A\sum_{k=1}^{\infty}\frac{A^{k-1}t^{k-1}}{(k-1)!} = Ae^{tA}. \tag{14.433}$$

Notez le jeu au niveau du point départ de la somme : elle passe de 0 à 1 parce que le terme zéro est nul, mais la simplification $\frac{k}{k!} = \frac{1}{(k-1)!}$ n'a pas de sens pour $k = 0$. □

Lemme 14.119 ([308]).
Soit $A \in \mathrm{End}(V)$ où V est un espace vectoriel réel de dimension finie. Si nous notons λ_i ($i = 1, \dots, r$) les valeurs propres distinctes de A alors il existe un polynôme $P \in \mathbb{R}[X]$ tel que

$$\|e^{tA}\| \leqslant P\big(|t|\big)\sum_{i=1}^{r} e^{t\,\mathrm{Re}(\lambda_i)}. \tag{14.434}$$

Démonstration. Le polynôme caractéristique de A se note, d'après le corolaire 9.142 de la façon suivante :

$$\chi_A(X) = \prod_{i=1}^{r}(X-\lambda_i)^{m_i} \tag{14.435}$$

où m_i est la multiplicité de la valeur propre λ_i. Le lemme des noyaux 9.88 nous dit qu'en posant

$$V_i = \ker(A - \lambda_i \mathbb{1})^{m_i} \tag{14.436}$$

nous avons $V = \bigoplus_{i=1}^{r} V_i$. Nous nommons $p_i\colon V \to V$ la projection canonique de E sur V_i ainsi que x_i la composante de $x \in V$ dans l'espace caractéristique V_i et nous posons $A_i = p_i \circ A$. Les espaces caractéristiques sont stables par A (lemme 9.193), donc $(Ax_i)_i = Ax_i$. Par conséquent $\sum_i A_i p_i = A$ parce que

$$\big(\sum_i p_i A p_i\big)(x) = \sum_i (Ax_i)_i = \sum_i Ax_i = A\sum_i x_i = Ax. \tag{14.437}$$

En ce qui concerne les puissances de A nous avons de même

$$A_i^n x_I = A_i \underbrace{A_i^{n-1}x_i}_{\in V_i} = AA_i^{n-1}x_i = A^n x_i, \tag{14.438}$$

et donc

$$\sum_{i=1}^{r} A_i^n p_i = A^n. \tag{14.439}$$

En particulier,

$$e^{tA} = \sum_i e^{tA_i} p_i. \tag{14.440}$$

C'est de cette exponentielle de matrice que nous devons étudier la norme.

La décomposition de Dunford du théorème 9.196 est toujours un bon plan pour traiter avec les exponentielles : nous avons $A = s + n$ avec

$$s = \sum_k \lambda_k p_k, \quad n = \sum_k (A - \lambda_k \mathbb{1}) p_k. \tag{14.441}$$

Nous montrons que la décomposition de Dunford de $p_i A$ est $p_i A = p_i s + p_i n$. Nous avons

$$p_i s = \sum_k \lambda_k p_i p_k = \lambda_i p_i \tag{14.442}$$

qui est bien diagonalisable. De plus les espaces caractéristiques sont stables par n, donc $p_i n$ est nilpotent. Enfin ils commutent :

$$[p_i s, p_i n] = \lambda_i (p_i n - p_i n p_i). \tag{14.443}$$

Vu que n préserve les espaces caractéristiques, lorsque $v \in V_k$ avec $k \neq i$ nous avons $p_i n p_i v = 0$ et $p_i n v = 0$. Mais si $v \in V_i$ alors

$$p_i n p_i v = p_i n v = n v \tag{14.444}$$

et $p_i n v = n v$, donc les opérateurs $p_i n$ et $p_i n p_i$ sont égaux et (14.443) donne bien zéro. En ce qui concerne l'exponentielle de A_i nous avons

$$e^{p_i A} = e^{p_i s} e^{p_i n} = e^{\lambda_i p_i} \exp\big((A - \lambda_i \mathbb{1}) p_i\big). \tag{14.445}$$

Nous pouvons maintenant sérieusement nous attaquer à la norme de e^{tA} de l'équation (14.440). D'abord nous avons $\|p_i\| = 1$ parce que l'opérateur p_i est l'identité sur au moins un vecteur (en fait tout ceux de l'espace caractéristique V_i). En utilisant les propriétés de la norme opérateur [34], nous trouvons dans un premier temps [35] :

$$\|e^{tA}\| \leqslant \sum_{i=1}^{r} \|e^{tA_i}\| \leqslant \sum_{i=1}^{r} |e^{t\lambda_i}| \underbrace{\sum_{k=0}^{m_i} \frac{|t|^k}{k!} \|A - \lambda_i \mathbb{1}_i\|^k}_{=P_i(|t|}} \tag{14.446}$$

où $\mathbb{1}_i$ est l'opérateur identité sur V_i. Petit détail dans le calcul :

$$\|e^{\lambda_i p_i}\| \leqslant \sum_{l=0}^{\infty} \frac{\lambda_i^l}{k!} \|p_i\|^l = e^{\lambda_i}. \tag{14.447}$$

Notons que tous les termes de $P_i(|t|)$ et $P_i\big(|t|\big)$ sont positifs, de telle sorte que nous pouvons majorer en ajoutant des termes partout. À la place d'avoir $P_i(|t|)$ comme coefficient de $|e^{t\lambda_i}|$ nous majorons en mettant $\sum_{j=1}^{r} P_j(|t|)$ comme coefficient :

$$\|e^{tA}\| \leqslant \sum_{i=1}^{r} |e^{t\lambda_i}| P_i\big(|t|\big) = \sum_{i=1}^{r} |e^{t\lambda_i}| \sum_{j=1}^{r} P_j\big(|t|\big) = P\big(|t|\big) \sum_{i=1}^{r} e^{t\,\mathrm{Re}(\lambda_i)}. \tag{14.448}$$

L'arrivée de la partie réelle est une égalité usuelle pour les nombres complexes : $|e^{a+bi}| = e^a |e^{bi}| = e^a$. □

34. Surtout le fait que ce soit une norme d'algèbre, lemme 10.13.
35. Si les valeurs propres de A sont λ_i, celles de tA sont $t\lambda_i$.

14.12.4 Calcul effectif de l'exponentielle d'une matrice

Nous reprenons l'exemple de [309]. Soit A une matrice dont le polynôme minimum s'écrit

$$P(X) = (X-1)^2(X-2). \tag{14.449}$$

Par le théorème 9.88 de décomposition des noyaux nous avons

$$E = \ker(A-1)^2 \oplus \ker(A-2). \tag{14.450}$$

En suivant les notations de ce théorème nous avons $P_1(X) = (X-1)^2$, $P_2(X) = X-2$ et

$$Q_1(X) = X-2 \tag{14.451a}$$
$$Q_2(X) = (X-1)^2. \tag{14.451b}$$

Les polynômes R_i dont l'existence est assurée par le théorème de Bézout sont

$$\begin{aligned} R_1(X) &= -X \\ R_2(X) &= 1. \end{aligned} \tag{14.452}$$

Nous avons

$$R_1Q_1 + R_2Q_2 = 1. \tag{14.453}$$

Le projecteur p_i sur $\ker P_i$ est R_iQ_i :

$$\begin{aligned} p_1 &= -A(A-2) = \mathtt{proj}_{\ker(u-1)^2} \\ p_2 &= (A-1)^2 = \mathtt{proj}_{\ker(u-2)}. \end{aligned} \tag{14.454}$$

Passons maintenant au calcul de l'exponentielle[36]. Nous avons évidemment

$$e^A = e^Ap_1 + e^Ap_2. \tag{14.455}$$

Étant donné que p_1 est le projecteur sur le noyau de $(A-1)^2$, nous avons

$$e^Ap_1 = ee^{A-1}p_1 = ep_1 + e(u-1)1 = ep_1 = -Ae(A-2). \tag{14.456}$$

En effet $e^{A-1}p_1 = \sum_{k=0}^{\infty}(A-1)^k \circ p_1$. De la même façon nous avons

$$e^Ap_2 = e^2e^{A-2}p_2 = e^2p_2 = e^2(A-1)^2. \tag{14.457}$$

Au final,

$$e^A = -Ae(A-2) + e^2(A-1)^2. \tag{14.458}$$

14.13 Lemme de Borel

14.13.1 Fonctions plateaux

Vous voulez une fonction de classe C^∞ qui n'est pas nulle partout ? En voici une.

Lemme 14.120.
La fonction

$$\varphi(x) = \begin{cases} e^{-1/x} & \text{si } x > 0 \\ 0 & \text{sinon.} \end{cases} \tag{14.459}$$

est de classe C^∞.

36. Définition 14.42. Thème 58

Démonstration. Pour tout polynôme P nous avons la limite[37]

$$\lim_{x\to 0^+} \frac{e^{-1/x}}{P(x)} = 0. \tag{14.460}$$

De là, en écrivant les dérivées successives de φ, il est facile de voir qu'elles sont continues en $x = 0$. □

Proposition 14.121.
Soient $a < b < c < d$ dans $\mathbb{R}$. Il existe une fonction $f \in C^\infty(\mathbb{R})$ à valeurs positives telle que

(1) $f(x) = 1$ *si* $x \in [b, c]$

(2) $\operatorname{supp}(f) \subset [a, d]$.

Démonstration. Nous partons de la fonction φ du lemme 14.120. Ensuite nous considérons

$$\psi_m(x) = 1 - \frac{\int_0^x \varphi(t)dt}{\int_0^m \varphi(t)dt} \tag{14.461}$$

Cette fonction est encore de classe C^∞. En effet, $\int_0^m$ est un simple nombre strictement positif sans histoires tandis que la proposition 13.235 dit que $x \mapsto \int_0^x \varphi(t)dt$ est une primitive de φ. Vu que φ est déjà de classe C^∞, sa primitive l'est également.

La fonction ψ_m vérifie

$$\psi_m(x) = \begin{cases} 1 & \text{si } x < 0 \\ 0 & \text{si } x > m \\ \text{positive} & \text{si } x \in [0, m]. \end{cases} \tag{14.462}$$

À partir de là nous considérons les fonctions

$$f_1(x) = \psi_{d-c}(x - c) = \begin{cases} 1 & \text{si } x < c \\ 0 & \text{si } x > d \\ \text{positive} & \text{si } x \in [c, d]. \end{cases} \tag{14.463a}$$

$$f_2(x) = \psi(b - a)(b - x) = \begin{cases} 0 & \text{si } x < a \\ 1 & \text{si } x > b \\ \text{positive} & \text{si } x \in [a, b]. \end{cases}, \tag{14.463b}$$

et finalement la fonction suivante répond à la question des fonctions plateaux sur $\mathbb{R}$:

$$f(x) = f_1(x)f_2(x). \tag{14.464}$$

□

Une variation sur le même thème est l'existence de fonctions infiniment dérivables à support compact, c'est-à-dire des fonctions dans $C_c^\infty(\mathbb{R}^d) = \mathscr{D}(\mathbb{R}^d)$.

Proposition 14.122.
La fonction

$$\xi(x) = \begin{cases} e^{-1/(1-|x|^2)} & \textit{si } x \in B(0, 1) \\ 0 & \textit{sinon.} \end{cases} \tag{14.465}$$

est de classe C^∞ et à support compact.

Le lemme d'Urysohn comprend de nombreuses variantes plus ou moins générales. En voici une parmi les plus simples.

37. Voir la proposition 14.78(3).

Lemme 14.123 (Lemme d'Urysohn)**.**
Soit un ouvert $\mathcal{U}$ de $\mathbb{R}$ et un compact K inclus dans $\mathcal{U}$. Il existe une fonction $f\colon \mathbb{R} \to [0,1]$ de classe C^∞ telle que

(1) $f(x) = 1$ *pour* $x \in K$,

(2) $\operatorname{supp}(f) \subset \mathcal{U}$.

Démonstration. Vu que K est compact, il est borné (théorème 7.127). Nous posons $b = \min(K)$ et $c = \max(K)$. En particulier b et c sont des éléments de K et donc de $\mathcal{U}$. Comme $\mathcal{U}$ est ouvert, il existe des boules centrées en b et c contenues dans $\mathcal{U}$. Soit r le rayon de telles boules :

$$B(b,r) \subset \mathcal{U} \tag{14.466a}$$
$$B(c,r) \subset \mathcal{U}. \tag{14.466b}$$

Nous posons $a = b - r \in \mathcal{U}$ et $d = c + r \in \mathcal{U}$. Nous considérons à présent la fonction plateau f de la proposition 14.121. Elle est de classe C^∞ et vérifie $f(x) = 1$ pour $x \in [b,c] = K$ ainsi que $f(x) = 0$ hors de $[a,d] \subset \mathcal{U}$. □

Notons que la fonction du lemme 14.123 n'épouse pas spécialement très bien la forme de K. Si par exemple $K = [0,1] \cup [10,11]$, la fonction f sera égale à 1 au moins sur $[0,11]$.

14.13.2 Le lemme de Borel

Lemme 14.124 (Lemme de Borel[54])**.**
Soit (a_n) une suite dans $\mathbb{R}$. Il existe une fonction $u \in C^\infty(\mathbb{R})$ telle que $u^{(k)}(0) = a_k$ pour tout $k \geqslant 0$.

Démonstration. Soit $\varphi \in C_c^\infty(\mathbb{R})$ une fonction telle que $\varphi(x) = 1$ si $|x| \leqslant \frac{1}{2}$ et telle que $\operatorname{supp}(\varphi) \subset]-1,1[$.

Nous commençons par considérer une suite de réels strictement positifs (λ_k) dont nous fixerons une valeur précise plus tard, et nous posons

$$f_k(x) = \varphi(\lambda_k x)\frac{a_k}{k!}x^k. \tag{14.467}$$

Nous allons étudier la convergence et les propriétés de $u(x) = \sum_{k=0}^\infty f_k(x)$.

Calculons (formellement) la m^{e}dérivée de f_k :

$$f_k^{(m)}(x) = \frac{a_k}{k!}\sum_{l=0}^m \binom{m}{l}\lambda_k^{m-l}\varphi^{(m-l)}(\lambda_k x)(x^k)^{(l)} \tag{14.468a}$$
$$= a_k \sum_{l=0}^m \binom{l}{m}\lambda_k^{m-l}\varphi^{(m-l)}(\lambda_k x)\frac{x^{k-l}}{(k-l)!}. \tag{14.468b}$$

Notons que nous travaillons à m fixé et que nous ne nous intéressons qu'aux termes avec k assez grand ; nous pouvons donc supposer $k \geqslant m$. De toutes façons pour $\sum_{k=0}^m f_k$, on a la classe C^∞, et la permutation de la somme avec tout ce qu'on veut. Vu que φ est continue à support compact nous pouvons poser

$$M_m = \max_{0\leqslant j\leqslant m}\|\varphi^j\|_\infty = \max_{0\leqslant j\leqslant m}\max_{x\in\mathbb{R}}|\varphi^{(j)}(x)|. \tag{14.469}$$

Nous continuons en nous fixant un $x \in \mathbb{R}$ et un $k \geqslant m$.

Si $|x| > \frac{1}{\lambda_k}$, alors $\varphi^{(m)}(\lambda_k x) = 0$ parce que $\lambda_k x$ est strictement hors du support de φ qui est $]-1,1[$. Donc pour $|x| > \frac{1}{\lambda_k}$.

Si par contre $|x| \leqslant \frac{1}{\lambda_k}$, nous avons les majorations

$$|f_k^{(m)}(x)| \leqslant |a_k| \sum_{l=0}^{m} \binom{l}{m} |\lambda_k|^{m-l} \underbrace{\varphi^{(m-l)}(\lambda_k x)}_{\leqslant M_m} \frac{1}{(k-l)!} \underbrace{|x|^{k-l}}_{\leqslant (1/\lambda_k)^{k-l}} \tag{14.470a}$$

$$\leqslant |a_k| M_m |\lambda_k|^{m-k} \frac{1}{(k-m)!} \sum_{l=0}^{m} \binom{m}{l} \tag{14.470b}$$

$$\leqslant \frac{|a_k| M_m |\lambda_k|^{m-k} 2^m}{(k-m)!} \tag{14.470c}$$

$$= \frac{|a_k| M_m 2^m}{(k-m)! |\lambda_k|^{k-m}} \tag{14.470d}$$

où pour faire disparaitre la somme de coefficients binomiaux, nous avons remarqué que $\sum_{l=0}^{m} \binom{m}{l}$ est le nombre total de termes dans le développement de $(a+b)^m$, c'est-à-dire 2^m. Nous voulons, pour m fixé, étudier la convergence de la somme de cela. Notons que le 2^m n'a en particulier strictement aucune importance parce qu'on travaille à m fixé.

Nous fixons maintenant la valeur des λ_k :

$$\lambda_k = \max\{|a_k|, 1\}. \tag{14.471}$$

Avec cela, en nous souvenant que nous n'étudions que les termes $k > m$, le dénominateur de (14.470d) est réellement croissant en k, donc nous avons la majoration

$$|f_k^{(m)}(x)| \leqslant \frac{M_m 2^m}{(k-m)!}. \tag{14.472}$$

Au final nous avons

$$\|f_k^{(m)}\|_\infty \leqslant \frac{2^m M_m}{(k-m)!}. \tag{14.473}$$

Et la somme de cela converge sans difficultés. Donc la série

$$u(x) = \sum_{k=0}^{\infty} f_k^{(m)}(x) \tag{14.474}$$

converge normalement et donc uniformément sur $\mathbb{R}$. Nous pouvons alors permuter la somme et la dérivation par le théorème 14.3. Donc

$$u^{(m)} = \sum_{k=0}^{\infty} f_k^{(m)} \tag{14.475}$$

est continue. En particulier, pour évaluer en zéro, on peut faire

$$u^{(m)}(0) = \sum_{k=0}^{\infty} f_k^{(m)}(0). \tag{14.476}$$

Nous avons

$$f_k(x) = \varphi(\lambda_k x) \frac{a_k}{k!} x^k. \tag{14.477}$$

Pour calculer la dérivée en zéro, il suffit de la calculer sur un voisinage sur lequel $\varphi(\lambda_k x)$ est la constante 1 ; un tel voisinage existe pour tout k. À ce moment le calcul est classique :

$$f_k^{(m)}(x) = \begin{cases} a_k & \text{si } k = m \\ 0 & \text{sinon.} \end{cases} \tag{14.478}$$

Finalement nous avons bien

$$u^{(m)}(0) = \sum_{k=0}^{\infty} f_k^{(m)}(0) = a_k. \tag{14.479}$$

□

Remarque 14.125.
Pour prouver le lemme de Borel, la première chose qui passe par la tête est la fonction toute simple

$$u(x) = \sum_{k=0}^{\infty} \frac{a_k}{k!} x^k. \tag{14.480}$$

Évidemment si on calcule les dérivées successives de cette fonction, nous trouvons les bons résultats. Le problème est la convergence. Rien qu'en prenant $a_k = k!k^k$, la série ne converge pour aucun x positif. L'idée de multiplier chacun de f_k par une fonction plateau sur un petit intervalle autour de zéro a plusieurs avantages. D'abord on conserve les dérivées correctes parce qu'on ne touche pas à la valeur des f_k sur un petit voisinage. Ensuite cela ne modifie pas la continuité ; et enfin en multipliant par $\varphi(\lambda_k x)$, ça calme méchamment les divergences parce que $\lambda_k x$ passe vite au dessus de 1 (et donc en dehors du support de φ) si λ_k est grand. D'où le fait qu'il soit normal que les λ_k soient de l'ordre des a_k.

14.14 Nombres de Bell

Théorème 14.126 (Nombres de Bell[54])**.**
Soient $n \geqslant 1$ et B_n le nombre de partitions distinctes de l'ensemble $\{1, \dots, n\}$ avec la convention que $B_0 = 0$. Alors

(1) La série entière

$$\sum_{n=0}^{\infty} \frac{B_n}{n!} x^n \tag{14.481}$$

a un rayon de convergence $R > 0$ et sa somme est donnée par

$$f(x) = e^{e^x - 1} \tag{14.482}$$

pour tout $x \in]-R, R[$.

(2) Pour tout $k \in \mathbb{N}$,

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}. \tag{14.483}$$

(3) Le rayon de convergence de la série (14.481) est en réalité infini : $R = \infty$.

Démonstration. (1) Soient $n \geqslant 1$ et $0 \leqslant k \leqslant n$. Nous notons E_k l'ensemble des partitions de $\{1, \dots, n+1\}$ pour lesquelles le « paquet » contenant $n+1$ soit de cardinal $k+1$. Calculons le cardinal de E_k.

Pour construire un élément de E_k, il faut d'abord prendre le nombre $n+1$ et lui adjoindre k éléments choisis dans $\{1, \dots, n\}$, ce qui donne $\binom{n}{k}$ possibilités. Ensuite il faut trouver une partition des $(n+1) - (k+1) = n - k$ éléments restants, ce qui fait B_{n-k} possibilités. Donc

$$\mathrm{Card}(E_k) = \binom{n}{k} B_{n-k}. \tag{14.484}$$

L'intérêt des ensembles E_k est que $\{E_0, \dots, E_n\}$ est une partition de l'ensemble des partitions de $\{1, \dots, n+1\}$, c'est-à-dire que $B_{n+1} = \sum_{k=0}^{n} \mathrm{Card}(E_k)$, ce qui va nous donner une relation de récurrence pour les B_n :

$$B_{n+1} = \sum_{k=0}^{n} \mathrm{Card}(E_k) = \sum_{k=0}^{n} \binom{n}{k} B_{n-k} = \sum_{l=0}^{n} \binom{n}{n-l} B_l = \sum_{l=0}^{n} \binom{n}{l} B_l. \tag{14.485}$$

où nous avons utilisé un petit changement de variables $l = n-k$. Afin d'étudier la convergence de la série (14.481), nous allons montrer par récurrence que pour tout n, $B_n < n!$. D'abord

pour $n = 0$ c'est bon : $B_1 = 1$ parce que la seule partition de $\{1\}$ est $\{1\}$. Supposons que l'inégalité soit vraie pour une certaine valeur k, et montrons qu'elle est vraie pour la valeur $k+1$:

$$B_{k+1} = \sum_{l=0}^{n} \binom{n}{k} B_k \leqslant \sum_{l=0}^{n} \binom{n}{k} k! = k! \sum_{l=0}^{k} \underbrace{\frac{1}{(n-k)!}}_{\leqslant 1} \leqslant n!(n+1) = (n+1)! \tag{14.486}$$

où nous avons utilisé la formule $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
Donc pour tout $x \in \mathbb{R}$ nous avons

$$0 \leqslant \frac{B_n}{n!} |x^n| \leqslant |x|^n, \tag{14.487}$$

et donc la série a un rayon de convergence au moins aussi grand que celui de la série géométrique, c'est-à-dire que 1. Donc $R \geqslant 1$. Nous nommons R ce rayon de convergence.

(2) Soit $x \in \,]-R, R[$. Pour une telle valeur de x à l'intérieur du disque de convergence, la proposition 14.25 nous permet de dériver terme à terme la série [38]

$$f(x) = \sum_{k=0}^{\infty} \frac{B_k}{k!} x^k = 1 + \sum_{k=0}^{\infty} \frac{B_{k+1}}{(k+1)!} x^{k+1}, \tag{14.488}$$

pour obtenir

$$f'(x) = \sum_{k=0}^{\infty} \frac{B_{k+1}}{(k+1)} (k+1) x^k = \sum_{k=0}^{\infty} \frac{B_{k+1}}{k!} x^k \tag{14.489a}$$

$$= \sum_{k=0}^{\infty} \left(\sum_{l=0}^{k} \binom{k}{l} B_l \right) \frac{x^k}{k!} = \sum_{k=0}^{\infty} \left(\sum_{l=0}^{k} \frac{B_l}{l!(l-k)!} \right) x^k. \tag{14.489b}$$

<++> En cette expression, nous reconnaissons un produit de Cauchy (proposition 14.15) avec $a_l = \frac{B_l}{l!}$ et $b_n = \frac{1}{n!}$. Vu que ce sont deux séries ayant un rayon de convergence plus grand que zéro, le produit a encore un rayon de convergence plus grand que zéro et nous pouvons prendre le produit des séries :

$$f'(x) = \left(\sum_{l=0}^{\infty} \frac{B_l}{l!} x^l \right) \left(\sum_{k=0}^{\infty} \frac{1}{k!} x^k \right) = f(x) e^x. \tag{14.490}$$

Étudions l'équation différentielle $y' = ye^x$. D'abord par un argument en lacet de chaussure [39], une solution est de classe C^∞. Ensuite si une solution est non nulle, elle est de signe constant. En effet si $y(x_0) < 0$ et $y(x_1) = 0$ (on choisit x_1 minimum pour cette propriété parmi les nombres plus grands que x_0) alors il existe [40] un $t \in \,]x_0, x_1[$ tel que $y'(t) > 0$, ce qui donnerait $y(t) > 0$, ce qui contredirait la minimalité de x_1.
Nous prétendons [41] que cette équation différentielle a un espace de solutions de dimension 1. En effet, si $y' = ye^x$ et $g' = ge^x$ alors en posant $\varphi = y/g$ nous obtenons tout de suite $\varphi' = 0$, ce qui signifie que φ est constante, ou encore que y et g sont multiples l'un de l'autre.
Si nous en trouvons une non nulle par n'importe quel moyen, c'est bon. Une solution étant dérivable est continue, donc l'équation $f' = fe^x$ nous indique que f' est continue. Une solution non nulle va automatiquement accepter un petit voisinage sur lequel la manipulation suivante a un sens :

$$\frac{f'(x)}{f(x)} = e^x, \tag{14.491}$$

38. C'est ici qu'on utilise la convention $B_0 = 0$ et ça aura une influence sur le choix de la constante K plus bas.
39. Genre ce qui est fait pour prouver 14.62(4).
40. Théorème de Rolle 11.127.
41. Ou alors on utilise le théorème 31.14 avec $M(x) = e^x$ dans les cas $n = 1$ et $I = \,]-R, R[$.

donc $\ln\big(|f(x)|\big) = e^x + C$ et $f(x) = Ke^{e^x}$ pour une certaine constante. Il est vite vérifié que cette fonction est une solution de l'équation différentielle $y'(x) = y(x)e^x$ et par unicité, toutes les solutions sont de cette forme. Autrement dit, l'espace des solutions est l'espace vectoriel $\mathrm{Span}\{x \mapsto e^{e^x}\}$. Étant donné que $f(0) = 0$, nous devons choisir $K = \frac{1}{e}$ et donc

$$f(x) = \frac{1}{e}e^{e^x} = e^{e^x-1}. \tag{14.492}$$

(3) Nous commençons par écrire la fonction f comme une série de puissance. La partie simple du calcul : pour $x \in]-R, R[$, nous avons

$$e^{e^x} = \sum_{k=0}^{\infty} \frac{(e^x)^k}{k!} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{l=0}^{\infty} \frac{(kx)^l}{l!} = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{k^l}{k!} \frac{x^l}{l!}. \tag{14.493}$$

Notons que cela n'est pas une série de puissance en x parce qu'il y a la double somme. Nous allons inverser les sommes au moyen du théorème de Fubini sous la forme du corolaire 13.261. Pour cela nous considérons la fonction

$$\begin{aligned} a\colon \mathbb{N} \times \mathbb{N} &\to \mathbb{R} \\ (k, l) &\mapsto \frac{(kx)^l}{k!l!} \end{aligned} \tag{14.494}$$

et nous mettons la mesure de comptage [42] sur $\mathbb{N}$ et $\mathbb{N}^2$. Nous commençons donc à vérifier l'intégrabilité variable par variable de $|a|$:

$$\int_{\mathbb{N}} \left(\int_{\mathbb{N}} |a(k, l)| dm(l) \right) dm(k) = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{(k|x|)^l}{l!} \tag{14.495a}$$

$$= \sum_{k=0}^{\infty} \frac{1}{k!} e^{k|x|}. \tag{14.495b}$$

Nous devons montrer que cette dernière somme va bien. Pour cela nous posons $u_k = \frac{e^{k|x|}}{k!}$ et nous remarquons que $\frac{u_{k+1}}{u_k} \to 0$. Donc la double intégrale (14.495) converge, ergo $a \in L^1(\mathbb{N} \times \mathbb{N})$, ce qui nous permet d'utiliser le théorème de Fubini 13.262 pour inverser les ~~sommes~~ ~~intégrales~~ sommes dans l'équation (14.493) :

$$\frac{1}{e}e^{e^x} = \frac{1}{e} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{1}{k!} \frac{1}{l!} (kx)^l = \sum_{l=0} \frac{1}{e} \frac{1}{l!} \left(\sum_{k=0}^{\infty} \frac{k^l}{k!} \right) x^l. \tag{14.496}$$

Cela est un développement en série entière pour la fonction $\frac{1}{e}e^{e^x}$, dont nous savions déjà le développement (14.481) ; par unicité du développement nous pouvons identifier les coefficients :

$$B_l = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^l}{k!}. \tag{14.497}$$

(4) Le développement (14.493) étant en réalité valable pour tout x et tous les calculs subséquents l'étant aussi, le développement

$$e^{e^x-1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} x^n \tag{14.498}$$

est en fait valable pour tout x, ce qui donne à la série entière un rayon de convergence infini. □

42. Nous passons outre les avertissements et menaces de Arnaud Girand.

Chapitre 15

Représentations et caractères

15.1 Représentations et caractères

Définition 15.1.
Si G est un groupe, l'ensemble des homomorphismes $\mathrm{Hom}(G, \mathbb{C}^*)$ est un groupe pour la multiplication. Un élément de $\mathrm{Hom}(G, \mathbb{C}^*)$ est un ***caractère abélien****. Le nom « abélien » vient du fait que le caractère prenne ses valeurs dans $\mathbb{C}^*$. Nous notons $\hat{G} = \mathrm{Hom}(G, \mathbb{C}^*)$.*

Théorème 15.2.
Soit G un groupe abélien fini. Alors G est isomorphe à $\hat{G}$.

L'isomorphisme n'est pas canonique.

Démonstration. Étant donné la structure des groupes abéliens finis donnée par le théorème 5.23, nous commençons par nous concentrer sur $G = \mathbb{Z}/n\mathbb{Z}$. Nous allons montrer que

$$\mathrm{Hom}(\mathbb{Z}/n\mathbb{Z}) \simeq \mathrm{U}_n = \{\xi \in \mathbb{C} \text{ tel que } \xi^n = 1\}. \tag{15.1}$$

Pour cela nous avons l'isomorphisme

$$\begin{aligned} \psi\colon \mathrm{Hom}(\mathbb{Z}, \mathbb{C}^*) &\to \mathbb{C}^* \\ f &\mapsto f(1). \end{aligned} \tag{15.2}$$

Notons que si $f \in \mathrm{Hom}(\mathbb{Z}, \mathbb{C}^*)$, alors $f(k) = f(1)^k$, donc ψ est bien un isomorphisme. Cela nous amène à définir

$$\begin{aligned} \varphi\colon \mathrm{Hom}\Big((\mathbb{Z}/n\mathbb{Z}, +), (\mathbb{C}^* n \ \cdot\ )\Big) &\to \mathrm{U}_n \\ g &\mapsto f(1). \end{aligned} \tag{15.3}$$

Remarquons que pour tout $f \in \mathrm{Hom}(\mathbb{Z}/n\mathbb{Z}, \mathbb{C}^*)$ on a bien $f(1)^n = 1$. En effet si $[k] \in \mathbb{Z}/n\mathbb{Z}$, alors $f\big([k]\big) = f(1)^k$ et en particulier

$$f(1)^n = f([n]) = f(0) = 1. \tag{15.4}$$

Donc $f(1) \in \mathrm{U}_n$. Le φ est injective parce que si $f(1) = g(1)$ alors $f = g$ du fait que $f(k) = f(1)^k = g(1)^k = g(k)$.

Nous en sommes à avoir prouvé que $\mathrm{Hom}(\mathbb{Z}/n\mathbb{Z}, \mathbb{C}^*) \simeq \mathrm{U}_n$ (introduit au lemme 18.2). Il faudrait encore montrer que $\mathrm{U}_n \simeq \mathbb{Z}/n\mathbb{Z}$. Pour cela nous nous rappelons du lemme 18.3 nous ayant raconté que le groupe U_n des racines de l'unité était cyclique et d'ordre n. Il est donc bien isomorphe à $\mathbb{Z}/n\mathbb{Z}$.

Passons au cas où

$$G \simeq \mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_2\mathbb{Z} \times \ldots \times \mathbb{Z}/n_k\mathbb{Z}. \tag{15.5}$$

Dans ce cas nous montrons que

$$\begin{aligned}\alpha\colon \bigtimes_{i=1}^{k} \operatorname{Hom}(\mathbb{Z}/d_i\mathbb{Z}, \mathbb{C}^*) &\to \operatorname{Hom}(G, \mathbb{C}^*)\\ \alpha(\chi_1, \dots, \chi_k)(g_1, \dots, g_k) &= \chi_1(g_1)\dots\chi_k(g_k).\end{aligned} \tag{15.6}$$

Ce α est injectif parce qu'en appliquant l'égalité

$$\alpha(\chi_1, \dots, \chi_k) = \alpha(\chi'_1, \dots, \chi'_k) \tag{15.7}$$

à l'élément $g = (9, \dots, 1, \dots, 0)$ alors nous trouvons $\chi_i(1) = \chi'_i(1)$ parce que $\chi_j(0) = 1$. Du coup $\chi_i = \chi'_i$.

L'application α est en plus surjective. En effet si $\chi \in \operatorname{Hom}(G, \mathbb{C}^*)$, alors nous définissons

$$\chi_i(g_i) = \chi(0, \dots, g_i, \dots, 0), \tag{15.8}$$

et nous avons alors $\alpha(\chi_1, \dots, \chi_k) = \chi$.

Nous devons encore montrer que α est un homomorphisme. Si $\chi, \chi' \in \bigtimes_{i=1}^{k} \operatorname{Hom}(\mathbb{F}_{d_i}, \mathbb{C}^*)$, alors

$$\begin{aligned}\alpha(\chi\chi')(g_1, \dots, g_k) &= (\chi_1\chi'_1)(g_1)\dots(\chi_k\chi'_k)(g_k) &(15.9a)\\ &= \chi_1(g_1)\dots\chi_k(g_k)\chi'_1(g_1)\dots\chi'_k(g_k) &(15.9b)\\ &= \alpha(\chi)(g_1, \dots, g_k)\alpha(\chi')(g_1, \dots, g_k) &(15.9c)\\ &= \big(\alpha(\chi)\alpha(\chi')\big)(g_1, \dots, g_k). &(15.9d)\end{aligned}$$

Donc $\alpha(\chi\chi') = \alpha(\chi)\alpha(\chi')$. □

Théorème 15.3.
Soit G un groupe abélien fini. Les groupes G et $\hat{\hat{G}}$ sont isomorphes et un isomorphisme canonique est donné par $\alpha\colon g \mapsto f_g$ donné par

$$f_g(\chi) = \chi(g). \tag{15.10}$$

Démonstration. D'abord f_g est bien un caractère de $\hat{G}$ parce que

$$f_g(\chi\chi') = (\chi\chi')(g) = \chi(g)\chi'(g) = f_g(\chi)f_g(\chi'). \tag{15.11}$$

Le fait que α soit un homomorphisme de groupes est direct :

$$f_{gg'}(\chi) = \chi(gg') = \chi(g)\chi(g') = f_g(\chi)f_{g'}(\chi) = (f_gf_{g'})(\chi). \tag{15.12}$$

D'autre part nous savon que G et $\hat{\hat{G}}$ ont le même cardinal. Il suffit donc de prouver l'injectivité de α pour être sûr de la bijectivité. Pour cela nous devons prouver que si $g \neq e$ alors $f_g \neq f_e$. Nous savons que pour tout caractère $\chi \in \hat{G}$, $f_e(\chi) = \chi(e) = 1$. Donc pour tout $g \in G\backslash\{e\}$, nous devons trouver $\chi \in \hat{G}$ tel que $\chi(g) \neq 1$.

En vertu de ce que nous connaissons sur la structure des groupes abéliens finis (théorème 5.23), nous commençons $G = \mathbb{Z}/n\mathbb{Z}$ et considérons le caractère donné par $\chi([1]) = e^{2i\pi/n}$. Ce χ est un isomorphisme entre G et $\mathrm{U}(n)$; nous n'avons $\chi([k]) = 0$ que si $[k] = [n] = [0]$. Pour rappel dans $\mathbb{Z}/n\mathbb{Z}$, le neutre est $e = 0$ et non $e = 1$.

Passons au cas général :

$$G \simeq \mathbb{Z}/n_1\mathbb{Z} \times \dots \times \mathbb{Z}/n_k\mathbb{Z} \tag{15.13}$$

Si $g = (g_1, \dots, g_k)$ est non nul dans G, alors il existe i tel que $g_i \neq 0$ et on prend

$$\chi(g_1, \dots, g_k) = \chi_i(g_i) \tag{15.14}$$

où χ_i est le caractère $\chi_i([1]) = e^{2\pi i/n_i}$. Ce χ est alors un caractère non trivial de G.

□

15.1.1 Crochet de dualité et transformée de Fourier

Si G est un groupe abélien, nous définissons le crochet de dualité entre G et $\hat{G}$ par

$$\begin{aligned}\langle .,. \rangle \colon G \times \hat{G} &\to \mathbb{C}^* \\ \langle g, \chi \rangle &= \chi(g).\end{aligned} \tag{15.15}$$

Notons que l'image de ce crochet n'est pas $\mathbb{C}^*$ entier, mais seulement le groupe unitaire $\mathrm{U}(n)$ où n est l'exposant [1] de G.

Si f, g sont des applications de G dans $\mathbb{C}$, alors on leur associe le produit scalaire

$$\langle f, g \rangle = \frac{1}{|G|} \sum_{s \in G} f(s)\overline{g(s)}. \tag{15.16}$$

Lemme 15.4.
Les caractères de G forment une base orthonormée de $\mathbb{C}^G$ pour ce produit scalaire.

Démonstration. Étant donné que les $\chi(s)$ sont des nombres complexe de module 1, nous avons $\chi(s)\overline{\chi(s)} = 1$ et par conséquent $\langle \chi, \chi \rangle = 1$.

Si par contre $\chi \neq \chi'$, alors il existe $s_\in G$ tel que $\chi(s_0) \neq \chi'(s_0)$. Dans ce cas en effectuant un changement de variable $s \to s_0 s$ dans la sommation,

$$\langle \chi, \chi' \rangle = \frac{1}{|G|} \sum_{s \in G} \chi(s)\overline{\chi'(s)} \tag{15.17a}$$

$$= \frac{1}{|G|} \sum_{s \in G} \chi(s_0 s)\overline{\chi'(s_0 s)} \tag{15.17b}$$

$$= \frac{1}{|G|} \chi(s_0)\chi'(s_0) \sum_{s \in G} \chi(s)\overline{\chi'(s)}. \tag{15.17c}$$

Donc nous avons trouvé

$$\langle \chi, \chi' \rangle \big(1 - \chi(s_0)\overline{\chi'(s_0)}\big) = 0. \tag{15.18}$$

Mais vu que $\chi(s_0) \neq \chi(s_0')$, la parenthèse est non nulle (pour rappel $\chi(s_0)$ est un complexe de module 1) et par conséquent $\langle \chi, \chi' \rangle = 0$.

Nous déduisons immédiatement que les caractères forment une famille libre parce que si $\sum_i \chi_i = 0$ (la somme est sur tous les caractères), alors en prenant le produit scalaire avec χ_k,

$$\sum_i a_i \langle \chi_k, \chi_i \rangle = 0, \tag{15.19}$$

et donc $a_k = 0$.

Les caractères forment donc un système libre orthonormé. De plus l'espace engendré à la bonne dimension parce que le cardinal de l'ensemble des caractères est la dimension (complexe) de l'espace des fonctions de G dans $\mathbb{C}$ parce que, en utilisant l'isomorphisme entre G et $\hat{G}$,

$$\operatorname{Card} \hat{G} = \operatorname{Card}(G) = \dim_{\mathbb{C}} \mathbb{C}^G. \tag{15.20}$$

La première □

Du fait que les caractères forment une base orthonormée, nous pouvons écrire, pour toute application $f \colon G \to \mathbb{C}$,

$$f = \sum_{\chi \in \hat{G}} \langle \chi, f \rangle \chi. \tag{15.21}$$

À une fonction $f \colon G \to \mathbb{C}$ nous associons la **transformée de Fourier**

$$\begin{aligned}\hat{f} \colon \hat{G} &\to \mathbb{C} \\ \chi &\mapsto \langle \chi, f \rangle.\end{aligned} \tag{15.22}$$

1. Définition 2.18.

Nous avons donc aussi une espèce de formule d'inversion

$$f = \sum_{\chi \in \hat{G}} \hat{f}(\chi)\chi \tag{15.23}$$

qui n'est qu'une réécriture de 15.21.

15.1.2 Groupes non abéliens

Nous avons vu que le groupe des caractères $\hat{G}$ contenait toute l'information sur un groupe abélien. Malheureusement, pour les groupes non abéliens, ça ne va pas suffire, et nous allons introduire la notion de représentations, dont les caractères seront un cas particulier de dimension un.

Proposition 15.5.
Soit G un groupe (pas spécialement abélien). Nous avons

$$\hat{G} \simeq \operatorname{Hom}\big(G/D(G), \mathbb{C}^*\big). \tag{15.24}$$

Démonstration. Ce qui fait fonctionner la preuve est le fait que si $f\colon G \to \mathbb{C}^*$ est un homomorphisme, alors f s'annule sur $D(G)$. L'isomorphisme est

$$\begin{aligned} \psi\colon \hat{G} &\to \operatorname{Hom}\big(G/D(G), \mathbb{C}^*\big) \\ \psi(f)[g] &= f(g). \end{aligned} \tag{15.25}$$

Cette application est bien définie parce que si f est un homomorphisme,

$$f(gklk^{-1}l^{-1}) = f(g). \tag{15.26}$$

D'autre part ψ est un homomorphisme de groupe parce que

$$\psi(f_1f_2)[g] = (f_1f_2)(g) = f_1(g)f_2(g) = \psi(f_1)[g]\psi(f_2)[g] = \big(\psi(f_1)\psi(f_2)\big)[g]. \tag{15.27}$$

Pour l'injectivité de ψ, soit f_1 et f_2 telles que $\psi(f_1) = \psi(f_2)$. Alors pour tout $g \in G$ nous avons

$$\psi(f_1)[g] = \psi(f_2)[g] \tag{15.28}$$

et donc $f_1(g) = f_2(g)$.

Enfin ψ est surjective. En effet, soit $\bar{f} \in \operatorname{Hom}\big(G/D(G), \mathbb{C}^*\big)$. Alors nous obtenons $\psi(f) = \bar{f}$ en posant

$$f(g) = \bar{f}[g]. \tag{15.29}$$

Il faut juste vérifier que le f ainsi défini est dans $\hat{G}$, c'est-à-dire que $f(g_1g_2) = f(g_1)f(g_2)$. □

Cette proposition nous montre que

$$\hat{G} = \widehat{G/D(G)}, \tag{15.30}$$

alors que $G/D(G)$ est abélien ; il n'est donc pas tellement possible que $\hat{G}$ contienne beaucoup d'informations intéressantes sur G.

15.1.3 Représentations linéaires des groupes finis

Si $\dim V = 1$, alors $\mathrm{GL}(V) = \mathbb{C}^*$ et les représentation sont les caractères abéliens.

Exemple 15.6.
Considérons le triangle équilatéral A, B, C, par exemple donné par les points

$$\begin{cases} A = 1 & (15.31a) \\ B = (-\frac{1}{2}, \frac{\sqrt{3}}{2}) & (15.31b) \\ C = (-\frac{1}{2}, -\frac{\sqrt{3}}{2}) & (15.31c) \\ & (15.31d) \end{cases}$$

Dans la base (pas orthonormée) $\{A, B\}$ de $\mathbb{R}^2$, ces trois points sont donnés par

$$A = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad C = \begin{pmatrix} -1 \\ -1 \end{pmatrix}. \tag{15.32}$$

Le groupe symétrique S_3 agit sur le triangle par permutation des sommets. Vues dans la base $\{A, B\}$, les transpositions correspondent aux matrices

$$(A, B) \to \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{15.33a}$$

$$(A, C) \to \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix} \tag{15.33b}$$

$$(B, C) \to \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix}. \tag{15.33c}$$

La permutation (A, B, C) s'écrit comme $(A, B, C) = (A, C)(A, B)$ et on lui associe la matrice

$$(A, B, C) \to \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}. \tag{15.34}$$

C'est bien le produit des matrices de (A, C) et de (A, B). De la même façon nous avons

$$(BAC) < ++ > \tag{15.35}$$

<++> △

Si (V, ρ) et (V', ρ') sont deux représentations du groupe G, alors nous définissons la **somme directe** par $\big(V \oplus V', \rho \oplus \rho'\big)$ donné par

$$(\rho \oplus \rho')(g) = \begin{pmatrix} \rho(g) & 0 \\ 0 & \rho'(g) \end{pmatrix} \in \mathrm{GL}(V \oplus V'). \tag{15.36}$$

Nous noterons souvent $2V$ pour la représentations $(V, \rho) \oplus (V, \rho)$ et plus généralement l'écriture

$$V = \bigoplus_i k_i W_i \tag{15.37}$$

signifiera la représentation somme de k_i termes de la représentation W_i. Ici encore un abus est commis entre la représentation (ρ_i, W_i) et l'espace W_i.

15.1.4 Module

Nous considérons la $\mathbb{C}$-algèbre $G[\mathbb{C}]$ des combinaisons (formelles) d'éléments de G à coefficients dans G, c'est-à-dire l'ensemble

$$\mathbb{C}[G] = \{\sum_{s \in G} a_s s\} \tag{15.38}$$

avec le produit hérité de la bilinéarité :

$$\sum_{s \in G} \sum_{t \in G} a_s b_t st = \sum_s \sum_t a_s b_{s^{-1}t} t, \tag{15.39}$$

et la somme

$$(\sum_s a_s s) + \sum_t b_t t = \sum_{s\in G}(a_s + b_s)s. \tag{15.40}$$

Le tout est une $\mathbb{C}$-algèbre agissant sur V par

$$\left(\sum_s a_s s\right) v = \sum_{s\in G} a_s \rho(s) v \in V \tag{15.41}$$

Les sous-modules indécomposables seront les représentations irréductibles.

Définition 15.7.
*La représentation (V, ρ) du groupe G est **irréductible** si les seuls sous-espaces invariants de V sous $\rho(G)$ sont V et $\{0\}$.*

Exemple 15.8.
La représentation de S_3 sur $\mathbb{R}^2$ donnée par les permutations des sommets d'un triangle équilatéral donnée dans l'exemple 15.6 est irréductible. △

La question qui vient est de savoir si une représentation possédant des sous-espaces invariants peut être écrite comme la somme de représentations irréductibles.

Proposition 15.9.
Soit (V, ρ) une représentation linéaire de dimension finie d'un groupe fini[2]*. Si W_1 est un sous-espace stable*[3]*, alors il existe un sous-espace W_2 également stable et tel que $V = W_1 \oplus W_2$.*

Toute représentation linéaire est décomposable en représentations irréductibles.

Démonstration. Soit $P\colon V \to V$ un projecteur sur W_1, c'est-à-dire que $P^2 = P$ et $P(V) = W_1$. Pour construire un tel projecteur, on peut par exemple prendre un supplémentaire de W_1 dans V puis utiliser la décomposition[4]. Nous considérons l'opérateur

$$P_G = \frac{1}{|G|}\sum_{g\in G} \rho(g) \circ P \circ \rho(g)^{-1}. \tag{15.42}$$

Prouvons que ce P_G est encore un projecteur. D'abord pour tout $g \in G$ nous avons

$$\rho(g) P_G \rho(g)^{-1} = \frac{1}{|G|}\sum_{s\in G} \rho(gs) P \rho(gs)^{-1} = P_G. \tag{15.43}$$

La dernière égalité est un changement de variables dans la somme[5]. Cela signifie que $P_G\rho = \rho P_G$. Nous avons même $P_G P = P$ parce que si $v \in W_1$, alors

$$P_G(v) = \frac{1}{|G|}\sum_{s\in G} \rho(s) P \underbrace{\rho(s)^{-1} v}_{\in W_1} \tag{15.44a}$$

$$= \frac{1}{|G|}\sum_s \rho(s)\rho(s)^{-1} v \tag{15.44b}$$

$$= v. \tag{15.44c}$$

2. La démonstration marche aussi pour les groupes compacts, mais il faudrait des intégrales.
3. c'est-à-dire si ρ n'est pas irréductible.
4. Ou encore prendre une base de W_1, l'étendre en une base de V et définir P comme l'annulation des coefficients des vecteurs « complétant » la base.
5. Et c'est ça qui demande un peu de technique pour écrire la preuve dans le cas d'un groupe compact : il faut une mesure de Haar.

Avec cela nous pouvons conclure que $P_G^2 = P_G$ parce que

$$P_G \circ P_G = \frac{1}{|G|}\sum_g P_G\rho(g)P\rho(g)^{-1} \tag{15.45a}$$
$$= \frac{1}{|G|}\sum_g \rho(g)P_GP\rho(g)^{-1} \tag{15.45b}$$
$$= \frac{1}{|G|}\sum_g \rho(g)P\rho(g)^{-1} \tag{15.45c}$$
$$= P_G. \tag{15.45d}$$

Donc P_G est un projecteur, est stable sous les conjugaisons par $\rho(g)$ et commute avec $\rho(g)$. Nous décomposant Id de façon évidente en

$$\mathrm{Id} = P_G + (\mathrm{Id} - P_G). \tag{15.46}$$

Étant donné que l'opérateur P_G commute avec tous les $\rho(g)$, les noyaux de P_G et $\mathrm{Id} - P_G$ sont des sous-espaces invariants. Vu que P_G est un projecteur, nous avons $q(P_G) = 0$ avec $q(X) = X^2 - X$. Pour appliquer le lemme des noyaux (théorème 9.88), nous remarquons que $q(X) = X(X-1)$ et donc

$$V = \ker P_G \oplus \ker(P_G - \mathbb{1}). \tag{15.47}$$

Si nous posons $W_2 = \ker P_G$, il reste à voir que $\ker(P_G\mathbb{1}) = W_1$. D'abord $W_1 \subset \ker(P_G - \mathrm{Id})$ parce que si $w \in W_1$, ce dernier étant stable,

$$P_Gw = \frac{1}{|G|}\sum_{g\in G}\rho(g)P\underbrace{\rho(g)^{-1}w}_{\in W_1} \tag{15.48a}$$
$$= \frac{1}{|G|}\sum_{g\in G} w \tag{15.48b}$$
$$= w. \tag{15.48c}$$

Pour prouver l'inclusion inverse, nous savons que P_G et P sont des projecteurs tels que $P_GP = P$, ce qui signifie que l'image de P_G est inclue à celle de P, c'est-à-dire à W_1. Mais $\mathrm{Image}(P_G) = \ker(\mathbb{1} - P_G)$, donc

$$\ker(\mathbb{1} - P_G) = \mathrm{Image}(P_G) \subset \mathrm{Image}(P) = W_1. \tag{15.49}$$

La représentation ρ se décompose donc en deux sous-représentations (ρ, W_1) et ρ, W_2. Si l'une des deux n'est pas irréductible, le processus peut recommencer. Vu que la dimension de V est finie, toute représentation se décompose en une somme finie de représentation irréductibles. □

15.1.5 Structure hermitienne

Soit (ρ, V) une représentation de G sur un espace vectoriel complexe V. Nous voulons munir V d'un produit scalaire hermitien (définition 7.189) tel que les opérateurs $\rho(g)$ soient tous des isométries. C'est-à-dire que nous voudrions définir $\langle u, v\rangle_G$ de telle sorte à avoir

$$\langle \rho(g)u, \rho(g)v\rangle_G = \langle u, v\rangle_G \tag{15.50}$$

pour tout $g \in G$. Nous commençons par considérer un produit hermitien $\langle .,. \rangle$ quelconque et puis nous définissons

$$\langle u, v\rangle_G = \frac{1}{|G|}\sum_{g\in G}\langle \rho(g)u, \rho(g)v\rangle. \tag{15.51}$$

Nous devons vérifier que c'est un produit. La seule des conditions dont la vérification n'est pas immédiate est celle de positivité. Pour tout $g \in G$ et tout $v \in V$, nous avons $\langle \rho(g)v, \rho(g)v\rangle$ est

positif et nul si et seulement si $\rho(g)v = 0$. Étant donné que $\rho(e)v = v$, parmi les termes de la somme

$$\langle u, u\rangle_G = \frac{1}{|G|}\sum_{g\in G}\langle \rho(g)v, \rho(g)v\rangle, \tag{15.52}$$

au moins un est strictement positif (pourvu que $v \neq 0$) ; les autres sont positifs ou nuls. Par conséquent $\langle v, v\rangle_G = 0$ si et seulement si $v = 0$.

Donc les groupes finis peuvent être vus comme des parties de groupes d'isométrie. De la même façon, en utilisant une mesure de Haar pour faire la moyenne, nous pouvons plonger les groupes compacts dans des groupes unitaires.

15.1.6 Caractères

Définition 15.10.
*Soit (V, ρ) une représentation linéaire du groupe G. Le **caractère** de ρ est la fonction*

$$\begin{aligned} \chi_\rho \colon G &\to \mathbb{C} \\ s &\mapsto \operatorname{Tr}\big(\rho(s)\big). \end{aligned} \tag{15.53}$$

Par invariance cyclique de la trace, nous avons

$$\chi_\rho(sts^{-1}) = \chi_\rho(t), \tag{15.54}$$

ce qui fait que le caractère est une fonction constante sur les classes de conjugaison.

D'après sa fiche wikipédia, le marquis de Sade, passionné de théâtre, faisait des représentations qui avaient du caractère.

Un **caractère irréductible** est un caractère d'une représentation irréductible.

Définition 15.11.
*Une application $f \colon G \to \mathbb{C}$ est **centrale** si elle est constante sur les classes de conjugaison.*

Les traces sont des applications centrales.

L'ensemble des fonctions centrales sur un groupe fini (ou tout au moins ayant un nombre fini de classes de conjugaison) est un $\mathbb{C}$-espace vectoriel de dimension égale au nombre de classes, et nous pouvons mettre le produit scalaire

$$\langle f, g\rangle = \frac{1}{|G|}\sum_{s\in G} f(s)\overline{g(s)}. \tag{15.55}$$

C'est une forme hermitienne sur l'espace des fonctions centrales.

15.2 Équivalence de représentations et caractères

Cette section prend des éléments des articles lemme de Schur, caractère d'une représentation, fonction centrale et trace de wikipédia.

Définition 15.12.
*Nous disons que les deux représentations (V, ρ) et (V', ρ') sont **équivalentes** s'il existe une bijection linéaire $f \colon V \to V'$ telle que*

$$f \circ \rho = \rho' \circ f. \tag{15.56}$$

*Nous disons alors que f **entrelace** ρ et ρ'.*

Théorème 15.13 (Théorème de Schur)**.**
Si (V, ρ) et (V', ρ') sont des représentations irréductibles non équivalentes alors la seule application linéaire $f \colon V \to V'$ entrelaçant ρ et ρ' est la fonction nulle.

En d'autres termes, soit les représentations sont équivalentes (et il y a un isomorphisme), soit il n'y a même pas un homomorphisme.

Démonstration. Soit $f \in \mathcal{L}(V, V')$ telle que $f \circ \rho = \rho' \circ f$. Alors $\ker f$ est un sous-espace stable sous $\rho(G)$, et $\text{Image}(f)$ est un sous-espace de V' stable par $\rho'(G)$. Par irréductibilité, nous avons que $\ker(f) = \{0\}$ ou V. Même chose pour $\text{Image}(f)$. Il y a deux possibilités.

(1) Si $\ker(f) = \{0\}$, alors $\text{Image}(f) \neq \{0\}$ et alors $\text{Image}(f) = V'$. Du coup f est injective et surjective, c'est-à-dire est un isomorphisme.

(2) Si $\ker(f) = V$, alors $f = 0$.

□

Corolaire 15.14 (Schur pour les représentations sur $\mathbb{C}$)**.**
Soit (V, ρ) une représentation irréductible, alors l'ensemble

$$\text{End}_G(V, \rho) = \{f \in \text{End}(V) \text{ tel que } \rho \circ f = f \circ \rho\} \tag{15.57}$$

est l'ensemble des homothéties.

Démonstration. Soit $f \in \text{End}_G(V, \rho)$. Vu que l'espace est sur $\mathbb{C}$, l'endomorphisme f a une valeur propre λ. L'opérateur $g = f - \lambda \mathbb{1}$ est aussi un opérateur d'entrelacement de ρ alors que $\ker(g) \neq \{0\}$ par définition de valeur propre. Du coup $\ker(g) = V$, ce qui signifie que f est l'isométrie de rapport λ : $f = \lambda \,\text{Id}$. □

Lemme 15.15.
Si (ρ, V) et (ρ', V') sont des représentations équivalentes de caractères χ et χ', alors $\chi = \chi'$.

Démonstration. Si $A \colon V \to V'$ est un isomorphisme d'espace vectoriel entrelaçant ρ et ρ', c'est-à-dire si pour tout g, $\rho'(g)A = A\rho(g)$, alors $\rho'(g) = A\rho(g)A^{-1}$ et

$$\chi'(g) = \text{Tr}\left(\rho'(g)\right) = \text{Tr}\left(A\rho(g)A^{-1}\right) = \text{Tr}\left(\rho(g)\right) \tag{15.58}$$

parce que la trace est un invariant de similitude (lemme 9.120). □

Lemme 15.16.
Si χ est le caractère de la représentation complexe (V, ρ) du groupe fini G, alors pour tout $g \in G$ nous avons $\chi(g^{-1}) = \overline{\chi(g)}$.

Démonstration. Par le corolaire 2.35 au théorème de Lagrange, nous avons $g^{|G|} = e$ et donc en tant qu'opérateur, $\rho(g)^{|G|} = \mathbb{1}$. Les valeurs propres de $\rho(g)$ sont donc des racines de l'unité. Si nous notons λ_i ces valeurs propres, alors $\chi(g) = \sum_i \lambda_i$, et en considérant la matrice dans sa base de diagonalisation (lemme de Schur complexe, 9.137), nous voyons que

$$\chi(g^{-1}) = \text{Tr}\left(\rho(g)^{-1}\right) = \sum_i \frac{1}{\lambda_i}. \tag{15.59}$$

Mais λ_i étant une racine de l'unité nous avons $\frac{1}{\lambda_i} = \bar{\lambda}_i$, ce qui fait que

$$\chi(g^{-1}) = \sum_i \bar{\lambda}_i = \overline{\chi(g)}. \tag{15.60}$$

□

Proposition 15.17.
Soient deux représentations irréductibles complexes (V, ρ) et (V', ρ') du même groupe fini G, et χ et χ' leurs caractères respectifs. Nous avons

(1) $\langle \chi, \chi' \rangle = 0$ si ρ et ρ' ne sont pas équivalentes.

(2) $\langle \chi, \chi' \rangle = 1$ si les représentations sont équivalentes.

Démonstration. Nous considérons les bases $\{e_1, \dots, e_n\}$ de V et $\{f_1, \dots, f_m\}$ de V'. Puis nous considérons la matrice $F(k,l) = E_{kl} \in \mathbb{M}_{m,n}(\mathbb{C})$ où pour rappel, E_{kl} est la matrice de composantes $(E_{kl})_{ij} = \delta_{ki}\delta_{lj}$. Nous posons

$$F_G(k,l) = \frac{1}{|G|}\sum_{g\in G}\rho(g)\circ F(k,l)\circ\rho'(g)^{-1}. \tag{15.61}$$

En nous permettant de ne pas réécrire les indices k et l de F et F_G, nous montrons que F_G entrelace ρ et ρ' :

$$\begin{aligned}
F_G\circ\rho'(t) &= \frac{1}{|G|}\sum_{s\in G}\rho(s)\circ F\circ\rho'(s^{-1})\circ\rho(t) && (15.62a)\\
&= \frac{1}{|G|}\sum_{s}\rho(s)F\rho'(s^{-1}t) && (15.62b)\\
&= \frac{1}{|G|}\sum_{k}\rho(tk)F\rho'(k^{-1}) && (15.62c)\\
&= \frac{1}{|G|}\rho(t)\sum_{k}\rho(k)F\rho'(k^{-1}) && (15.62d)\\
&= \rho(t)\circ F_G. && (15.62e)
\end{aligned}$$

Dans ce calcul nous avons effectué le changement de variables $k = (s^{-1}t)^{-1}$ qui donne $s = tk$.

Par ailleurs nous avons

$$\begin{aligned}
\Big(\rho(g)F(k,l)\rho'(g^{-1})\Big)_{ij} &= \sum_{r=1}^{n}\sum_{s=1}^{m}\rho(g)_{ir}F(k,l)_{rs}\rho'(g^{-1})_{sj} && (15.63a)\\
&= \sum_{rs}\rho(g)_{ir}\delta_{kr}\delta_{ls}\rho'(g^{-1})_{sj} && (15.63b)\\
&= \rho(g)_{ik}\rho'(g^{-1})_{lj}, && (15.63c)
\end{aligned}$$

et par conséquent

$$F_G(k,l)_{ij} = \frac{1}{|G|}\sum_{g\in G}\rho(g)_{ik}\rho'(g^{-1})_{lj}. \tag{15.64}$$

Si χ et χ' sont les caractères de ρ et ρ', alors nous avons le produit (15.55) qui donne

$$\begin{aligned}
\langle\chi,\chi'\rangle &= \frac{1}{|G|}\sum_{g\in G}\chi(g)\overline{\chi'(g)} && (15.65a)\\
&= \frac{1}{|G|}\sum_{g}\chi(g)\chi'(g^{-1}) && \text{lemme 15.16} && (15.65b)\\
&= \frac{1}{|G|}\sum_{g}\sum_{i=1}^{n}\sum_{j=1}^{m}\rho(g)_{ii}\rho'(g^{-1})_{jj} && (15.65c)\\
&= \sum_{ij}F_G(i,j)_{ij} && \text{par (15.64).} && (15.65d)
\end{aligned}$$

Si les représentations ρ et ρ' ne sont pas équivalentes, le fait que F_G en soit un opérateur d'entrelacement implique par le théorème de Schur 15.13 que $F_G = 0$ et donc $\langle\chi,\chi'\rangle = 0$.

Si au contraire les représentation sont équivalentes, alors le lemme 15.15 nous dit que $\chi = \chi'$ et nous reprenons la définition :

$$\langle\chi,\chi\rangle = \frac{1}{|G|}\sum_{g}\chi(g)\overline{\chi(g)} = \frac{1}{|G|}\sum_{g\in G}1 = 1 \tag{15.66}$$

parce que les nombres $\chi(g)$ sont des racines de l'unité. □

15.2.1 Représentation régulière

Nous notons λ la **représentation régulière gauche**, agissant sur le $\mathbb{K}$-espace vectoriel des fonctions $G \to \mathbb{K}$ par

$$\Big(\lambda(g)f\Big)(g) = f(g^{-1}h). \tag{15.67}$$

D'autre part nous considérons les fonctions $\delta_g \colon G \to \mathbb{K}$ (ici $\mathbb{K}$ est $\mathbb{R}$ ou $\mathbb{C}$ ou pire) définie par

$$\delta_g(h) = \begin{cases} 1 & \text{si } g = h \\ 0 & \text{sinon.} \end{cases} \tag{15.68}$$

La représentation régulière agit sur les fonctions δ_s de la façon suivante :

$$\lambda(g)\delta_s = \delta_{gs} \tag{15.69}$$

parce que $\big(\lambda(g)\delta_s\big)(h) = \delta_s(g^{-1}h) = \delta_{gs}(h)$.

Lemme 15.18.
Le caractère de la représentation régulière gauche est donné par

$$\chi_\lambda = |G|\delta_e. \tag{15.70}$$

Démonstration. Appliquer l'équation (15.70) fonctionne parce que $\chi_\lambda(e)$ est la dimension de l'espace des fonctions sur G, c'est-à-dire $|G|$. Si par contre $g \neq e$, alors $\lambda(g)$ est une matrice de permutation (dans la base des δ_h) et a donc tous ses éléments diagonaux nuls. □

Si ρ est une représentation et si f est une fonction sur le groupe, alors nous considérons l'opérateur

$$\rho_f = \sum_{g\in G} f(g)\rho(g). \tag{15.71}$$

Proposition 15.19 ([310])**.**
Si (ρ, V) est une représentation irréductible et si f est une fonction centrale sur G, alors l'opérateur ρ_f est une homothétie de V de rapport

$$\frac{1}{\dim V}\sum_{g\in G} f(g)\chi(g) \tag{15.72}$$

où χ est le caractère de ρ.

Démonstration. Nous commençons par voir que ρ_f entrelace ρ. En effet,

$$\begin{aligned}
\rho(t)^{-1}\circ\rho_f\circ\rho(t) &= \sum_g f(g)\rho(t^{-1}gt) && \text{(15.73a)}\\
&= \sum_h f(tht^{-1})\rho(g) \qquad h = t^{-1}gt && \text{(15.73b)}\\
&= \sum_h f(h)\rho(h) && \text{(15.73c)}\\
&= \rho_f && \text{(15.73d)}
\end{aligned}$$

où en écrivant $f(tht^{-1}) = f(h)$, nous avons utilisé le fait que f était centrale. Étant donné que ρ_f entrelace une représentation irréductible, le lemme de Schur (15.13) nous indique que ρ_f est une homothétie. Soit k le facteur d'homothétie. Alors d'une part $\mathrm{Tr}(\rho_f) = nk$. D'autre part,

$$\begin{aligned}
\mathrm{Tr}(\rho_f) &= \mathrm{Tr}\big(\sum_g f(g)\rho(g)\big) && \text{(15.74a)}\\
&= \sum_g f(g)\,\mathrm{Tr}\big(\rho(g)\big) && \text{(15.74b)}\\
&= \sum_g f(g)\chi(g). && \text{(15.74c)}
\end{aligned}$$

Du coup effectivement

$$k = \frac{1}{n} \sum_{g \in G} f(g)\chi(g). \tag{15.75}$$

□

15.2.2 Caractères et représentations : suite et fin

Lemme 15.20.
Un groupe fini n'a (à équivalence près) qu'un nombre fini de représentations irréductibles.

Démonstration. Les caractères irréductibles forment un système orthonormé (proposition 15.17) et donc libre parmi les fonctions centrales. Donc il y a au plus autant de caractères irréductibles que la dimension de l'espace des fonctions centrales ; et ce dernier est de dimension finie donnée par le nombre de classes de conjugaison de G. □

Nous savons que les caractères de deux représentations irréductibles sont égaux. Étant donné qu'il n'existe qu'un nombre fini de représentations irréductibles, il existe un nombre fini de caractères irréductibles. Nous pouvons donc fixer les notations suivantes. Les caractères irréductibles seront notés $\{\varphi_i\}_{i=1,\dots,h}$ et nous noterons (σ_i, W_i) une représentation ayant le caractère φ_i.

Théorème 15.21 ([310]).
Soit (ρ, V) une représentation de G de caractère χ. Alors sa décomposition en représentations irréductibles est donnée par

$$(V, \rho) = \bigoplus_{i=1}^{h} k_i (W_i, \sigma_i) \tag{15.76}$$

avec $k_i = \langle \chi, \varphi_i \rangle$. En particulier, à permutation près des facteurs, la décomposition d'une représentation en représentations irréductibles est unique.

Démonstration. La décomposition de χ en caractères irréductibles est donnée par $\chi = \sum_i k_i \varphi_i$; en prenant le produit de cette égalité avec φ_j et en tenant compte de l'othonormalité des caractères irréductibles,

$$\langle \chi, \varphi_j \rangle = \sum_i k_i \langle \varphi_i, \varphi_j \rangle = k_j. \tag{15.77}$$

□

Le théorème suivant est ce qui nous permet de dire que l'étude des caractères et l'étude des représentations, c'est la même chose.

Théorème 15.22.
Soit G un groupe fini[6].

(1) Deux représentations sont équivalentes si et seulement si elles ont même caractères.

(2) Si χ est le caractère d'une représentation, alors

(2a) $\langle \chi, \chi \rangle \in \mathbb{N}$

(2b) $\langle \chi, \chi \rangle = 1$ si et seulement si la représentation est irréductible.

Démonstration. Nous démontrons chaque point séparément.

(1) Le fait que deux représentations équivalentes aient même caractère est le lemme 15.15. Nous montrons l'autre sens. Si (ρ, V) et (ρ', V') sont deux représentations irréductibles de décompositions

$$V = \bigoplus_i k_i W_i \tag{15.78a}$$

$$V' = \bigoplus_i k'_i W_i, \tag{15.78b}$$

6. Nous sommes depuis longtemps dans l'étude des représentations des groupes finis.

alors si $\chi = \chi'$, nous avons $k_i = k'_i$ et les représentations sont identiques.

(2) Soit (ρ, V) une représentation ayant χ comme caractère. En posant $k_i = \langle \chi, \varphi_i \rangle$ nous avons la décomposition en représentations irréductibles

$$V = \bigoplus_i k_i W_i, \tag{15.79}$$

et aussi

$$\langle \chi, \chi \rangle = \langle \sum_i k_i \varphi_i, \sum_j k_j \varphi_j \rangle = \sum_i k_i^2 \in \mathbb{N}. \tag{15.80}$$

Ce nombre est de plus égal à 1 si et seulement si tous les termes de la somme sont nuls sauf un qui vaudrait 1. Ce cas donne une représentation irréductible.

□

Proposition 15.23.
Si (λ, R) est la représentation régulière gauche de décomposition en représentations irréductibles

$$R = \bigoplus_i k_i W_i, \tag{15.81}$$

alors

(1) $k_i = \dim W_i$,

(2) $\sum_i (\dim W_i)^2 = |G|$,

(3) pour tout $g \in G$, $\sum_i (\dim W_i)\varphi_i(g) = 0$ [7].

(4) Si $\{(n_i, \varphi_i)\}$ *est la liste des couples dimension,caractère des représentations irréductibles non équivalentes, alors pour tout* $s \in G \backslash \{e\}$ *nous avons* $\sum_{i=1}^p n_i \varphi_i(s) = 0$ *où la somme porte sur les représentations irréductibles non équivalentes.*

Démonstration. Nous notons r le caractère de la représentation régulière gauche. Nous avons

$$k_i = \langle r, \varphi_i \rangle = \frac{1}{|G|} \sum_{s \in G} r(s) \overline{\varphi_i(s)} = \overline{\varphi_i(e)}. \tag{15.82}$$

Mais $\varphi_i(e) = \dim W_i \in \mathbb{R}$, donc nous avons bien $k_i = \dim W_i$. Le caractère de la représentation régulière peut alors s'exprimer de deux façons :

$$|G| \delta_e = \sum_i (\dim W_i) \varphi_i. \tag{15.83}$$

En évaluant cette égalité en e nous trouvons directement

$$|G| = \sum_i (\dim W_i)^2, \tag{15.84}$$

et en l'évaluant en $s \neq e$, nous trouvons

$$0 = \sum_i (\dim W_i) \varphi_i(s). \tag{15.85}$$

□

Le théorème suivant est valable pour les groupes finis (comme toute cette section).

Théorème 15.24 ([310])**.**
Les caractères irréductibles $\chi_1, \ldots, \chi_h$ forment une base orthonormé des fonctions centrales sur G.

7. Cette propriété est appelée « orthogonalité des colonnes » pour une raison qui apparaîtra au moment de compléter le tableau (15.110).

Démonstration. Nous savons déjà qu'ils forment un système orthonormé. Considérons le sous-espace $H = \mathrm{Span}\{\varphi_i\}_{i=1,\dots,h}$ de l'espace des fonctions centrales sur G. En vertu de la proposition 4.135, il nous suffit de prouver que $H^\perp = 0$. Soit donc f, une fonction centrale appartenant à $H^\perp$. Pour tout i, nous avons $\langle f, \varphi_i\rangle = 0$ et donc aussi $\langle \bar{f}, \bar{\varphi}_i\rangle = 0$.

Considérant une représentation irréductible (σ, W) de caractère φ, nous savons par la proposition 15.19 que l'opérateur

$$\sigma_{\bar{f}} = \sum_g \bar{f}(g)\varphi(g) \tag{15.86}$$

est une homothétie de rapport $\langle \bar{f}, \bar{\varphi}\rangle / \dim W = 0$. Étant donné que toutes les représentations sont des sommes directes de représentations irréductibles, en réalité l'opérateur $\rho_{\bar{f}}$ est nul pour toute représentation ρ. En particulier pour la représentation régulière,

$$0 = \lambda_{\bar{f}}(\delta_t) = \sum_{g\in G} \bar{f}(g)\lambda(g)(\delta_t) = \sum_g \bar{f}(g)\delta_{ft}. \tag{15.87}$$

En écrivant cette égalité avec $t = e$ et puis en appliquant à $k \in G$ nous trouvons

$$0 = \sum_g \bar{f}(g)\delta_g(k) = \bar{f}(k). \tag{15.88}$$

Donc $\bar{f} = 0$ et f est nulle. □

Corolaire 15.25.
Le nombre de représentations irréductibles non équivalentes d'un groupe fini est égal à son nombre de classes de conjugaison.

Démonstration. Le nombre de classes de conjugaison est la dimension de l'espace des fonctions centrales qui elle-même est égale au nombre de caractères irréductibles par le théorème 15.24. Enfin deux caractères irréductibles sont égaux si et seulement si les représentations sous-jacentes sont équivalentes. □

Corolaire 15.26.
Toutes les représentations irréductibles d'un groupe abélien sont de dimension 1.

Démonstration. Le corolaire 15.25 nous dit qu'il y a autant de représentations unitaires qu'il n'y a de représentations irréductibles (non équivalentes). Mais les classes de conjugaisons sont des singletons (lemme 2.22). Nous avons donc exactement $|G|$ représentations irréductibles lorsque G est abélien.

Mais d'autre part la proposition 15.23(2) donne $\sum_i (\dim W_i)^2 = |G|$ lorsque la somme parcours les représentations irréductibles. Il y a $|G|$ termes à la somme, donc tous les termes doivent être 1. □

15.3 Représentation produit tensoriel

Soient ρ et ϕ, deux représentations d'un groupe G sur des espaces vectoriels V et W. La représentation **produit tensoriel** est la représentation

$$\begin{aligned} \rho\otimes\phi\colon G &\to \mathrm{GL}(V\otimes W)\\ (\rho\otimes\phi)(g)(v\otimes w) &= \rho(g)v\otimes\phi(g)w. \end{aligned} \tag{15.89}$$

Pour trouver son caractère, nous considérons une base $\{e_i\}$ de V et une base $\{e_\alpha\}$ de W, et la base $\{e_i\otimes e_\alpha\}$ de $V\otimes W$. Donc

$$(\rho\otimes\phi)(g)(e_i\otimes e_\alpha) = \rho(g)e_i\otimes\phi(g)e_\alpha. \tag{15.90}$$

Nous devons savoir quelle est la composante « $e_i\otimes e_\alpha$ » de cette dernière expression, et c'est évidemment

$$\rho(g)_{ii}\rho_{\alpha\alpha}, \tag{15.91}$$

ce qui nous amène à dire que

$$\operatorname{Tr}(\rho\otimes\phi)(g)=\sum_i\sum_\alpha\rho(g)_{ii}\phi(g)_{\alpha\alpha}=\operatorname{Tr}\big(\rho(g)\big)\operatorname{Tr}\big(\phi(g)\big),\tag{15.92}$$

c'est-à-dire au final que

$$\chi_{\rho\otimes\phi}=\chi_\rho\chi_\phi.\tag{15.93}$$

15.4 Exemple sur le groupe symétrique

Soit $G=S_3$, un des premiers groupes finis non abéliens. On en a une représentation de dimension deux en tant que permutation des sommets d'un triangle équilatéral, donnée dans l'exemple 15.6 ; nous notons ρ cette représentation.

Nous y avons aussi la représentation de signature donnée par

$$\begin{aligned}\epsilon\colon S_3&\to\mathrm{GL}(\mathbb{C})\\ \sigma&\mapsto\epsilon(\sigma)\operatorname{Id}.\end{aligned}\tag{15.94}$$

Et enfin il y a la représentation triviale. Ce sont les trois représentations irréductibles ; pour rappel il y a autant de représentations irréductibles que de classes de conjugaison (corolaire 15.25).

Classe de conjugaison	taille	χ_1	χ_ϵ	χ_ρ
Id	1	1	1	2
(A,B)	3	1	-1	0
(A,B,C)	2	1	1	-1

Nous calculons par exemple le produit scalaire

$$\langle\chi_1,\chi_\epsilon\rangle=\frac{1}{6}\big(1\cdot\chi_1(\mathrm{Id})\overline{\chi_\epsilon(\mathrm{Id})}+3\cdot\chi_1(A,B)\overline{\chi_\epsilon(A,B)}+2\cdot\chi_1(A,B,C)\overline{\chi_\epsilon(A,B,C)}\big)\tag{15.95a}$$

$$=0.\tag{15.95b}$$

D'autre part nous avons aussi

$$\langle\chi_\rho,\chi_\rho\rangle=\frac{1}{6}(1\cdot2\cdot2+3\cdot0+2\cdot1)=1.\tag{15.96}$$

15.5 Table des caractères du groupe symétrique S_4

Pour la table des caractères de S_4, voir [54]. Et si vous voulez la table des caractères du groupe diédral, vu que ce sont de isométries de $\mathbb{R}^n$, il faudra voir plus bas en la section 17.16.

15.5.1 Calculs à partir de rien ou presque

Nous savons que les classes de conjugaison dans S_4 sont caractérisées par la structure des décompositions en cycles (proposition 2.70). Elles sont données dans l'exemple 2.72.

Nous avons donc 5 classes de conjugaison, et il nous faut donc 5 représentations irréductibles non équivalentes (corolaire 15.25) dont nous allons chercher les caractères.

La première est la représentation triviale de dimension 1 ; nous notons χ_1 son caractère et nous avons la ligne

	dimension	Id	(12)	(123)	(1234)	(12)(34)
χ_1	1	1	1	1	1	1

(15.97)

Ensuite nous avons la signature qui est un morphisme non trivial $\epsilon\colon S_n\to\{-1,1\}$. Nous avons alors la ligne

	dimension	Id	(12)	(123)	(1234)	(12)(34)
χ_ϵ	1	1	-1	1	-1	1

(15.98)

Une troisième représentation pas trop compliquée à trouver est celle

$$\begin{aligned} \rho_p \colon S_4 &\to \mathrm{GL}(4,\mathbb{C}) \\ \rho_p(\sigma)e_i &= e_{\sigma(i)}. \end{aligned} \tag{15.99}$$

Cela n'est pas une représentation irréductible parce que $\mathbb{C}^4$ se décompose en deux sous-espaces stables :

$$D = \mathrm{Span}(1,1,1,1) \tag{15.100a}$$

$$H = \{x \in \mathbb{C}^4 \text{ tel que } x_1 + x_2 + x_3 + x_4 = 0\}. \tag{15.100b}$$

La représentation induite sur D est la représentation triviale. Puis sur H, elle induit une autre représentations que nous allons noter ρ_s.

Nous allons à présent déduire le caractère de la représentation ρ_s et prouver qu'elle est irréductible. Il est cependant possible de sauter cette étape en échange d'un certain travail sur les isométries du tétraèdre. Voir la proposition 17.164 et ensuite

Nous avons la décomposition $\rho_p = \rho_1 \oplus \rho_s$ et donc

$$\chi_p = \chi_1 + \chi_s. \tag{15.101}$$

Nous savons déjà χ_1. Le caractère χ_p n'est pas très compliqué parce que $\chi_p(\sigma)$ est une matrice de permutations des vecteurs de base. Donc la matrice $\rho_p(\sigma)$ a un 1 sur la diagonale pour les i tels que $\sigma(i) = i$. Nous avons donc

$$\chi_p(\mathrm{Id}) = 4 \qquad \chi_p(12) = 2 \tag{15.102a}$$

$$\chi_p\big((12)(34)\big) = 0 \qquad \chi_p(123) = 1 \tag{15.102b}$$

$$\chi_p(1234) = 0. \tag{15.102c}$$

Le caractère χ_s peut être calculé par simple soustraction :

	dimension	Id	(12)	(123)	(1234)	(12)(34)
χ_s	3	3	1	0	-1	-1

(15.103)

Avant d'ajouter cette ligne au tableau des représentations irréductibles nous devons savoir si ρ_s en est une. Pour cela, tant que nous avons son caractère nous pouvons utiliser le critère du théorème 15.22 :

$$\langle \chi_s, \chi_s \rangle = \frac{1}{|S_4|} \sum_{\sigma \in S_4} \chi_s(\sigma)^2. \tag{15.104}$$

Nous avons tout de suite $|S_4| = 4 \cdot 3 \cdot 2 = 24$ et puis

$$24\langle \chi_s, \chi_s \rangle = 3^2 + 6 \cdot 1^2 + 8 \cdot 0^2 + 6 \cdot (-1)^2 + 3 \cdot (-1)^2 = 24, \tag{15.105}$$

donc oui, le caractère est irréductible parce que $\langle \chi_s, \chi_s \rangle = 1$. Et nous pouvons donc ajouter la ligne (15.103) à notre tableau. Par ailleurs, nous notons qu'elle est de dimension 3.

Pour le reste nous savons qu'il y a autant de représentations irréductibles que de classes de conjugaison, de telle sorte qu'il ne manque que deux représentations irréductibles. De plus la proposition 15.23 nous dit que si n_i est la dimension de la i^{e} représentation irréductible, alors

$$|S_4| = \sum_i n_i^2. \tag{15.106}$$

Dans notre situation, si nous nommons n_1 et n_2 les dimensions des deux représentations qui nous manquent, nous avons $24 = n_1^2 + n_2^2 + (1^2 + 1^2 + 3^2)$, c'est-à-dire $n_1^2 + n_2^2 = 13$. Il n'y a pas des tonnes de sommes de deux carrés qui font 13. Il y a $n_1 = 2$ et $n_2 = 3$, et c'est tout.

Nous recherchons donc encore une représentation de dimension 2 et une de dimension 3. Pour cela nous allons un peu regarder les produits tensoriels qui s'offrent à nous. Pour faire une dimension

3, il faut faire le produit d'une de dimension 1 par une de dimension 3. Là encore le choix est très limité et nous demande d'essayer

$$\rho_W = \rho_s \otimes \rho_\epsilon \tag{15.107}$$

qui agit sur l'espace $V_2 \otimes V_\epsilon$ par

$$\rho_W(g)(v \otimes x) = \rho_s(g)v \otimes \rho_\epsilon(g)x. \tag{15.108}$$

Pour savoir son caractère nous utilisons la petite formule toute simple (15.93) : nous multiplions case par case les tableaux (15.103) et (15.98) :

	dimension	Id	(12)	(123)	(1234)	(12)(34)
χ_W	3	3	-1	0	1	-1

(15.109)

Avant de réellement ajouter cette ligne au tableau, nous devons nous assurer qu'elle est bien irréductible. Nous utilisons le même critère : $\langle \chi_W, \chi_W \rangle = 1$, donc c'est bon.

Pour trouver le dernier caractère, que nous nommerons χ_u, il ne faut pas beaucoup d'imagination. Il suffit d'utiliser les relations d'orthogonalité du théorème 15.24, en sachant que la dimension est 2 et qu'alors $\chi_W(\mathrm{Id}) = 2$, c'est pas trop compliqué :

	dimension	Id	(12)	(123)	(1234)	(12)(34)
χ_1	1	1	1	1	1	1
χ_ϵ	1	1	-1	1	-1	1
χ_s	3	3	1	0	-1	-1
χ_W	3	3	-1	0	1	-1
χ_u	2	2	b	c	d	e

(15.110)

Les relations d'orthogonalité des colonnes de la propriété 15.23 nous permettent de calculer les coefficients manquants. En pratique, il suffit de prendre le produit scalaire de chaque ligne avec la première et d'égaler avec zéro. Nous trouvons $b = 0$, $c = 1$, $d = 0$, et $e = 2$. Le tableau final est :

	dimension	Id	(12)	(123)	(1234)	(12)(34)
χ_1	1	1	1	1	1	1
χ_ϵ	1	1	-1	1	-1	1
χ_s	3	3	1	0	-1	-1
χ_W	3	3	-1	0	1	-1
χ_u	2	2	0	-1	0	2

(15.111)

Notons que nous sommes parvenus à remplir la dernière ligne sans rien savoir de la représentation qui va avec. Nous allons cependant donner une interprétation géométrique et fixer cette représentation comme agissant sur le triangle équilatéral en 15.30.

15.5.2 Représentation de S_4 via les isométries du tétraèdre

Une des représentations trouvées (la représentation ρ_s) peut être vue comme le groupe $\mathrm{Iso}(T)$ des isométries affine du tétraèdre grâce à la proposition 17.164 qui donne un isomorphisme de groupe $S_4 \simeq \mathrm{Iso}(T)$ lorsque T est un tétraèdre régulier de $\mathbb{R}^3$.

Nous verrons donc ça plus en détail dans la section 17.9.16.

15.5.3 À propos de la représentation ρ_u

Nous nous penchons à présent sur la représentations ρ_u dont nous ne savons rien à part qu'elle est de dimension 2 et son caractère.

Lemme 15.27.
Nous avons $\rho_u(s) = \mathrm{Id}$ pour tout $s \in V_4$.

Démonstration. Tous les éléments de V_4 sont conjugués (à part l'identité, mais pour elle le résultat est clair), donc il suffit de prouver le résultat pour un élément quelconque.

L'endomorphisme $\rho_u\big((12)(34)\big)$ est un endomorphisme d'ordre 2 sur $\mathbb{R}^2$ dont la trace est 2. Imposons donc

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{15.112}$$

sous la contrainte $a + d = 2$. La résolution est assez rapide et donne $b = c = 0$, $a = d = 1$.

Vous voulez une démonstration plus technologique ? Oui ? Alors commencez par remarquer que l'opérateur $A = \rho_u\big((12)(34)\big)$ vérifie $A^2 = 1$, donc le polynôme $X^2 - 1$ est un polynôme annulateur de A. Il est peut-être minimal ou peut être pas, mais en tout cas le polynôme minimal divise celui-là et donc est soit $X - 1$ soit $X + 1$ soit $X^2 - 1$. Dans les trois cas il est scindé à racines simples, et l'endomorphisme A est diagonalisable par le théorème 9.128(3).

Mais comme $A^2 = 1$, les valeurs propres (ce qui est sur la diagonale) de A ne peuvent être que ± 1. La trace étant 2, les éléments diagonaux ne peuvent être que 1. Et $A = \mathrm{Id}$. □

Le groupe V_4 définit en 5.41 est normal dans S_4, donc le quotient S_4/V_4 est un groupe par le lemme 2.31.

Lemme 15.28.
L'application

$$\begin{aligned} \tilde{\rho}_u \colon S_4/V_4 &\to \mathrm{GL}(2, \mathbb{R}) \\ [g] &\mapsto \rho_u(g) \end{aligned} \tag{15.113}$$

est bien définie et donne une représentation irréductible de S_4/V_4.

Démonstration. Montrons que c'est bien définit. Si $s \in V_4$ nous devons prouver que $\rho_u(gs) = \rho_u(g)$. Vu que ρ_u est un homomorphisme (c'est une représentation), et que $\rho_u(s) = \mathrm{Id}$ nous avons directement

$$\rho_u(gs) = \rho_u(g)\rho_u(s) = \rho_u(g). \tag{15.114}$$

Nous devons prouver que la représentation $\tilde{\rho}_u$ est irréductible. Si un sous-espace non trivial $\mathrm{Span}(x)$ était stabilisé par $\tilde{\rho}_u$, il serait également stabilisé par ρ_u. Mais comme ρ_u est irréductible, elle ne stabilise personne. □

Lemme 15.29.
Le groupe S_4/V_4 est un groupe non-abélien, isomorphe à S_3.

Démonstration. Le groupe S_4/V_4 a une représentation irréductible de dimension 2, et n'est donc pas abélien par le corolaire 15.26.

Il contient $|S_4|/|V_4| = 24/4 = 6$ éléments (théorème de Lagrange 2.34). Or $6 = 3 \times 2$, donc le groupe S_4/V_4 est dans le cas non-abélien du théorème 5.25(2). Cette partie parle d'unicité du groupe non-abélien d'ordre 6. Or S_3 est un groupe non-abélien d'ordre 6, donc S_4/V_4 est isomorphe à S_3. □

Attention : il n'est pas correct de dire que S_4/V_4 est un sous-groupe de S_4 juste parce que c'est un quotient de S_4 ; ce n'est en général pas vrai (exemple 2.32).

15.30.
Nous sommes maintenant aptes à identifier la représentation ρ_u. D'abord nous nous rappelons de la représentation $\rho_s \colon S_4 \to \mathrm{Iso}(T)$ de S_4 sur le tétraèdre. Ensuite si A est un sommet dudit tétraèdre et que $S_3 \subset S_4$ est la partie qui fixe A alors nous avons une représentation

$$\rho_s \colon S_3 \to \mathrm{Iso}(T) \tag{15.115}$$

qui agit en réalité sur le triangle équilatéral T' opposé au sommet A.

Nous avons finalement la chaine d'homomorphismes de groupes

$$S_4 \xrightarrow{\texttt{proj}} S_4/V_4 \xrightarrow{\simeq} S_3 \xrightarrow{\rho_s} \mathrm{Iso}(T') \tag{15.116}$$

Cela est donc une représentation $S_4 \to \mathrm{Iso}(T')$. Elle est de dimension 2 et est irréductible (elle contient les rotations d'angle $2\pi/3$ qui ne fixent aucune direction). Elle est donc la représentation ρ_u qui est la seule irréductible de dimension 2.

Nous avons donc montré que la représentation ρ_u dont nous ne savions rien est la représentation de S_4 sur un triangle équilatéral obtenue à partir de celle de S_4 sur le tétraèdre, en fixant un point.

Chapitre 16

Encore de l'analyse (et c'est pas fini)

16.1 Densité des polynômes

Corolaire 16.1.
Si $X \subset \mathbb{R}$ est compact et de mesure finie[1]*, alors l'ensemble des polynômes est denses dans* $\big(C(X,\mathbb{R}), \|.\|_2\big)$.

Démonstration. Si f est une fonction dans $C(X,\mathbb{R})$ et si $\epsilon \geqslant 0$ est donné alors nous pouvons considérer un polynôme P tel que $\|f-P\|_\infty \leqslant \epsilon$. Dans ce cas nous avons

$$\|f-P\|_2^2 = \int_X |f(x)-P(x)|^2 dx \leqslant \int_X \epsilon^2 dx = \epsilon^2 \mu(X) \tag{16.1}$$

où $\mu(X)$ est la mesure de X (finie par hypothèse). □

16.2 Primitive et intégrale

Nous avons déjà parlé de primitive de fonction continue en la proposition 11.322.

Proposition 16.2.
Soit I un intervalle borné ouvert de $\mathbb{R}$. Une fonction $h \in C_c^\infty(I)$ admet une primitive dans $C_c^\infty(I)$ si et seulement si $\int_I h = 0$.

Démonstration. Si une primitive H de h est à support compact, alors

$$\int_I h = H(b) - H(a) = 0 - 0 = 0. \tag{16.2}$$

Pas de problèmes dans ce sens.

Supposons maintenant que $\int_I h = 0$. Le fait que h admette une primitive dans $C^\infty(I)$ est évident : toute fonction continue admet une primitive[2]. Soit H une telle primitive et $\tilde{H} = H - H(b)$. Alors $\tilde{H}(b) = 0$ et

$$\tilde{H}(a) = H(a) - H(b) = -\int_I h = 0. \tag{16.3}$$

Nous rappelons que le support d'une fonction est *la fermeture* de l'ensemble des points de non-annulation.

Supposons que le support de h soit inclus dans $[m,M] \subset]a,b[$. En prenant des nombres m' et M' tels que $a < m' < m$ et $M < M' < b$ (nous insistons sur le caractère strict de ces inégalités), la fonction h est nulle sur $[a,m']$ et sur $[M',b]$; la fonction $\tilde{H}$ doit donc y être constante. Mais nous avons déjà vu que $\tilde{H}(a) = \tilde{H}(b) = 0$. Donc l'ensemble des points sur lesquels $\tilde{H}$ n'est pas nul est inclus dans $]m',M'[$ et donc est strictement (des deux côtés) inclus dans I. □

1. Dans $\mathbb{R}$ cette hypothèse est évidemment superflue par rapport à l'hypothèse de compacité ; mais ça suggère des généralisations ...
2. Théorème 11.322.

16.2.1 Théorème taubérien de Hardy-Littlewood

Un théorème **taubérien** est un théorème qui compare les modes de convergence d'une série.

Lemme 16.3.
Si f et g sont des fonctions continues, alors $s(x) = \max\{f(x), g(x)\}$ est également une fonction continue.

Démonstration. Soit x_0 et prouvons que s est continue en x_0. Si $f(x_0) \neq g(x_0)$ (supposons $f(x_0) > g(x_0)$ pour fixer les idées), alors nous avons un voisinage de x_0 sur lequel $f > g$ et alors $s = f$ sur ce voisinage et la continuité provient de celle de f.

Si au contraire $f(x_0) = g(x_0) = s(x_0)$ alors si (a_n) est une suite tendant vers x_0, nous prenons N tel que $\big|f(a_n) - f(x_0)\big| \leqslant \epsilon$ pour tout $n > N$ et M tel que $\big|g(a_n) - g(x_0)\big| \leqslant \epsilon$ pour tout $n > M$. Alors pour tout $n > \max\{N, M\}$ nous avons

$$\big|s(a_n) - s(x_0)\big| \leqslant \epsilon, \tag{16.4}$$

d'où la continuité de s en x_0. □

La proposition suivante dit que si une fonction connaît un saut, alors on peut le lisser par une fonction continue.

Proposition 16.4.
Soit f continue sur $[a, x_0[$ et sur $[x_0, b]$ avec $f(x_0^-) < f(x_0)$. En particulier nous supposons que $f(x^-)$ existe et est finie. Alors pour tout $\epsilon > 0$, il existe une fonction continue s telle que sur $[a, b]$ on ait $s \leqslant f$ et

$$\int_a^b s(x) - f(x)\,dx \leqslant \epsilon. \tag{16.5}$$

Démonstration. Nous notons A la taille du saut :

$$A = f(x_0) - f(x_0^-). \tag{16.6}$$

Quitte à changer a et b, nous pouvons supposer que

$$f(x) < f(x_0) + \frac{A}{3} \tag{16.7}$$

pour $x \in [a, x_0[$ et

$$f > f(x_0) + \frac{2A}{3} \tag{16.8}$$

pour $x \in [x_0, b]$. C'est le théorème des valeurs intermédiaires qui nous permet de faire ce choix.

Soit $m(x)$ la droite qui joint le point $\big(x_0 - \epsilon, f(x_0 - \epsilon)\big)$ au point $\big(x_0, f(x_0^+)\big)$. Nous posons

$$s(x) = \begin{cases} f(x) & \text{si } x < x_0 - \epsilon \\ \max\{m(x), f(x)\} & \text{si } x_0 - \epsilon \leqslant x \leqslant x_0 \\ f(x) & \text{si } x > x_0. \end{cases} \tag{16.9}$$

En vertu des différents choix effectués, c'est une fonction continue. En effet

$$s(x_0 - \epsilon) = \max\{f(x_0 - \epsilon), f(x_0, \epsilon)\} = f(x_0 - \epsilon) \tag{16.10}$$

et

$$s(x_0) = \max\{m(x_0), f(x_0^+)\} = f(x_0^+) \tag{16.11}$$

parce que $m(x_0) = f(x_0^+)$. En ce qui concerne l'intégrale, si nous posons

$$M = \sup_{x,y \in [a,b]} |f(x) - f(y)|, \tag{16.12}$$

nous avons

$$\int_a^b s - f = \int_{x_0 - \epsilon}^{x_0} s - f \leqslant \epsilon M. \tag{16.13}$$

□

Lemme 16.5.
Pour tout polynôme P, nous avons la formule

$$\lim_{x\to 1^-}(1-x)\sum_{n=0}^{\infty}x^nP(x^n)=\int_0^1 P(x)dx. \tag{16.14}$$

Démonstration. D'abord pour $P=1$, la formule se réduit à la série harmonique connue. Ensuite nous prouvons la formule pour le polynôme $P=X^k$ et la linéarité fera le reste pour les autres polynômes. Nous avons

$$(1-x)\sum_n x^nx^{kn}=(1-x)\sum_n(x^{1+k})^n=\frac{1-x}{1-x^{1+k}}=\frac{1}{1+x+\cdots+x^k}. \tag{16.15}$$

Donc

$$\lim_{x\to 1^-}(1-x)\sum_n x^nP(x^n)=\frac{1}{1+k}. \tag{16.16}$$

Par ailleurs, c'est vite vu que

$$\int_0^1 x^kdx=\frac{1}{k+1}. \tag{16.17}$$

□

Théorème 16.6 (Hardy-Littlewood[311])**.**
Soit (a_n) une suite réelle telle que

(1) $\frac{a_n}{n}$ tend vers une constante,
(2) $F(x)=\sum_{n=0}^{\infty}a_nx^n$ a un rayon de convergence $\geqslant 1$,
(3) $\lim_{x\to 1^-}F(x)=l$.

Alors $\sum_{n=0}^{\infty}a_n=l$.

Démonstration. Quitte à prendre la suite $b_0=a_0-l$ et $b_n=a_n$, on peut supposer $l=0$.

Soit Γ l'ensemble des fonctions

$$\gamma\colon[0,1]\to\mathbb{R} \tag{16.18}$$

telles que

(1) $\sum_{n=0}^{\infty}a_n\gamma(x^n)$ converge pour $0\leqslant x<1$,
(2) $\lim_{x\to 1^-}\sum_{n\geqslant 0}a_n\gamma(^n)=0$.

Ce Γ est un espace vectoriel.

Les polynômes sont dans Γ Soit $\gamma(t)=t^s$. Pour $0\leqslant x<1$ nous avons

$$\sum_{n=0}^{\infty}a_n\gamma(x^n)=\sum_{n=0}^{\infty}a_nx^{ns}<\sum_{n=0}^{\infty}a_nx^n. \tag{16.19}$$

Donc la condition de convergence est vérifiée. En ce qui concerne la limite,

$$\lim_{x\to 1^-}\sum_{n=0}^{\infty}a_nx^{ns}=\lim_{x\to 1^-}F(x^s)=0 \tag{16.20}$$

parce que par hypothèse, $\lim_{x\to 1^-}F(x)=0$.

Définition de la fonction qui va donner la réponse Nous considérons la fonction

$$g(t)=\begin{cases}0 & \text{si } 0\leqslant t<1/2\\ 1 & \text{si } 1/2\leqslant t\leqslant 1,\end{cases} \tag{16.21}$$

c'est-à-dire $g = \mathbb{1}_{[\frac{1}{2},1]}$. Nous montrons que si $g \in \gamma$, alors le théorème est terminé. Si $0 \leqslant x \leqslant 1$, on a $0 \leqslant x^n < 1/2$ dès que

$$n > -\frac{\ln(2)}{\ln(x)} \tag{16.22}$$

avec une note comme quoi $\ln(x) < 0$, donc la fraction est positive. Nous désignons par N_x la partie entière de ce n adapté à x. L'idée est que la fonction $g(x^n)$ est la fonction indicatrice de $0 \leqslant n \leqslant N_x$, et donc

$$\sum_{n\geqslant 0} a_n g(x^n) = \sum_{n=0}^{N_x} a_n. \tag{16.23}$$

Mais si $x \to 1^-$, alors $N_x \to \infty$, donc

$$\lim_{N\to\infty} \sum_{n=0}^{N} a_n = \lim_{x\to 1^-} \sum_{n=0}^{N_x} a_n = \lim_{x\to 1^-} \sum_{n\in\mathbb{N}} a_n g(x^n), \tag{16.24}$$

et cela fait zéro si $g \in \Gamma$.

Approximation de g par des polynômes Nous considérons la fonction

$$h(t) = \frac{g(t) - t}{t(1-1)} = \begin{cases} \frac{1}{t-1} & \text{si } t \in [0, 1/2[\\ \frac{1}{t} & \text{si } t \in [1/2, 1]. \end{cases} \tag{16.25}$$

La seconde égalité est au sens du prolongement par continuité. La fonction h est une fonction non continue qui fait un saut de -2 à 2 en $x = 1/2$. En vertu de la proposition 16.4 (un peu adaptée), nous pouvons considérer deux fonctions continues s_1 et s_2 telles que

$$s_1 \leqslant h \leqslant s_2 \tag{16.26}$$

et

$$\int_0^1 s_2 - s_1 \leqslant \epsilon. \tag{16.27}$$

Notons que l'inégalité $s_1 \leqslant s_2$ doit être stricte sur au moins un petit intervalle autour de $x = 1/2$. Soient P_1 et P_2, deux polynômes tels que $\|P_1 - s_1\|_\infty \leqslant \epsilon$ et $\|P_2 - s_2\|_\infty \leqslant \epsilon$ (ici la norme supremum est prise sur $[0, 1]$). C'est le théorème de Stone-Weierstrass (11.319) qui nous permet de le faire.
Nous posons aussi [3]

$$Q_1 = P_1 + \epsilon \tag{16.28a}$$
$$Q_2 = P_2 - \epsilon. \tag{16.28b}$$

Nous avons

$$\int_0^1 Q_1 - Q_2 \leqslant \int_0^1 Q_1 - P_1 + P_1 - P_2 + P_2 - Q_2. \tag{16.29}$$

Pour majorer cela, d'abord $Q_1 - P_1 = P_2 - Q2 = \epsilon$, ensuite,

$$P_1 - P_2 = P_1 - s_1 + s_1 - s_2 + s_2 - P_2 \tag{16.30}$$

dans lequel nous avons $P_1 - s_1 \leqslant \epsilon$, $s_2 - P_2 \leqslant \epsilon$ et $\int_0^1 s_1 - s_2 \leqslant \epsilon$. Au final, nous posons $q = Q_2 - Q_1$ et nous avons

$$\int_0^1 q \leqslant 5\epsilon. \tag{16.31}$$

Enfin nous posons aussi

$$R_i(x) = x + x(1-x)Q_i. \tag{16.32}$$

3. À ce niveau, je crois qu'il y a une faute de frappe dans [311].

Ces polynômes vérifient $R_i(0) = 0$, $R_i(1) = 1$ et

$$R_1 \leqslant g \leqslant R_2 \tag{16.33}$$

parce que

$$Q_1 \leqslant P_1 \leqslant h \leqslant P_2 \leqslant Q_2 \tag{16.34}$$

et

$$t + t(1-t)Q_1 \leqslant \underbrace{t + t(1-t)h(t)}_{g(t)} \leqslant t + t(1-t)Q_2. \tag{16.35}$$

Preuve que g est dans Γ D'abord si $0 \leqslant x < 1$, $x^N < \frac{1}{2}$ pour un certain N, et alors $g(x^N) = 0$. Du coup la série

$$\sum_{n=0}^{\infty} a_n g(x^n) = \sum_{n=0}^{N} a_n \tag{16.36}$$

est une somme finie qui converge donc.

D'autre part nous prenons M tel que $|a_n| < \frac{M}{n}$ pour tout n. Nous majorons $\sum_{n\in\mathbb{N}} a_n g(x^n)$ en utilisant R_1. Mais vu que R_1 est un polynôme, nous pouvons dire que $|\sum_{n=0}^{\infty} a_n R_1(x^n)| \leqslant \epsilon$ en prenant $x \in [\lambda, 1[$ et λ assez grand. Nous avons :

$$\left|\sum_{n=0}^{\infty} a_n g(x^n)\right| \leqslant \left|\sum_{n=0}^{\infty} a_n g(x^n) - \sum_{n=0}^{\infty} a_n R_1(x^n)\right| + \underbrace{\left|\sum_{n=0}^{\infty} a_n R_1(x^n)\right|}_{\leqslant\epsilon} \tag{16.37a}$$

$$\leqslant \epsilon + \sum_{n=0}^{\infty} |a_n|(g - R_1)(x^n) \tag{16.37b}$$

$$\leqslant \epsilon + \sum_{n=0}^{\infty} |a_n|(R_2 - R_1)(x^n) \tag{16.37c}$$

$$\leqslant \epsilon + M \sum_{n=0}^{\infty} \frac{x^n(1-x^n)}{n}(Q_2 - Q_1)(x^n) \tag{16.37d}$$

$$= \epsilon + M \sum_{n=0}^{\infty} \frac{x^n(1-x^n)}{n} q(x^n) \tag{16.37e}$$

$$\leqslant \epsilon + M(1-x)\sum_{n} x^n q(x^n). \tag{16.37f}$$

Justifications :

— La ligne (16.37d) est par le fait que $R_2 - R_1 = x(1-x)(Q_2 - Q_1)$.

— La ligne (16.37f) provient d'une majoration sauvage de $1/n$ par 1 et de $1 - x^n$ par $1 - x$.

Par le lemme 16.5, nous avons alors

$$\lim_{x\to 1^-} |\sum_{n} a_n g(x^n)| \leqslant \epsilon + M \int_0^1 q \leqslant 6\epsilon. \tag{16.38}$$

□

16.2.2 Théorème de Müntz

Théorème 16.7 (Théorème de Müntz[312, 313, 314])**.**
Soit $C_0([0,1])$, l'espace des fonctions continues sur $[0,1]$ muni de la norme $\|.\|_\infty$ ou $\|.\|_2$ et une suite (α_n) strictement croissante de nombres positifs. Nous notons ϕ_λ la fonction $x \mapsto x^\lambda$.

Alors

$$\overline{\mathrm{Span}\{1, \phi_{\alpha_n}\}} \tag{16.39}$$

est dense dans $C_0\big([0,1]\big)$ *si et seulement si*

$$\sum_{n=2}^{\infty} \frac{1}{\alpha_n} = +\infty. \tag{16.40}$$

Nous prouvons le théorème pour la norme $\|.\|_2$.

Démonstration. Soit $m \in \mathbb{R}^+$; nous notons $\Delta_N(m)$ la distance entre ϕ_m et $\mathrm{Span}\{\phi_{\alpha_1}, \dots, \phi_{\alpha_N}\}$. Cette distance peut être évaluée avec le déterminant de Gram (proposition 9.15)

$$\Delta_N(m)^2 = \frac{G(\phi_m, \phi_{\alpha_1}, \dots, \phi_{\alpha_N})}{G(\phi_{\alpha_1}, \dots, \phi_{\alpha_N})}. \tag{16.41}$$

Pour calculer cela nous avons besoin des produits scalaires [4]

$$\langle \phi_a, \phi_b \rangle = \int_0^1 x^{a+b} dx = \frac{1}{a+b+1}. \tag{16.42}$$

Pour avoir des notations plus compactes, nous notons $\alpha_0 = m$. Donc nous avons à calculer le déterminant

$$G(\phi_m, \phi_{\alpha_1}, \dots, \phi_{\alpha_N}) = \det\Big(\tfrac{1}{\alpha_i+\alpha_j+1}\Big) \tag{16.43}$$

où $i, j = 0, \dots, N$. Nous reconnaissons un déterminant de Cauchy (proposition 9.16) en posant, dans $\frac{1}{\alpha_i+\alpha_j+1}$, $a_i = \alpha_i$ et $b_j = \alpha_j + 1$. Étant donné que $b_j - b_i = a_j - a_i$, nous avons

$$G(\phi_m, \phi_{\alpha_1}, \dots, \phi_{\alpha_N}) = \frac{\prod_{0\leqslant i<j\leqslant N}(\alpha_j - \alpha_i)^2}{\prod_{i=0}^N \prod_{j=0}^N (\alpha_i + \alpha_j + 1).} \tag{16.44}$$

Nous séparons maintenant les termes où i ou j sont nuls. En ce qui concerne le dénominateur, il faut prendre tous les couples (i, j) avec i et j éventuellement égaux à zéro. Nous décomposant cela en trois paquets. Le premier est $(0, 0)$; le second est $(0, i)$ (chaque couple arrive en fait deux fois parce qu'il y a aussi $(i, 0)$) ; et le troisième sont les i, j tous deux différents de zéro :

$$(2m+1)\prod_{ij}(\alpha_i + \alpha_j + 1)\prod_i(\alpha_i + m + 1)^2. \tag{16.45}$$

Notons que dans le produit central, le carré est contenu dans le fait qu'on écrit $\prod_{ij}$ et non $\prod_{i<j}$. Nous avons donc

$$G(\phi_m, \phi_{\alpha_1}, \dots, \phi_{\alpha_N}) = \frac{\prod_{i<j}(\alpha_i - \alpha_j)^2 \prod_i(\alpha_i - m)^2}{(2m+1)\prod_{ij}(\alpha_i + \alpha_j + 1)\prod_i(\alpha_i + m + 1)^2}. \tag{16.46}$$

Le calcul de $G(\phi_{\alpha_1}, \dots, \phi_{\alpha_N})$ est plus simple [5] :

$$G(\phi_{\alpha_1}, \dots, \phi_{\alpha_N}) = \frac{\prod_{i<j}(\alpha_i - \alpha_j)^2}{\prod_{ij}(\alpha_i + \alpha_j + 1)}. \tag{16.47}$$

En divisant l'un par l'autre il ne reste que les facteurs comprenant m et en prenant la racine carrée,

$$\Delta_N(m) = \frac{1}{\sqrt{2m+1}} \prod_{i=1}^N \left| \frac{\alpha_i - m}{\alpha_i + m + 1} \right|. \tag{16.48}$$

Nous passons maintenant à la preuve proprement dite. Supposons que $V = \mathrm{Span}\{\phi_{\alpha_i}, i \in \mathbb{N}\}$ est dense. Si m est un des α_i, il peut évidemment être approché par les ϕ_{α_i}. Mais vue la densité

4. C'est ici qu'on se particularise à la norme $\|.\|_2$.
5. Je crois qu'il y a une faute de frappe dans le dénominateur de [312].

de V, un ϕ_m avec $m \neq \alpha_i$ (pour tout i) alors ϕ_m peut également être arbitrairement approché par les ϕ_{α_i}, c'est-à-dire que

$$\lim_{N\to\infty} \Delta_N(m) = 0. \tag{16.49}$$

Nous posons

$$u_n = \ln\left(\frac{\alpha_n - m}{\alpha_n + m + 1}\right) \tag{16.50}$$

et nous prouvons que la série $\sum_n u_n$ diverge. En effet nous nous souvenons de la formule $\ln(ab) = \ln(a) + \ln(b)$, de telle sorte que la N^{e}somme partielle de $\sum_n u_n$ est

$$\ln\left(\frac{\alpha_1 - m}{\alpha_1 + m + 1} \cdot \ldots \cdot \frac{\alpha_N - m}{\alpha_N + m + 1}\right) = \ln\left(\sqrt{2m+1}\Delta_N(m)\right), \tag{16.51}$$

qui tend vers $-\infty$ lorsque $N \to \infty$.

Si la suite (α_n) est majorée et plus généralement si nous n'avons pas $\alpha_n \to \infty$, alors évidemment la série $\sum_n \frac{1}{\alpha_n}$ diverge. Nous supposons donc que $\lim_{n\to\infty} \alpha_n = \infty$. Nous avons aussi [6]

$$u_n = \ln\left(\frac{\alpha_n - m}{\alpha_n + m + 1}\right) = \ln\left(1 - \frac{2m+1}{\alpha_n + m + 1}\right) \sim -\frac{2m+1}{\alpha_n}. \tag{16.52}$$

Une justification est donné à l'équation (14.267). Ce que nous avons surtout est

$$\sum_n u_n \sim -(2m+1)\sum_n \frac{1}{\alpha_n}. \tag{16.53}$$

Étant donné que la série de gauche diverge, celle de droite diverge [7].

Nous faisons maintenant le sens opposé : nous supposons que la série $\sum_n 1/\alpha_n$ diverge et nous nous posons

$$V = \text{Span}\{\phi_{\alpha_n} \text{ tel que } n \in \mathbb{N}\}. \tag{16.54}$$

Il suffit de prouver que $\phi_m \in \bar{V}$ pour tout m parce qu'un corolaire du théorème de Stone-Weierstrass 16.1 montre que $\text{Span}\{\phi_k \text{ tel que } k \in \mathbb{N}\}$ est dense dans C pour la norme $\|.\|_2$.

Si $\alpha_n \to \infty$, nous avons :

$$u_n \sim \frac{2m+1}{\alpha_n} \to 0 \tag{16.55}$$

et alors $\Delta_N(m) \to 0$. Dans ce cas nous avons immédiatement $\phi_m \in \bar{V}$.

Si par contre α_n ne tend pas vers l'infini, nous repartons de l'expression (16.48), nous posons $0 < \alpha = \sup_i \alpha_i$ et nous calculons :

$$\begin{aligned}
\sqrt{2m+1}\Delta_N(m) &= \prod_{i=1}^N \frac{|\alpha_i - m|}{\alpha_i + m + 1} && (16.56a)\\
&\leqslant \prod_{i=1}^N \frac{\alpha_i + m}{\alpha_i + m + 1} && (16.56b)\\
&= \prod_{i=1}^N \left(1 - \frac{1}{\alpha_i + m + 1}\right) && (16.56c)\\
&\leqslant \prod_{i=1}^N \left(1 - \frac{1}{\alpha + m + 1}\right) && (16.56d)\\
&= \left(1 - \frac{1}{\alpha + m + 1}\right)^N. && (16.56e)
\end{aligned}$$

Cette dernière expression tend vers 0 lorsque $N \to \infty$. □

6. Je crois qu'il y a une faute de signe dans la dernière expression de [313].
7. Nous utilisons le fait que si $u_n = \sum v_n$ en tant que suites et si $\sum_n u_n$ diverge, alors $\sum_n v_n$ diverge.

Remarque 16.8.
Certaines sources[8] citent le théorème de Müntz comme ceci (avec un implicite que $\alpha_i \neq 0$) :

$$\overline{\mathrm{Span}\{1, \phi_{\alpha_i}\}} = C\big([0,1]\big) \Leftrightarrow \sum_{i\geqslant 1} \frac{1}{\alpha_i} = +\infty. \tag{16.57}$$

Que penser de la présence explicite du 1 (c'est-à-dire de ϕ_0) ou non dans l'ensemble ?

Première chose : la présence éventuelle de ϕ_0 est la raison pour laquelle nous faisons commencer la somme à $i = 2$ et non $i = 1$. Dans le même ordre d'idée, si $\mathrm{Span}\{\phi_{\alpha_i}\}$ est dense, alors en prenant n'importe quelle queue de suite, ça reste dense.

Prouvons donc l'énoncé (16.57). Si $\mathrm{Span}\{1, \phi_{\alpha_i}\}$ est dense, alors en posant $\beta_1 = 0$, $\beta_i = \alpha_{i-1}$ notre théorème prouve que $\sum_{\beta=2}^{\infty} \frac{1}{\beta_i} = +\infty$, cela est exactement que $\sum_{i=1}^{\infty} \frac{1}{\alpha_i} = +\infty$. Dans l'autre sens, si $\sum_{i\geqslant 1} \frac{1}{\alpha_i} = +\infty$, alors nous avons aussi $\sum_{i\geqslant 2} \frac{1}{\alpha_i} = +\infty$ et notre théorème dit que $\mathrm{Span}\{\phi_{\alpha_i}\}$ est dense. A fortiori, $\mathrm{Span}\{1, \phi_{\alpha_i}\}$ est dense.

Exemple 16.9.
Nous savons depuis le théorème 14.85 que la somme des inverses des nombres premiers diverge. △

16.3 Intégrales convergeant uniformément

16.3.1 Définition et propriété

Définition 16.10.
Soit (Ω, μ) un espace mesuré. Nous disons que l'intégrale

$$\int_\Omega f(x,\omega) d\mu(\omega) \tag{16.58}$$

converge uniformément *en x si pour tout $\epsilon > 0$, il existe un compact K_ϵ tel que pour tout compact K tel que $K_\epsilon \subset K$ nous avons*

$$\left| \int_{\Omega\setminus K} f(x,\omega) d\mu(\omega) \right| \leqslant \epsilon. \tag{16.59}$$

Le point important est que le choix de K_ϵ ne dépend pas de x.

Lemme 16.11.
Soit

$$F(x) = \int_\Omega f(x,\omega) d\mu(\omega), \tag{16.60}$$

une intégrale uniformément convergente. Pour chaque $k \in \mathbb{N}$ nous considérons un compact K_k tel que

$$\left| \int_{\Omega\setminus K_k} f(x,\omega) d\mu(\omega) \right| \leqslant \frac{1}{k}. \tag{16.61}$$

Alors la suite de fonctions F_k définie par

$$F_k(x) = \int_{K_k} f(x,\omega) d\mu(\omega) \tag{16.62}$$

converge uniformément vers F.

8. Dont le rapport du jury 2014

Démonstration. Nous avons

$$\begin{aligned}\big|F_k(x) - F(x)\big| &= \left|\int_{K_k} f(x,\omega)d\mu(\omega) - \int_\Omega f(x,\omega)d\mu(\omega)\right| && (16.63a)\\ &= |\int_{\Omega\setminus K_k} f(x,\omega)d\mu(\omega)| && (16.63b)\\ &\leqslant \frac{1}{k}. && (16.63c)\end{aligned}$$

□

16.3.2 Critères de convergence uniforme

Afin de tester l'uniforme convergence d'une intégrale, nous avons le **critère de Weierstrass** :

Théorème 16.12.
Soit $f(x,t)\colon [\alpha,\beta]\times[a,\infty[\to\mathbb{R}$, une fonction dont la restriction à toute demi-droite $x = cst$ est mesurable. Si $|f(x,t)| < \varphi(t)$ et $\int_a^\infty \varphi(t)dt$ existe, alors l'intégrale

$$\int_0^\infty f(x,t)dt \tag{16.64}$$

est uniformément convergente.

Le théorème suivant est le **critère d'Abel** :

Théorème 16.13.
Supposons que $f(x,t) = \varphi(x,t)\psi(x,t)$ où φ et ψ sont bornée et intégrables en t au sens de Riemann sur tout compact $[a,b]$, $b \geqslant a$. Supposons que :

(1) $|\int_a^T \varphi(x,t)dt| \leqslant M$ où M est indépendant de T et de x,

(2) $\psi(x,t) \geqslant 0$,

(3) pour tout $x \in [\alpha,\beta]$, $\psi(x,t)$ est une fonction décroissante de t,

(4) les fonctions $x \mapsto \psi(x,t)$ convergent uniformément vers 0 lorsque $t \to \infty$.

Alors l'intégrale

$$\int_a^\infty f(x,t)dt \tag{16.65}$$

est uniformément convergente.

Remarque 16.14.
Étant donné que la fonction sinus est bornée, il est tentant de l'utiliser comme φ dans le critère d'Abel (théorème 16.13). Hélas,

$$\int_0^T \sin(xt) = -\frac{1}{x}\big(\cos(xT) - \cos(x)\big), \tag{16.66}$$

qui n'est pas bornée pour tout x ! Poser $\varphi(x,t) = \sin(xt)$ *ne fonctionne pas* pour assurer la convergence uniforme sur un intervalle qui contient des x arbitrairement proches de 0. Le critère d'Abel avec $\varphi(x,t) = \sin(xt)$ ne permet que de conclure à l'uniforme convergence *sur tout compact* ne contenant pas 0. Cela est toutefois souvent suffisant pour étudier la continuité ou la dérivabilité en se servant du fameux coup du compact.

16.4 Fonctions définies par une intégrale

Soit (Ω,μ) un espace mesuré. Nous nous demandons dans quel cas l'intégrale

$$F(x) = \int_\Omega f(x,\omega)d\omega \tag{16.67}$$

définit une fonction F continue, dérivable ou autre.

Dans la suite nous allons considérer des fonctions f à valeurs réelles. Quitte à passer aux composantes, nous pouvons considérer des fonctions à valeurs vectorielles. Par contre le fait que x soit dans $\mathbb{R}$ ou dans $\mathbb{R}^n$ n'est pas spécialement une chose facile à traiter.

16.4.1 Continuité sous l'intégrale

Nous allons présenter deux théorèmes donnant la continuité de F.

(1) Si f est majorée par une fonction ne dépendant pas de x, nous avons le théorème 16.15,

(2) si l'intégrale est uniformément convergente, nous avons le théorème 16.16.

Théorème 16.15.
Soit (Ω, μ) est un espace mesuré, soit $x_0 \in \mathbb{R}^m$ et $f\colon U \times \Omega \to \mathbb{R}$ où U est ouvert dans $\mathbb{R}^m$. Nous supposons que

(1) Pour chaque $x \in \mathbb{R}^m$, la fonction $\omega \mapsto f(x, \omega)$ est dans $L^1(\Omega, \mu)$.

(2) Pour chaque $\omega \in \Omega$, la fonction $x \mapsto f(x, \omega)$ est continue en x_0.

(3) Il existe une fonction $G \in L^1(\Omega)$ telle que

$$|f(x, \omega)| \leqslant G(\omega) \tag{16.68}$$

pour tout $x \in U$.

Alors la fonction

$$\begin{aligned} F\colon U &\to \mathbb{R} \\ x &\mapsto \int_\Omega f(x, \omega) d\mu(\omega) \end{aligned} \tag{16.69}$$

est continue en x_0.

Démonstration. Soit (x_n) une suite convergente vers x_0. Nous considérons la suite de fonctions $f_n\colon \Omega \to \mathbb{R}$ définies par

$$f_n(\phi) = f(x_n, \omega). \tag{16.70}$$

sur qui nous pouvons utiliser le théorème de la convergence dominée (théorème 13.184) pour obtenir

$$\lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} \int_\Omega f(x_n, \omega) d\mu(\omega) \tag{16.71a}$$

$$= \int_\Omega \lim_{n\to\infty} f(x_n, \omega) d\mu(\omega) \tag{16.71b}$$

$$= \int_\Omega f(x, \omega) d\mu(\omega) \tag{16.71c}$$

$$= F(x). \tag{16.71d}$$

Nous avons utilisé la continuité de $f(., \omega)$. □

Si nous avons un peu de compatibilité entre la topologie et la mesure, alors nous pouvons utiliser l'uniforme convergence d'une intégrale pour obtenir la continuité d'une fonction définie par une intégrale.

Théorème 16.16.
Soit (Ω, μ) un espace topologique mesuré tel que tout compact est de mesure finie. Soit une fonction $f\colon \mathbb{R} \times \Omega \to \mathbb{R}$ telle que

(1) Pour chaque $x \in \mathbb{R}$, la fonction $f(x, .)$ est $L^1(\Omega, \mu)$.

(2) Pour chaque $\omega \in \Omega$, la fonction $f(., \omega)$ est continue en x_0.

(3) L'intégrale

$$F(x) = \int_\Omega f(x,\omega)d\mu(\omega) \tag{16.72}$$

est uniformément convergente[9].

Alors la fonction F est continue en x_0.

Démonstration. Nous reprenons les notations du lemme 16.11. Les fonctions

$$F_k(x) = \int_{K_k} f(x,\omega)d\mu(\omega) \tag{16.73}$$

existent parce que les fonctions $f(x,.)$ sont dans $L^1(\Omega)$. Montrons que les fonctions F_k sont continues. Soit une suite $x_k \to x_0$ nous avons

$$\lim_{n\to\infty} F_k(x_n) = \lim_{n\to\infty} \int_{K_k} f(x_n,\omega)d\mu(\omega). \tag{16.74}$$

Nous pouvons inverser la limite et l'intégrale en utilisant le théorème de la convergence dominée. Pour cela, la fonction $f(x_n,\omega)$ étant continue sur le compact K_k, elle y est majorée par une constante. Le fait que les compacts soient de mesure finie (hypothèse) implique que les constantes soient intégrales sur K_k. Le théorème de la convergence dominée implique alors que

$$\lim_{n\to\infty} F_k(x_n) = \int_{K_k} \lim_{n\to\infty} f(x_n,\omega)d\mu(\omega) = \int_{K_k} f(x_0,\omega)d\mu(\omega) = F_k(x_0). \tag{16.75}$$

Nous avons utilisé le fait que $f(.,\omega)$ était continue en x_0.

Le lemme 16.11 nous indique alors que la convergence $F_k \to F$ est uniforme. Les fonctions F_k étant continues, la fonction F est continue. □

Pour finir, citons ce résultat concernant les fonctions réelles.

Théorème 16.17.
Nous considérons $F(x) = \int_a^\infty f(x,t)dt$. Si f est continue sur $[\alpha,\beta] \times [a,\alpha[$ et l'intégrale converge uniformément, alors $F(x)$ est continue.

16.4.2 Le coup du compact

Nous avons vu des fonctions définies par toute une série de processus de limite (suites, séries, intégrales). Une des questions centrales est de savoir si la fonction limite est continue, dérivable, intégrale, etc. étant donné que les fonctions sont continues.

Pour cela, nous inventons le concept de *convergence uniforme*. Si la limite (série, intégrale) est uniforme, alors la fonction limite sera continue. Il arrive qu'une limite ne soit pas uniforme sur un intervalle ouvert $]0,1]$, et que nous voulions quand même prouver la continuité sur cet intervalle. C'est à cela que sert la notion de convergence uniforme *sur tout compact*. En effet, la notion de continuité est une notion locale : savoir ce qu'il se passe dans un petit voisinage autour de x est suffisant pour savoir la continuité en x (idem pour sa dérivée).

Si nous avons uniforme convergence sur tout compact de $]0,1]$, mais pas uniforme convergence sur cet intervalle, la limite sera quand même continue sur $]0,1]$. En effet, si $x \in]0,1]$, il existe un ouvert autour de x contenu dans un compact contenu dans $]0,1]$. L'uniforme convergence sur ce compact suffit à prouver la continuité en x.

Déduire la continuité sur un ouvert à partir de l'uniforme convergence sur tout compact de l'ouvert est appelé faire le **coup du compact**.

9. Définition 16.10.

16.4.3 Dérivabilité sous l'intégrale

Nous traitons à présent de la dérivabilité de la fonction F définie comme intégrale de f.

Théorème 16.18 (Dérivation sous le signe intégral, formule de Leibnitz[298, 315])**.**
Soit (Ω, μ) un espace mesuré et une fonction $f\colon \mathbb{R}\times\Omega \to \mathbb{R}$ dont nous voulons étudier la dérivabilité en $a \in \mathbb{R}$. Nous supposons qu'il existe $\delta > 0$, A mesurable de mesure nulle dans Ω tels que

(1) Pour tout $x \in \mathbb{R}$, la fonction $\omega \mapsto f(x,\omega)$ soit dans $L^1(\Omega)$.

(2) L'application $x \mapsto f(x,\omega)$ est dérivable pour tout $x \in B(a,\delta)$ et pour tout $\omega \in \complement A$.

(3) Il existe une fonction G intégrable sur Ω telle que

$$\left|\frac{\partial f}{\partial x}(x,\omega)\right| \leqslant G(\omega) \tag{16.76}$$

pour tout $x \in B(a,\delta)$ et pour tout $\omega \in \complement A$.

Alors la fonction

$$F(x) = \int_\Omega f(x,\omega)d\mu(\omega) \tag{16.77}$$

est dérivable en a et nous pouvons permuter la dérivée et l'intégrale :

$$F'(a) = \int_\Omega \frac{\partial f}{\partial x}(a,\omega)d\mu(\omega). \tag{16.78}$$

Démonstration. Soit une suite (x_n) dans $B(a,\delta)$ telle que $x_n \neq a$ et $x_n \to a$. Si la limite

$$\lim_{n\to\infty} \frac{F(a) - F(x_n)}{a - x_n} \tag{16.79}$$

existe et ne dépend pas de la suite choisie, alors la fonction F est dérivable en a et sa dérivée vaut cette limite. Autrement dit, nous nous mettons en devoir d'étudier la limite

$$\lim_{n\to\infty} \int_\Omega \frac{f(a,\omega) - f(x_n,\omega)}{a - x_n} d\omega. \tag{16.80}$$

montrer qu'elle existe, ne dépend pas de la suite choisie et vaut $\int_\Omega \partial_x f(a,\omega)d\omega$. On y va.

<u>La bonne suite de fonctions</u> D'abord nous posons

$$g_n(\omega) = \frac{f(x_n,\omega) - f(a,\omega)}{x_n - a}. \tag{16.81}$$

Nous montrons à présent que cette suite vérifie les hypothèses du théorème de la convergence dominée 13.184.

— Chacune des fonctions g_n est dans $L^1(\Omega)$ parce que, a étant fixé, l'élément x_n est dans $B(a,\delta)\backslash\{a\}$; le dénominateur n'a donc aucun rôle. L'hypothèse (1) montre que $\omega \mapsto f(x_n\omega)$ et $\omega \mapsto f(a,\omega)$ sont dans L^1. La somme est donc dans L^1 (proposition 13.173).

— Par l'hypothèse (2), pour chaque ω nous avons une fonction dérivable. Nous pouvons donc passer à la limite :

$$\lim_{n\to\infty} g_n(x) = \frac{\partial f}{\partial x}(a,\omega). \tag{16.82}$$

— En ce qui concerne la majoration de g_n, nous utilisons le théorème des accroissements finis 11.130(2). Pour chaque ω, ce théorèmes peut être utilisé sur la fonction $x \mapsto f(x,\omega)$ qui est dérivable. Nous avons :

$$\left|\frac{f(x_n,\omega) - f(a,\omega)}{x_n - a}\right| \leqslant \sup_{x\in[a,x_n]} \left|\frac{\partial f}{\partial x}(x,\omega)\right| \leqslant G(\omega). \tag{16.83}$$

Nous avons utilisé l'hypothèse (3).

Les hypothèses de la convergence dominée sont satisfaites.

Convergence dominée Le théorème de la convergence dominée de Lebesgue (théorème 13.184) nous permet alors de calculer la limite (16.80) :

$$\lim_{n\to\infty}\int_\Omega g_n(\omega)d\omega = \int_\Omega \lim_{n\to\infty} g_n(\omega)d\omega = \int_\Omega \frac{\partial f}{\partial x}(a,\omega)d\omega. \tag{16.84}$$

Notons que l'existence de la dernière intégrale fait partie du théorème de la convergence dominée.

Nous avons donc prouvé que la limite de gauche existait et ne dépendant pas de la suite choisie. Donc F est dérivable en a et la dérivée vaut cette limite :

$$F'(a) = \int_\Omega \frac{\partial f}{\partial x}(a,\omega)d\mu(\omega). \tag{16.85}$$

□

Théorème 16.19.
Supposons f continue et sa dérivée partielle $\frac{\partial f}{\partial x}$ continue sur $[\alpha,\beta]\times[a,\alpha[$. Supposons que $F(x) = \int_a^\infty f(x,t)dt$ converge et que $\int_a^\infty \frac{\partial f}{\partial x}dt$ converge uniformément. Alors F est C^1 sur $[\alpha,\beta]$ et

$$\frac{dF}{dx} = \int_a^\infty \frac{\partial f}{\partial x}dt. \tag{16.86}$$

En ce qui concerne les fonctions dans $\mathbb{R}^n$, il y a les propositions 16.25 et 16.26 qui parlent de différentiabilité sous l'intégrale.

16.4.4 Absolue continuité

Définition 16.20.
*Une fonction $F\colon \mathbb{R}\to\mathbb{R}$ est **absolument continue** sur $[a,b]$ s'il existe une fonction f sur $[a,b]$ telle que*

$$F(x) = \int_a^x f(t)dt \tag{16.87}$$

pour tout $x\in[a,b]$.

Théorème 16.21.
Soient A un ouvert de $\mathbb{R}$ et Ω un espace mesuré. Soient une fonction $f\colon A\times\Omega\to\mathbb{R}$ et

$$F(x) = \int_\Omega f(x,\omega)d\omega. \tag{16.88}$$

Nous supposons les points suivants.

(1) La fonction f est mesurable en tant que fonction $A\times\Omega\to\mathbb{R}$. Pour chaque $x\in A$, la fonction $f(x,\ \cdot\ )$ est intégrable sur Ω.

(2) Pour presque tout $\omega\in\Omega$, la fonction $f(x,\omega)$ est une fonction absolument continue de x.

(3) La fonction $\frac{\partial f}{\partial x}$ est localement intégrable, c'est-à-dire que pour tout $[a,b]\subset A$,

$$\int_a^b\int_\Omega \left|\frac{\partial f}{\partial x}(x,\omega)\right| d\omega\, dx < \infty. \tag{16.89}$$

Alors la fonction F est absolument continue et pour presque tout $x\in A$, la dérivée est donné par

$$\frac{d}{dx}\int_\Omega f(x,\omega)d\omega = \int_\Omega \frac{\partial f}{\partial x}(x,\omega)d\omega. \tag{16.90}$$

La proposition suivante sera utilisée entre autres pour montrer que sous l'hypothèse d'une densité continue, la loi exponentielle est sans mémoire, proposition 35.106.

Proposition 16.22.
Soit $f(x,t)$ une fonction continue sur $[\alpha,\beta]\times[a,b]$, telle que $\frac{\partial f}{\partial x}$ existe et soit continue sur $]\alpha,\beta[\times[a,b]$. Soient $\varphi(x)$ et $\psi(x)$, des fonctions continues de $[\alpha,\beta]$ dans $\mathbb{R}$ et admettant une dérivée continue sur $]\alpha,\beta[$. Alors la fonction

$$F(x)=\int_{\varphi(x)}^{\psi(x)} f(x,t)dt \tag{16.91}$$

admet une dérivée continue sur $]\alpha,\beta[$ et

$$\frac{dF}{dx}=\int_{\varphi(x)}^{\psi(x)}\frac{\partial f}{\partial x}(x,t)dt+f\big(x,\psi(x)\big)\cdot\frac{d\psi}{dx}-f\big(x,\varphi(x)\big)\cdot\frac{d\varphi}{dx}. \tag{16.92}$$

L'exemple qui suit devrait pouvoir être rendu rigoureux en utilisant des distributions correctement.

Exemple 16.23.
Si g est une fonction continue, la fonction suivante est une primitive de g :

$$\int_0^x f(t)dt=\int_0^\infty f(t)\mathbb{1}_{t<x}(t)dt. \tag{16.93}$$

Nous nous proposons de justifier *de façon un peu heuristique* le fait que ce soit bien une primitive de g en considérant la fonction

$$f(t,x)=g(t)\mathbb{1}_{t<x}(t). \tag{16.94}$$

Nous posons

$$F(x)=\int_0^\infty f(x,t)dt, \tag{16.95}$$

et nous calculons F' en permutant la dérivée et l'intégrale [10]. D'abord,

$$f(t,x)=\begin{cases} g(t) & \text{si } t\in[0,x]\\ 0 & \text{sinon.}\end{cases} \tag{16.96}$$

La dérivée de f par rapport à x est donnée par la distribution

$$\frac{\partial f}{\partial x}(t_0,x_0)=g(t_0)\delta(t_0-x_0). \tag{16.97}$$

Donc

$$F'(x_0)=\int_0^\infty\frac{\partial f}{\partial x}(t,x_0)dt=\int_0^\infty g(t)\delta(t-x_0)=g(x_0), \tag{16.98}$$

comme attendu. △

Cet exemple est rendu rigoureux par la proposition suivante.

Proposition 16.24.
Si $f\in L^1(\mathbb{R})$, alors la fonction

$$F(x)=\int_{-\infty}^x f(t)dt \tag{16.99}$$

est presque partout dérivable et pour les points où elle l'est nous avons $F'(x)=f(x)$.

10. Ceci n'est pas rigoureux : il faudrait avoir un théorème à propos de distributions qui permet de le faire.

16.4.5 Différentiabilité sous l'intégrale

Le théorème suivant est restrictif sur l'ensemble d'intégration (qui doit être compact), mais accepte des fonctions de plusieurs variables, ce qui est un premier pas vers la différentiabilité.

Proposition 16.25 (Dérivation sous l'intégrale)**.**
Supposons $A \subset \mathbb{R}^m$ ouvert et $B \subset \mathbb{R}^n$ compact. Nous considérons une fonction $f\colon A \times B \to \mathbb{R}$. Si pour un $i \in \{i, \dots, n\}$, la dérivée partielle $\frac{\partial f}{\partial x_i}$ existe dans $A \times B$ et est continue, alors la fonction

$$F(x) = \int_B f(x,t)dt \tag{16.100}$$

admet une dérivée partielle dans la direction x_i sur A. Cette dérivée partielle y est continue et

$$\frac{\partial F}{\partial x_i}(a) = \int_B \frac{\partial f}{\partial x_i}(a,t)dt, \tag{16.101}$$

pour tout a dans l'ouvert A.

Démonstration. Nous procédons en plusieurs étapes.

<u>**F est dérivable**</u> Nous voulons prouver que $\frac{\partial F}{\partial x_i}(a,t)$ existe. Pour cela nous posons

$$g_l(t) = \frac{f(a_1, \dots, a_i + \epsilon_l, \dots, a_n, t) - f(a_1, \dots, a_i, \dots, a_n, t)}{\epsilon_l} \tag{16.102}$$

où ϵ_l est une suite de nombres tendant vers zéro. La fonction f est dérivable dans la direction x_i si et seulement si $\lim_{l\to\infty} g_l(t)$ existe et ne dépend pas du choix de la suite. À ce moment, la valeur de la dérivée partielle sera cette limite. Dans notre cas, nous savons que f admet une dérivée partielle dans la direction x_i et donc nous avons

$$\frac{\partial f}{\partial x_i}(a,t) = \lim_{l\to\infty} g_l(t). \tag{16.103}$$

De la même façon pour F nous avons

$$\frac{\partial F}{\partial x_i} = \lim_{l\to\infty} \int_B g_l(t)dt. \tag{16.104}$$

Sous-entendu : si la limite de droite ne dépend pas de la suite choisie, alors $\frac{\partial F}{\partial x_i}$ existe et vaut cette limite.

Vu la continuité de f, le seul point à vérifier pour le théorème de la convergence dominée de Lebesgue est l'existence d'une fonction intégrable de t majorant g_l. Pour cela le théorème de accroissements finis (théorème 11.130) appliqué à la fonction $\epsilon \mapsto f(a_n, \dots, a_i + \epsilon, \dots, a_n)$ nous dit que

$$f(a_1, \dots, a_i + \epsilon_l, \dots, a_n, t) - f(a_1, \dots, a_i, \dots, a_n, t) = \epsilon_l \frac{\partial f}{\partial x_i}(a_1, \dots, \theta, \dots, a_n, t) \tag{16.105}$$

pour un certain $\theta \in B(a_i, \epsilon_l)$. Notons que ce θ dépend de t mais pas de l. Vu que $\partial_i f$ est continue par rapport à ses deux variables, si K est un voisinage compact autour de a, il existe $M > 0$ tel que

$$\left| \frac{\partial f}{\partial x_i}(x,t) \right| < M \tag{16.106}$$

pour tout $x \in K$ et tout $t \in B$. La valeur de $\frac{\partial f}{\partial x_i}(a_1, \dots, \theta, \dots, a_n, t)$ est donc bien majorée par rapport à θ et par rapport à t en même temps par une constante qui n'a pas de mal à être intégrée sur le compact B.

Le théorème de la convergence dominée (théorème 13.184) s'applique donc bien et nous avons

$$\lim_{l\to\infty}\int_B g_l(t)dt = \int_B \lim_{l\to\infty} g_l(t) = \int_B \frac{\partial f}{\partial x_i}(a,t)dt. \tag{16.107}$$

Le membre de droite ne dépendant pas de la suite ϵ_l choisie, le membre de gauche est bien la dérivée de F par rapport à x_i et nous avons

$$\frac{\partial F}{\partial x_i}(a) = \int_B \frac{\partial f}{\partial x_i}(a,t)dt. \tag{16.108}$$

Cela prouve la première partie de la proposition.

La dérivée est continue Soit K un voisinage compact autour de a et U' un ouvert tel que $a \in U' \subset K$. Nous avons encore la majoration (16.106) sur U' et donc le théorème de continuité sous l'intégrale 16.15 nous indique que la fonction

$$\begin{aligned} U' &\to \mathbb{R} \\ x &\mapsto \int_B \frac{\partial f}{\partial x_i}(x,t)dt \end{aligned} \tag{16.109}$$

est continue en a.

□

Une conséquence de la proposition 16.25 est que si elle fonctionne pour tous les i, alors F est différentiable et même de classe C^1, et la différentielle de F s'obtient comme intégrale de la différentielle de f.

Proposition 16.26.
Supposons $A \subset \mathbb{R}^m$ ouvert et $B \subset \mathbb{R}^n$ compact. Si pour tout $i \in \{i,\dots,n\}$, la dérivée partielle $\frac{\partial f}{\partial x_i}$ existe dans $A \times B$ et est continue, alors F est de classe C^1 et

$$(dF)_a = \int_B (df_t)_a dt \tag{16.110}$$

où $f_t(x) = f(x,t)$.

Démonstration. En vertu de la proposition 16.25, toutes les dérivées partielles de F sont continues. Cela implique que F est de classe C^1 par le théorème 11.243 et que la différentielle s'écrive en termes des dérivées partielles avec la formule usuelle. Nous avons alors

$$(dF)_a(u) = \sum_k \frac{\partial F}{\partial x_k}(a)u_k \tag{16.111a}$$

$$= \int_B \sum_k \frac{\partial f}{\partial x_k}(a,t)dt \tag{16.111b}$$

$$= \int_B \sum_k \frac{\partial f_t}{\partial x_k}(a)u_k dt \tag{16.111c}$$

$$= \int_B (df_t)_a(u)dt. \tag{16.111d}$$

Cela est la formule annoncée. □

Un autre théorème tourne autour du pot, et me semble inutile.

Théorème 16.27.
Soit (Ω,μ) un espace mesuré, une fonction $f\colon \mathbb{R}^n\times\Omega \to \mathbb{R}$ et $a \in \mathbb{R}^n$. Nous considérons la fonction

$$F(x) = \int_\Omega f(x,\omega)d\mu(\omega). \tag{16.112}$$

Pour chaque $k = 1, \dots, n$ nous supposons avoir

$$\frac{\partial F}{\partial x_k}(a) = F'_{|_k}(a) = \int_\Omega \frac{\partial f_{|_k}}{\partial t}(a_k, \omega) d\mu(\omega) \tag{16.113}$$

où $F_{|_k}(t) = F(a_1, \dots, t, \dots, a_n)$ et $f_{|_k}$ est définie de façon similaire.

Nous supposons de plus que les fonctions $\partial_{x_k} F$ sont continues.

Alors F est de classe C^1 et sa différentielle est donnée par

$$df_a = \int_\Omega (df_\omega)_a d\omega \tag{16.114}$$

où f_ω est définie par $f_\omega(x) = f(x, \omega)$.

Démonstration. Étant donné que les dérivées partielles de F en a existent et sont continues, le théorème 11.243 dit que F est différentiable et que

$$dF_a(u) = \sum_{k=1}^n \frac{\partial F}{\partial x_k}(a) u_k. \tag{16.115}$$

La linéarité de l'intégrale et les hypothèses nous donnent alors

$$\begin{aligned} df_a(u) &= \sum_{k=1}^n \frac{\partial F}{\partial x_k}(a) u_k && (16.116a)\\ &= \int_\Omega \sum_k \frac{\partial f_{|_k}}{\partial t}(a_k; \omega) u_k d\mu(\omega) && (16.116b)\\ &= \int_\Omega \sum_k \frac{\partial f}{\partial x_k}(a; \omega) u_k d\mu(\omega) && (16.116c)\\ &= \int_\Omega (df_\omega)_a(u) d\mu(\omega), && (16.116d) \end{aligned}$$

et donc $df_a = \int_\Omega (df_\omega)_a d\mu(\omega)$. □

Notons qu'en passant aux composantes, ce théorème fonctionne tout aussi bien pour des fonctions à valeurs dans un espace vectoriel normé de dimension finie plutôt que dans $\mathbb{R}$.

Lemme 16.28 (Hadamard[316])**.**
Soit une fonction $f\colon \mathbb{R}^n \to \mathbb{R}$ de classe C^p avec $p \geqslant 1$. Pour tout $a \in \mathbb{R}^n$ il existe des fonctions $g_1, \dots, g_n$ de classe C^{p-1} telles que

$$f(x) = f(a) + \sum_{i=1}^n (x_i - a_i) g_i(x). \tag{16.117}$$

Démonstration. Vu que f est de classe C^1, le théorème fondamental de l'analyse 13.236 fonctionne et

$$f(x) - f(a) = \int_0^1 \frac{d}{dt}\Big[f\big(a + t(x-a)\big)\Big] dt = \int_0^1 \sum_{i=1}^n \frac{\partial f}{\partial x_i}\big(a + t(x-a)\big)(x_i - a_i). \tag{16.118}$$

Plus de détails : la fonction $t \mapsto \frac{d}{dt}\Big[f\big(a + t(x-a)\big)\Big]$ possède comme primitive la fonction $F(t) = f\big(a + t(x-a)\big)$.

Nous posons

$$g_i(x) = \int_0^1 \frac{\partial f}{\partial x_i}\big(a + t(x-a)\big) dt \tag{16.119}$$

Le fait que l'intégrale existe est simplement le fait qu'il s'agit d'une fonction continue sur un compact et donc majorée par une constante. Pour voir que g_i est de classe C^{p-1} nous pouvons calculer $\frac{\partial g_i}{\partial x_k}$ en permutant dérivée et intégrale par la proposition 16.25 :

$$\frac{\partial g_i}{\partial x_k}(x) = \int_0^1 \frac{\partial}{\partial x_k}\left(\frac{\partial f}{\partial x_i}\big(a+t(x-a)\big)\right) dt = \int_0^1 t\frac{\partial^2 f}{\partial x_k \partial x_i}\big(a+t(x-a)\big). \tag{16.120}$$

Nous pouvons ainsi permuter $p-1$ dérivées tout en gardant une fonction continue dans l'intégrale. Le théorème 16.15 nous donne alors une fonction continue. Ainsi toutes les fonctions

$$\frac{\partial^{p-1} g_i}{\partial x_{i_1} \dots \partial x_{i_{p-1}}} \tag{16.121}$$

sont continues et g_i est de classe C^{p-1} par le théorème 11.273.

En repartant de (16.118) nous avons alors bien ce qui était annoncé :

$$f(x) = f(a) + \sum_{i=1}^n g_i(x)(x_i - a_i). \tag{16.122}$$

□

Corolaire 16.29.
Soit $\phi \in \mathscr{D}(\mathbb{R})$ tel que $\phi^{(k)}(x_0) = 0$ pour tout $k \leqslant n$. Alors il existe une fonction $\psi \in \mathscr{D}(\mathbb{R})$ telle que

$$\phi(x) = (x - x_0)^{n+1}\psi(x) \tag{16.123}$$

pour tout $x \in \mathbb{R}$.

Démonstration. En utilisant le lemme de Hadamard 16.28 avec $a = x_0$, $n = 1$ et $f(x_0) = 0$, nous avons une fonction g_1 à support compact telle que

$$\phi(x) = \phi(x_0) + (x - x_0)g_1(x). \tag{16.124}$$

Alors $\phi'(x) = g_1(x) + (x - x_0)g_1'(x)$, ce qui donne immédiatement $g_1(x_0) = 0$ et donc une fonction g_2 telle que $g_1(x) = (x - x_0)g_2(x)$. En injectant dans (16.124) nous avons

$$\phi(x) = (x - x_0)^2 g_2(x). \tag{16.125}$$

Il suffit de continuer ainsi tant que les dérivées de ϕ s'annulent. □

16.5 Deux théorèmes de point fixe

Nous allons voir Picard. Les autres théorème de point fixe que sont Brouwer, Schauder et Markov-Kakutani sont plus bas [11] parce qu'ils utilisent de l'intégration. Voir le thème 19 pour les retrouver.

16.5.1 Points fixes attractifs et répulsifs

Définition 16.30.
*Soit I un intervalle fermé de $\mathbb{R}$ et $\varphi\colon I \to I$ une application C^1. Soit a un point fixe de φ. Nous disons que a est **attractif** s'il existe un voisinage V de a tel que pour tout $x_0 \in V$ la suite $x_{n+1} = \varphi(x_n)$ converge vers a. Le point a sera dit **répulsif** s'il existe un voisinage V de a tel que pour tout $x_0 \in V$ la suite $x_{n+1} = \varphi(x_n)$ diverge.*

Lemme 16.31 ([317])**.**
Soit a un point fixe de φ.

11. Dans la section 19.5.

(1) Si $|\varphi'(a)| < 1$ alors a est attractif et la convergence est au moins exponentielle.

(2) Si $|\varphi'(a)| > 1$ alors a est répulsif et la divergence est au moins exponentielle.

Démonstration. Si $|\varphi'(a) < 1|$ alors il existe k tel que $|\varphi'(a)| < k < 1$ et par continuité il existe un voisinage V de a dans lequel $|\varphi'(x)| < k$ pour tout $x \in V$. En utilisant le théorème des accroissements finis nous avons

$$|x_n - a| = \big|f(x_{n-1} - a)\big| \leqslant k|x_{n-1} - a| \tag{16.126}$$

et par récurrence

$$|x_n - a| \leqslant k^n|x_0 - a|. \tag{16.127}$$

Le cas $|\varphi'(a) > 1|$ se traite de façon similaire. □

Remarque 16.32.
Dans le cas $|\varphi'(a)| = 1$, nous ne pouvons rien conclure. Si $\varphi(x) = \sin(x)$ nous avons $\sin(x) < x$ et le point $a = 0$ est attractif. A contrario, si $\varphi(x) = \sinh(x)$ nous avons $|\sinh(x)| > |x|$ et le point $a = 0$ est répulsif.

16.5.2 Picard

Définition 16.33.
Une application $f\colon (X, \|.\|_X) \to (Y, \|.\|_Y)$ entre deux espaces métriques est une ***contraction*** *si elle est k-**Lipschitz** pour un certain $0 \leqslant k < 1$, c'est-à-dire si pour tout $x, y \in X$ nous avons*

$$\|f(x) - f(y)\|_Y \leqslant k\|x - y\|_X. \tag{16.128}$$

Théorème 16.34 (Picard [318, 319] [12].)**.**
Soit X un espace métrique complet et $f\colon X \to X$ une application contractante, de constante de Lipschitz k. Alors f admet un unique point fixe, nommé ξ. Ce dernier est donné par la limite de la suite définie par récurrence

$$\begin{cases} x_0 \in X & \text{(16.129a)} \\ x_{n+1} = f(x_n). & \text{(16.129b)} \end{cases}$$

De plus nous pouvons majorer l'erreur par

$$\|x_n - x\| \leqslant \frac{k^n}{1-k}\|x_n - x_{n-1}\| \leqslant \frac{k^n}{1-k}\|x_1 - x_0\|. \tag{16.130}$$

Soit $r > 0$, $a \in X$ tels que la fonction f laisse la boule $K = \overline{B(a, r)}$ invariante (c'est-à-dire que f se restreint à $f\colon K \to K$). Nous considérons les suites (u_n) et (v_n) définies par

$$\begin{cases} u_0 = v_0 \in K & \text{(16.131a)} \\ u_{n+1} = f(v_n), v_{n+1} \in B(u_n, \epsilon). & \text{(16.131b)} \end{cases}$$

Alors le point fixe ξ de f est dans K et la suite (v_n) satisfait l'estimation

$$\|v_n - \xi\| \leqslant \frac{k^n}{1-k}\|u_1 - u_0\| + \frac{\epsilon}{1-k}. \tag{16.132}$$

La première inégalité (16.130) donne une estimation de l'erreur calculable en cours de processus ; la seconde donne une estimation de l'erreur calculable avant de commencer.

12. Il me semble qu'à la page 100 de [319], l'hypothèse H1 qui est prouvée ne prouve pas Hn dans le cas $n = 1$. Merci de m'écrire si vous pouvez confirmer ou infirmer. La preuve donnée ici ne contient pas cette « erreur ».

Démonstration. Nous commençons par l'unicité du point fixe. Si a et b sont des points fixes, alors $f(a) = a$ et $f(b) = b$. Par conséquent

$$\|f(a) - f(b)\| = \|a - b\|, \tag{16.133}$$

ce qui contredit le fait que f soit une contraction.

En ce qui concerne l'existence, notons que si la suite des x_n converge dans X, alors la limite est un point fixe. En effet en prenant la limite des deux côtés de l'équation $x_{n+1} = f(x_n)$, nous obtenons $\xi = f(\xi)$, c'est-à-dire que ξ est un point fixe de f. Notons que nous avons utilisé ici la continuité de f, laquelle est une conséquence du fait qu'elle soit Lipschitz. Nous allons donc porter nos efforts à prouver que la suite est de Cauchy (et donc convergente parce que X est complet). Nous commençons par prouver que $\|x_{n+1} - x_n\| \leqslant k^n\|x_0 - x_1\|$. En effet pour tout n nous avons

$$\|x_{n+1} - x_n\| = \|f(x_n) - f(x_{n-1})\| \leqslant k\|x_n - x_{n-1}\|. \tag{16.134}$$

La relation cherchée s'obtient alors par récurrence. Soient $q > p$. En utilisant une somme télescopique,

$$\begin{aligned}
\|x_q - x_p\| &\leqslant \sum_{l=p}^{q-1} \|x_{l+1} - x_l\| && (16.135a)\\
&\leqslant \left(\sum_{l=p}^{q-1} k^l\right) \|x_1 - x_0\| && (16.135b)\\
&\leqslant \left(\sum_{l=p}^{\infty} k^l\right) \|x_1 - x_0\|. && (16.135c)
\end{aligned}$$

Étant donné que $k < 1$, la parenthèse est la queue d'une série qui converge, et donc tend vers zéro lorsque p tend vers l'infini.

En ce qui concerne les inégalités (16.130), nous refaisons une somme télescopique :

$$\begin{aligned}
\|x_{n+p} - x_n\| &\leqslant \|x_{n+p} - x_{n+p-1}\| + \cdots + \|x_{n+1} - x_n\| && (16.136a)\\
&\leqslant k^p\|x_n - x_{n-1}\| + k^{p-1}\|x_n - x_{n-1}\| + \cdots + k\|x_n - x_{n-1}\| && (16.136b)\\
&= k(1 + \cdots + k^{p-1})\|x_n - x_{n-1}\| && (16.136c)\\
&\leqslant \frac{k}{1-k}\|x_n - x_{n-1}\|. && (16.136d)
\end{aligned}$$

En prenant la limite $p \to \infty$ nous trouvons

$$\|\xi - x_n\| \leqslant \frac{k}{1-k}\|x_n - x_{n-1}\| \leqslant \frac{k}{1-k}\|x_1 - x_0\|. \tag{16.137}$$

Nous passons maintenant à la seconde partie du théorème en supposant que f se restreigne en une fonction $f\colon K \to K$. D'abord K est encore un espace métrique complet, donc la première partie du théorème s'y applique et f y a un unique point fixe.

Nous allons montrer la relation par récurrence. Tout d'abord pour $n = 1$ nous avons

$$\|v_1 - \xi\| \leqslant \|v_1 - u_1\| + \|u_1 - \xi\| \leqslant \epsilon + \frac{k}{1-k}\|u_1 - u_0\| \tag{16.138}$$

où nous avons utilisé l'estimation (16.137), qui reste valable en remplaçant x_1 par u_1 [13]. Nous pouvons maintenant faire la récurrence :

$$\begin{aligned}
\|v_{n+1} - \xi\| &\leqslant \|v_{n+1} - u_{n+1}\| + \|u_{n+1} - \xi\| && (16.139a)\\
&\leqslant \epsilon + k\|v_n - \xi\| && (16.139b)\\
&\leqslant \epsilon + k\left(\frac{k^n}{1-k}\|u_1 - u_0\| + \frac{\epsilon}{1-k}\right) && (16.139c)\\
&= \frac{\epsilon}{1-k} + \frac{k^{n+1}}{1-k}\|u_1 - u_0\|. && (16.139d)
\end{aligned}$$

13. Elle n'est cependant pas spécialement valable si on remplace x_n par u_n.

□

Remarque 16.35.
Ce théorème comporte deux parties d'intérêts différents. La première partie est un théorème de point fixe usuel, qui sera utilisé pour prouver l'existence de certaines équations différentielles.

La seconde partie est intéressante d'un point de vie numérique. En effet, ce qu'elle nous enseigne est que si à chaque pas de calcul de la récurrence $x_{n+1} = f(x_n)$ nous commettons une erreur d'ordre de grandeur ϵ, alors le procédé (la suite (v_n)) ne converge plus spécialement vers le point fixe, mais tend vers le point fixe avec une erreur majorée par $\epsilon/(k-1)$.

Remarque 16.36.
Au final l'erreur minimale qu'on peut atteindre est de l'ordre de ϵ. Évidemment si on commet une faute de calcul de l'ordre de ϵ à chaque pas, on ne peut pas espérer mieux.

Remarque 16.37.
Si f elle-même n'est pas contractante, mais si f^p est contractante pour un certain $p \in \mathbb{N}$ alors la conclusion du théorème de Picard reste valide et f a le même unique point fixe que f^p. En effet nommons x le point fixe de f : $f^p(x) = x$. Nous avons alors

$$f^p\big(f(x)\big) = f\big(f^p(x)\big) = f(x), \tag{16.140}$$

ce qui prouve que $f(x)$ est un point fixe de f^p. Par unicité nous avons alors $f(x) = x$, c'est-à-dire que x est également un point fixe de f.

Si la fonction n'est pas Lipschitz mais presque, nous avons une variante.

Proposition 16.38.
Soit E un ensemble compact[14] *et si $f\colon E \to E$ est une fonction telle que*

$$\|f(x) - f(y)\| < \|x - y\| \tag{16.141}$$

pour tout $x \neq y$ dans E alors f possède un unique point fixe.

Démonstration. La suite $x_{n+1} = f(x_n)$ possède une sous-suite convergente. La limite de cette sous-suite est un point fixe de f parce que f est continue. L'unicité est due à l'aspect strict de l'inégalité (16.141). □

Théorème 16.39 (Équation de Fredholm)**.**
Soit $K\colon [a,b] \times [a,b] \to \mathbb{R}$ et $\varphi\colon [a,b] \to \mathbb{R}$, deux fonctions continues. Alors si λ est suffisamment petit, l'équation

$$f(x) = \lambda \int_a^b K(x,y) f(y) dy + \varphi(x) \tag{16.142}$$

admet une unique solution qui sera de plus continue sur $[a,b]$.

Démonstration. Nous considérons l'ensemble $\mathcal{F}$ des fonctions continues $[a,b] \to [a,b]$ muni de la norme uniforme. Le lemme 11.297 implique que $\mathcal{F}$ est complet. Nous considérons l'application $\Phi\colon \mathcal{F} \to \mathcal{F}$ donnée par

$$\Phi(f)(x) = \lambda \int_a^b K(x,y) f(y) dy + \varphi(x). \tag{16.143}$$

Nous montrons que Φ^p est une application contractante pour un certain p. Pour tout $x \in [a,b]$ nous avons

$$\|\Phi(f) - \Phi(g)\|_\infty \leqslant \|\Phi(f)(x) - \Phi(g)(x)\| \tag{16.144a}$$

$$= |\lambda| \Big\| \int_a^b K(x,y)\big(f(y) - g(y)\big) dy \Big\| \tag{16.144b}$$

$$\leqslant |\lambda| \|K\|_\infty |b-a| \|f - g\|_\infty \tag{16.144c}$$

14. Notez cette hypothèse plus forte

Si λ est assez petit, et si p est assez grand, l'application Φ^p est donc une contraction. Elle possède donc un unique point fixe par le théorème de Picard 16.34. □

16.6 Théorèmes de point fixes et équations différentielles

16.6.1 Théorème de Cauchy-Lipschitz

Nous démontrons ici deux théorèmes de Cauchy-Lipschitz. De nombreuses propriétés annexes seront démontrées dans le chapitre sur les équations différentielles, section 31.8.

Le théorème de Cauchy-Arzella 19.33 sera pour plus tard parce qu'il utilise Schauder 19.32.

Théorème 16.40 (Cauchy-Lipschitz[320, 321])**.**
Nous considérons l'équation différentielle

$$\begin{cases} y'(t) = f\big(t, y(t)\big) & \text{(16.145a)} \\ y(t_0) = y_0 & \text{(16.145b)} \end{cases}$$

avec $f\colon U = I \times \Omega \to \mathbb{R}^n$ où I est ouvert dans $\mathbb{R}$ et Ω ouvert dans $\mathbb{R}^n$. Nous supposons que f est continue sur U et localement Lipschitz[15] *par rapport à y.*

Alors il existe un intervalle $J \subset I$ sur lequel la solution au problème est unique. De plus toute solution du problème est une restriction de cette solution à une partie de J. La solution sur J (dite « solution maximale ») est de classe C^1.

Démonstration. Nous divisions la preuve en plusieurs étapes (même pas toutes simples).

Cylindre de sécurité Précisons l'espace fonctionnel $\mathcal{F}$ adéquat. Soient V et W les voisinages de t_0 et y_0 sur lesquels f est localement Lipschitz. Nous considérons les quantités suivantes :

(1) $M = \sup_{V\times W} f$;

(2) $r > 0$ tel que $\overline{B(y_0, r)} \subset V$

(3) $T > 0$ tel que $\overline{B(t_0, T)} \subset W$ et $T < r/M$.

Nous considérons alors l'ensemble

$$\mathcal{F} = C^0\big(\overline{B(t_0, T)}, \overline{B(y_0, r)}\big) \tag{16.146}$$

que nous munissons de la norme uniforme. Par le lemme 11.297 l'espace $\big(\mathcal{F}, \|.\|_\infty\big)$ est complet.

Une application $\Phi\colon \mathcal{F} \to \mathcal{F}$ Si y est une solution de l'équation différentielle considérée, elle vérifie[16]

$$y(t) = y_0 + \int_{t_0}^{t} f\big(u, y(u)\big)du. \tag{16.147}$$

Ceci nous incite à considérer l'opérateur $\Phi\colon \mathcal{F} \to \mathcal{F}$ défini par

$$\Phi(y)(t) = y_0 + \int_{t_0}^{t} f\big(u, y(u)\big)du. \tag{16.148}$$

Pour que l'application Φ soit utile nous devons montrer que pour tout $y \in \mathcal{F}$,

— l'application $\Phi(y)$ est bien définie,

— pour tout $t \in \overline{B(y_0, r)}$ nous avons $\Phi(y)(t) \in \overline{B(t_0, T)}$,

— l'application $\Phi(y)\colon \overline{B(t_0, T)} \to \overline{B(y_0, r)}$ est continue.

Attention : nous ne prétendons pas que Φ elle-même soit continue. C'est parti.

15. Définition 11.262. Notons que nous ne supposons pas que f soit une contraction.
16. C'est le théorème fondamental du calcul intégral 13.236.

$\Phi(y)$ est bien définie Il faut montrer que l'intégrale converge. Le calcul de $\Phi(y)(t)$ ne se fait qu'avec $t \in \overline{B(t_0,T)}$. Vu que u prend ses valeurs dans $[t_0,t]$ et que $y \in \mathcal{F}$, le nombre $y(u)$ est toujours dans $\overline{B(y_0,r)}$. Ceci pour dire que dans l'intégrale, la fonction f n'est considérée que sur $[t_0,t] \times \overline{B(y_0,r)} \subset V \times W$. La fonction f est donc uniformément majorable, et l'intégrale ne pose pas de problèmes.

$\Phi(y)(t) \in \overline{B(t_0,T)}$ Prouvons que $\Phi(y)(t) \in \overline{B(y_0,r)}$. Pour cela, notons que

$$|\Phi(y)(t) - y_0| \leqslant \int_{t_0}^{t} |f\big(u,y(u)\big)|du \leqslant |t-t_0|\|f\|_\infty. \tag{16.149}$$

Étant donné que $t \in \overline{B(t_0,T)}$ nous avons $|t-t_0| \leqslant r/M$ et donc $|\Phi(y)(t) - y_0| \leqslant r$.

$\Phi(y)$ est continue Nous pourrions invoquer le théorème 16.15, mais nous allons le faire à la main. Soit $s_0 \in B(t_0,T)$ et prouvons que $\Phi(y)$ est continue en s_0. Pour cela nous prenons $s \in B(s_0,\delta)$ et nous calculons :

$$|\Phi(y)(s) - \Phi(y)(s_0)| \leqslant \int_{s_0}^{s} |f\big(u,y(u)\big)|du \leqslant |s_0 - s|\|f\|_\infty. \tag{16.150}$$

C'est le fait que f soit bornée dans le cylindre de sécurité qui fait en sorte que cela tende vers zéro lorsque $s \to s_0$.

L'équation (16.147) signifie que y est un point fixe de Φ. L'espace $\mathcal{F}$ étant complet le théorème de point fixe de Picard (théorème 16.34) s'applique. Nous allons montrer qu'il existe un $p \in \mathbb{N}$ tel que Φ^p soit contractante. Par conséquent Φ^p aura un unique point fixe qui sera également unique point fixe de Φ par la remarque 16.37.

Contractante Prouvons donc que Φ^p est contractante pour un certain p. Pour cela nous commençons par montrer la formule suivante par récurrence :

$$\big\|\Phi^p(x)(t) - \Phi^p(y)(t)\big\| \leqslant \frac{k^p|t-t_0|^p}{p!}\|x-y\|_\infty \tag{16.151}$$

pour tout $x,y \in \mathcal{F}$, et pour tout $t \in \overline{B(t_0,T)}$. Pour $p=0$ la formule (16.151) est vérifiée parce que $\|x-y\|_\infty$ est le supremum de $\|x(t)-y(t)\|$ pour $t \in \overline{B(t_0,T)}$. Supposons que la formule soit vraie pour p et calculons pour $p+1$. Pour tout $t \in \overline{B(t_0,T)}$ nous avons

$$\begin{aligned}
\big\|\Phi^{p+1}(x)(t) - \Phi^{p+1}(y)(t)\big\| &\leqslant \Big|\int_{t_0}^{t} \big\|f\big(u,\Phi^p(x)(u)\big) - f\big(u,\Phi^p(y)(u)\big)\big\|du\Big| && \text{(16.152a)}\\
&\leqslant \Big|\int_{t_0}^{t} k\|\Phi^p(x)(u) - \Phi^p(y)(u)\|du\Big| && \text{(16.152b)}\\
&\leqslant \Big|\int_{t_0}^{t} k\frac{k^p|t-t_0|}{p!}\|x-y\|_\infty\Big| && \text{(16.152c)}\\
&= \frac{k^{p+1}|t-t_0|^{p+1}}{(p+1)!}\|x-y\|_\infty. && \text{(16.152d)}
\end{aligned}$$

Justifications :

— (16.152b) parce que f est Lipschitz.

— (16.152c) par hypothèse de récurrence.

La formule (16.151) est maintenant établie. Nous pouvons maintenant montrer que Φ^p est une contraction pour un certain p. Pour tout $t \in \overline{B(t_0,T)}$ nous avons

$$\|\Phi^p(x)(t) - \Phi^p(y)(t)\| \leqslant \frac{k^p}{t!}|t-t_0|^p\|x-y\|_\infty \leqslant \frac{k^pT^p}{p!}\|x-y\|_\infty \tag{16.153}$$

où nous avons utilisé le fait que $|t-t_0|^p < T^p$. En prenant le supremum sur t des deux côtés il vient

$$\|\Phi^p(x) - \Phi^p(y)\|_\infty \leqslant \frac{k^pT^p}{p!}\|x-y\|_\infty. \tag{16.154}$$

Le membre de droite tend vers zéro lorsque $p \to \infty$ parce que $k < 1$ et $T^p/p! \to 0$[17]. Nous concluons donc que Φ^p est une contraction pour un certain p.

Conclusion L'unique point fixe de Φ est alors l'unique solution continue de l'équation différentielle (16.145). Par ailleurs l'équation elle-même $y' = f(t, y)$ demande implicitement que y soit dérivable et donc continue. Nous concluons que l'unique point fixe de Φ est l'unique solution de l'équation différentielle donnée. Cette dernière est automatiquement C^1 parce que si y est continue alors $u \mapsto f(u, y(u))$ est continue, c'est-à-dire que y' est continue.

Unicité Nous passons maintenant à la partie « prolongement maximum » du théorème. Soient x_1 et x_2 deux solutions maximales du problème (16.145) sur des intervalles I_1 et I_2 respectivement. Les intervalles I_1 et I_2 contiennent $\overline{B(t_0, r)}$ sur lequel $x_1 = x_2$ par unicité.

Nous allons maintenant montrer que pour tout $t \geqslant t_0$ pour lequel x_1 ou x_2 est défini, $x_1(t)$ et $x_2(t)$ sont définis et sont égaux. Le raisonnement sur $t \leqslant t_0$ est similaire.

Supposons que l'ensemble des $t \geqslant t_0$ tels que $x_1 = x_2$ soit ouvert à droite, c'est-à-dire soit de la forme $[t_0, b[$. Dans ce cas, soit x_1 soit x_2 (soit les deux) cesse d'exister en b. En effet si nous avions les fonctions x_i sur $[t_0, b + \epsilon[$ alors l'équation $x_1 = x_2$ définirait un fermé dans $[t_0, b + \epsilon[$. Supposons pour fixer les idées que x_1 cesse d'exister : le domaine de x_1 (parmi les $t \geqslant 0$) est $[t_0, b[$ et sur ce domaine nous avons $x_1 = x_2$. Dans ce cas x_1 pourrait être prolongé en x_2 au-delà de b. Si x_1 et x_2 s'arrêtent d'exister en même temps en b, alors nous avons bien $x_1 = x_2$.

Nous devons donc traiter le cas où $x_1 = x_2$ sur $[t_0, b]$ alors que x_1 et x_2 existent sur $[t_0, b + \epsilon[$ pour un certain ϵ.

Nous pouvons appliquer le théorème d'existence locale au problème

$$\begin{cases} y' = f(t, y) & \text{(16.155a)} \\ y(b) = x_1(b). & \text{(16.155b)} \end{cases}$$

Il existe un voisinage de b sur lequel la solution est unique. Sur ce voisinage nous devons donc avoir $x_1 = x_2$, ce qui contredit le fait que $x_1 \neq x_2$ en dehors de $[t_0, b]$.

Donc x_1 et x_2 existent et sont égaux sur au moins $I_1 \cup I_2$.

□

Le théorème de Cauchy-Lipschitz donne existence et unicité d'une solution maximale. Cependant cette solution peut ne pas exister partout où les hypothèses sur f sont remplies. En d'autres termes, il peut arriver que f soit Lipschitz jusqu'à t_1, mais que la solution maximale ne soit définie que jusqu'en $t_2 < t_1$. Ce cas fait l'objet du théorème d'explosion en temps fini 31.18.

Sous quelques hypothèses nous pouvons nous assurer de l'existence d'une solution unique sur tout $\mathbb{R}$.

Théorème 16.41 (Cauchy-Lipschitz global[322, 54]).
Soit un intervalle I de $\mathbb{R}$, $y_0 \in \mathbb{R}^n$, $t_0 \in I$ et une fonction continue $f\colon I \times \mathbb{R}^n \to \mathbb{R}^n$ telle que pour tout compact K dans I, il existe $k > 0$ tel que

$$\|f(t, y_1) - f(t, y_2)\| \leqslant k\|y_1 - y_2\| \tag{16.156}$$

pour tout $t \in K$ et $y_1, y_2 \in \mathbb{R}^n$.

Alors le problème

$$\begin{cases} y'(t) = f\big(t, y(t)\big) & \text{(16.157a)} \\ y(t_0) = y_0 & \text{(16.157b)} \end{cases}$$

possède une unique solution $y\colon I \to \mathbb{R}^n$ sur I.

17. C'est le terme général du développement de e^T qui est une série convergente.

Démonstration. Soit un intervalle compact K dans I et contenant t_0. Nous notons ℓ le diamètre de K. Sur l'espace $E = C^0(K, \mathbb{R}^n)$ nous considérons la topologie uniforme : $(E, \|.\|_\infty)$. C'est un espace complet par le lemme 11.297 (nous utilisons le fait que $\mathbb{R}^n$ soit complet, proposition 1.115). Nous allons utiliser l'application suivante :

$$\begin{aligned} \Phi\colon E &\to E \\ \Phi(y)(t) &= y_0 + \int_{t_0}^{t} f\big(s, y(s)\big) ds \end{aligned} \tag{16.158}$$

Démontrons quelques faits à propos de Φ.

La définition fonctionne bien Nous devons commencer par prouver que cette application est bien définie. Si $y \in E$ alors f et y sont continues ; l'application $s \mapsto f\big(s, y(s)\big)$ est donc également continue. L'intégrale de cette fonction sur le compact $[t_0, t]$ ne pose alors pas de problèmes. En ce qui concerne la continuité de $\phi(y)$ sous l'hypothèse que y soit continue,

$$\|\Phi(y)(t) - \Phi(y)(t')\| \leqslant \int_{t}^{t'} \|f(s, y(s))\| ds \leqslant M|t - t'| \tag{16.159}$$

où M est une majoration de $\|s \mapsto f\big(s, y(s)\big)\|_{\infty,K}$.

Si y est solution alors $\Phi(y) = y$ Supposons que y soit une solution de l'équation différentielle (16.157). Alors, vu que $y'(t) = f\big(t, y(t)\big)$ nous avons :

$$y(t) = y_0 + \int_{t_0}^{t} y'(s) ds = y_0 + \int_{t_0}^{t} f\big(s, y(s)\big) ds = \Phi(y)(t). \tag{16.160}$$

Si $\Phi(y) = y$ alors y est solution Nous avons, pour tout t :

$$y(t) = y_0 + \int_{t_0}^{t} f\big(s, y(s)\big) ds. \tag{16.161}$$

Le membre de droite est dérivable par rapport à t, et la dérivée fait $f\big(t, y(t)\big)$. Donc le membre de gauche est également dérivable et nous avons bien

$$y'(t) = f\big(t, y(t)\big). \tag{16.162}$$

De plus $y(t_0) = y_0 + \int_{t_0}^{t_0} \ldots = y_0$.

Nous sommes encore avec K compact et $E = C^0(K, \mathbb{R}^n)$ muni de la norme uniforme. Nous allons montrer que Φ est une contraction de E pour une norme bien choisie.

Une norme sur E Pour $y \in E$ nous posons

$$\|y\|_k = \max_{t \in K} \big(e^{-k|t-t_0|} \|y(t)\|\big). \tag{16.163}$$

Ce maximum est bien définit et fini parce que la fonction de t dedans est une fonction continue sur le compact K. Cela est également une norme parce que si $\|y\|_k = 0$ alors $e^{-k|t-t_0|}\|y(t)\| = 0$ pour tout t. Étant donné que l'exponentielle ne s'annule pas, $\|y(t)\| = 0$ pour tout t.

Équivalence de norme Nous montrons que les normes $\|.\|_k$ et $\|.\|_\infty$ sont équivalentes[18] :

$$\|y\|_\infty e^{-k\ell} \leqslant \|y\|_k \leqslant \|y\|_\infty \tag{16.164}$$

pour tout $y \in E$. Pour la première inégalité, $\ell \geqslant |t - t_0|$ pour tout $t \in K$, et $k > 0$, donc

$$\|y(t)\| e^{-k\ell} \leqslant e^{-k|t-t_0|} \|y(t)\|. \tag{16.165}$$

En prenant le maximum des deux côtés, $\|y\|_\infty e^{-k\ell} \leqslant \|y\|_k$.

En ce qui concerne la seconde inégalité dans (16.164), $k|t - t_0| \geqslant 0$ et donc $e^{-k|t-t_0|} < 1$.

18. Définition 7.221

Vu que les normes $\|.\|_\infty$ et $\|.\|_k$ sont équivalentes, l'espace $(E, \|.\|_k)$ est tout autant complet que $(E, \|.\|_\infty)$. Nous démontrons à présent que Φ est une contraction dans $(E, \|\|_k)$.

Soient $y, z \in E$. Si $t \geqslant t_0$ nous avons

$$\|\Phi(y)(t) - \Phi(z)(t)\| \leqslant \int_{t_0}^{t} \|f\big(s, y(s)\big) - f\big(s, z(s)\big)\| ds \tag{16.166a}$$

$$\leqslant k \int_{t_0}^{t} \|y(s) - z(s)\| ds. \tag{16.166b}$$

Il convient maintenant de remarquer que

$$\|y(t)\| = e^{-k|t-t_0|} e^{k|t-t_0|} \|y(t)\| \leqslant \|y\|_k e^{k|t-t_0|}. \tag{16.167}$$

Nous pouvons avec ça prolonger les inégalités (16.166) par

$$\|\Phi(y)(t) - \Phi(z)(t)\| \leqslant k\|y - z\|_k \int_{t_0}^{t} e^{k|s-t_0|} ds = k\|y - z\|_k \int_{t_0}^{t} e^{k(s-t_0)} ds \tag{16.168}$$

où nous avons utilisé notre supposition $t \geqslant t_0$ pour éliminer les valeurs absolues. L'intégrale peut être faite explicitement, mais nous en sommes arrivés à un niveau de fainéantise tellement inconcevable que

```

 SageMath version 7.3, Release Date: 2016-08-04
 Type "notebook()" for the browser-based notebook interface.
 Type "help()" for help.

sage: var('a,b,k')
(a, b, k)
sage: f(x)=exp(-k*x)
sage: f.integrate(x,a,b)
e^(-a*k)/k - e^(-b*k)/k
```

tex/sage/sageSnip014.sage

Au final, si $t \geqslant t_0$,

$$\|\Phi(y)(t) - \Phi(z)(t)\| \leqslant \|y - z\|_k \big(e^{k(t-t_0)} - 1\big). \tag{16.169}$$

Si $t \leqslant t_0$, il faut retourner les bornes de l'intégrale avant d'y faire rentrer la norme parce que $\|\int_0^1 f\| \leqslant \int_0^1 \|f\|$, mais ça ne marche pas avec $\|\int_1^0 f\|$. Pour $t \leqslant t_0$ tout le calcul donne

$$\|\Phi(y)(t) - \Phi(z)(t)\| \leqslant \|y - z\|_k \big(e^{k(t_0-t)} - 1\big). \tag{16.170}$$

Les deux inéquations sont valables a fortiori en mettant des valeurs absolues dans l'exponentielle, de telle sorte que pour tout $t \in K$ nous avons

$$e^{-k|t_0-t|} \|\phi(y)(t) - \Phi(z)(t)\| \leqslant \|y - z\|_k \big(1 - e^{-k|t_0-t|}\big). \tag{16.171}$$

En prenant le supremum sur t,

$$\|\Phi(y) - \Phi(z)\|_k \leqslant \|y - z\|_k (1 - e^{-k\ell}), \tag{16.172}$$

mais $0 < (1 - e^{e-k\ell}) < 1$, donc Φ est contractante pour la norme $\|.\|_k$. Vu que $(E, \|.\|_k)$ est complet, l'application Φ y a un unique point fixe par le théorème de Picard 16.34.

Ce point fixe est donc l'unique solution de l'équation différentielle de départ.

Existence et unicité sur I Il nous reste à prouver que la solution que nous avons trouvée existe sur I : jusqu'à présent nous avons démontré l'existence et l'unicité sur n'importe quel compact dans I.

Soit une suite croissante de compacts K_n contenant t_0 (par exemple une suite exhaustive comme celle du lemme 7.312). Nous avons en particulier

$$I = \bigcup_{n=0}^{\infty} K_n. \tag{16.173}$$

Existence sur I Soit y_n l'unique solution sur K_n. Il suffit de poser

$$y(t) = y_n(t) \tag{16.174}$$

pour n tel que $t \in K_n$. Cette définition fonctionne parce que si $t \in K_n \cap K_m$, il y a forcément un des deux qui est inclus dans l'autre et le résultat d'unicité sur le plus grand des deux donne $y_n(t) = y_m(t)$.

Unicité sur I Soient y et z des solutions sur I ; vu que I n'est pas spécialement compact, le travail fait plus haut ne permet pas de conclure que $y = z$.

Soit $t \in I$. Alors $t \in K_n$ pour un certain n et y et z sont des solutions sur K_n qui est compact. L'unicité sur K_n donne $y(t) = z(t)$.

□

16.42.
Il y a d'autres moyens de prouver qu'une solution existe globalement sur $\mathbb{R}$. Si f est globalement bornée, le théorème d'explosion en temps fini donne quelques garanties, voir 31.20.

Le théorème suivant donne une version du théorème de Cauchy-Lipschitz lorsque la fonction f dépend d'un paramètre. Ce théorème n'utilise rien de fondamentalement nouveau. Nous le donnons seulement pour montrer que l'on peut choisir l'espace $\mathcal{F}$ de façon un peu maligne pour élargir le résultat. Si vous voulez un théorème de Cauchy-Lipschitz avec paramètre vraiment intéressant, allez voir le théorème 31.31.

Théorème 16.43 (Cauchy-Lipschitz avec paramètre[1, 323])**.**
Soit un intervalle ouvert I de $\mathbb{R}$, un connexe ouvert Ω de $\mathbb{R}^n$ et un intervalle ouvert Λ de $\mathbb{R}^d$. Soit une fonction $f\colon I \times \Omega \times \Lambda \to \mathbb{R}^n$ continue et localement Lipschitz en Ω. Soient $t_0 \in I$, $y_0 \in \Omega$ et $\lambda_0 \in \Lambda$. Il existe un voisinage compact de (t_0, y_0, λ_0) sur lequel le problème

$$\begin{cases} y'_\lambda(t) = f\big(t, y_\lambda(t), \lambda\big) & \text{(16.175a)} \\ y_\lambda(t_0) = y_0 & \text{(16.175b)} \end{cases}$$

possède une unique solution. De plus $(t, \lambda) \mapsto y_\lambda(t)$ est continue[19].

Démonstration. **Problèmes et choses à faire**
Ceci est une idée de la preuve. Je n'ai pas vérifié toutes les étapes. Soyez prudent.

D'abord nous avons un voisinage compact $V \times \overline{B(y_0, r)} \times \Lambda_0$ de (t_0, y_0, λ_0) sur lequel f est bornée. Ensuite nous récrivons l'équation différentielle sous la forme

$$\begin{cases} \dfrac{\partial y}{\partial t}(t, \lambda) = f\big(t, y(t, \lambda), \lambda\big) & \text{(16.176a)} \\ y(t_0, \lambda) = y_0. & \text{(16.176b)} \end{cases}$$

pour une fonction $y\colon V \times \Lambda_0 \to \mathbb{R}^n$.

19. Ici, la surprise est que ce soit continu par rapport à λ. Le fait qu'elle le soit par rapport à t est clair depuis le départ parce que c'est finalement rien d'autre que le Cauchy-Lipschitz vieux et connu.

Nous posons $\mathcal{F} = C^0\big(V \times \Lambda_0, \mathbb{R}^n\big)$ et nous y définissons l'application

$$\begin{aligned} \Phi\colon \mathcal{F} &\to \mathcal{F} \\ \Phi(y)(t,\lambda) &= y_0 + \int_{t_0}^{t} f\big(s, y(s,\lambda), \lambda\big) ds. \end{aligned} \tag{16.177}$$

Il y a plein de vérifications à faire[323], mais je parie que Φ est bien définie, et que une de ses puissances est une contraction de $(\mathcal{F}, \|.\|_\infty)$. L'unique point fixe est une solution de notre problème et est dans C^0, donc $(t,\lambda) \mapsto y(t,\lambda) = y_\lambda(t)$ est de classe C^0, c'est-à-dire continue. □

16.44.
Ce théorème marque un peu la limite de ce que l'on peut faire avec la méthode des points fixes dans le cadre de Cauchy-Lipschitz : nous sommes limités à la continuité de la solution parce que les espaces C^p ne sont pas complets[20]. Il n'y a donc pas d'espoir d'adapter la méthode pour prouver que si f est de classe C^p alors $(t,\lambda) \mapsto y_\lambda(t)$ est de classe C^p. On peut, à λ fixé prouver que $t \mapsto y_\lambda(t)$ est de classe C^p (utiliser une récurrence), mais pas plus.

La régularité C^1 de y par rapport à la condition initiale sera l'objet du théorème 31.28. Ce résultat n'est vraiment pas facile et utilise des ingrédients bien autres qu'un point fixe. Ensuite la régularité C^p par rapport à la condition initiale et par rapport à un paramètre seront presque des cadeaux (proposition 31.29 et 31.31).

Exemple 16.45 ([324])**.**
Nous savons que le théorème de Picard permet de trouver le point fixe par itération de la contraction à partir d'un point quelconque. Tentons donc de résoudre

$$\begin{cases} y'(t) = y(t) & \text{(16.178a)} \\ y(0) = 1 & \text{(16.178b)} \end{cases}$$

dont nous savons depuis l'enfance que la solution est l'exponentielle[21]. Partons donc de la fonction constante $y_0 = 1$, et appliquons la contraction (16.158) :

$$u_1 = 1 + \int_0^1 u_0(s) ds = 1 + t. \tag{16.179}$$

Ensuite

$$u_2 = 1 + \int_0^t (1+s) ds = 1 + t + \frac{t^2}{2}. \tag{16.180}$$

Et on voit que les itérations suivantes vont donner l'exponentielle.

Nous sommes évidemment en droit de se dire que nous avons choisi un bon point de départ. Tentons le coup avec une fonction qui n'a rien à voir avec l'exponentielle : $u_0(x) = \sin(x)$.

Le programme suivant permet de faire de belles investigations numériques en partant d'à peu près n'importe quelle fonction :

```
#! /usr/bin/sage -python
# -*- coding: utf8 -*-

from sage.all import *

x=var('x')

def Phi(f):
    prim=f.integrate()
```

20. Par exemple, le théorème de Stone-Weierstrass 11.319 nous dit que la limite uniforme de polynômes (de classe C^∞) peut n'être que continue. Voir aussi le thème 25.
21. Voir par exemple le théorème 14.54.

```
    return 1+prim(x)-prim(0)

f=sin(x)

for i in range(1,30):
    print(i,f)
    f=Phi(f)

g=f(x)-exp(x)
plot(g,(x,-10,10)).show()
```

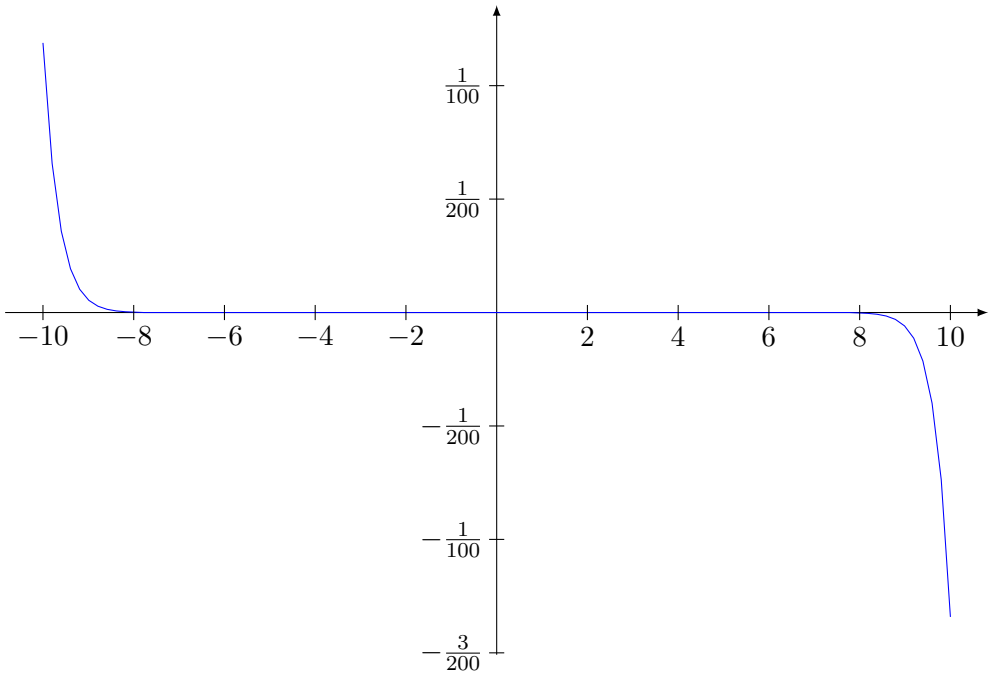
tex/sage/picard_exp.py

Ce programme fait 30 itérations depuis la fonction $\sin(x)$ pour tenter d'approximer $\exp(x)$. Pour donner une idée, après 7 itérations nous avons la fonction suivante :

$$\frac{1}{60}x^5 + \frac{1}{24}x^4 + \frac{1}{2}x^2 + 2x - \sin(x) + 1. \tag{16.181}$$

Nous voyons que les coefficients sont des factorielles, mais pas toujours celles correspondantes à la puissance, et qu'il manque certains termes par rapport au développement de l'exponentielle que nous connaissons. Bref, le polynôme qui se met en face de $\sin(x)$ s'adapte tout seul pour compenser.

Et après 30 itérations, ça donne quoi ? Voici un graphe de l'erreur entre $u_{30}(x)$ et $\exp(30)$:



Pour donner une idée, $\exp(10) \simeq 22000$. Donc il y a une faute de 0.01 sur 22000. Pas mal.

△

16.7 Théorèmes d'inversion locale et de la fonction implicite

16.7.1 Mise en situation

Dans un certain nombre de situation, il n'est pas possible de trouver des solutions explicites aux équations qui apparaissent. Néanmoins, l'existence « théorique » d'une telle solution est souvent déjà suffisante. C'est l'objet du théorème de la fonction implicite.

Prenons par exemple la fonction sur $\mathbb{R}^2$ donnée par

$$F(x, y) = x^2 + y^2 - 1. \tag{16.182}$$

Nous pouvons bien entendu regarder l'ensemble des points donnés par $F(x,y)=0$. C'est le cercle dessiné à la figure 16.1.

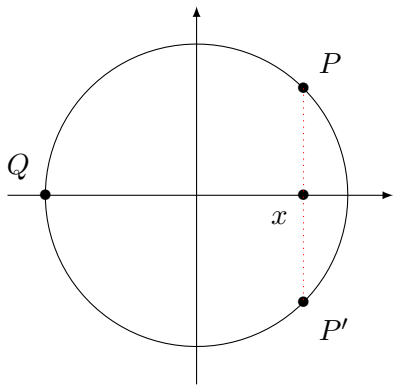


FIGURE 16.1 – Un cercle pour montrer l'intérêt de la fonction implicite. Si on donne x, nous ne pouvons pas savoir si nous parlons de P ou de P'.

Nous ne pouvons pas donner le cercle sous la forme $y=y(x)$ à cause du $\pm$ qui arrive quand on prend la racine carrée. Mais si on se donne le point P, nous pouvons dire que *autour de* P, le cercle est la fonction

$$y(x)=\sqrt{1-x^2}. \tag{16.183}$$

Tandis que autour du point P', le cercle est la fonction

$$y(x)=-\sqrt{1-x^2}. \tag{16.184}$$

Autour de ces deux points, donc, le cercle est donné par une fonction. Il n'est par contre pas possible de donner le cercle autour du point Q sous la forme d'une fonction.

Ce que nous voulons faire, en général, est de voir si l'ensemble des points tels que

$$F(x_1,\dots,x_n,y)=0 \tag{16.185}$$

peut être donné par une fonction $y=y(x_1,\dots,x_n)$. En d'autre termes, est-ce qu'il existe une fonction $y(x_1,\dots,x_n)$ telle que

$$F\big(x_1,\dots,x_n,y(x_1,\dots,x_n)\big)=0. \tag{16.186}$$

Plus généralement, soit une fonction

$$\begin{aligned} F\colon D\subset\mathbb{R}^n\times\mathbb{R}^m &\to \mathbb{R}^m\\ (x,y)&\mapsto \big(F_1(x,y),\dots,F_m(x,y)\big) \end{aligned} \tag{16.187}$$

avec $x=(x_1,\dots,x_n)$ et $y=(y_1,\dots,y_m)$. Pour chaque x fixé, on s'intéresse aux solutions du système de m équations $F(x,y)=0$ pour les inconnues y ; en particulier, on voudrait pouvoir écrire $y=\varphi(x)$ vérifiant $F(x,\varphi(x))=0$.

16.7.2 Théorème d'inversion locale

Lemme 16.46 ([248]).
Soit E un espace de Banach (métrique complet) et $\mathcal{O}$ un ouvert de E. Nous considérons une λ-contraction $\varphi\colon\mathcal{O}\to E$. Alors l'application

$$f\colon x\mapsto x+\varphi(x) \tag{16.188}$$

est un homéomorphisme entre $\mathcal{O}$ et un ouvert de E. De plus f^{-1} est Lipschitz de constante plus petite ou égale à $(1-\lambda)^{-1}$.

Cette proposition utilise le théorème de point fixe de Picard 16.34, et sera utilisée pour démontrer le théorème d'inversion locale 16.48.

Démonstration. Soient $x_1, x_2 \in \mathcal{O}$. Nous posons $y_1 = f(x_1)$ et $y_2 = f(x_2)$. En vertu de l'inégalité de la proposition 7.114 nous avons

$$\begin{aligned}\|f(x_2) - f(x_1)\| &= \|x_2 + \varphi(x_2) - x_1 - \varphi(x_1)\| && (16.189a)\\ &\geqslant \Big|\|x_2 - x_1\| - \|\varphi(x_2) - \varphi(x_1)\|\Big| && (16.189b)\\ &\geqslant (1-\lambda)\|x_2 - x_1\|. && (16.189c)\end{aligned}$$

À la dernière ligne les valeurs absolues sont enlevées parce que nous savons que ce qui est à l'intérieur est positif. Cela nous dit d'abord que f est injective parce que $f(x_2) = f(x_1)$ implique $x_2 = x_1$. Donc f est inversible sur son image. Nous posons $A = f(\mathcal{O})$ et nous devons prouver que que $f^{-1}\colon A \to \mathcal{O}$ est continue, Lipschitz de constante majorée par $(1-\lambda)^{-1}$ et que A est ouvert.

Les inéquations (16.189) nous disent que

$$\|f^{-1}(y_1) - f^{-1}(y_2)\| \leqslant \frac{\|y_1 - y_2\|}{1-\lambda}, \tag{16.190}$$

c'est-à-dire que

$$f^{-1}\big(B(y,r)\big) \subset B\big(f^{-1}(y), \frac{r}{1-\lambda}\big), \tag{16.191}$$

ce qui signifie que f^{-1} est Lipschitz de constante souhaitée et donc continue.

Il reste à prouver que $f(\mathcal{O})$ est ouvert. Pour cela nous prenons $y_0 = f(x_0)$ dans $f(\mathcal{O})$ est nous prouvons qu'il existe ϵ tel que $B(y_0, \epsilon)$ soit dans $f(\mathcal{O})$. Il faut donc que pour tout $y \in B(y_0, \epsilon)$, l'équation $f(x) = y$ ait une solution. Nous considérons l'application

$$L_y\colon x \mapsto y - \varphi(x). \tag{16.192}$$

Ce que nous cherchons est un point fixe de L_y parce que si $L_y(x) = x$ alors $y = x + \varphi(x) = f(x)$. Vu que

$$\|L_y(x) - L_y(x')\| = \|\varphi(x) - \varphi(x')\| \leqslant \lambda\|x - x'\|, \tag{16.193}$$

l'application L_y est une contraction de constante λ. Par ailleurs x_0 est un point fixe de L_{y_0}, donc en vertu de la caractérisation (11.655) des fonctions Lipschitziennes,

$$L_{y_0}\big(\overline{B(x_0,\delta)}\big) \subset \overline{B\big(L_{y_0}(x_0), \lambda\delta\big)} = \overline{B(x_0, \lambda\delta)}. \tag{16.194}$$

Vu que pour tout y et x nous avons $L_y(x) = L_{y_0}(x) + y - y_0$,

$$L_y\big(\overline{B(x_0,\delta)}\big) = L_{y_0}\big(\overline{B(x_0,\delta)}\big) + (y - y_0) \subset \overline{B(x_0,\lambda\delta)} + (y - y_0) \subset \overline{B(x_0), \lambda\delta + \|y - y_0\|}. \tag{16.195}$$

Si $\epsilon < (1-\lambda)\delta$ alors $\lambda\delta + \|y - y_0\| < \delta$. Un tel choix de $\epsilon > 0$ est possible parce que $\lambda < 1$. Pour une telle valeur de ϵ nous avons

$$L_y\big(\overline{B(x_0,\delta)}\big) \subset \overline{B(x_0,\delta)}. \tag{16.196}$$

Par conséquent L_y est une contraction sur l'espace métrique complet $\overline{B(x_0,\delta)}$, ce qui signifie que L_y y possède un point fixe par le théorème de Picard 16.34. □

Le théorème d'inversion locale s'énonce de la façon suivante dans $\mathbb{R}^n$:

Théorème 16.47 (Inversion locale dans $\mathbb{R}^n$)**.**
Soit $f \in C^k(\mathbb{R}^n, \mathbb{R}^n)$ et $x_0 \in \mathbb{R}^n$. Si df_{x_0} est inversible, alors il existe un voisinage ouvert U de x_0 et V de $f(x_0)$ tels que $f\colon U \to V$ soit un C^k-difféomorphisme. (c'est-à-dire que f^{-1} est également de classe C^k)

Nous allons le démontrer dans le cas un peu plus général (mais pas plus cher [22]) des espaces de Banach en tant que conséquence du théorème de point fixe de Picard 16.34.

Théorème 16.48 (Inversion locale dans un espace de Banach[325, 248])**.**
Soit une fonction $f \in C^p(E,F)$ avec $p \geqslant 1$ entre deux espaces de Banach. Soit $x_0 \in E$ tel que df_{x_0} soit une bijection bicontinue [23]. Alors il existe un voisinage ouvert V de x_0 et W de $f(x_0)$ tels que

(1) $f\colon V \to W$ soit une bijection,

(2) $f^{-1}\colon W \to V$ soit de classe C^p.

Démonstration. Nous commençons par simplifier un peu le problème. Pour cela, nous considérons la translation $T\colon x \mapsto x + x_0$ et l'application linéaire

$$\begin{aligned} L\colon \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto (df_{x_0})^{-1}x \end{aligned} \tag{16.197}$$

qui sont tout deux des difféomorphismes (L en est un par hypothèse d'inversibilité). Quitte à travailler avec la fonction $k = L \circ f \circ T$, nous pouvons supposer que $x_0 = 0$ et que $df_{x_0} = \mathbb{1}$. Pour comprendre cela il faut utiliser deux fois la formule de différentielle de fonction composée de la proposition 11.559 :

$$dk_0(u) = dL_{(f\circ T)(0)}\Big(df_{T(0)}dT_0(u)\Big). \tag{16.198}$$

Vu que L est linéaire, sa différentielle est elle-même, c'est-à-dire $dL_{(f\circ T)(0)} = (df_{x_0})^{-1}$, et par ailleurs $dT_0 = \mathbb{1}$, donc

$$dk_0(u) = (df_{x_0})^{-1}\Big(df_{x_0}(u)\Big) = u, \tag{16.199}$$

ce qui signifie bien que $dk_0 = \mathbb{1}$. Pour tout cela nous avons utilisé en plein le fait que df_{x_0} était inversible.

Nous posons $g = f - \mathbb{1}$, c'est-à-dire $g(x) = f(x) - x$, qui a la propriété $dg_0 = 0$. Étant donné que g est de classe C^1, l'application [24]

$$\begin{aligned} dg\colon E &\to \mathrm{GL}(F) \\ x &\mapsto dg_x \end{aligned} \tag{16.200}$$

est continue. En conséquence de quoi nous avons un voisinage U' de 0 pour lequel

$$\sup_{x\in U'} \|dg_x\| < \frac{1}{2}. \tag{16.201}$$

Maintenant le théorème des accroissements finis 10.145 (11.257 pour la dimension finie) nous indique que pour tout $x, x' \in U'$ nous avons [25]

$$\|g(x') - g(x)\| \leqslant \sup_{a\in[x,x']} \|dg_a\| \cdot \|x - x'\| \leqslant \frac{1}{2}\|x - x'\|, \tag{16.202}$$

ce qui prouve que g est une contraction au moins sur l'ouvert U'. Nous allons aussi donner une idée de la façon dont f fonctionne : si $x_1, x_2 \in U'$ alors

$$\|x_1 - x_2\| = \|g(x_1) - f(x_1) - g(x_2) + f(x_2)\| \tag{16.203a}$$

$$\leqslant \|g(x_1) - g(x_2)\| + \|f(x_1) - f(x_2)\| \tag{16.203b}$$

$$\leqslant \frac{1}{2}\|x_1 - x_2\| + \|f(x_1) - f(x_2)\|, \tag{16.203c}$$

22. Sauf la justification de la régularité de l'application $A \mapsto A^{-1}$

23. En dimension finie, une application linéaire est toujours continue et d'inverse continu.

24. Ici $\mathrm{GL}(F)$ est l'ensemble des applications linéaires, inversibles et continues de F dans lui-même. Ce ne sont pas spécialement des matrices parce que nous n'avons pas d'hypothèses sur la dimension de F, finie ou non.

25. Ici nous supposons avoir choisi U' convexe afin que tous les $a \in [x, x']$ soient bien dans U' et donc soumis à l'inéquation (16.201), ce qui est toujours possible, il suffit de prendre une boule.

ce qui montre que

$$\|x_1 - x_2\| \leqslant 2\|f(x_1) - f(x_2)\|. \tag{16.204}$$

Maintenant que nous savons que g est contractante de constante $\frac{1}{2}$ et que $f = g + \mathbb{1}$ nous pouvons utiliser la proposition 16.46 pour conclure que f est un homéomorphisme sur un ouvert U (partie de U') de E et f^{-1} a une constante de Lipschitz plus petite ou égale à $(1 - \frac{1}{2})^{-1} = 2$.

Nous allons maintenant prouver que f^{-1} est différentiable et que sa différentielle est donnée par $(df^{-1})_{f(x)} = (df_x)^{-1}$.

Soient $a, b \in U$ et $u = b - a$. Étant donné que f est différentiable en a, il existe une fonction $\alpha \in o(\|u\|)$ telle que

$$f(b) - f(a) - df_a(u) = \alpha(u). \tag{16.205}$$

En notant $y_a = f(a)$ et $y_b = f(b)$ et en appliquant $(df_a)^{-1}$ à cette dernière équation,

$$(df_a)^{-1}(y_b - y_a) - u = (df_a)^{-1}\big(\alpha(u)\big). \tag{16.206}$$

Vu que df_a est bornée (et son inverse aussi), le membre de droite est encore une fonction β ayant la propriété $\lim_{u\to 0} \beta(u)/\|u\| = 0$; en réordonnant les termes,

$$b - a = (df_a)^{-1}(y_b - y_a) + \beta(u) \tag{16.207}$$

et donc

$$f^{-1}(y_b) - f^{-1}(y_a) - (df_a)^{-1}(y_b - y_a) = \beta(u), \tag{16.208}$$

ce qui prouve que f^{-1} est différentiable et que $(df^{-1})_{y_a} = (df_a)^{-1}$.

La différentielle df^{-1} est donc obtenue par la chaine

$$df^{-1}\colon f(U) \xrightarrow{f^{-1}} U' \xrightarrow{df} \mathrm{GL}(F) \xrightarrow{\mathrm{Inv}} \mathrm{GL}(F) \tag{16.209}$$

où l'application $\mathrm{Inv}\colon \mathrm{GL}(F) \to \mathrm{GL}(F)$ est l'application $X \mapsto X^{-1}$ qui est de classe C^∞ par le théorème 10.153. D'autre part, par hypothèse df est une application de classe C^{k-1} et donc au minimum C^0 parce que $k \geqslant 1$. Enfin, l'application $f^{-1}\colon f(U) \to U$ est continue (parce que la proposition 16.46 précise que f est un homéomorphisme). Donc toute la chaine est continue et df^{-1} est continue. Cela entraine immédiatement que f^{-1} est C^1 et donc que toute la chaine est C^1.

Par récurrence nous obtenons la chaine

$$df^{-1}\colon f(U) \xrightarrow[C^{k-1}]{f^{-1}} U' \xrightarrow[C^{k-1}]{df} \mathrm{GL}(F) \xrightarrow[C^\infty]{\mathrm{Inv}} \mathrm{GL}(F) \tag{16.210}$$

qui prouve que df^{-1} est C^{k-1} et donc que f^{-1} est C^k. La récurrence s'arrête ici parce que df n'est pas mieux que C^{k-1}. □

16.7.3 Théorème de la fonction implicite

Nous énonçons et le démontrons le théorème de la fonction implicite dans le cas d'espaces de Banach.

Théorème 16.49 (Théorème de la fonction implicite dans Banach[226])**.**
Soient E, F et G des espaces de Banach et des ouverts $U \subset E$, $V \subset F$. Nous considérons une fonction $f\colon U \times V \to G$ de classe C^r telle que[26]

$$d_y f_{(x_0,y_0)}\colon F \to G \tag{16.211}$$

soit un isomorphisme pour un certain $(x_0, y_0) \in U \times V$.

26. La notation d_y est la différentielle partielle de la définition 10.150.

Alors nous avons des voisinages U_0 de x_0 dans E et W_0 de $f(x_0,y_0)$ dans G et une fonction de classe C^r

$$g\colon U_0\times W_0\to V \tag{16.212}$$

telle que

$$f\big(x,g(x,w)\big)=w \tag{16.213}$$

pour tout $(x,w)\in U_0\times W_0$.

Cette fonction g est unique au sens suivant : il existe un voisinage V_0 de y_0 tel que si $(x,y)\in U_0\times V_0$ et $w\in W_0$ satisfont à $f(x,y)=w$ alors $y=g(x,w)$. Autrement dit, la fonction $g\colon U_0\times W_0\to V_0$ est unique.

Démonstration. Nous commençons par considérer la fonction

$$\begin{aligned}\Phi\colon U\times V&\to E\times G\\ (x,y)&\mapsto \big(x,f(x,y)\big)\end{aligned} \tag{16.214}$$

et sa différentielle

$$\begin{aligned}d\Phi_{(x_0,y_0)}(u,v)&=\frac{d}{dt}\Big[\big(x_0+tu,f(x_0+tu,y_0+tv)\big)\Big]_{t=0} &(16.215a)\\ &=\left(\frac{d}{dt}\Big[x_0+tu\Big]_{t=0},\frac{d}{dt}\Big[f(x_0+tu,y_0+tv)\Big]_{t=0}\right) &(16.215b)\\ &=\big(u,df_{(x_0,y_0)}(u,v)\big)\,. &(16.215c)\end{aligned}$$

Nous utilisons alors la proposition 10.151 pour conclure que

$$d\Phi_{(x_0,y_0)}(u,v)=\big(u,(d_1f)_{(x_0,y_0)}(u)+(d_2f)_{(x_0,y_0)}(v)\big), \tag{16.216}$$

mais comme par hypothèse $(d_2f)_{(x_0,y_0)}\colon F\to G$ est un isomorphisme, l'application $d\Phi_{(x_0,y_0)}\colon E\times F\to E\times G$ est également un isomorphisme. Par conséquent le théorème d'inversion locale 16.48 nous indique qu'il existe un voisinage $\mathcal{O}$ de (x_0,y_0) et $\mathcal{P}$ de $\Phi(x_0,y_0)$ tels que $\Phi\colon\mathcal{O}\to\mathcal{P}$ soit une bijection et $\Phi^{-1}\colon\mathcal{P}\to\mathcal{O}$ soit de classe C^r. Vu que $\mathcal{P}$ est un voisinage de

$$\Phi(x_0,y_0)=\big(x_0,f(x_0,y_0)\big), \tag{16.217}$$

nous pouvons par 10.37 le choisir un peu plus petit de telle sorte à avoir $\mathcal{P}=U_0\times W_0$ où U_0 est un voisinage de x_0 et W_0 un voisinage de $f(x_0,y_0)$. Dans ce cas nous devons obligatoirement aussi restreindre $\mathcal{O}$ à $U_0\times V_0$ pour un certain voisinage V_0 de y_0. L'application Φ^{-1} a obligatoirement la forme

$$\begin{aligned}\Phi^{-1}\colon U_0\times W_0&\to U_0\times V_0\\ (x,w)&\mapsto\big(x,g(x,w)\big)\end{aligned} \tag{16.218}$$

pour une certaine fonction $g\colon U_0\times W_0\to V$. Cette fonction g est la fonction cherchée parce qu'en appliquant Φ à (16.218),

$$(x,w)=\Phi\big(x,g(x,w)\big)=\Big(x,f\big(x,g(x,w)\big)\Big), \tag{16.219}$$

qui nous dit que pour tout $x\in U_0$ et tout $w\in W_0$ nous avons

$$f\big(x,g(x,w)\big)=w. \tag{16.220}$$

Si vous avez bien suivi le sens de l'équation (16.218) alors vous avez compris l'unicité. Sinon, considérez $(x,y)\in U_0\times V_0$ et $w\in W_0$ tels que $f(x,y)=w$. Alors $\big(x,f(x,y)\big)=(x,w)$ et

$$\Phi(x,y)=(x,w). \tag{16.221}$$

Mais vu que $\Phi\colon U_0\times V_0\to U_0\times W_0$ est une bijection, cette relation définit de façon univoque l'élément (x,y) de $U_0\times V_0$, qui ne sera autre que $g(x,w)$. □

Le théorème de la fonction implicite s'énonce de la façon suivante pour des espaces de dimension finie.

Théorème 16.50 (Théorème de la fonction implicite en dimension finie)**.**
Soit une fonction $F\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ de classe C^k et $(\alpha,\beta) \in \mathbb{R}^n \times \mathbb{R}^m$ tels que

(1) $F(\alpha,\beta) = 0$,

(2) $\frac{\partial(F_1,\dots,F_m)}{\partial(y_1,\dots,y_m)} \neq 0$, *c'est-à-dire que* $(d_yF)_{(\alpha,\beta)}$ *est inversible.*

Alors il existe un voisinage ouvert V de α dans $\mathbb{R}^n$, un voisinage ouvert W de β dans $\mathbb{R}^m$ et une application $\varphi\colon V \to W$ de classe C^k telle que pour tout $x \in V$ on ait

$$F\big(x,\varphi(x)\big) = 0. \tag{16.222}$$

De plus si $(x,y) \in V \times W$ satisfait à $F(x,y) = 0$, alors $y = \varphi(x)$.

Remarque 16.51.
Notons que cet énoncé est tourné un peu différemment en ce qui concerne le nombre de variables dont dépend la fonction implicite : comparez

$$f\big(x,g(x,w)\big) = w \tag{16.223a}$$
$$F\big(x,\varphi(x)\big) = 0. \tag{16.223b}$$

Le deuxième est un cas particulier du premier en posant

$$F(x,y) = f(x,y) - f(x_0,y_0) \tag{16.224}$$

et donc en considérant w comme valant la constante $f(x_0,y_0)$; dans ce cas la fonction g ne dépend plus que de la variable x.

Exemple 16.52.
La remarque 16.51 signifie entre autres que le théorème 16.49 est plus fort que 16.50 parce que le premier permet de choisir la valeur d'arrivée. Parlons de l'exemple classique du cercle et de la fonction $f(x,y) = x^2 + y^2$. Nous savons que

$$f(\alpha,\beta) = 1. \tag{16.225}$$

Alors le théorème 16.49 nous donne une fonction g telle que

$$f(x,g(x,r)) = r \tag{16.226}$$

tant que x est proche de α, que r est proche de 1 et que g donne des valeurs proches de β.

L'énoncé 16.50 nous oblige à travailler avec la fonction $F(x,y) = x^2 + y^2 - 1$, de telle sorte que

$$F(\alpha,\beta) = 0, \tag{16.227}$$

et que nous ayons une fonction φ telle que

$$F(x,\varphi(x)) = 0. \tag{16.228}$$

La fonction φ ne permet donc que de trouver des points sur le cercle de rayon 1. △

16.7.4 Exemple

Le théorème de la fonction implicite a pour objet de donner l'existence de la fonction φ. Maintenant nous pouvons dire beaucoup de choses sur les dérivées de φ en considérant la fonction

$$x \mapsto F\big(x,\varphi(x)\big). \tag{16.229}$$

Par définition de φ, cette fonction est toujours nulle. En particulier, nous pouvons dériver l'équation

$$F\big(x,\varphi(x)\big) = 0, \tag{16.230}$$

et nous trouvons plein de choses.

Exemple 16.53.
Prenons par exemple la fonction

$$F\big((x,y),z\big) = ze^z - x - y, \tag{16.231}$$

et demandons nous ce que nous pouvons dire sur la fonction $z(x,y)$ telle que

$$F\big(x,y,z(x,y)\big) = 0, \tag{16.232}$$

c'est-à-dire telle que

$$z(x,y)e^{z(x,y)} - x - y = 0. \tag{16.233}$$

pour tout x et $y \in \mathbb{R}$. Nous pouvons facilement trouver $z(0,0)$ parce que

$$z(0,0)e^{z(0,0)} = 0, \tag{16.234}$$

donc $z(0,0) = 0$.

Nous pouvons dire des choses sur les dérivées de $z(x,y)$. Voyons par exemple $(\partial_x z)(x,y)$. Pour trouver cette dérivée, nous dérivons la relation (16.233) par rapport à x. Ce que nous trouvons est

$$(\partial_x z)e^z + ze^z(\partial_x z) - 1 = 0. \tag{16.235}$$

Cette équation peut être résolue par rapport à $\partial_x z$:

$$\frac{\partial z}{\partial x}(x,y) = \frac{1}{e^z(1+z)}. \tag{16.236}$$

Remarquez que cette équation ne donne pas tout à fait la dérivée de z en fonction de x et y, parce que z apparaît dans l'expression, alors que z est justement la fonction inconnue. En général, c'est la vie, nous ne pouvons pas faire mieux.

Dans certains cas, on peut aller plus loin. Par exemple, nous pouvons calculer cette dérivée au point $(x,y) = (0,0)$ parce que $z(0,0)$ est connu :

$$\frac{\partial z}{\partial x}(0,0) = 1. \tag{16.237}$$

Cela est pratique pour calculer, par exemple, le développement en Taylor de z autour de $(0,0)$. △

Exemple 16.54.
Est-ce que l'équation $e^y + xy = 0$ définit au moins localement une fonction $y(x)$? Nous considérons la fonction

$$f(x,y) = \begin{pmatrix} x \\ e^y + xy \end{pmatrix} \tag{16.238}$$

La différentielle de cette application est

$$df_{(0,0)}(u) = \frac{d}{dt}\Big[f(tu_1,tu_2)\Big]_{t=0} = \frac{d}{dt}\begin{pmatrix} tu_1 \\ e^{tu_2} + t^2u_1u_2 \end{pmatrix}_{t=0} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}. \tag{16.239}$$

L'application f définit donc un difféomorphisme local autour des points (x_0,y_0) et $f(x_0,y_0)$. Soit $(u,0)$ un point dans le voisinage de $f(x_0,y_0)$. Alors il existe un unique (x,y) tel que

$$f(x,y) = \begin{pmatrix} x \\ e^y + xy \end{pmatrix} = \begin{pmatrix} u \\ 0 \end{pmatrix}. \tag{16.240}$$

Nous avons automatiquement $x = u$ et $e^y + xy = 0$. Notons toutefois que pour que ce procédé donne effectivement une fonction implicite $y(x)$ nous devons avoir des points de la forme $(u,0)$ dans le voisinage de $f(x_0,y_0)$. △

16.8 Décomposition polaire (régularité)

16.55.
Nous allons montrer que l'application

$$\begin{aligned} f\colon S^{++}(n,\mathbb{R}) &\to S^{++}(n,\mathbb{R}) \\ A &\mapsto \sqrt{A} \end{aligned} \tag{16.241}$$

est une difféomorphisme.

Cependant $S^{++}(n,\mathbb{R})$ n'est pas un ouvert de $\mathbb{M}(n,\mathbb{R})$ et nous ne savons pas ce qu'est la différentielle d'une application non définie sur un ouvert. Nous allons donc en réalité montrer que l'application racine carrée existe sur un voisinage de chacun des points de $S^{++}(n,\mathbb{R})$. Et comme une union quelconque d'ouverts est un ouvert, la fonction f sera bien définie sur un ouvert de $\mathbb{M}(n,\mathbb{R})$.

Lemme 16.56.
L'application

$$\begin{aligned} f\colon S^{++}(n,\mathbb{R}) &\to S^{++}(n,\mathbb{R}) \\ A &\mapsto A^2 \end{aligned} \tag{16.242}$$

est un C^∞-difféomorphisme.

Démonstration. Prouvons d'abord que f prend ses valeurs dans $S^{++}(n,\mathbb{R})$. Si $A \in S^{++}(n,\mathbb{R})$ alors par la diagonalisation 9.150 elle s'écrit $A = QDQ^{-1}$ où D est diagonale avec des nombres strictement positifs sur la diagonale. Avec cela, $A^2 = QD^2Q^{-1}$ où D^2 contient encore des nombres strictement positifs sur la diagonale.

L'application f étant essentiellement des polynômes en les entrées de A, elle est de classe C^∞.

Passons à l'étude de la différentielle. Comme mentionné en 16.55 nous allons en réalité voir f sur un ouvert de $\mathbb{M}(n,\mathbb{R})$ autour de $A \in S^{++}(n,\mathbb{R})$. Par conséquent si $A \in S^{++}(n,\mathbb{R})$,

$$df\colon S^{++}(n,\mathbb{R}) \to \mathcal{L}\big(\mathbb{M}(n,\mathbb{R}),\mathbb{M}(n,\mathbb{R})\big) \tag{16.243a}$$
$$df_A\colon \mathbb{M}(n,\mathbb{R}) \to \mathbb{M}(n,\mathbb{R}). \tag{16.243b}$$

Le calcul de df_A est facile. Soit $u \in \mathbb{M}(n,\mathbb{R})$ et faisons le calcul en utilisant la formule du lemme (11.196) :

$$df_A(u) = \frac{d}{dt}\Big[f(A+tu)\Big]_{t=0} \tag{16.244a}$$
$$= \frac{d}{dt}\Big[A^2 + tAu + tuA + t^2u^2\Big]_{t=0} \tag{16.244b}$$
$$= Au + uA. \tag{16.244c}$$

Nous allons utiliser le théorème d'inversion locale 16.48 à la fonction f. Dans la suite, A est une matrice de $S^{++}(n,\mathbb{R})$.

df_A est injective Soit $M \in \mathbb{M}(n,\mathbb{R})$ dans le noyau de df_A. En posant $M' = A^{-1}MQ$ nous avons $M = QM'Q^{-1}$ et on applique df_A à $QM'Q^{-1}$:

$$df_A(QM'Q^{-1}) = Q\big(DM + MD\big)Q^{-1}. \tag{16.245}$$

où $D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$ avec $\lambda_i > 0$. La matrice D est inversible. Nous avons $M' =$

$-DM'D^{-1}$, et en coordonnées,

$$M'_{ij} = -\sum_{kl} D_{ikM'_{kl}} D^{-1}_{lj} \tag{16.246a}$$

$$= -\sum_{kl} \lambda_i \delta_{ik} M'_{kl} \frac{1}{\lambda_j} \delta_{lj} \tag{16.246b}$$

$$= -\frac{\lambda_i}{\lambda_i} M'_{ij}. \tag{16.246c}$$

C'est-à-dire que $M'_{ij} = -\frac{\lambda_i}{\lambda_j} M'_{ij}$ avec $-\frac{\lambda_i}{\lambda_j} < 0$. Cela implique $M' = 0$ et par conséquent $M = 0$.

df_A est surjective Soit $N \in \mathbb{M}(n, \mathbb{R})$; nous cherchons $M \in \mathbb{M}(n, \mathbb{R})$ tel que $df_A(M) = N$. Nous posons $N' = Q^{-1}NQ$ et $M = QM'Q^{-1}$, ce qui nous donne à résoudre $df_D(M') = N'$. Passons en coordonnées :

$$(DM' + M'D)_{ij} = \sum_k (\delta_{ik}\lambda_i M'_{kj} + M'_{ik}\delta_{kj}\lambda_j) = M'_{ij}(\lambda_i + \lambda_j) \tag{16.247}$$

où $\lambda_i + \lambda_j \neq 0$. Il suffit donc de prendre la matrice M' donnée par

$$M'_{ij} = \frac{1}{\lambda_i + \lambda_j} N'_{ij} \tag{16.248}$$

pour que $df_A(M') = N'$.

Le théorème d'inversion locale donne un voisinage V de A dans $\mathbb{M}(n, \mathbb{R})$ et un voisinage W de A^2 dans $\mathbb{M}(n, \mathbb{R})$ tels que $f\colon V \to W$ soit une bijection et $f^{-1}\colon W \to V$ soit de même régularité, en l'occurrence C^∞. □

Remarque 16.57.
Oui, il y a des matrices non symétriques qui ont une unique racine carrée.

La proposition suivante, qui dépend du le théorème d'inversion locale par le lemme 16.56, donne plus de régularité à la décomposition polaire donnée dans le théorème 12.30.

Proposition 16.58 (Décomposition polaire : cas réel (suite))**.**
L'application

$$\begin{aligned} f\colon \mathrm{O}(n, \mathbb{R}) \times S^{++}(n, \mathbb{R}) &\to \mathrm{GL}(n, \mathbb{R}) \\ (Q, S) &\mapsto SQ \end{aligned} \tag{16.249}$$

est un difféomorphisme de classe C^∞.

Démonstration. Si M est donnée dans $\mathrm{GL}(n, \mathbb{R})$ alors la décomposition polaire[27] $M = QS$ est donnée par $S = \sqrt{MM^t}$ et $Q = MS^{-1}$. Autrement dit, si nous considérons la fonction de décomposition polaire

$$f\colon \mathrm{O}(n, \mathbb{R}) \times S^{++}(n, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R}) \tag{16.250}$$

alors

$$f^{-1}(M) = \big(M(\sqrt{MM^t})^{-1}, \sqrt{MM^t}\big). \tag{16.251}$$

Nous avons vu dans le lemme 16.56 que la racine carrée était un C^∞-difféomorphisme. Le reste n'étant que des produits de matrices, la régularité est de mise. □

27. Proposition 12.30.

16.9 Théorème de Von Neumann

Lemme 16.59 ([54]).
Soit G, un sous-groupe fermé de $\mathrm{GL}(n,\mathbb{R})$ et

$$\mathcal{L}_G = \{m \in \mathbb{M}(n,\mathbb{R}) \text{ tel que } e^{tm} \in G \,\forall t \in \mathbb{R}\}. \tag{16.252}$$

Alors $\mathcal{L}_G$ est un sous-espace vectoriel de $\mathbb{M}(n,\mathbb{R})$.

Démonstration. Si $m \in \mathcal{L}_G$, alors $\lambda m \in \mathcal{L}_G$ par construction. Le point délicat à prouver est le fait que si $a, b \in \mathcal{L}_G$, alors $a + b \in \mathcal{L}_G$. Soit $a \in \mathbb{M}(n,\mathbb{R})$; nous savons qu'il existe une fonction $\alpha_a \colon \mathbb{R} \to \mathbb{M}$ telle que

$$e^{ta} = \mathbb{1} + ta + \alpha_a(t) \tag{16.253}$$

et

$$\lim_{t\to 0} \frac{\alpha(t)}{t} = 0. \tag{16.254}$$

Si a et b sont dans $\mathcal{L}_G$, alors $e^{ta}e^{tb} \in G$, mais il n'est pas vrai en général que cela soit égal à $e^{t(a+b)}$. Pour tout $k \in \mathbb{N}$ nous avons

$$e^{a/k}e^{b/k} = \left(\mathbb{1} + \frac{a}{k} + \alpha_a(\frac{1}{k})\right)\left(\mathbb{1} + \frac{b}{k} + \alpha_b(\frac{1}{k})\right) = \mathbb{1} + \frac{a+b}{2} + \beta\left(\frac{1}{k}\right) \tag{16.255}$$

où $\beta \colon \mathbb{R} \to \mathbb{M}$ est encore une fonction vérifiant $\beta(t)/t \to 0$. Si k est assez grand, nous avons

$$\left\| \frac{a+b}{k} + \beta(\frac{1}{k}) \right\| < 1, \tag{16.256}$$

et nous pouvons profiter du lemme 14.117 pour écrire alors

$$\left(e^{a/k}e^{b/k}\right)^k = e^{k\ln\left(\mathbb{1} + \frac{a+b}{k} + \beta(\frac{1}{k})\right)}. \tag{16.257}$$

Ce qui se trouve dans l'exponentielle est

$$k\left[\frac{a+b}{k} + \alpha(\frac{1}{k}) + \sigma\left(\frac{a+b}{k} + \alpha(\frac{1}{k})\right)\right]. \tag{16.258}$$

Les diverses propriétés vues montrent que le tout tend vers $a + b$ lorsque $k \to \infty$. Par conséquent

$$\lim_{k\to\infty}\left(e^{a/k}e^{b/k}\right)^k = e^{a+b}. \tag{16.259}$$

Ce que nous avons prouvé est que pour tout t, $e^{t(a+b)}$ est une limite d'éléments dans G et est donc dans G parce que ce dernier est fermé. □

Vu que $\mathcal{L}_G$ est un sous-espace vectoriel de $\mathbb{M}(n,\mathbb{R})$, nous pouvons considérer un supplémentaire M.

Lemme 16.60.
Il n'existe pas se suites (m_k) dans $M\backslash\{0\}$ convergeant vers zéro et telle que $e^{m_k} \in G$ pour tout k.

Démonstration. Supposons que nous ayons $m_k \to 0$ dans $M\backslash\{0\}$ avec $e^{m_k} \in G$. Nous considérons les éléments $\epsilon_k = \frac{m_k}{\|m_k\|}$ qui sont sur la sphère unité de $\mathrm{GL}(n,\mathbb{R})$. Quitte à prendre une sous-suite, nous pouvons supposer que cette suite converge, et vu que M est fermé, ce sera vers $\epsilon \in M$ avec $\|\epsilon\| = 1$. Pour tout $t \in \mathbb{R}$ nous avons

$$e^{t\epsilon} = \lim_{k\to\infty} e^{t\epsilon_k}. \tag{16.260}$$

En vertu de la décomposition d'un réel en partie entière et décimale, pour tout k nous avons $\lambda_k \in \mathbb{Z}$ et $|\mu_k| \leqslant \frac{1}{2}$ tel que $t/\|m_k\| = \lambda_k + \mu_k$. Avec ça,

$$e^{t\epsilon} = \lim_{k\to\infty} \exp\left(\frac{t}{m_k}m_k\right) = \lim_{k\to\infty} e^{\lambda_k m_k}e^{\mu_k m_k}. \tag{16.261}$$

Pour tout k nous avons $e^{\lambda_k m_k} \in G$. De plus $|\mu_k|$ étant borné et m_k tendant vers zéro nous avons $e^{\mu_k m_k} \to 1$. Au final

$$e^{t\epsilon} = \lim_{k\to\infty} e^{t\epsilon_k} \in G \tag{16.262}$$

Cela signifie que $\epsilon \in \mathcal{L}_G$, ce qui est impossible parce que nous avions déjà dit que $\epsilon \in M\backslash\{0\}$. □

Lemme 16.61.
L'application

$$\begin{aligned} f\colon \mathcal{L}_G \times M &\to \mathrm{GL}(n,\mathbb{R}) \\ l, m &\mapsto e^l e^m \end{aligned} \tag{16.263}$$

est un difféomorphisme local entre un voisinage de $(0,0)$ *dans* $\mathbb{M}(n,\mathbb{R})$ *et un voisinage de* $\mathbb{1}$ *dans* $\exp\big(\mathbb{M}(n,\mathbb{R})\big)$.

Notons que nous ne disons rien de $e^{\mathbb{M}(n,\mathbb{R})}$. Nous n'allons pas nous embarquer à discuter si ce serait tout $\mathrm{GL}(n,\mathbb{R})$ [28] ou bien si ça contiendrait ne fut-ce que G.

Démonstration. Le fait que f prenne ses valeurs dans $\mathrm{GL}(n,\mathbb{R})$ est simplement dû au fait que les exponentielles sont toujours inversibles. Nous considérons ensuite la différentielle : si $u \in \mathcal{L}_G$ et $v \in M$ nous avons

$$df_{(0,0)}(u,v) = \frac{d}{dt}\Big[f\big(t(u,v)\big)\Big]_{t=0} = \frac{d}{dt}\Big[e^{tu}e^{tv}\Big]_{t=0} = u + v. \tag{16.264}$$

L'application df_0 est donc une bijection entre $\mathcal{L}_G \times M$ et $\mathbb{M}(n,\mathbb{R})$. Le théorème d'inversion locale 16.48 nous assure alors que f est une bijection entre un voisinage de $(0,0)$ dans $\mathcal{L}_G \times M$ et son image. Mais vu que df_0 est une bijection avec $\mathbb{M}(n,\mathbb{R})$, l'image en question contient un ouvert autour de $\mathbb{1}$ dans $\exp\big(\mathbb{M}(n,\mathbb{R})\big)$. □

Théorème 16.62 (Von Neumann[54, 326, 327])**.**
Tout sous-groupe fermé de $\mathrm{GL}(n,\mathbb{R})$ *est une sous-variété de* $\mathrm{GL}(n,\mathbb{R})$.

Démonstration. Soit G un tel groupe ; nous devons prouver que c'est localement difféomorphe à un ouvert de $\mathbb{R}^n$. Et si on est pervers, on ne va pas faire localement difféomorphe à un ouvert de $\mathbb{R}^n$, mais à un ouvert d'un espace vectoriel de dimension finie. Nous allons être pervers.

Étant donné que pour tout $g \in G$, l'application

$$\begin{aligned} L_g\colon G &\to G \\ h &\mapsto gh \end{aligned} \tag{16.265}$$

est de classe C^∞ et d'inverse C^∞, il suffit de prouver le résultat pour un voisinage de $\mathbb{1}$.

Supposons d'abord que $\mathcal{L}_G = \{0\}$. Alors 0 est un point isolé de $\ln(G)$; en effet si ce n'était pas le cas nous aurions un élément m_k de $\ln(G)$ dans chaque boule $B(0,r_k)$. Nous aurions alors $m_k = \ln(a_k)$ avec $a_k \in G$ et donc

$$e^{m_k} = a_k \in G. \tag{16.266}$$

De plus m_k appartient forcément à M parce que $\mathcal{L}_G$ est réduit à zéro. Cela nous donnerait une suite $m_k \to 0$ dans M dont l'exponentielle reste dans G. Or cela est interdit par le lemme 16.60. Donc 0 est un point isolé de $\ln(G)$. L'application ln étant continue [29], nous en déduisons que $\mathbb{1}$ est isolé dans G. Par le difféomorphisme L_g, tous les points de G sont isolés ; ce groupe est donc discret et par voie de conséquence une variété.

Nous supposons maintenant que $\mathcal{L}_G \neq \{0\}$. Nous savons par la proposition 14.116 que

$$\exp\colon \mathbb{M}(n,\mathbb{R}) \to \mathbb{M}(n,\mathbb{R}) \tag{16.267}$$

28. Vu les dimensions y'a tout de même peu de chance.
29. Par le lemme 14.117.

est une application C^∞ vérifiant $d\exp_0 = \mathrm{Id}$. Nous pouvons donc utiliser le théorème d'inversion locale 16.48 qui nous offre donc l'existence d'un voisinage U de 0 dans $\mathbb{M}(n,\mathbb{R})$ tel que $W = \exp(U)$ soit un ouvert de $\mathrm{GL}(n,\mathbb{R})$ et que $\exp\colon U \to W$ soit un difféomorphisme de classe C^∞.

Montrons que quitte à restreindre U (et donc W qui reste par définition l'image de U par exp), nous pouvons avoir $\exp\big(U \cap \mathcal{L}_G\big) = W \cap G$. D'abord $\exp(\mathcal{L}_G) \subset G$ par construction. Nous avons donc $\exp\big(U \cap \mathcal{L}_G\big) \subset W \cap G$. Pour trouver une restriction de U pour laquelle nous avons l'égalité, nous supposons que pour tout ouvert $\mathcal{O}$ dans U,

$$\exp\colon \mathcal{O} \cap \mathcal{L}_G \to \exp(\mathcal{O}) \cap G \tag{16.268}$$

ne soit pas surjective. Cela donnerait un élément de $\mathcal{O} \cap \complement\mathcal{L}_G$ dont l'image par exp n'est pas dans G. Nous construisons ainsi une suite en considérant une boule $B(0,\frac{1}{k})$ inclue à U et $x_k \in B(0,\frac{1}{k}) \cap \complement\mathcal{L}_G$ vérifiant $e^{x_k} \in G$. Vu le choix des boules nous avons évidemment $x_k \to 0$.

L'élément e^{x_k} est dans $e^{\mathbb{M}(n,\mathbb{R})}$ et le difféomorphisme du lemme 16.61 [30] nous donne $(l_k, m_k) \in \mathcal{L}_G \times M$ tel que $e^{l_k}e^{m_k} = e^{x_k}$. À ce point nous considérons k suffisamment grand pour que e^{x_k} soit dans la partie de l'image de f sur lequel nous avons le difféomorphisme. Plus prosaïquement, nous posons

$$(l_k, m_k) = f^{-1}(e^{x_k}) \tag{16.269}$$

et nous profitons de la continuité pour permuter la limite avec f^{-1} :

$$\lim_{k\to\infty}(l_k, m_k) = f^{-1}\big(\lim_{k\to\infty} e^{x_k}\big) = f^{-1}(\mathbb{1}) = (0,0). \tag{16.270}$$

En particulier $m_k \to 0$ alors que $e^{m_k} = e^{x_k}e^{-l_k} \in G$. La suite m_k viole le lemme 16.60. Nous pouvons donc restreindre U de telle façon à avoir

$$\exp\big(U \cap \mathcal{L}_G\big) = W \cap G. \tag{16.271}$$

Nous avons donc un ouvert de $\mathcal{L}_G$ (l'ouvert $U \cap \mathcal{L}_G$) qui est difféomorphe avec l'ouvert $W \cap G$ de G. Donc G est une variété et accepte $\mathcal{L}_G$ comme carte locale.

□

Remarque 16.63.
En termes savants, nous avons surtout montré que si G est un groupe de Lie d'algèbre de Lie $\mathfrak{g}$, alors l'exponentielle donne un difféomorphisme local entre $\mathfrak{g}$ et G.

16.10 Recherche d'extrémums

16.10.1 Extrema à une variable

Définition 16.64.
Soit $f\colon A \subset \mathbb{R} \to \mathbb{R}$ et $a \in A$. Le point a est un ***maximum local*** *de f s'il existe un voisinage $\mathcal{U}$ de a tel que $f(a) \geqslant f(x)$ pour tout $x \in \mathcal{U} \cap A$. Le point a est un* ***maximum global*** *si $f(a) \geqslant g(x)$ pour tout $x \in A$.*

La proposition basique à utiliser lors de la recherche d'extrémums est la suivante :

Proposition 16.65.
Soit $f\colon A \subset \mathbb{R} \to \mathbb{R}$ et $a \in \mathrm{Int}(A)$. Supposons que f est dérivable en a. Si a est un extrémum local, alors $f'(a) = 0$.

La réciproque n'est pas vraie, comme le montre l'exemple de la fonction $x \mapsto x^3$ en $x = 0$: sa dérivée est nulle et pourtant $x = 0$ n'est ni un maximum ni un minimum local.

Cette proposition ne sert donc qu'à sélectionner des *candidats* extrémum. Afin de savoir si ces candidats sont des extrémums, il y a la proposition suivante.

30. Il me semble que l'utilisation de ce lemme manque à l'avant-dernière ligne de la preuve chez [54].

Proposition 16.66.
Soit $f\colon I \subset \mathbb{R} \to \mathbb{R}$, une fonction de classe C^k au voisinage d'un point $a \in \operatorname{Int} I$. Supposons que

$$f'(a) = f''(a) = \ldots = f^{(k-1)}(a) = 0, \tag{16.272}$$

et que

$$f^{(k)}(a) \neq 0. \tag{16.273}$$

Dans ce cas,

(1) Si k est pair, alors a est un point d'extrémum local de f, c'est un minimum si $f^{(k)}(a) > 0$, et un maximum si $f^{(k)}(a) < 0$,

(2) Si k est impair, alors a n'est pas un extrémum local de f.

Note : jusqu'à présent nous n'avons rien dit des extrémums *globaux* de f. Il n'y a pas grand chose à en dire. Si un point d'extrémum global est situé dans l'intérieur du domaine de f, alors il sera extrémum local (a fortiori). Ou alors, le maximum global peut être sur le bord du domaine. C'est ce qui arrive à des fonctions strictement croissantes sur un domaine compact.

Une seule certitude : si une fonction est continue sur un compact, elle possède une minimum et un maximum global par le théorème 7.154.

Soit une fonction $f\colon I \to \mathbb{R}$, et soit $a \in I$. Si $f'(a) > 0$, alors la tangente au graphe de f au point $\big(a, f(a)\big)$ sera une droite croissante (coefficient directeur positif). Cela ne veut pas spécialement dire que la fonction elle-même sera croissante, mais en tout cas cela est un bon indice.

Exemple 16.67.
Si $f(x) = x^2$, il est connu que $f'(x) = 2x$. Nous avons donc que f' est positive si $x \geqslant 0$ et $f' >$ est négative si $x < 0$. Cela correspond bien au fait que x^2 est décroissante sur $]-\infty, 0[$ et croissante sur $]0, \infty[$. △

Sur la figure 16.2, nous avons dessiné la fonction $f(x) = x\cos(x)$ et sa dérivée. Nous voyons que partout où la dérivée est négative, la fonction est décroissante tandis que, inversement, partout où la dérivée est positive, la fonction est croissante.

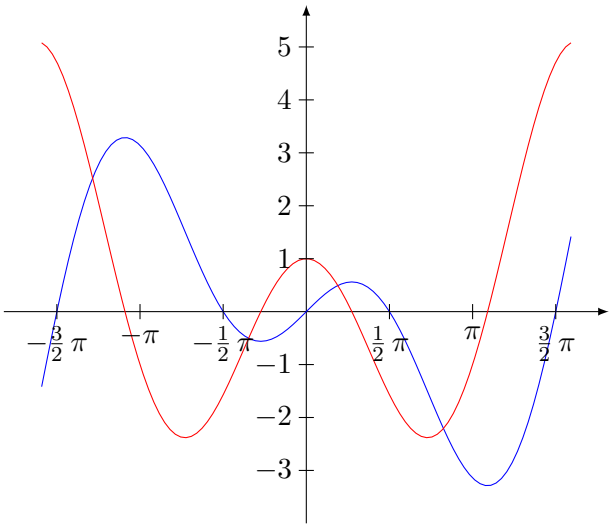


FIGURE 16.2 – La fonction $f(x) = x\cos(x)$ en bleu et sa dérivée en rouge.

Les extrémums de la fonction f sont donc placés là où f' change de signe. En effet si $f'(x) < 0$ pour $x < a$ et $f'(x) > 0$ pour $x > a$, la fonction est décroissante jusqu'à a et est ensuite croissante. Cela signifie que la fonction connait un creux en a. Le point a est donc un minimum de la fonction.

Attention cependant. Le fait que $f'(a) = 0$ ne signifie pas automatiquement que f a un maximum ou un minimum en a. Nous avons par exemple tracé sur la figure 16.3 les fonctions x^3 et sa dérivée. Il est à noter que, conformément à ce que l'on pense, certes la dérivée s'annule en $x = 0$, mais elle ne change pas de signe.

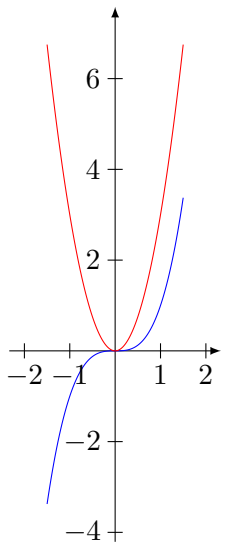


FIGURE 16.3 – La dérivée de x^3 s'annule en $x = 0$, mais ce n'est ni un minimum ni un maximum.

16.10.2 Extrema libre

Définition 16.68.
Un point a à l'intérieur du domaine d'une fonction $f\colon A \subset \mathbb{R}^n \to \mathbb{R}$ est un ***point critique*** *de f lorsque $df(a) = 0$.*

Ces points sont analogues aux points où la dérivée d'une fonction sur $\mathbb{R}$ s'annule. Les points critiques de f sont dons les candidats à être des points d'extrémum.

Dans le cas d'une fonction de deux variables,l la proposition 11.286 nous permet de voir $(d^2 f)_a$ comme étant la matrice

$$d^2 f(a) = \begin{pmatrix} \frac{d^2 f}{dx^2}(a) & \frac{d^2 f}{dx\,dy}(a) \\ \frac{d^2 f}{dy\,dx}(a) & \frac{d^2 f}{dy^2}(a) \end{pmatrix}. \tag{16.274}$$

Dans le cas d'une fonction C^2, cette matrice est symétrique.

Proposition 16.69 ([328])**.**
Soit un ouvert Ω de $\mathbb{R}^n$ et $a \in \Omega$. Soit une fonction $f\colon \Omega \to \mathbb{R}$ différentiable en a. Si a est un extrémum local de f, alors a est un point critique de f.

Démonstration. Nous supposons que a est un maximum local (ce sera la même chose si a est un minimum). Soit $r > 0$ tel que $f(x) \leqslant f(a)$ pour tout $x \in B(a, r)$ (et tel que cette boule reste dans Ω). Soit $u \in \mathbb{R}^n$ assez petit pour que $a \pm u \in B(a, r)$ de sorte que la définition suivante ait un sens :

$$\begin{aligned} g\colon [-1, 1] &\to \mathbb{R} \\ t &\mapsto f(a + tu) \end{aligned} \tag{16.275}$$

Cette fonction est différentiable en $t = 0$ (composée de fonctions différentiables, proposition 11.208) et a un maximum local en $t = 0$. Donc $g'(0) = 0$ par la proposition 16.65. Donc

$$0 = \frac{d}{dt}\Big[f(a + tu)\Big]_{t=0} = df_a(u). \tag{16.276}$$

□

16.10.3 Extrema et Hessienne

Proposition 16.70 ([1, 329, 330])**.**
Soit un ouvert Ω de $\mathbb{R}^n$ et une fonction $f\colon \Omega \to \mathbb{R}$ deux fois différentiable ainsi que $a \in \Omega$.

(1) Si a est un point critique de f et si il existe r tel que $(d^2 f_x)$ est semi-définie positive pour tout $x \in B(a,r)$ alors f possède un minimum local en a.

(2) Si a est un point critique[31] *de f, et si $d^2 f_a$ est strictement définie positive*[32]*, alors a est un minimum local strict de f,*

(3) Si a est un minimum local, alors $(d^2 f)_a$ est semi-définie positive.

Démonstration. Nous subdivisons la preuve.

(1) Soit h tel que $a + h \in B(a,r)$. Nous allons montrer que $f(a) \leqslant f(a+h)$; cela montrera que $x = a$ est un minimum local. Pour cela nous utilisons un développement de Taylor[33] : il existe $c \in]a, a+h[$ tel que

$$f(a+h) = f(a) + df_a(h) + \frac{1}{2}(d^2 f)_c(h,h) \geqslant f(a) \tag{16.277}$$

parce que par hypothèse $(d^2 f)_c$ est définie positive et parce que $df_a = 0$.

(2) La forme bilinéaire $d^2 f_a$ est strictement définie positive, donc il existe $\alpha > 0$ tel que

$$d^2 f_a(h,h) > \alpha \|h\|^2 \tag{16.278}$$

pour tout h. Nous écrivons encore Taylor : il existe une fonction ϵ telle que $\lim_{h\to 0} \epsilon(h) = 0$ et

$$f(a+h) = g(a) + df_a(h) + \frac{1}{2}(d^2 f)_a(h,h) + \|h\|^2 \epsilon(h). \tag{16.279}$$

En tenant compte du fait que $df_a = 0$,

$$f(a+h) > f(a) + \|h\|^2 \big(\frac{1}{2}\alpha + \epsilon(h)\big). \tag{16.280}$$

La limite de ϵ nous dit qu'il existe $r > 0$ tel que $\|\epsilon(h)\| < \frac{1}{2}\alpha$ pour tout $h \in B(0,r)$. Pour ces valeurs de h nous avons

$$f(a+h) > f(a). \tag{16.281}$$

Donc a est un minimum local strict de f.

(3) Si a est un minimum local, nous savons déjà que $df_a = 0$ par la proposition 16.69. Nous écrivons le développement de Taylor de f à l'ordre 2 de la proposition 11.382 :

$$f(a+h) = f(a) + df_a(h) + \frac{1}{2}(d^2 f)_a(h,h) + \|h\|^2 \alpha(\|h\|). \tag{16.282}$$

En prenant h assez petit pour que $a + h$ ne sorte pas de la boule dans laquelle a est un minimum, nous avons $f(a+h) - f(a) > 0$. Donc

$$\frac{1}{2}(d^2 f)_a(h,h) + \|h\|^2 \alpha(\|h\|) > 0 \tag{16.283}$$

Nous divisons cela par $\|h\|^2$ et notons $e_h = h/\|h\|$:

$$\frac{1}{2}(d^2 f)_a(e_h, e_h) + \alpha(\|h\|) > 0. \tag{16.284}$$

À la limite $h \to 0$, le premier terme est constant tandis que le deuxième tend vers zéro. À la limite,

$$(d^2 f)_a(e_h, e_h) \geqslant 0. \tag{16.285}$$

La caractérisation du lemme 9.157(2) nous dit alors que $(d^2 f)_a$ est semi-définie positive.

31. Définition 16.68.
32. La fonction f est de classe C^2, donc les dérivées croisées sont égales et $d^2 f$ est symétrique. La définition 9.152 s'applique donc.
33. Proposition 11.385.

□

La partie (3) est tout à fait comparable au fait bien connu que, pour une fonction $f\colon \mathbb{R} \to \mathbb{R}$, si le point a est minimum local, alors $f'(a) = 0$ et $f''(a) > 0$.

Notons que le point (3) ne parle pas de minimum strict, et donc pas de matrice *strictement* définie positive.

Exemple 16.71 (Proposition 16.70(2) sans point critique)**.**
L'hypothèse de point critique pour l'utilisation de la stricte définition positive de $d^2 f_a$ est nécessaire. Soit en effet la fonction

$$f(x) = x^2 + x. \tag{16.286}$$

Elle vérifie $f''(0) = 2$, de telle sorte que sa différentielle seconde en zéro soit strictement définie positive. Le point $x = 0$ n'est cependant même pas un minimum local. Entre autres parce que $f'(0) = 1 \neq 0$. △

La méthode pour chercher les extrémums de f est donc de suivre le points suivants :

(1) Trouver les candidats extrémums en résolvant $\nabla f = (0, 0)$,
(2) écrire $d^2 f(a)$ pour chacun des candidats
(3) calculer les valeurs propres de $d^2 f(a)$, déterminer si la matrice est définie positive ou négative,
(4) conclure.

Une conséquence de la proposition 9.155(3) [34] est que si $\det M < 0$, alors le point a n'est pas un extrémum dans le cas où $M = d^2 f(a)$ par le point (3) de la proposition 16.70.

Exemple 16.72.
Soit la fonction $f(x, y) = x^4 + y^4 - 4xy$. C'est une fonction différentiable sans problèmes. D'abord sa différentielle est

$$df = \big(4x^3 - 4y; 4y^3 - 4x\big), \tag{16.287}$$

et la matrice des dérivées secondes est

$$M = d^2 f(x, y) = \begin{pmatrix} 12x^2 & -4 \\ -4 & 12y^2 \end{pmatrix}. \tag{16.288}$$

Nous avons $fd = 0$ pour les trois points $(0, 0)$, $(1, 1)$ et $-1, -1$.

Pour le point $(0, 0)$ nous avons

$$M = \begin{pmatrix} 0 & -4 \\ -4 & 0 \end{pmatrix}, \tag{16.289}$$

dont les valeurs propres sont 4 et -4. Elle n'est donc semi-définie ou définie rien du tout. Donc $(0, 0)$ n'est pas un extrémum local.

Au contraire pour les points $(1, 1)$ et $(-1, -1)$ nous avons

$$M = \begin{pmatrix} 12 & -4 \\ -4 & 12 \end{pmatrix}, \tag{16.290}$$

dont les valeurs propres sont 16 et 8. La matrice $d^2 f$ y est donc définie positive. Ces deux points sont donc extrémums locaux. △

16.10.4 Un peu de recettes de cuisine

(1) Rechercher les points critiques, càd les (x, y) tels que

$$\begin{cases} \frac{\partial f}{\partial x}(x, y) = 0 \\ \frac{\partial f}{\partial y}(x, y) = 0 \end{cases}$$

En effet, si (x_0, y_0) est un extrémum local de f, alors $\frac{\partial f}{\partial x}(x_0, y_0) = 0 = \frac{\partial f}{\partial y}(x_0, y_0)$.

34. La matrice $d^2 f(a)$ est toujours symétrique quand f est de classe C^2.

(2) Déterminer la nature des points critiques : « test » des dérivées secondes :

$$\text{On pose } H(x_0,y_0) = \frac{\partial^2 f}{\partial x^2}(x_0,y_0)\frac{\partial f^2}{\partial y^2}(x_0,y_0) - \left(\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0)\right)^2$$

(2a) Si $H(x_0,y_0) > 0$ et $\frac{\partial^2 f}{\partial x^2}(x_0,y_0) > 0 \Longrightarrow (x_0,y_0)$ est un minimum local de f.

(2b) Si $H(x_0,y_0) > 0$ et $\frac{\partial^2 f}{\partial x^2}(x_0,y_0) < 0 \Longrightarrow (x_0,y_0)$ est un maximum local de f.

(2c) Si $H(x_0,y_0) < 0 \Longrightarrow f$ a un point de selle en (x_0,y_0).

(2d) Si $H(x_0,y_0) = 0 \Longrightarrow$ on ne peut rien conclure.

16.10.5 Extrema liés

Soit f, une fonction sur $\mathbb{R}^n$, et $M \subset \mathbb{R}^n$ une variété de dimension m. Nous voulons savoir quelles sont les plus grandes et plus petites valeurs atteintes par f sur M.

Pour ce faire, nous avons un théorème qui permet de trouver des extrémums *locaux* de f sur la variété. Pour rappel, $a \in M$ est une **extrémum local de f relativement** à l'ensemble M s'il existe une boule $B(a,\epsilon)$ telle que $f(a) \leqslant f(x)$ pour tout $x \in B(a,\epsilon) \cap M$.

Théorème 16.73 (Extrema lié [311])**.**
Soit A, un ouvert de $\mathbb{R}^n$ et

(1) une fonction (celle à minimiser) $f \in C^1(A,\mathbb{R})$,

(2) des fonctions (les contraintes) $G_1,\dots,G_r \in C^1(A,\mathbb{R})$,

(3) $M = \{x \in A$ tel que $G_i(x) = 0\,\forall i\}$,

(4) un extrémum local $a \in M$ de f relativement à M.

Supposons que les gradients $\nabla G_1(a)$, $\dots$,$\nabla G_r(a)$ soient linéairement indépendants. Alors $a = (x_1,\dots,x_n)$ est une solution de $\nabla L(a) = 0$ où

$$L(x_1,\dots,x_n,\lambda_1,\dots,\lambda_r) = f(x_1,\dots,x_n) + \sum_{i=1}^{r} \lambda_i G_i(x_1,\dots,x_n). \tag{16.291}$$

Autrement dit, si a est un extrémum lié, alors $\nabla f(a)$ est une combinaisons des $\nabla G_i(a)$, ou encore il existe des λ_i tels que

$$df(a) = \sum_i \lambda_i dG_i(a). \tag{16.292}$$

La fonction L est le **lagrangien** du problème et les variables λ_i sont les **multiplicateurs de Lagrange**.

Démonstration. Si $r = n$ alors les vecteurs linéairement indépendantes $\nabla G_i(a)$ forment une base de $\mathbb{R}^n$ et donc évidemment les λ_i existent. Nous supposons donc maintenant que $r < n$. Nous notons $(z_i)_{i=1\dots n}$ les coordonnées sur $\mathbb{R}^n$.

La matrice

$$\begin{pmatrix} \frac{\partial G_1}{\partial z_1}(a) & \cdots & \frac{\partial G_1}{\partial z_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial G_r}{\partial z_1}(a) & \cdots & \frac{\partial G_r}{\partial z_n}(a) \end{pmatrix} \tag{16.293}$$

est de rang r parce que les lignes sont par hypothèses linéairement indépendantes. Nous nommons $(y_i)_{i=1,\dots,r}$ un choix de r parmi les (z_i) tels que

$$\det \begin{pmatrix} \frac{\partial G_1}{\partial y_1} & \cdots & \frac{\partial G_1}{\partial y_r} \\ \vdots & \ddots & \vdots \\ \frac{\partial G_r}{\partial y_1} & \cdots & \frac{\partial G_r}{\partial y_r} \end{pmatrix} \neq 0. \tag{16.294}$$

Nous identifions $\mathbb{R}^n$ à $\mathbb{R}^s \times \mathbb{R}^r$ dans lequel $\mathbb{R}^r$ est la partie générée par les $(y_i)_{i=1,\dots,r}$. Les coordonnées sur $\mathbb{R}^s$ seront nommées $(x_j)_{j=1,\dots,s}$, de telle sorte que les coordonnées sur $\mathbb{R}^n$ setont $x_1,\dots,x_s,y_1,\dots,y_r$. Dans ces coordonnées, nous nommons $a=(\alpha,\beta)$ avec $\alpha\in\mathbb{R}^s$ et $\beta\in\mathbb{R}^r$.

Si nous notons $G=(G_1,\dots,G_r)$, le théorème de la fonction implicite (théorème 16.49) nous dit qu'il existe un voisinage U' de $\alpha\in\mathbb{R}^n$, un voisinage V' de $\beta\in\mathbb{R}^r$ et une fonction $\varphi\colon U'\to V'$ de classe C^1 telle que si $(x,y)\in U'\times V'$, alors

$$g(x,y)=0 \tag{16.295}$$

si et seulement si $y=\varphi(x)$. Nous posons maintenant

$$\psi(x)=(x,\varphi(x)) \tag{16.296a}$$
$$h(x)=f\big(\psi(x)\big). \tag{16.296b}$$

Nous avons $\psi(\alpha)=a$ et $\psi(x)\in M$ pour tout $x\in U'$. La fonction h a donc un extrémum local en α et donc les dérivées partielles de h y sont nulles. Cela signifie que

$$0=\frac{\partial h}{\partial x_i}(\alpha)=\sum_{j=1}^n\frac{\partial f}{\partial x_j}\frac{\partial x_j}{\partial x_i}+\sum_{k=1}^r\frac{\partial f}{\partial y_k}\frac{\partial\varphi_k}{\partial x_i}, \tag{16.297}$$

c'est-à-dire

$$\frac{\partial f}{\partial x_i}(\alpha)+\sum_{k=1}^r\frac{\partial f}{\partial y_k}(a)\frac{\partial\varphi_k}{\partial x_i}(\alpha)=0 \tag{16.298}$$

pour tout $i=1,\dots,s$. D'autre part pour tout k, la fonction $l_k(x)=G_k\big(x,\varphi(x)\big)$ est constante et vaut zéro ; ses dérivées partielles sont donc nulles :

$$\frac{\partial l}{\partial x_i}(\alpha)=\frac{\partial G_k}{\partial x_i}(\alpha)+\sum_{k=1}^r\frac{\partial G_k}{\partial y_k}(a)\frac{\partial\varphi_k}{\partial x_i}(\alpha)=0 \tag{16.299}$$

pour tout $i=1,\dots,s$ et $k=1,\dots,r$.

Les s premières colonnes de la matrice

$$\begin{pmatrix}\frac{\partial f}{\partial x_1} & \dots & \frac{\partial f}{\partial x_s} & \frac{\partial f}{\partial y_1} & \dots & \frac{\partial f}{\partial y_r}\\ \frac{\partial G_1}{\partial x_1} & \dots & \frac{\partial G_1}{\partial x_s} & \frac{\partial G_1}{\partial y_1} & \dots & \frac{\partial G_1}{\partial y_r}\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ \frac{\partial G_r}{\partial x_1} & \dots & \frac{\partial G_r}{\partial x_s} & \frac{\partial G_r}{\partial y_1} & \dots & \frac{\partial G_r}{\partial y_r}\end{pmatrix} \tag{16.300}$$

s'expriment en termes des r dernières. La matrice est donc au maximum de rang r. Notons que la première ligne est ∇f et les r suivantes sont les ∇G_i. Vu que ces lignes sont des vecteurs liés, il existe $\mu_0,\dots,\mu_r$ tels que

$$\mu_0\nabla f+\sum_{i=1}^r\mu_i\nabla G_i=0. \tag{16.301}$$

Par hypothèse les ∇G_i sont linéairement indépendants, ce qui nous dit que $\mu_0\neq 0$. Donc nous avons ce qu'il nous faut :

$$\nabla f(a)=\sum_i\frac{\mu_i}{\mu_0}\nabla G_i(a). \tag{16.302}$$

Notons qu'au vu de l'expression (16.292), le fait que les formes $\{dG_i(a)\}_{1\leqslant i\leqslant r}$ forment une partie libre dans $(\mathbb{R}^n)^*$ implique que les λ_i sont uniques. □

La proposition suivante est la même que 16.73.

Proposition 16.74.
Soit U, un ouvert de $\mathbb{R}^n$ et des fonctions de classe C^1 $f, g_1, \dots, g_r \colon U \to \mathbb{R}$. Nous considérons

$$\Gamma = \{x \in U \text{ tel que } g_1(x) = \dots = g_r(x) = 0\}. \tag{16.303}$$

Soit a un extrémum de $f|_\Gamma$. Supposons que les formes $dg_1, \dots, dg_r$ soient linéairement indépendantes en a. Alors il existe $\lambda_1, \dots, \lambda_r$ dans $\mathbb{R}$ tel que

$$df_a = \sum_{i=1}^{r} \lambda_i (dg_i)_a. \tag{16.304}$$

En pratique les candidats extrémums locaux sont tous les points où les gradients ne sont pas linéairement indépendants, plus tous les points donnés par l'équation $\nabla L = 0$. Parmi ces candidats, il faut trouver lesquels sont maximums ou minima, locaux ou globaux.

L'existence d'extrémums locaux se prouve généralement en invoquant de la compacité, et en invoquant le lemme suivant qui permet de réduire le problème à un compact.

Lemme 16.75.
Soit S, une partie de $\mathbb{R}^n$ et C, un ouvert de $\mathbb{R}^n$. Si $a \in \operatorname{Int} S$ est un minimum local relatif à $S \cap C$, alors il est un minimum local par rapport à S.

Démonstration. Nous avons que $\forall x \in B(a, \epsilon_1) \cap S \cap C$, $f(x) \geqslant f(x)$. Mais étant donné que C est ouvert, et que $a \in C$, il existe un ϵ_2 tel que $B(a, \epsilon_2) \subset C$. En prenant $\epsilon = \min\{\epsilon_1, \epsilon_2\}$, nous trouvons que $f(x) \geqslant f(a)$ pour tout $x \in B(a, \epsilon) \cap (S \cap C) = B(a, \epsilon) \cap S$. □

16.11 Fonctions convexes

Définition 16.76 ([331])**.**
*Une fonction f d'un intervalle I de $\mathbb{R}$ vers $\mathbb{R}$ est dite **convexe** lorsque, pour tous x_1 et x_2 de I et tout λ dans $[0, 1]$ nous avons*

$$f\big(\lambda x_1 + (1-\lambda) x_2\big) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2) \tag{16.305}$$

Si pour tout $\theta \in]0, 1[$ et pour tout $x \neq y$ dans I nous avons

$$f\big(\lambda x_1 + (1-\lambda) x_2\big) < \lambda f(x_1) + (1-\lambda) f(x_2) \tag{16.306}$$

*alors nous disons que la fonction f est **strictement convexe** sur I.*

*Une fonction est **concave** si son opposée est convexe.*

16.77 ([331])**.**
Les différents résultats pour les fonctions convexes s'adaptent généralement sans mal aux fonctions strictement convexes. Une nuance cependant : de même que les fonctions dérivables convexes sont celles qui ont une dérivée croissante, les fonctions dérivables strictement convexes sont celles qui ont une dérivée strictement croissante (proposition 16.81). En revanche, il ne faudrait pas croire que la dérivée seconde d'une fonction dérivable strictement convexe est nécessairement une fonction à valeurs strictement positives (voir théorème 16.82) : la dérivée d'une fonction strictement croissante peut s'annuler occasionnellement, ou plus exactement peut s'annuler sur un ensemble de points d'intérieur vide. Penser à $x \mapsto x^4$ pour un exemple de fonction strictement convexe dont la dérivée seconde s'annule.

16.11.1 Inégalité des pentes

Dans l'étude des fonctions convexes nous allons souvent utiliser la fonction **taux d'accroissement** qui est, pour α dans le domaine de convexité de f définie par

$$\begin{aligned} \tau_\alpha \colon I \backslash \{\alpha\} &\to \mathbb{R} \\ x &\mapsto \frac{f(x) - f(\alpha)}{x - \alpha}. \end{aligned} \tag{16.307}$$

Proposition 16.78 (Inégalité des pentes[332]).
Soit f une fonction convexe sur un intervalle $I \subset \mathbb{R}$. Alors pour tout $a < b < c$ dans I nous avons[35]

$$\frac{f(b)-f(a)}{b-a} \leqslant \frac{f(c)-f(a)}{c-a} \leqslant \frac{f(c)-f(b)}{c-b}. \tag{16.308}$$

En d'autres termes,

$$\tau_a(b) \leqslant \tau_a(c) \leqslant \tau_b(c), \tag{16.309}$$

c'est-à-dire que τ est croissante en ses deux arguments.

Démonstration. D'abord les inégalités $a < b < c$ impliquent $0 < b-a < c-a$ et donc

$$\lambda = \frac{b-a}{c-a} < 1. \tag{16.310}$$

L'astuce est de remarquer que $(1-\lambda)a + \lambda c = b$. Donc λ a toutes les bonnes propriétés pour être utilisé dans la définition de la convexité :

$$f\big((1-\lambda)a + \lambda c\big) \leqslant \lambda f(c) + (1-\lambda)f(a), \tag{16.311}$$

c'est-à-dire

$$f(b) - f(a) \leqslant \lambda\big(f(c) - f(a)\big) \tag{16.312}$$

ou encore, en remplaçant λ par sa valeur :

$$\frac{f(b)-f(a)}{b-a} \leqslant \frac{f(c)-f(a)}{c-a}. \tag{16.313}$$

Cela fait déjà une des inégalités à savoir.

D'autre part en partant de $-a < -b < -c$ nous posons

$$0 < \lambda = \frac{c-b}{c-a}. \tag{16.314}$$

Nous avons à nouveau $b = (1-\lambda)c + \lambda a$ et nous pouvons obtenir la seconde inégalité

$$\frac{f(c)-f(a)}{c-a} \leqslant \frac{f(c)-f(b)}{c-b}. \tag{16.315}$$

□

Géométriquement, l'inégalité des pentes se comprend facilement : le coefficient angulaire de la corde du graphe augmente. Donc si $x < y < z$, le coefficient moyen entre x et y est plus petit que celui entre x et z qui est plus petit que celui entre y et z.

Donc si le coefficient angulaire moyen entre a et $b+u$ vaut celui entre a et b, ce coefficient ne peut qu'être constant entra a et b : sinon il serait plus grand entre b et $b+u$ et la moyenne sur $a \to b+u$ serait plus grande que sa moyenne sur $a \to b$. Mais avoir un coefficient angulaire constant signifie être une droite.

En résumé, si une fonction est convexe et non strictement convexe, alors son graphe est une droite. C'est en gros cela que la proposition 16.86 clarifiera.

35. Les inégalités sont strictes si la fonction f est strictement convexe.

16.11.2 Convexité et régularité

Lemme 16.79 ([331]).
Une fonction convexe sur un ouvert

(1) y admet des dérivées à gauche et à droite en chaque point,

(2) y est continue.

Démonstration. Soit $I =]a, b[$ un intervalle sur lequel f est convexe et $\alpha \in I$. Nous allons prouver que f est continue en α. Nous considérons τ_α le taux d'accroissement définit par (16.307) ; c'est une fonction croissante comme précisé dans l'inégalité des trois pentes 16.78 et de plus $\tau_\alpha(x)$ est bornée supérieurement par $\tau_\alpha(b)$ pour $x < \alpha$ et inférieurement par $\tau_\alpha(a)$ pour $x > \alpha$. Les limites existent donc et sont finies par la proposition 11.66. Autrement dit les limites

$$\lim_{x\to\alpha+} \frac{f(x) - f(\alpha)}{x - \alpha} = \lim_{x\to\alpha^+} \tau_\alpha(x) = \inf_{t>\alpha} \tau_\alpha(t) \tag{16.316a}$$

$$\lim_{x\to\alpha^-} \frac{f(x) - f(\alpha)}{x - \alpha} = \lim_{x\to\alpha^-} \tau_\alpha(x) = \sup_{t<\alpha} \tau_\alpha(t). \tag{16.316b}$$

existent et sont finies, c'est-à-dire que la fonction f admet une dérivée à gauche et à droite.

Pour tout x nous avons les inégalités

$$\tau_\alpha(a) \leqslant \frac{f(x) - f(\alpha)}{x - \alpha} \leqslant \tau_\alpha(b). \tag{16.317}$$

En posant $k = \max\{\tau_\alpha(a), \tau_\alpha(b)\}$ nous avons

$$\big|f(x) - f(\alpha)\big| \leqslant k|x - \alpha|. \tag{16.318}$$

La fonction est donc Lipschitzienne et par conséquent continue par la proposition 11.261. □

Remarque 16.80.
Les dérivées à gauche et à droite ne sont à priori pas égales. Penser par exemple à une fonction affine par morceaux dont les pentes augmentent à chaque morceau.

16.11.3 Dérivées d'une fonction convexe

Proposition 16.81 ([333, 334, 1]).
Une fonction dérivable sur un intervalle I de $\mathbb{R}$

(1) est convexe si et seulement si sa dérivée est croissante sur I.

(2) est strictement convexe si et seulement si sa dérivée est strictement croissante sur I

Démonstration. Pour la preuve de (1) et (2), nous allons démontrer les énoncés « non stricts » et indiquer ce qu'il faut changer pour obtenir les énoncés « stricts ».

Sens direct Nous supposons que f est convexe. Soient $a < b$ dans I et $x \in]a, b[$. D'après l'inégalité des pentes 16.78,

$$\frac{f(x) - f(a)}{x - a} \leqslant \frac{f(b) - f(a)}{b - a} \leqslant \frac{f(b) - f(x)}{b - x}. \tag{16.319}$$

En faisant la limite $x \to a$ nous avons

$$f'(a) \leqslant \frac{f(b) - f(a)}{b - a} \tag{16.320}$$

et la limite $x \to b$ donne

$$\frac{f(b) - f(a)}{b - a} \leqslant f'(b). \tag{16.321}$$

Ici les inégalités sont non à priori strictes, même si f est strictement convexe : même avec des inégalités strictes dans (16.319), le passage à la limite rend l'inégalité non stricte. Quoi qu'il en soit nous avons

$$f'(a) \leqslant f'(b). \tag{16.322}$$

Sens direct : strict Nous savons déjà que f' est croissante. Si (16.322) était une égalité, alors f' serait constante sur $]a,b[$ parce qu'en prenant c entre a et b nous aurions $f'(a) \leqslant f'(c) \leqslant f'(b)$ avec $f'(a) = f'(b)$. Donc $f'(a) = f'(c)$. Avoir f' constante sur un intervalle est contraire à la stricte convexité.

Sens réciproque Nous supposons que f' est croissante et nous considérons $a < b$ dans I ainsi que $\lambda \in [0,1]$. Nous posons $x = \lambda a + (1-\lambda)b$, et nous savons que $a \leqslant x \leqslant b$. Le théorème des accroissements finis 11.130 donne $c_1 \in]a,x[$ et $c_2 \in]x,b[$ tels que

$$f'(c_1) = \frac{f(x) - f(a)}{x - a} \tag{16.323}$$

et

$$f'(c_2) = \frac{f(b) - f(x)}{b - x}. \tag{16.324}$$

Et en plus $c_1 < c_2$. Vu que f' est croissante nous avons $f'(c_1) \leqslant f'(c_2)$ et donc

$$\frac{f(x) - f(a)}{x - a} \leqslant \frac{f(b) - f(x)}{b - x}. \tag{16.325}$$

En remplaçant x par sa valeur en termes de λ, a et b nous avons $x - a = (1-\lambda)(b-a)$ et $b - x = \lambda(b-a)$, et l'inégalité (16.325) nous donne

$$f(x) \leqslant \lambda f(a) + (1-\lambda) f(b). \tag{16.326}$$

Sens réciproque : strict Si f' est strictement croissante, nous avons $f'(c_2) < f'(c_2)$ et les inégalité suivantes sont strictes, ce qui donne

$$f(x) < \lambda f(a) + (1-\lambda) f(b). \tag{16.327}$$

□

Théorème 16.82 ([333]).
Soit une fonction f de classe C^2.

(1) Est convexe si et seulement si f'' est positive.

(2) Si f'' est strictement positive, elle est strictement convexe.

Démonstration. En deux parties.

Pour (1) La fonction est C^2, donc f'' est positive si et seulement si f' est croissante (proposition 11.125) alors que la proposition 16.81 nous jure que f sera convexe si et seulement si f' est croissante.

Pour (2) Si f'' est strictement positive, f' sera strictement croissante et donc f strictement convexe (proposition 16.81).

□

Remarque 16.83.
Une fonction peut être strictement convexe sans que sa dérivée seconde ne soit toujours strictement positive. En exemple : $x \mapsto x^4$ est strictement convexe alors que sa dérivée seconde s'annule en zéro.

Exemple 16.84.
Quelques exemples utilisant le théorème 16.82

(1) La fonction $x \mapsto x^2$ est convexe parce que sa dérivée seconde est la constante (positive) 2.

(2) La fonction $x \mapsto \frac{1}{x}$ est convexe sur $\mathbb{R}^+\backslash\{0\}$ (sa dérivée seconde est $2x^{-3}$).

(3) La fonction exponentielle est strictement convexe par le théorème 16.82.

(4) La fonction ln est concave parce que la dérivée seconde de $-\ln$ est $\frac{1}{x^2}$ qui est strictement positif.

△

Nous en faisons une en détail ; elle sera utile en analyse fonctionnelle, lors de l'étude des espaces L^p. Voir par exemple le théorème de la projection 26.127.

Lemme 16.85.
Soient $p > 1$ et la fonction

$$\begin{aligned} f\colon]0,\infty[&\to \mathbb{R} \\ x &\mapsto x^p \end{aligned} \tag{16.328}$$

est strictement convexe.

Démonstration. La proposition 13.245 nous permet de dire que la fonction f est de classe C^∞ et que la dérivée seconde est donnée par

$$f''(x) = p(p-1)x^{p-2}. \tag{16.329}$$

Cela est strictement positif pour tous les x considérée, le théorème 16.82 conclu. □

16.11.4 Graphe d'une fonction convexe

L'idée principale du graphe d'une fonction convexe est qu'il est toujours au dessus du graphe de ses tangentes (lorsqu'elles existent). Lorsqu'elles n'existent pas, le lemme 16.79 donne des coefficients directeurs de droites qui vont rester en dessous du graphe de la fonction.

Proposition 16.86 ([335])**.**
Une fonction convexe est strictement convexe si et seulement s'il n'existe aucun intervalle de longueur non nulle sur lequel elle coïncide avec une fonction affine.

Démonstration. Si sur l'intervalle (non réduit à un point) $[x, y]$, la fonction convexe f coïncide avec une fonction affine, alors $f(t) = at + b$ et pour $\lambda \in]0, 1[$ nous avons

$$f\big(\lambda x + (1-\lambda)y\big) = a\lambda x + a(1-\lambda)y + b = \lambda f(x) + (1-\lambda)f(y) \tag{16.330}$$

où nous avons remplacé b par $\lambda b + (1-\lambda)b$. Par conséquent la fonction n'est pas strictement convexe.

Nous supposons maintenant que la fonction convexe f n'est pas strictement convexe sur l'intervalle I. Il existe $x \neq y \in I$ et $\lambda \in]0, 1[$ tels que

$$f\big(\lambda x + (1-\lambda)y\big) = \lambda f(x) + (1-\lambda)f(y). \tag{16.331}$$

Nous posons $z = \lambda x + (1-\lambda)y$ et $u \in]x, z[$ pour écrire des inégalités des pentes entre $x < u < z < y$. Plus précisément si nous notons $a \to b$ la pente de a à b, c'est-à-dire $a \to b = \frac{f(b)-f(a)}{b-a}$, alors les inégalités des pentes pour $x < u < z$ puis $u < z < y$ donnent

$$x \to z \leqslant u \to z \leqslant z \to y. \tag{16.332}$$

Voyons maintenant qu'en réalité $z \to y = x \to z$. En effet en replaçant

$$f(y) = \frac{f(z) - \lambda f(x)}{1-\lambda} \tag{16.333}$$

et

$$y = \frac{\lambda x}{1-\lambda} \tag{16.334}$$

dans l'expression $z \to y = \frac{f(y)-f(z)}{y-z}$ nous obtenons

$$z \to y = \frac{f(y)-f(z)}{y-z} = \frac{f(z)-f(x)}{z-x} = x \to z. \tag{16.335}$$

Les inégalités (16.332) sont donc des égalités :

$$\frac{f(z)-f(x)}{z-x}=\frac{f(z)-f(u)}{z-u}=\frac{f(y)-f(z)}{y-z}. \tag{16.336}$$

Nous avons donc montré que le nombre $a=\frac{f(z)-f(u)}{z-u}$ ne dépend pas de u. Nous avons alors

$$f(z)-f(u)=a(z-u) \tag{16.337}$$

ou encore :

$$f(u)=f(z)-a(z-u), \tag{16.338}$$

ce qui signifie que sur $]x,z[$, la fonction f est affine. □

Proposition 16.87.
Une fonction dérivable sur un intervalle I de $\mathbb{R}$ est convexe si et seulement si son graphe est au dessus de chacune de ses tangentes.

Démonstration. En deux parties.

Sens direct Soient $x,y\in I$. Nous voulons :

$$f(y)\geqslant f(x)+f'(x)(y-x). \tag{16.339}$$

Étant donné que nous aurons besoin, dans le quotient différentiel de quelque chose comme $f(x+t)-f(x)$ nous écrivons la définition (16.305) de la convexité en inversant les rôles de x et y et en manipulant un peu :

$$f\big(ty+(1-t)x\big)\leqslant tf(y)+(1-t)f(x) \tag{16.340a}$$
$$f\big(x+t(y-x)\big)\leqslant tf(y)+(1-t)f(x) \tag{16.340b}$$
$$f\big(x+t(y-x)\big)=f(x)\leqslant tf(y)-tf(x) \tag{16.340c}$$

Nous divisons par t :

$$\frac{f\big(x+t(y-x)\big)-f(x)}{t}\leqslant f(y)-f(x). \tag{16.341}$$

Le passage à la limite $t\to 0$ donne

$$(y-x)f'(x)\leqslant f(y)-f(x), \tag{16.342}$$

ce qu'il fallait.

Sens inverse Pour tout $x,y\in I$ nous supposons avoir

$$f(y)\geqslant f(x)+f'(x)(y-x). \tag{16.343}$$

Si nous supposons $x\neq y$ et si nous posons $z=\lambda x+(1-\lambda)y$ nous voulons prouver que

$$f(z)\leqslant \lambda f(x)+(1-\lambda)f(y). \tag{16.344}$$

Pour cela nous écrivons l'inégalité (16.343) avec les couples (x,z) et (y,z) :

$$f(x)\geqslant f(z)+f'(z)'(x-z) \tag{16.345a}$$
$$f(y)\geqslant f(z)+f'(z)'(y-z) \tag{16.345b}$$

En multipliant la première par λ et la seconde par $(1-\lambda)$ et en sommant,

$$\lambda f(x)+(1-\lambda)f(y)\geqslant \lambda f(z)+\lambda f'(z)(x-z)+(1-\lambda)f(z)+(1-\lambda)f'(z)(y-z) \tag{16.346a}$$
$$=f(z)+f'(z)\big(\lambda(x-z)+(1-\lambda)(y-z)\big) \tag{16.346b}$$
$$=f(z). \tag{16.346c}$$

□

Proposition 16.88 ([1]).
Soit $f\colon \mathbb{R} \to \mathbb{R}$ une fonction convexe et $a \in \mathbb{R}$. Il existe une constante $c_a \in \mathbb{R}$ telle que pour tout x nous ayons

$$f(x) - f(a) \geqslant c_a(x - a). \tag{16.347}$$

Autrement dit, le graphe de la fonction f est toujours au dessus de la droite d'équation

$$y = f(a) + c_a(x - a). \tag{16.348}$$

Démonstration. Les dérivées à gauche et à droite de f données par le lemme 16.79 sont les candidats tout cuits pour être coefficient directeur de la droite que l'on cherche. Nous allons prouver qu'en posant

$$c_a = \inf_{t>a} \tau_a(t), \tag{16.349}$$

la droite $y = f(a) + c_a(x - a)$ répond à la question [36].

Nous devons prouver que le nombre $\Delta_x = f(x) - \big(f(a) + c_a(x - a)\big)$ est positif pour tout x.

Si $x > a$ Nous divisons par $x - a$ et nous devons prouver que $\frac{\Delta_x}{x-a}$ est positif :

$$\begin{aligned}
\frac{\Delta_x}{x-a} &= \frac{f(x) - f(a)}{x - a} - c_a && (16.350a)\\
&= \tau_a(x) - \inf_{t>a} \tau_a(t) && (16.350b)\\
&\geqslant 0 && (16.350c)
\end{aligned}$$

parce que $t \to \tau_a(t)$ est croissante et que $x > a$.

Si $x < a$ Nous divisons par $x - a$ et nous devons prouver que $\frac{\Delta_x}{x-a}$ est négatif :

$$\begin{aligned}
\frac{\Delta_x}{x-a} &= \frac{f(x) - f(a)}{x - a} - c_a && (16.351a)\\
&= \tau_a(x) - \inf_{t>a} \tau_a(t) && (16.351b)\\
&\leqslant 0 && (16.351c)
\end{aligned}$$

parce que $t \to \tau_a(t)$ est croissante et que $x < a$.

□

Proposition 16.89 ([1]).
Si g est une fonction convexe, il existe deux suites réelles (a_n) et (b_n) telles que

$$g(x) = \sup_{n\in\mathbb{N}}(a_n x + b_n). \tag{16.352}$$

Démonstration. Pour $u \in \mathbb{R}$ nous considérons $a(u)$ et $b(u)$ tels que la droite $y(x) = a(u)x + b(u)$ vérifie $y(u) = g(u)$ et $y(x) \leqslant g(x)$ pour tout x. Cela est possible par la proposition 16.88. Il s'agit d'une droite coupant le graphe de g en $x = u$ et restant en dessous. Nous considérons alors (u_n) une suite quelconque dense dans $\mathbb{R}$ (disons les rationnels pour fixer les idées) et nous posons

$$\begin{cases}
a_n = a(u_n) & (16.353a)\\
b_n = b(u_n). & (16.353b)
\end{cases}$$

Si $q \in \mathbb{Q}$ alors $a_n x + b_n \leqslant g(x)$ pour tout n et $g(q)$ est le supremum qui est atteint pour le n tel que $u_n = q$. Si maintenant x n'est pas dans $\mathbb{Q}$ il faut travailler plus.

36. En prenant l'autre, $c_a' = \sup_{t<a} \tau_a(t)$, ça fonctionne aussi. En pensant à une fonction affine par morceaux, on remarque qu'en choisissant un nombre entre les deux, nous avons plus facilement une inégalité stricte dans (16.347).

Nous prenons $(\tilde{q}_n)$, une sous-suite de (q_n) convergeant vers x et N suffisamment grand pour que pour tout $n \geqslant N$ on ait $|\tilde{q}_n - x| \leqslant \epsilon$ et $|g(\tilde{q}_n) - g(x)| \leqslant \epsilon$; cela est possible grâce à la continuité de g (lemme 16.79). Ensuite les sous-suites $(\tilde{a}_n)$ et $(\tilde{b}_n)$ sont celles qui correspondent :

$$\tilde{a}_n \tilde{q}_n + \tilde{b}_n = g(\tilde{q}_n). \tag{16.354}$$

Nous considérons la majoration

$$|\tilde{a}_n x + \tilde{b}_n - g(x)| \leqslant |\tilde{a}_n x + \tilde{b}_n - (\tilde{a}_n \tilde{q}_n + \tilde{b}_n)| + \underbrace{|\tilde{a}_n \tilde{q}_n + \tilde{b}_n - g(\tilde{q}_n)|}_{=0} + \underbrace{|g(\tilde{q}_n) - g(x)|}_{\leqslant \epsilon} \tag{16.355a}$$

$$\leqslant |\tilde{a}_n||x - \tilde{q}_n| + \epsilon \tag{16.355b}$$

$$= \epsilon\big(|\tilde{a}_n| + 1\big). \tag{16.355c}$$

Il nous reste à montrer que $|\tilde{a}_n|$ est borné par un nombre ne dépendant pas de n (pour les $n > N$).

Vu que la droite de coefficient directeur $\tilde{a}_n$ et passant par le point $\big(\tilde{q}_n, g(\tilde{q}_n)\big)$ reste en dessous du graphe de g, nous avons pour tout n et tout $y \in \mathbb{R}$ l'inégalité

$$g(y) \geqslant \tilde{a}_n(y - \tilde{q}_n) + g(\tilde{q}_n) \in \tilde{a}_n B(y - x, \epsilon) + B\big(g(x), \epsilon\big). \tag{16.356}$$

Si $\tilde{a}_n$ n'est pas borné vers le haut, nous prenons y tel que $B(y - x, \epsilon)$ soit minoré par un nombre k strictement positif et nous obtenons

$$g(y) \geqslant k\tilde{a}_n + l \tag{16.357}$$

avec k et l indépendants de n. Cela donne $g(y) = \infty$. Si au contraire $\tilde{a}_n$ n'est pas borné vers le bas, nous prenons y tel que $B(y - x, \epsilon)$ est majoré par un nombre k strictement négatif. Nous obtenons encore $g(y) = \infty$.

Nous concluons que $|\tilde{a}_n|$ est bornée. □

Lemme 16.90 ([54]).
L'application

$$\begin{aligned} \phi\colon S^{++}(n, \mathbb{R}) &\to \mathbb{R} \\ A &\mapsto \det(A) \end{aligned} \tag{16.358}$$

est ***log-convave****, c'est-à-dire que l'application* $\ln \circ \phi$ *est concave*[37]*. De façon équivalente, si* $A, B \in S^{++}$ *et si* $\alpha + \beta = 1$*, alors*

$$\det(\alpha A + \beta B) \geqslant \det(A)^\alpha \det(B)^\beta. \tag{16.359}$$

Ici S^{++} est l'ensemble des matrices symétriques strictement définies positives, définition 9.152.

Démonstration. En plusieurs étapes.

Pseudo-réduction Le théorème de pseudo-réduction simultanée, corolaire 9.254, appliqué aux matrices A et B nous donne une matrice inversible Q telle que

$$\begin{cases} B = Q^t D Q & (16.360a) \\ A = Q^t Q & (16.360b) \end{cases}$$

avec

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \tag{16.361}$$

$\lambda_i > 0$. Nous avons alors

$$\det(A)^\alpha \det(B)^\beta = \det(Q)^{2\alpha} \det(Q)^{2\beta} \det(D)^\beta = \det(Q)^2 \det(D)^\beta \tag{16.362}$$

37. La définition 14.58 du logarithme ne fonctionne que pour les réels strictement positifs. C'est le cas du déterminant d'une matrice réelle symétrique strictement définie positive.

(parce que $\alpha + \beta = 1$) et

$$\det(\alpha A + \beta B) = \det(\alpha Q^t Q + \beta Q^t D Q) \tag{16.363a}$$
$$= \det\big(Q^t(\alpha \mathbb{1} + \beta D)Q\big) \tag{16.363b}$$
$$= \det(Q)^2 \det(\alpha \mathbb{1} + \beta D). \tag{16.363c}$$

Ré-expression L'inégalité (16.359) qu'il nous faut prouver se réduit donc à

$$\det(\alpha \mathbb{1} + \beta D) \geqslant \det(D)^\beta. \tag{16.364}$$

Vue la forme de D nous avons

$$\det(\alpha \mathbb{1} + \beta D) = \prod_{i=1}^n (\alpha + \beta \lambda_i) \tag{16.365}$$

et

$$\det(D)^\beta = \big(\prod_{i=1}^n \lambda_i\big)^\beta. \tag{16.366}$$

Il faut donc prouver que

$$\prod_{i=1}^n (\alpha + \beta \lambda_i) \geqslant \big(\prod_{i=1}^n \lambda_i\big)^\beta. \tag{16.367}$$

Cette dernière égalité de produit sera prouvée en passant au logarithme.

Logarithme Vu que le logarithme est concave par l'exemple 16.84, nous avons pour chaque i que

$$\ln(\alpha + \beta \lambda_i) \geqslant \alpha \ln(1) + \beta \ln(\lambda_i) = \beta \ln(\lambda_i). \tag{16.368}$$

En sommant cela sur i et en utilisant les propriétés de croissance et de multiplicativité du logarithme nous obtenons successivement

$$\sum_{i=1}^n \ln(\alpha + \beta \lambda_i) \geqslant \beta \sum_i \ln(\lambda_i) \tag{16.369a}$$
$$\ln\big(\prod_i (\alpha + \beta \lambda_i)\big) \geqslant \ln\Big(\big(\prod_i \lambda_i\big)^\beta\Big) \tag{16.369b}$$
$$\prod_i (\alpha + \beta \lambda_i) \geqslant \big(\prod_i \lambda_i\big)^\beta, \tag{16.369c}$$

ce qui est bien (16.367).

□

Rappel de notations : $\mathbb{R}^+ = [0, \infty[$. Voir la remarque 1.137.

Lemme 16.91 ([1]).
Soit une fonction strictement convexe $g\colon \mathbb{R}^+ \to \mathbb{R}^+$. Soit une fonction $f\colon \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ vérifiant

(1) $f(0,0) = 0$,
(2) $f(tx, ty) = tf(x,y)$ *pour tout t (tant que ça ne déborde pas du domaine)*
(3) $f(1,y) = g(y)$ *pour tout y*
(4) $f(0,y) = f(y,0)$.

Alors f est convexe.

Démonstration. Nous devons prouver que pour toute paire de points A, B sur le graphe de f, le segment $[A, B]$ est au-dessus du graphe de f. Ledit graphe étant d'ailleurs constitué de droites joignant $(0,0,0)$ et les points du graphe de g (situé en $x = 1$).

Nous notons $\mathcal{C}$ le graphe de f.

Une corde alignée à O Soient deux point A et B alignés à l'origine O. Un point quelconque de $[A,B]$ (et même de toute la droite) s'écrit

$$\big(tA_x, tA_y, tf(A_x,A_y)\big) = \big(tA_x, tA_y, f(tA_x,tA_y)\big), \tag{16.370}$$

et donc est sur le graphe de f.

Autre corde Nous prouvons que $[A,B]\cap\mathcal{C}=\{A,B\}$.

Si A et B sont dans le plan $x=0$ alors c'est d'accord parce que le graphe de f dans le plan est le même que celui de la fonction strictement convexe g.

Si A et B ne sont pas alignés à O et si ils ne sont pas dans le plan $x=0$ alors le plan AOB coupe le plan $x=1$ en une droite.

Nous supposons l'existence d'un point $C\in\,]A,B[\,\cap\mathcal{C}$.

Nous considérons la droite (OA) qui est contenue dans ce plan et dans $\mathcal{C}$ (au moins la partie positive) et nous notons A' son intersection avec le plan $x=1$. Même chose pour B et C qui donnent B' et C'.

Cela nous donne des points A', B' et C' qui sont alignés dans le graphe de f en $x=1$. Or le graphe de f en $x=1$ est le graphe de la fonction g qui est strictement convexe et qui ne contient donc pas de points alignés.

Nous en concluons que si $A,B\in\mathcal{C}$ alors $]A,B[$ est soit complètement strictement au-dessus de $\mathcal{C}$ soit complètement strictement en-dessous de $\mathcal{C}$.

Nous prouvons à présent que toutes les cordes sont au-dessus de $\mathcal{C}$. Pour cela, soient $A,B\in\,]0,\infty[\times]0,\infty[$, deux points non alignés à $O=(0,0)$. Nous considérons les points A',B' qui sont les intersections entre les droites (AO) et (BO) et la droite $x=1$ ainsi que le chemin σ qui parcours le segment $[A,A']$ et le chemin γ qui parcours le segment $[B,B']$:

$$\begin{aligned}\sigma(0)&=A,\quad \gamma(0)=B,\\ \sigma(1)&=A',\quad \gamma(1)=B'.\end{aligned}\tag{16.371}$$

Pour tout u, la seule droite passant par O et par $\sigma(u)$ passe également par A, et pas par B. En conséquence de quoi, pour tout $u_1,u_2\in[0,1]$, la droite $\big(\sigma(u_1)\sigma(u_2)\big)$ ne passe pas par $(0,0)$.

Nous considérons à présent non seulement la corde joignant $\big(A,f(A)\big)$ à $\big(B,f(B)\big)$ et la corde joignant $\big(A',f(A')\big)$ à $\big(B',f(B')\big)$ mais également toutes les cordes intermédiaires (si vous aimez les gros mots, vous pouvez parler d'homotopie) :

$$c(u,t)=t\Big(\sigma(t),f\big(\sigma(u)\big)\Big)+(1-t)\Big(\gamma(t),f\big(\gamma(t)\big)\Big)\tag{16.372}$$

Pour chaque $u\in[0,1]$, cela représente une corde entre deux points non alignés à $(0,0,0)$ et donc une corde qui est soit strictement au-dessus de $\mathcal{C}$ soit strictement en-dessous (à par les points correspondant à $t=0$ et $t=1$ qui, eux, sont sur $\mathcal{C}$).

Soit $t_0\in\,]0,1[$. La courbe $c(u,t_0)$ avec $u\in[0,1]$ ne touche jamais $\mathcal{C}$. Or le point $c(1,t_0)$ est au-dessus de $\mathcal{C}$, donc le point $c(0,t_0)$ est également au-dessus de $\mathcal{C}$.

Nous en concluons que toutes les cordes entre $(A,f(A))$ et $(B,f(B))$ est située au-dessus de $\mathcal{C}$ et non en-dessous de $\mathcal{C}$.

□

16.11.5 Convexité et hessienne

Définition 16.92.
Soit une partie convexe U de $\mathbb{R}^n$ et une fonction $f\colon U\to\mathbb{R}$.

*(1) La fonction f est **convexe** si pour tout $x,y\in U$ avec $x\neq y$ et pour tout $\theta\in\,]0,1[$ nous avons*

$$f\big(\theta x+(1-\theta)y\big)\leqslant\theta f(x)+(1-\theta)f(y).\tag{16.373}$$

*(2) Elle est **strictement convexe** si nous avons l'inégalité stricte.*

Proposition 16.93 ([330]).
Soit Ω ouvert dans $\mathbb{R}^n$ et U convexe dans Ω, et une fonction différentiable $f\colon U \to \mathbb{R}$.

(1) La fonction f est convexe sur U si et seulement si pour tout $x, y \in U$,

$$f(y) \geqslant f(x) + df_x(y - x). \tag{16.374}$$

(2) La fonction f est strictement convexe sur U si et seulement si pour tout $x, y \in U$ avec $x \neq y$,

$$f(y) > f(x) + df_x(y - x). \tag{16.375}$$

Démonstration. Nous avons quatre petites choses à démontrer.

(1) sens direct Soit une fonction convexe f. Nous avons :

$$f\big((1-\theta)x + \theta y\big) \leqslant (1-\theta)f(x) + \theta f(y), \tag{16.376}$$

donc

$$f\big(x + \theta(y - x)\big) - f(x) \leqslant \theta\big(f(y) - f(x)\big) \tag{16.377}$$

Vu que $\theta > 0$ nous pouvons diviser par θ sans changer le sens de l'inégalité :

$$\frac{f\big(x + \theta(y - x)\big) - f(x)}{\theta} \leqslant f(y) - f(x). \tag{16.378}$$

Nous prenons la limite $\theta \to 0^+$. Cette limite est égale à a limite simple $\theta \to 0$ et vaut (parce que f est différentiable) :

$$\frac{\partial f}{\partial(y - x)}(x) \leqslant f(y) - f(x), \tag{16.379}$$

et aussi

$$df_x(y - x) \leqslant f(y) - f(x) \tag{16.380}$$

par le lemme 11.196.

(1) sens inverse Pour tout $a \neq b$ dans U nous avons

$$f(b) \geqslant f(a) + df_a(b - a). \tag{16.381}$$

Pour $x \neq y$ dans U et pour $\theta \in \,]0, 1[$ nous écrivons (16.381) pour les couples $\big(\theta x + (1-\theta)y, y\big)$ et $\big(\theta x + (1-\theta)y, x\big)$. Ça donne :

$$f(y) \geqslant f\big(\theta x + (1-\theta)y\big) + df_{\theta x + (1-\theta)y}\big(\theta(y - x)\big), \tag{16.382}$$

et

$$f(x) \geqslant f\big(\theta x + (1-\theta)y\big) + df_{\theta x + (1-\theta)y}\big((1-\theta)(x - y)\big). \tag{16.383}$$

La différentielle est linéaire ; en multipliant la première par $(1-\theta)$ et la seconde par θ et en la somme, les termes en df se simplifient et nous trouvons

$$\theta f(x) + (1-\theta)f(y) \geqslant f\big(\theta x + (1-\theta)y\big). \tag{16.384}$$

(2) sens direct Nous avons encore l'équation (16.378), avec une inégalité stricte. Par contre, ça ne va pas être suffisant parce que le passage à la limite ne conserve pas les inégalités strictes. Nous devons donc être plus malins.

Soient $0 < \theta < \omega < 1$. Nous avons $(1-\theta)x + \theta y \in [x, (1-\omega)x + \omega y]$, donc nous pouvons écrire $(1-\theta)x + \theta y$ sous la forme $(1-s)x + s\big((1-\omega)x + \omega y\big)$. Il se fait que c'est bon pour $s = \theta/\omega$ (et aussi que nous avons $\theta/\omega < 1$). Donc nous avons

$$f\big((1-\theta)x + \theta y\big) = f\Big((1 - \frac{\theta}{\omega})x + \frac{\theta}{\omega}\big((1-\omega)x + \omega y\big)\Big) \tag{16.385a}$$

$$< (1 - \frac{\theta}{\omega})f(x) + \frac{\theta}{\omega}f\big((1-\omega)x + \omega y\big). \tag{16.385b}$$

Cela nous permet d'écrire

$$\frac{f\big((1-\theta)x+\theta y\big)-f(x)}{\theta}<\frac{f\big((1-\omega)x+\omega y\big)}{\omega}<f(y)-f(x). \tag{16.386}$$

Le seconde inégalité est le pendant de (16.378). Maintenant en passant à la limite pour θ nous conservons une inégalité stricte par rapport à $f(y)-f(x)$:

$$df_x(y-x)<f(y)-f(x). \tag{16.387}$$

□

Avant de lire la proposition suivante, il faut relire la proposition 11.286 et ce qui s'y rapporte. Lire aussi la remarque 16.83 qui indique qu'il n'y a pas de réciproque dans l'énoncé (2).

Proposition 16.94 ([330]).
Soit une fonction $f\colon\Omega\to\mathbb{R}$ deux fois différentiable sur l'ouvert Ω de $\mathbb{R}^n$ et un convexe $U\subset\Omega$.

(1) La fonction f est convexe sur U si et seulement si

$$(d^2f)_x(y-x,y-x)\geqslant 0 \tag{16.388}$$

pour tout $x,y\in U$.

(2) Si pour tout $x\neq y$ dans U nous avons

$$(d^2f)_x(y-x,y-x)>0 \tag{16.389}$$

alors la fonction f est strictement convexe sur U.

Remarque 16.95.
Notons que la condition (16.388) n'est pas équivalente à demander $(d^2f)_x(h,h)\geqslant 0$ pour tout h. En effet nous ne demandons la positivité que dans les directions atteignables comme différence de deux éléments de U. La partie U n'est pas spécialement ouverte ; elle pourrait n'être qu'une droite dans $\mathbb{R}^3$. Dans ce cas, demander que f (qui est C^2 sur l'ouvert Ω) soit convexe sur U ne demande que la positivité de $(d^2f)_x$ appliqué à des vecteurs situés sur la droite U.

Démonstration. Il y a trois parties à démontrer.

(1) sens direct Soit une fonction convexe f sur U. Soient aussi $x,y\in U$ et $h=y-x$. Nous utilisons ma version préférée de Taylor [38] : celui de la proposition 11.382 :

$$f(x+th)=f(x)+tdf_x(h)+\frac{t^2}{2}(d_x^2)(h,h)+t^2\|h\|^2\alpha(th) \tag{16.390}$$

avec $\lim_{s\to 0}\alpha(s)=0$. Le fait que f soit convexe donne

$$0\leqslant f(x+th)-f(x)-tdf_x(h), \tag{16.391}$$

et donc

$$0\leqslant\frac{t^2}{2}(d^2f)_x(h,h)+f^2\|h\|^2\alpha(th). \tag{16.392}$$

En multipliant par 2 et en divisant par t^2,

$$0\leqslant(d^2f)_x(h,h)+2\|h\|^2\alpha(th). \tag{16.393}$$

En prenant $t\to 0$ nous avons bien $(d^2f)_x(y-x,y-x)\geqslant 0$.

38. Si vous présentez ceci au jury d'un concours, vous devriez être capable de raconter ce que signifie d^2f, et pourquoi nous l'utilisons comme une 2-forme.

(1) sens inverse Soient $x, y \in U$. Nous écrivons Taylor en version de la proposition 11.383 :

$$f(y) = f(x) + df_x(y-x) + \frac{1}{2}(d^2f)_z(y-x, y-x) \tag{16.394}$$

pour un certain $z \in]x, y[$. En vertu de ce qui a été dit dans la remarque 16.95 nous ne pouvons pas évoquer l'hypothèse (16.388) pour conclure que $(d^2f)_z(y-x, y-x) \geqslant 0$. Il y a deux manières de nous sortir du problème :

— Trouver $s \in U$ tel que $y - x = s - z$.

— Trouver un multiple de $y - x$ qui soit de la forme $y - x$.

La première approche ne fonctionne pas parce que $s = y - x + z$ n'est pas garanti d'être dans U ; par exemple avec $x = 1$, $z = 2$, $y = 3$ et $U = [0, 3]$. Dans ce cas $s = 4 \notin U$.

Heureusement nous avons $z = \theta x + (1-\theta)y$, donc $z - x = (1-\theta)(y-x)$. Dans ce cas la bilinéarité de $(d^2f)_z$ donne[39]

$$f(y) = f(x) + df_x(y-x) + \underbrace{\frac{1}{2}\frac{1}{(1-\theta)^2}(d^2f)_z(z-x, z-x)}_{\geqslant 0}. \tag{16.395}$$

Nous en déduisons que f est convexe par la proposition 16.93(1).

(2) Le raisonnement que nous venons de faire pour le sens inverse de (1) tient encore, et nous avons

$$f(y) = f(x) + df_x(y-x) + \underbrace{\frac{1}{2}\frac{1}{(1-\theta)^2}(d^2f)_z(z-x, z-x)}_{>0} \tag{16.396}$$

d'où nous déduisons la stricte convexité de f par la proposition 16.93(2).

□

Corolaire 16.96.
Soit un ouvert Ω de $\mathbb{R}^n$ et une fonction deux fois différentiable f sur Ω.

(1) La fonction f est convexe si et seulement si pour tout x, la matrice hessienne d^2f_x est semi-définie positive.

(2) Si pour tout x de Ω, la matrice hessienne d^2f_x est strictement définie positive, alors f est strictement convexe.

Démonstration. Nous pouvons voir ce résultat comme une conséquence directe de la proposition 16.94 en posant $U = \Omega$. Nous allons cependant en donner une démonstration directe.

Soit $a \in \Omega$ et posons la fonction

$$\begin{aligned} g\colon \Omega &\to \mathbb{R} \\ x &\mapsto f(x) - f(a) - (df)_a(x-a). \end{aligned} \tag{16.397}$$

Nous allons calculer des différentielles de f, et une chose importante à comprendre est que la différentielle de la fonction $x \mapsto df_a(x-a)$ ne fait pas intervenir la différentielle seconde de f ; c'est la différentielle de $a \mapsto df_a(x)$ qui demanderait la différentielle seconde de f. Ici la point a étant donné, df_a est une application linéaire sans histoires. En particulier, $df_a(x-a) = df_a(x) - df_a(a)$.

La fonction g vérifie :

(1) $g(a) = 0$,

(2) $dg_x = df_x - df_a$, parce que la différentielle de $x \mapsto df_a(x)$ est $x \mapsto df_a(x)$ en vertu du lemme 10.131.

(3) $dg_a = 0$. Le point a est un point critique de g.

39. Si vous avez bien suivi, la bilinéarité est contenue dans la proposition 11.286.

(4) $d^2 g_x = d^2 f_x$ parce que la différentielle de $x \mapsto df_a$ est nulle.

Ceci étant dit, nous pouvons commencer avec la preuve.

(1) sens direct Nous supposons que f est convexe. Alors $g(x) \geqslant 0$ pour tout x par la caractérisation 16.93(1). Cela signifie que $x = 0$ est un minimum global de g. Par conséquent la proposition 16.70(3) nous dit que la Hessienne $d^2 f_a$ est semi-définie positive.

(1) sens inverse Nous sommes dans le cas de la proposition 16.70(1). Le point $x = a$ est un minimum local de g, ce qui signifie que $g(x) \geqslant 0$ pour tout x de Ω. Encore une fois la caractérisation 16.93(1) nous permet de conclure.

(2) La fonction g vérifie les conditions de 16.70(2), donc $x = 0$ est un minimum local strict de g. La caractérisation 16.93(2) nous fait conclure que f est strictement convexe.

□

16.97.
Nous rappelons que, avec $p > 1$, la fonction

$$\begin{aligned} f\colon [0, \infty[&\to \mathbb{R} \\ x &\mapsto x^p \end{aligned} \tag{16.398}$$

est strictement croissante. La proposition 11.346 est formelle sur ce point.

Lemme 16.98.
Soient un réel a, et $p > 1$. L'équation

$$|x|^p = ax \tag{16.399}$$

possède au plus deux solutions réelles.

Démonstration. En deux parties suivant le signe de a.

Si $a = 0$ Alors l'équation est $|x|^p = 0$, et l'unique solution est $x = 0$

Si $a > 0$ Si $x < 0$ alors $ax < 0$ et $|x|^p - ax > |x|^p > 0$. Cela prouve que notre équation n'a pas de solutions $x < 0$ lorsque $a > 0$.

Cherchons donc des solutions avec $x \geqslant 0$. D'abord $|x| = x$ et ensuite, en posant $p = 1 + \delta$ ($\delta > 0$), nous avons la factorisation

$$x^p - ax = x(x^\delta - a) = 0. \tag{16.400}$$

Poser $x = 0$ est clairement une solution. Si $x \neq 0$, nous avons le raisonnement suivant. Les réels formant un corps, c'est un anneau intègre qui vérifie alors la règle du produit nul [40]. Cela pour dire que si $x \neq 0$, alors

$$x^\delta = a. \tag{16.401}$$

La proposition 11.350 nous dit que cela a une unique solution dans les réels positifs.

Si $a < 0$ Ce cas se traite de façons similaire [41].

□

Proposition 16.99.
Soit $1 < p < \infty$. La fonction

$$\begin{aligned} f\colon \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto \|x\|^p \end{aligned} \tag{16.402}$$

est strictement convexe.

40. Voir le lemme 1.100 et la définition 1.88.
41. Vérifiez par vous-même et écrivez-moi si il y a un problème (personnellement, je n'ai pas vérifié.).

Démonstration. La preuve va se diviser en deux parties. D'abord nous allons utiliser la matrice Hessienne pour démontrer le résultat sur l'ouvert $\mathbb{R}^n\backslash\{0\}$, et ensuite nous allons un peu bricoler pour ajouter 0 au domaine de stricte convexité [42].

La Hessienne Nous notons

$$f_p(x) = \|x\|^p = \big(\sum_i x_i^2\big)^{p/2}, \tag{16.403}$$

et nous dérivons :

$$\frac{\partial f_p}{\partial x_i}(x) = \frac{p}{2}2x_i\big(\sum_i x_i^2\big)^{(p-2)/2} = px_i f_{p-2}(x). \tag{16.404}$$

Cela n'est déjà pas bien défini en $x = 0$ lorsque $p < 2$, mais qu'importe ? Nous dérivons encore en utilisant entre autres la formule (16.404) elle-même avec $p \to p-2$:

$$\frac{\partial^2 f_p}{\partial x_j \partial x_i}(x) = p_{\delta_{ij}}\|x\|^{p-2} + p(p-2)x_i x_j \|p-4\|. \tag{16.405}$$

Nous avons donc la matrice Hessienne

$$H_{ij}(x) = p\delta_{ij}\|x\|^{p-2} + p(p-2)x_i x_j \|x\|^{p-4}. \tag{16.406}$$

Pour prouver que cette matrice est strictement définie positive, nous avons le choix entre la proposition 9.155(1) ou le lemme 9.157(1). Nous utilisons le second. Nous avons [43] :

$$\begin{aligned}
y \cdot H(x)y &= \sum_{kl} y_k H(x)_{kl} y_l && (16.407a)\\
&= p\sum_l y_l^2\|x\|^{p-2} + p(p-2)\|x\|^{p-4}\sum_{kl} x_k y_k x_l y_l && (16.407b)\\
&= p\|y\|^2\|x\|^{p-2} + p(p-2)\|x\|^{p-4}(x\cdot y)^2 && (16.407c)\\
&= p\|x\|^{p-4}\big(\|y\|^2\|x\|^2 + (p-2)(x\cdot y)^2\big) && (16.407d)\\
&> p\|x\|^{p-4}\big(\|y\|^2\|x\|^2 - (x\cdot y)^2\big) && (16.407e)\\
&\geqslant 0. && (16.407f)
\end{aligned}$$

Justifications :

— Pour 16.407e. Vu que $p > 1$ nous avons $p - 2 > -1$. Là, l'inégalité est stricte et c'est important.

— Pour 16.407f. C'est l'inégalité de Cauchy-Schwarz du théorème 7.191.

Voila. La matrice Hessienne est strictement définie positive par le lemme 9.157(1) sur $\mathbb{R}^n\backslash\{0\}$. Le corolaire 16.96(1) nous indique que pour tout $x, y \in \mathbb{R}^n\backslash\{0\}$ et pour tout $\theta \in]0,1[$,

$$f_p\big(\theta x + (1-\theta)y\big) < \theta f_p(x) + (1-\theta)f_p(y) \tag{16.408}$$

pourvu que $\theta x + (1-\theta)y \neq 0$ pour tout θ.

La suite Nous devons prouver que l'inéquation (16.408) tient également lorsque $\theta x + (1-\theta)y = 0$ pour une certaine valeur $\theta = \theta_0 \in]0,1[$.

Les cordes passant par zéro Pour ce faire, nous allons montrer que le segment de droite joignant $\big(a, f_p(a)\big)$ à $(b, f_p(b))$ est toujours au-dessus de la courbe $\big(x, f_p(x)\big)$. Nous commençons par $b = 0$. Vu que $p > 1$ et que $\theta \in]0,1[$ nous avons

$$\|\theta a\|^p = |\theta|^p\|a\|^p < \theta\|a\|^p = \theta f_p(a). \tag{16.409}$$

42. Si quelqu'un sait comment éviter ce bricolage en deux parties, je suis preneur.
43. Si vous ne savez pas où placer les indices, voyez la proposition 4.123.

Les cordes passant au-dessus de zéro Dans le cas $b \neq 0$ nous considérons

$$\begin{aligned} l_1 \colon [a,b] &\to \mathbb{R} \\ x &\mapsto l_1(x) \end{aligned} \tag{16.410}$$

tel que $\big(x, l_1(x)\big)$ soit (le segment) la droite joignant $\big(a, \|a\|^p\big)$ à $\big(b, \|b\|^p\big)$.

Nous considérons aussi, pour $x \in [0,a]$ la fonction l_2 telle que $\big(x, l_2(x)\big)$ soit la droite joignant $(0,0)$ à $\big(a, \|a\|^p\big)$. Nous avons déjà vu que pour $x \in]0,a[$ nous avons $l_1(x) > \|x\|^p$.

Nous avons $l_1(a) = l_2(a) = \|a\|^p$. Donc $l_2(x) - l_1(x)$ ne change pas de signe sur $[0,a]$. Mais comme $l_1(0) > 0 = l_2(0)$ nous avons

$$l_2(x) > l_1(x) \tag{16.411}$$

pour tout $x \in [0,a]$.

Au final, $l_2(x) > l_1(x) > f_p(x)$.

Pour la partie $[b,0]$ nous faisons de même en considérant l_3 de telle sorte que $\big(x, l_3(x)\big)$ soit le segment joignant $(0,0)$ à $\big(b, \|b\|^p\big)$.

□

16.11.6 Quelques inégalités

16.11.6.1 Inégalité de Jensen

Proposition 16.100 (Inégalité de Jensen)**.**
Soit $f \colon \mathbb{R} \to \mathbb{R}$ une fonction convexe et des réels $x_1, \dots, x_n$. Soient des nombres positifs $\lambda_1, \dots, \lambda_n$ formant une combinaison convexe[44]*. Alors*

$$f\big(\sum_i \lambda_i x_i\big) \leqslant \sum_i \lambda_i f(x_i). \tag{16.412}$$

Démonstration. Nous procédons par récurrence sur n, en sachant que $n = 2$ est la définition de la convexité de f. Vu que

$$\sum_{k=1}^{n} \lambda_k x_k = \lambda_n x_n + (1-\lambda_n) \sum_{k=1}^{n-1} \frac{\lambda_k x_k}{1-\lambda_n}, \tag{16.413}$$

nous avons

$$f\big(\sum_{k=1}^{n} \lambda_k x_k\big) \leqslant \lambda_n f(x_n) + (1-\lambda_n) f\big(\sum_{k=1}^{n-1} \frac{\lambda_k x_k}{1-\lambda_n}\big). \tag{16.414}$$

La chose à remarquer est que les nombres $\frac{\lambda_k}{1-\lambda_n}$ avec k allant de 1 à $n-1$ forment eux-mêmes une combinaison convexe. L'hypothèse de récurrence peut donc s'appliquer au second terme du membre de droite :

$$f\big(\sum_{k=1}^{n} \lambda_k x_k\big) \leqslant \lambda_n f(x_n) + (1-\lambda_n) \sum_{k=1}^{n-1} \frac{\lambda_k}{1-\lambda_n} f(x_k) = \lambda_n f(x_n) + \sum_{k=1}^{n-1} \lambda_k f(x_k). \tag{16.415}$$

□

44. Définition 8.29.

16.11.6.2 Inégalité arithmético-géométrique

La proposition suivante dit que la moyenne arithmétique de nombres strictement positifs est supérieure ou égale à la moyenne géométrique.

Proposition 16.101 (Inégalité arithmético-géométrique[336]).
Soient $x_1, \dots, x_n$ des nombres strictement positifs. Nous posons

$$m_a = \frac{1}{n}(x_1 + \cdots + x_n) \tag{16.416}$$

et

$$m_g = \sqrt[n]{x_1 \dots x_n} \tag{16.417}$$

Alors $m_g \leqslant m_a$ et $m_g = m_a$ si et seulement si $x_i = x_j$ pour tout i, j.

Démonstration. Par hypothèse les nombres m_a et m_g sont tout deux strictement positifs, de telle sorte qu'il est équivalent de prouver $\ln(m_g) \leqslant \ln(m_a)$ ou encore

$$\frac{1}{n}\big(\ln(x_1) + \cdots + \ln(x_n)\big) \leqslant \ln\left(\frac{x_1 + \cdots + x_n}{n}\right). \tag{16.418}$$

Cela n'est rien d'autre que l'inégalité de Jensen de la proposition 16.100 appliquée à la fonction ln et aux coefficients $\lambda_i = \frac{1}{n}$. □

16.11.6.3 Inégalité de Kantorovitch

Proposition 16.102 (Inégalité de Kantorovitch[337]).
Soit A une matrice symétrique strictement définie positive dont les plus grandes et plus petites valeurs propres sont λ_{min} et λ_{max}. Alors pour tout $x \in \mathbb{R}^n$ nous avons

$$\langle Ax, x\rangle\langle A^{-1}x, x\rangle \leqslant \frac{1}{4}\left(\frac{\lambda_{min}}{\lambda_{max}} + \frac{\lambda_{max}}{\lambda_{min}}\right)^2 \|x^4\|. \tag{16.419}$$

Démonstration. Sans perte de généralité nous pouvons supposer que $\|x\| = 1$. Nous diagonalisons [45] la matrice A par la matrice orthogonale $P \in \mathrm{O}(n, \mathbb{R})$: $A = PDP^{-1}$ et $A^{-1} = PD^{-1}P^{-1}$ où D est une matrice diagonale formée des valeurs propres de A.

Nous posons $\alpha = \sqrt{\lambda_{min}\lambda_{max}}$ et nous regardons la matrice

$$\frac{1}{\alpha}A + \alpha A^{-1} \tag{16.420}$$

dont les valeurs propres sont

$$\frac{\lambda_i}{\alpha} + \frac{\alpha}{\lambda_i} \tag{16.421}$$

parce que les vecteurs propres de A et de A^{-1} sont les mêmes (ce sont les valeurs de la diagonale de D). Nous allons quelque peu étudier la fonction

$$\theta(x) = \frac{x}{\alpha} + \frac{\alpha}{x}. \tag{16.422}$$

Elle est convexe en tant que somme de deux fonctions convexes. Elle a son minimum en $x = \alpha$ et ce minimum vaut $\theta(\alpha) = 2$. De plus

$$\theta(\lambda_{max}) = \theta(\lambda_{min}) = \sqrt{\frac{\lambda_{min}}{\lambda_{max}}} + \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}. \tag{16.423}$$

45. Théorème spectral 9.150.

Une fonction convexe passant deux fois par la même valeur doit forcément être plus petite que cette valeur entre les deux[46] : pour tout $x \in [\lambda_{min}, \lambda_{max}]$,

$$\theta(x) \leqslant \sqrt{\frac{\lambda_{min}}{\lambda_{max}}} + \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}. \tag{16.424}$$

Nous sommes maintenant en mesure de nous lancer dans l'inégalité de Kantorovitch.

$$\sqrt{\langle Ax, x\rangle\langle A^{-1}x, x\rangle} \leqslant \frac{1}{2}\left(\frac{\langle Ax, x\rangle}{\alpha} + \alpha\langle A^{-1}x, x\rangle\right) \tag{16.425a}$$

$$= \frac{1}{2}\langle(\frac{A}{\alpha} + \alpha A^{-1})x, x\rangle \tag{16.425b}$$

$$\leqslant \frac{1}{2}\|(\frac{A}{\alpha} + \alpha A^{-1})x\|\|x\| \tag{16.425c}$$

$$\leqslant \frac{1}{2}\|\frac{A}{\alpha} + \alpha A^{-1}\| \tag{16.425d}$$

Justifications :

— 16.425a par l'inégalité arithmético-géométrique, proposition 16.101. Nous avons aussi inséré $\alpha\frac{1}{\alpha}$ dans le produit sous la racine.

— 16.425c par l'inégalité de Cauchy-Schwarz, théorème 7.191.

— 16.425d par la définition de la norme opérateur de la proposition 10.3

La norme opérateur est la plus grande des valeurs propres. Mais les valeurs propres de $A/\alpha + \alpha A^{-1}$ sont de la forme $\theta(\lambda_i)$, et tous les λ_i sont entre λ_{min} et λ_{max}. Donc la plus grande valeur propre de $A/\alpha + \alpha A^{-1}$ est $\theta(x)$ pour un certain $x \in [\lambda_{min}, \lambda_{max}]$. Par conséquent

$$\sqrt{\langle Ax, x\rangle\langle A^{-1}x, x\rangle} \leqslant \frac{1}{2}\|\frac{A}{\alpha} + \alpha A^{-1}\| \leqslant \sqrt{\frac{\lambda_{min}}{\lambda_{max}}} + \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}. \tag{16.426}$$

□

16.11.6.4 Inégalité de Hölder

Si vous cherchiez l'inégalité de Hölder dans L^p, c'est la proposition 26.33. Les normes ℓ^p sont définies dans 7.219.

Lemme 16.103 ([338]).
Soient $x, y > 0$ ainsi que $\alpha, \beta > 0$ tels que $\alpha + \beta = 1$. Alors

$$xy \leqslant \alpha x^{1/\alpha} + \beta e^{1/\beta}. \tag{16.427}$$

Nous avons

$$xy = \alpha x^{1/\alpha} + \beta e^{1/\beta} \tag{16.428}$$

si et seulement si $x^{1/\alpha} = y^{1/\beta}$.

Démonstration. Nous utilisons le logarithme[47] et ses propriétés (surtout la proposition 14.60). D'abord

$$xy = e^{\ln(xy)} = e^{\ln(x)+\ln(y)} = e^{\alpha\frac{\ln(x)}{\alpha}+\frac{\ln(y)}{\beta}} \tag{16.429}$$

Vu que l'exponentielle est strictement convexe (exemple 16.84(3)) et vu que $\alpha + \beta = 1$, nous avons

$$xye^{\alpha\frac{\ln(x)}{\alpha}+\frac{\ln(y)}{\beta}} \leqslant \alpha e^{\ln(x)/\alpha} + \beta e^{\ln(y)/\beta} = \alpha x^{1/\alpha} + \beta y^{1/\beta}. \tag{16.430}$$

46. Je ne suis pas certain que cette phrase soit claire, non ?
47. Définition 14.58

Vu que α et β ne sont pas nuls, l'inégalité (16.430) est une égalité si et seulement si

$$\frac{\ln(x)}{\alpha} = \frac{\ln(y)}{\beta}. \tag{16.431}$$

Cela signifie $\ln(x^{1/\alpha}) = \ln(y^{1/\beta})$, qui implique $x^{1/\alpha} = y^{1/\beta}$ parce que le logarithme est une bijection. □

Corolaire 16.104.
Si $p, q > 0$ vérifient $\frac{1}{p} + \frac{1}{q} = 1$ et si $p > 1$ alors nous avons

$$xy \leqslant \frac{1}{p}x^p + \frac{1}{q}y^q \tag{16.432}$$

pour tout $x, y \geqslant 0$, avec une égalité si et seulement si $x^p = y^q$.

Démonstration. Il suffit de poser $\alpha = 1/p$ et $\beta = 1/q$ et appliquer le lemme 16.103. □

Théorème 16.105 (Inégalité de Hölder[338])**.**
Soient $p, q > 0$ tels que $\frac{1}{p} + \frac{1}{q} = 1$. Pour tout $x, y \in \mathbb{R}^n$ nous avons

$$|x \cdot y| = \sum_{i=1}^{n} |x_i y_i| \leqslant \|x\|_p \|x\|_q. \tag{16.433}$$

Il y a égalité si et seulement si $x_i y_i$ est de signe constant[48] *et les vecteurs $\sum_i |x_i|^p e_i$ et $\sum_i |y_i|^q e_i$ sont proportionnels.*

Démonstration. En plusieurs parties.

Le cas des vecteurs nuls Si x ou y est nul, les inégalités sont évidentes. Donc nous supposons que non.

Première inégalité En ce qui concerne la première inégalité,

$$|x \cdot y| = |\sum_i x_i y_i| \leqslant \sum_i |x_i||y_i| \leqslant \sum_i \left(\frac{1}{p}|x_i|^p + \frac{1}{q}|y_i|^q\right) \tag{16.434}$$

où nous avons utilisé le corolaire 16.104 dans chaque terme de la somme en tenant compte du fait que $|x_i|$ et $|y_i|$ sont positifs.

Seconde inégalité Pour la seconde inégalité, nous commençons avec $\|x\|_p = \|y\|_q = 1$. Utilisant encore le corolaire 16.104 pour chaque terme, nous avons

$$\sum_i |x_i y_i| \leqslant \frac{1}{p}\underbrace{\sum_i |x_i|^p}_{=1} + \frac{1}{q}\underbrace{\sum_i |y_i|^q}_{=1} = \frac{1}{p} + \frac{1}{q} = 1 = \|x\|_p \|y\|_q. \tag{16.435}$$

Si maintenant x et y sont arbitraires non nuls dans $\mathbb{R}^n$, nous posons $x' = x/\|x\|_p$ et $y' = y/\|y\|_q$; nous savons déjà que

$$\sum_i |x'_i y'_i| \leqslant 1. \tag{16.436}$$

En remplaçant x'_i par $x_i/\|x\|_p$ et y'_i par $y_i/\|y\|_q$, l'inégalité (16.436) devient

$$\sum_i \frac{|x_i|y_i}{\|x\|_p \|y\|_q} \leqslant 1, \tag{16.437}$$

ce qui signifie

$$\sum_i |x_i y_i| \leqslant \|x\|_p \|y\|_q. \tag{16.438}$$

48. Je n'utilie pas cette hypothèse de signe constant. Il doit y avoir une subtilité qui m'a échappée. Soyez prudente en lisant et écrivez-moi si vous trouvez une erreur.

Cas d'égalité, dans un sens Nous notons

$$\mathcal{D} = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n \text{ tel que } \sum_i |x_i y_i| = \|x\|_p \|y\|_q\}. \tag{16.439}$$

Multiplications D'abord si $(x,y) \in \mathcal{D}$, alors $(\mu x, \lambda y) \in \mathcal{D}$ pour tout $\mu, \lambda \in \mathbb{R}$. En effet,

$$\sum_i |(\mu x)_i (\lambda y)_i| = |\mu||\lambda| \sum_i |x_i y_i| = |\mu||\lambda| \|x\|_p \|y\|_q = \|\mu x\|_p \|\lambda y\|_q. \tag{16.440}$$

Avec normes égales à 1 Soit $(x,y) \in \mathcal{D}$ tels que $\|x\|_p = \|y\|_q = 1$. Nous avons en particulier,

$$1 = \sum_i |x_i y_i| \leqslant \frac{1}{p} \sum_i |x_i|^p + \frac{1}{q} \sum_i |y_i|^q \tag{16.441}$$

grâce à l'inégalité (16.432) appliquée à chaque terme. Vu que $\|x\|_p = 1$, nous avons $\sum_i |x_i|^p = 1$, de telle sorte que le membre de droite de (16.441) se réduise à $\frac{1}{p} + \frac{1}{q} = 1$. Nous pouvons donc écrire

$$1 = \sum_i |x_i y_i| \leqslant \frac{1}{p} \sum_i |x_i|^p + \frac{1}{q} \sum_i |y_i|^q = 1. \tag{16.442}$$

L'inégalité est donc une égalité :

$$\sum_i |x_i y_i| = \sum_i \left(\frac{1}{p} |x_i|^p + \frac{1}{q} |y_i|^q \right). \tag{16.443}$$

Mais chaque terme à gauche est en inégalité avec le terme correspondant à droite :

$$|x_i y_i| \leqslant \frac{1}{p} |x_i|^p + \frac{1}{q} |y_i|^q. \tag{16.444}$$

Pour que le tout soit une égalité, il faut que chaque inégalité (16.444) soit une égalité. Pour chaque i, nous avons

$$|x_i y_i| = \frac{1}{p} |x_i|^p + \frac{1}{q} |y_i|^q. \tag{16.445}$$

La condition d'égalité du corolaire 16.104 nous dit alors que $|x_i|^p = |y_i|^q$.

Avec normes arbitraires Soit donc $(x,y) \in \mathcal{D}$. Nous savons qu'en posant $x' = x/\|x\|_p$ et $y' = y/\|y\|_q$ nous avons $(x',y') \in \mathcal{D}$ et donc

$$\left(\frac{|x_i|}{\|x\|_p} \right)^p = \left(\frac{|y_i|}{\|y\|_q} \right)^q. \tag{16.446}$$

Cela donne tout de suite

$$|x_i|^p = \frac{\|x\|_p^p}{\|y\|_q^q} |y_i|^q, \tag{16.447}$$

ce qui est bien ce que nous voulions : le vecteur $\sum_i |x_i|^p e_i$ est proportionnel au vecteur $\sum_i |y_i|^q e_i$.

Cas d'égalité dans l'autre sens[1] Nous supposons que les vecteurs $\sum_i |x_i|^p e_i$ et $\sum_i |y_i|^q e_i$ sont proportionnels. Nous nommons c^q le facteur de proportionnalité, c'est-à-dire que nous posons

$$|x_i|^p = c^q |y_i|^q. \tag{16.448}$$

Dans ce cas, pour chaque i, les nombres $c|y_i|$ et $|x_i|$ sont dans le cas d'égalité du corolaire 16.104. Nous avons alors

$$\sum_i |x_i y_i| = \frac{1}{c}\sum_i c|x_i||y_i| \tag{16.449a}$$

$$= \frac{1}{c}\left(\frac{1}{p}|x_i|^p + \frac{1}{q}\big(c|y_i|\big)^q\right) \tag{16.449b}$$

$$= \frac{1}{c}\sum_i \left(\frac{1}{p}|x_i|^p + \frac{1}{q}|x_i|^p\right) \tag{16.449c}$$

$$= \frac{1}{c}\sum_i |x_i|^p. \tag{16.449d}$$

Et c'est maintenant que nous subdivisons.

Si $\|x\|_p = \|y\|_q = 1$ Dans ce cas, l'égalité (16.449) se réduisent à

$$\sum_i |x_i y_i| = \frac{1}{c}. \tag{16.450}$$

Mais l'hypothèse sur les normes donne

$$1 = \sum_i |x_i|^p = \sum_i c^q |y_i|^q = c^q \sum_i |y_i|^q = c^q. \tag{16.451}$$

Donc $c = 1$ et nous avons bien

$$\sum_i |x_i y_i| = \frac{1}{c} = 1 = \|x\|_p \|y\|_q. \tag{16.452}$$

Pour des normes arbitraires Soit $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ tels que $|x_i|^p = c^q |y_i|^q$. Nous posons comme d'habitude $x' = x/\|x\|_p$ et $y' = y/\|y\|_q$. En utilisant le cas « de norme 1 » nous avons

$$1 = \sum_i |x'_i y'_i| = \frac{1}{\|x\|_p \|y_q\|}\sum_i |x_i y_i|. \tag{16.453}$$

Donc $\sum_i |x_i y_i| = \|x\|_p \|y\|_q$ comme nous le voulions.

□

La majoration de la proposition suivante sera utile pour les inégalités de Clarkson du lemme 26.123. Pour d'autres inégalités (plus simples) autour des normes $\|.\|_p$, voir le thème 7.

Proposition 16.106 ([338]).
Si $x \in \mathbb{R}^n$ et si $0 < q < p$, alors

$$\|x\|_q \leqslant n^{\frac{1}{q}-\frac{1}{p}} \|x\|_p. \tag{16.454}$$

En particulier, si $0 < q < p$, alors

$$\|x\|_p \leqslant \|x\|_p. \tag{16.455}$$

Démonstration. Nous posons $P = p/q$ et $Q = P/(P-1)$. Les nombres P et Q sont des exposants conjugués, parce que

$$\frac{1}{P} + \frac{1}{Q} = \frac{q}{p} + \frac{p-q}{p} = 1. \tag{16.456}$$

Nous posons $y = (1, \dots, 1) \in \mathbb{R}^n$ ainsi que

$$v = \sum_i |x_i|^q e_i, \tag{16.457}$$

et nous écrivons l'inégalité de Hölder de la proposition 16.105 sur les vecteurs v et y :

$$\sum_i |v_i y_i| \leqslant \|v\|_P \|y\|_Q. \tag{16.458}$$

En déballant,

$$\sum_i |x_i|^q \leqslant \left(\sum_i (|x_i|^q)^P\right)^{1/P} \underbrace{\left(\sum_i 1^Q\right)}_{=n}^{1/Q} \tag{16.459a}$$

$$= \left(\sum_i |x_i|^p\right)^{q/p} n^{1/Q} \tag{16.459b}$$

$$= n^{1-q/p}\|x\|_p^q. \tag{16.459c}$$

Cela donne

$$\|x\|_q^q \leqslant n^{1-q/p}\|x\|_p^q. \tag{16.460}$$

En prenant la puissance $1/q$ des deux côtés,

$$\|x\|_q \leqslant n^{\frac{1}{q}-\frac{1}{p}}\|x\|_p \tag{16.461}$$

□

16.12 Trucs et astuces de calcul d'intégrales

Afin d'alléger le texte de calculs parfois un peu longs, nous regroupons ici les intégrales à une variable que nous devons utiliser dans les autres parties du cours.

16.12.1 Quelques intégrales « usuelles »

(1) L'intégrale

$$\boxed{I = \int x\ln(x)dx = \frac{x^2}{2}\big(\ln(x) - \frac{1}{2}\big)} \tag{16.462}$$

se fait par partie en posant

$$\begin{aligned} u &= \ln(x), \quad dv = x\,dx \\ du &= \frac{1}{x}\,dx, \qquad v = \frac{x^2}{2}, \end{aligned} \tag{16.463}$$

et ensuite

$$I = \ln(x)\frac{x^2}{2} - \int \frac{x}{2} = \frac{x^2}{2}\big(\ln(x) - \frac{1}{2}\big). \tag{16.464}$$

(2) L'intégrale

$$\boxed{I = \int x\ln(x^2)dx = x^2\ln(x) - \frac{x^2}{2}.} \tag{16.465}$$

En utilisant le fait que $\ln(u^2) = 2\ln(u)$, nous retombons sur une intégrale du type (1) :

$$I = x^2\ln(x) - \frac{x^2}{2}. \tag{16.466}$$

(3) L'intégrale

$$\boxed{I = \int x\ln(1+x^2)dx = \frac{1}{2}\ln(x^2+1)(x^2+1) - x^2 - \frac{1}{2}} \tag{16.467}$$

se traite en posant $v = 1 + x^2$ de telle sorte à avoir $dx = \frac{dv}{2x}$ et donc

$$I = \frac{1}{2}\ln(x^2+1)(x^2+1) - x^2 - \frac{1}{2}. \tag{16.468}$$

(4) L'intégrale

$$I = \int \cos(\theta)\sin(\theta)\ln\left(1+\frac{1}{\cos^2(\theta)}\right)d\theta \tag{16.469}$$

demande le changement de variable $u = \cos(\theta)$, $d\theta = -\frac{du}{\sin(\theta)}$. Nous tombons sur l'intégrale

$$I = -\int u\ln\left(\frac{1+u^2}{u^2}\right) = -\int u\ln(1+u^2) + \int u\ln(u^2), \tag{16.470}$$

qui sont deux intégrales déjà faites. Nous trouvons

$$I = -\frac{1}{2}\ln\left(\frac{\sin^2(\theta)-1}{\sin^2(\theta)-2}\right)\sin^2(\theta) - \ln\left(\sin^2(\theta)-2\right) + \frac{1}{2}\ln\left(\sin^2(\theta)-1\right) \tag{16.471}$$

(5) L'intégrale

$$\boxed{\int \frac{r^3}{1+r^2}dr = \frac{r^2}{2} - \frac{1}{2}\ln(r^2+1).} \tag{16.472}$$

commence par faire la division euclidienne de r^3 par r^2+1 ; ce que nous trouvons est $r^3 = (r^2+1)r - r$. Il reste à intégrer

$$\int \frac{r^3}{1+r^2}dr = \int r\,dr - \int \frac{r}{1+r^2}dr. \tag{16.473}$$

La fonction dans la seconde intégrale est $\frac{r}{1+r^2} = \frac{1}{2}\frac{f'(r)}{f(r)}$ où $f(r) = 1+r^2$, et donc $\int \frac{r}{1+r^2} = \frac{1}{2}\ln(1+r^2)$. Au final,

$$I = \frac{1}{2}r^2 - \frac{1}{2}\ln(r^2+1). \tag{16.474}$$

(6) L'intégrale

$$\boxed{I = \int \cos(\theta)\sin(\theta)d\theta = \frac{\sin^2(\theta)}{2}} \tag{16.475}$$

se traite par le changement de variable $u = \sin(\theta)$, $du = \cos(\theta)d\theta$, et donc

$$\int \cos(\theta)\sin(\theta)d\theta = \int u du = \frac{u^2}{2} = \frac{\sin^2(\theta)}{2}. \tag{16.476}$$

(7) L'intégrale

$$\boxed{\int \sqrt{1+x^2}dx = \frac{x}{2}\sqrt{1+x^2} + \frac{1}{2}\operatorname{arcsinh}(x)} \tag{16.477}$$

s'obtient en effectuant le changement de variable $u = \sinh(\xi)$.

(8) L'intégrale

$$\boxed{\int \cos^2(x)\sin^2(x)dx = \frac{x}{8} - \frac{\sin(4x)}{32}} \tag{16.478}$$

s'obtient à coups de formules de trigonométrie. D'abord, $\sin(t)\cos(t) = \frac{1}{2}\sin^2(2t)$ fait en sorte que la fonction à intégrer devient

$$f(x) = \frac{1}{4}\sin^2(x). \tag{16.479}$$

Ensuite nous utilisons le fait que $\sin^2(t) = (1-\cos(2t))/2$ pour transformer la formule à intégrer en

$$f(x) = \frac{1-\cos(4x)}{8}. \tag{16.480}$$

Cela s'intègre facilement en posant $u = 4x$, et le résultat est

$$\int f(x)dx = \frac{x}{8} - \frac{\sin(4x)}{32}. \tag{16.481}$$

(9) La fonction

$$\operatorname{sinc}(x) = \frac{\sin(x)}{x} \tag{16.482}$$

est le **sinus cardinal** de x. Nous allons montrer que

$$\boxed{\int_0^\infty \big|\operatorname{sinc}(x)\big| dx = \infty}. \tag{16.483}$$

D'abord nous avons

$$\int_{(n-1)\pi}^{n\pi} \frac{\big|\sin(t)\big|}{t} dt \geqslant \int_{(n-1)\pi}^{n\pi} \frac{\big|\sin(t)\big|}{n\pi} dt, \tag{16.484}$$

mais par périodicité,

$$\int_{(n-1)\pi}^{n\pi} \big|\sin(t)\big| dt = \int_0^\pi \sin(t) dt = 2. \tag{16.485}$$

Par conséquent

$$\int_0^{n\pi} \big|\operatorname{sinc}(t)\big| dt \geqslant \frac{2}{\pi} \sum_{k=1}^{n} \frac{1}{k}, \tag{16.486}$$

ce qui diverge lorsque $n \to \infty$.

(10) Les intégrales, pour $\epsilon > 0$,

$$\boxed{\int_0^\infty \cos(kx) e^{-\epsilon x} dx = \frac{\epsilon}{k^2 + \epsilon^2}} \tag{16.487}$$

et

$$\boxed{\int_0^\infty \sin(kx) e^{-\epsilon x} dx = \frac{k}{k^2 + \epsilon^2}} \tag{16.488}$$

se calculent deux fois par partie. Nous posons

$$I = \int_0^\infty \cos(kx) e^{-\epsilon x} dx \tag{16.489a}$$

$$J = \int_0^\infty \sin(kx) e^{-\epsilon x} dx. \tag{16.489b}$$

L'intégrale I s'effectue par partie en posant $u = \cos(kx)$ et $v' = e^{-\epsilon x}$. Un peu de calcul montre que

$$I = \frac{1}{\epsilon} - \frac{k}{\epsilon} J. \tag{16.490}$$

Par ailleurs l'intégrale J se fait également par partie pour obtenir

$$J = \frac{k}{\epsilon} I. \tag{16.491}$$

En résolvant pour I et J les deux équations déduites, nous trouvons

$$I = \frac{\epsilon}{k^2 + \epsilon^2} \tag{16.492a}$$

$$J = \frac{k}{k^2 + \epsilon^2}. \tag{16.492b}$$

16.12.2 Reformer un carré au dénominateur

Lorsqu'on a un second degré au dénominateur, le bon plan est de reformer un carré parfait. Par exemple :

$$x^2 + 2x + 2 = (x + 1)^2 + 1. \tag{16.493}$$

Ensuite, le changement de variable $t = x + 1$ est pratique parce que cela donne $t^2 + 1$ au dénominateur.

Cherchons

$$I = \int \frac{1-x}{x^2+2x+2}dx = \int \frac{1-x}{(x+1)^2+1}dx = \int \frac{1-(t-1)}{t^2+1} \tag{16.494}$$

où nous avons fait le changement de variable $t = x+1$, $dt = dx$. L'intégrale se coupe maintenant en deux parties :

$$I = \int \frac{-t}{t^2+1} + \int \frac{2}{t^2+1}. \tag{16.495}$$

La seconde est dans les formulaires et vaut

$$2\arctan(t) = 2\arctan(x+1), \tag{16.496}$$

tandis que la première est presque de la forme f'/f :

$$\int \frac{t}{t^2+1} = \frac{1}{2}\int \frac{2t}{t^2+1} = \frac{1}{2}\ln(t^1+1) = \frac{1}{2}\ln(u^2+2u+2). \tag{16.497}$$

16.12.3 Décomposition en fractions simples

La décomposition en fractions simples décrite en 18.1.6 permet d'intégrer des fractions rationnelles. Elle peut parfois être évitée par la méthode de Rothstein-Trager que nous expliquerons dans 19.14.4.

16.13 Algorithme du gradient à pas optimal

Une idée pour trouver un minimum à une fonction est de prendre un point p au hasard, calculer le gradient $\nabla f(p)$ et suivre la direction $-\nabla f(p)$ tant que ça descend. Une fois qu'on est « dans le creux », recalculer le gradient et continuer ainsi.

Nous allons détailler cet algorithme dans un cas très particulier d'une matrice A symétrique et strictement définie positive.

— Dans la proposition 16.108 nous montrons que résoudre le système linéaire $Ax = -b$ est équivalent à minimiser une certaine fonction.

— La proposition 16.109 donnera une méthode itérative pour trouver ce minimum.

Définition 16.107.
Si X est un espace vectoriel normé et $f\colon X \to \mathbb{R} \cup \{\pm\infty\}$ nous disons que f est ***coercive*** *sur le domaine non borné P de X si pour tout $M \in \mathbb{R}$, l'ensemble*

$$\{x \in P \text{ tel que } f(x) \leqslant M\} \tag{16.498}$$

est borné.

En langage imagé la coercivité de f s'exprime par la limite

$$\lim_{\substack{\|x\|\to\infty \\ x\in P}} f(x) = +\infty. \tag{16.499}$$

Nous rappelons que $S^{++}(n,\mathbb{R})$ est l'ensemble des matrices symétriques strictement définies positives définies en 9.153.

Proposition 16.108.
Soit $A \in S^{++}(n,\mathbb{R})$ et $b \in \mathbb{R}^n$. Nous considérons l'application

$$\begin{aligned} f\colon \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle. \end{aligned} \tag{16.500}$$

Alors :

(1) Il existe un unique $\bar{x} \in \mathbb{R}^n$ tel que $A\bar{x} = -b$.

(2) Il existe un unique $x^* \in \mathbb{R}^n$ minimisant f.

(3) Ils sont égaux : $\bar{x} = x^*$.

Démonstration. Une matrice symétrique strictement définie positive est inversible, entre autres parce qu'elle se diagonalise par des matrices orthogonales (qui sont inversibles) et que la matrice diagonalisée est de déterminant non nul : tous les éléments diagonaux sont strictement positifs. Voir le théorème spectral symétrique 9.150.

D'où l'unicité du $\bar{x}$ résolvant le système $Ax = -b$ pour n'importe quel b.

f est strictement convexe La fonction f s'écrit

$$f(x) = \frac{1}{2}\sum_{kl} A_{kl}x_l x_k + \sum_k b_k x_k. \tag{16.501}$$

Elle est de classe C^2 sans problèmes, et il est vite vu que $\frac{\partial^2 f}{\partial x_i \partial x_j} = A_{ij}$, c'est-à-dire que A est la matrice hessienne de f. Cette matrice étant strictement définie positive par hypothèse, la fonction f est strictement convexe par le corolaire 16.96(2).

f est coercive Montrons à présent que f est coercive. Nous avons :

$$|f(x)| = \left|\frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle\right| \tag{16.502a}$$

$$\geqslant \frac{1}{2}|\langle Ax, x\rangle| - |\langle b, x\rangle| \tag{16.502b}$$

$$\geqslant \frac{1}{2}\lambda_{max}\|x\|^2 - \|b\|\|x\| \tag{16.502c}$$

Pour la dernière ligne nous avons nommé λ_{max} la plus grande valeur propre de A et utilisé Cauchy-Schwarz pour le second terme. Nous avons donc bien $|f(x)| \to \infty$ lorsque $\|x\| \to \infty$ et la fonction f est coercive.

Soit M une valeur atteinte par f. L'ensemble

$$\{x \in \mathbb{R}^n \text{ tel que } f(x) \leqslant M\} \tag{16.503}$$

est fermé (parce que f est continue) et borné parce que f est coercive. Cela est donc compact [49] et f atteint un minimum qui sera forcément dedans. Cela est pour l'existence d'un minimum.

Pour l'unicité du minimum nous invoquons la convexité : si $\bar{x}_1$ et $\bar{x}_2$ sont deux points réalisant le minimum de f, alors

$$f\left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) < \frac{1}{2}f(\bar{x}_1) + \frac{1}{2}f(\bar{x}_2) = f(\bar{x}_1), \tag{16.504}$$

ce qui contredit la minimalité de $f(\bar{x}_1)$.

Nous devons maintenant prouver que $\bar{x}$ vérifie l'équation $A\bar{x} = -b$. Vu que $\bar{x}$ est minimum local de f qui est une fonction de classe C^2, le théorème des minimums locaux 16.69 nous indique que $\bar{x}$ est solution de $\nabla f(x) = 0$. Calculons un peu cela avec la formule

$$df_x(u) = \frac{d}{dt}\Big[f(x+tu)\Big]_{t=0} = \frac{1}{2}\big(\langle Ax, u\rangle + \langle Au, x\rangle\big) + \langle b, u\rangle = \langle Ax, u\rangle + \langle b, u\rangle = \langle Ax + b, u\rangle. \tag{16.505}$$

Donc demander $df_x(u) = 0$ pour tout u demande $Ax + b = 0$. □

Proposition 16.109 (Gradient à pas optimal)**.**
Soit $A \in S^{++}(n, \mathbb{R})$ (A est une matrice symétrique strictement définie positive) et $b \in \mathbb{R}^n$. Nous considérons l'application

$$\begin{aligned} f\colon \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle. \end{aligned} \tag{16.506}$$

49. Théorème 7.127

Soit $x_0 \in \mathbb{R}^n$. Nous définissons la suite (x_k) par

$$x_{k+1} = x_k + t_k d_k \tag{16.507}$$

où

— $d_k = -(\nabla f)(x_k)$

— *t_k est la valeur minimisant la fonction $t \mapsto f(x_k + td_k)$ sur $\mathbb{R}$.*

Alors pour tout $k \geqslant 0$ nous avons

$$\|x_k - \bar{x}\| \leqslant K \left(\frac{c_2(A) - 1}{c_2(A) + 1} \right)^k \tag{16.508}$$

où $c_2(A) = \frac{\lambda_{max}}{\lambda_{min}}$ est le rapport ente la plus grande et la plus petite valeur propre[50] *de la matrice A et $\bar{x}$ est l'unique élément de $\mathbb{R}^n$ à minimiser f.*

Démonstration. Décomposition en plusieurs points.

Existence de $\bar{x}$ Le fait que $\bar{x}$ existe et soit unique est la proposition 16.108.

Si $(\nabla f)(x_k) = 0$ D'abord si $\nabla f(x_k) = 0$, c'est que $x_{k+1} = x_k$ et l'algorithme est terminé : la suite est stationnaire. Pour dire que c'est gagné, nous devons prouver que $x_k = \bar{x}$. Pour cela nous écrivons (à partir de maintenant « x_k » est la k^ecomposante de x qui est une variable, et non le x_k de la suite)

$$f(x) = \frac{1}{2} \sum_{kl} A_{kl} x_l x_k + \sum_k b_k x_k \tag{16.509}$$

et nous calculons $\frac{\partial f}{\partial x_i}(a)$ en tenant compte du fait que $\frac{\partial x_k}{\partial x_i} = \delta_{ki}$. Le résultat est que $(\partial_i f)(a) = (Ax + b)_i$ et donc que

$$(\nabla f)(a) = Aa + b. \tag{16.510}$$

Vu que A est inversible (symétrique définie positive), il existe un unique $a \in \mathbb{R}^n$ qui vérifie cette relation. Par la proposition 16.108, cet élément est le minimum $\bar{x}$.

Cela pour dire que si $a \in \mathbb{R}^n$ vérifie $(\nabla f)(a) = 0$ alors $a = \bar{x}$. Nous supposons donc à partir de maintenant que $\nabla f(x_k) \neq 0$ pour tout k.

t_k est bien défini Pour $t \in \mathbb{R}$ nous avons

$$f(x_k + td_k) = f(x_k) + \frac{1}{2} t^2 \langle Ad_k, d_k \rangle + t \langle \underbrace{Ax_k + b}_{=-d_k}, d_k \rangle = \frac{1}{2} t^2 \langle Ad_k, d_k \rangle - t_k \|d_k\|^2 + f(x_k). \tag{16.511}$$

qui est un polynôme du second degré en t. Le coefficient de t^2 est $\frac{1}{2}\langle Ad_k, d_k \rangle > 0$ parce que $d_k \neq 0$ et A est strictement définie positive. Par conséquent la fonction $t \mapsto f(x_k + td_k)$ admet bien un unique minimum. Nous pouvons même calculer t_k parce que l'on connaît pas cœur le sommet d'une parabole :

$$t_k = -\frac{\langle Ax_k + b, d_k \rangle}{\langle Ad_k, d_k \rangle} = \frac{\|d_k\|^2}{\langle Ad_k, d_k \rangle} \tag{16.512}$$

parce que $d_k = -\nabla f(x_k) = -(Ax_k + b)$.

La valeur de d_{k+1} Par définition, $d_{k+1} = -\nabla f(x_{k+1}) = -(Ax_{k+1} + b)$. Mais $x_{k+1} = x_k + t_k d_k$, donc

$$d_{k+1} = -Ax_k - t_k Ad_k - b = d_k - t_k Ad_k \tag{16.513}$$

parce que $-Ax_k - b = d_k$.

Par ailleurs, $\langle d_{k+1}, d_k \rangle = 0$ parce que

$$\langle d_{k+1}, d_k \rangle = \langle d_k, d_k \rangle - t_k \langle d_k, Ad_k \rangle = \|d_k\|^2 - \frac{\|d_k\|^2}{\langle Ad_k, d_k \rangle} \langle d_k, Ad_k \rangle = 0 \tag{16.514}$$

où nous avons utilisé la valeur (16.512) de t_k.

50. Cela est certainement très lié au conditionnement de la matrice A, voir la proposition 33.108.

Calcul de $f(x_{k+1})$ Nous repartons de (16.511) où nous substituons la valeur (16.512) de t_k :

$$f(x_{k+1}) = f(x_k) + \frac{1}{2}\frac{\|d_k\|^4}{\langle Ad_k, d_k\rangle} - \frac{\|d_k\|^4}{\langle Ad_k, d_k\rangle} = f(x_k) - \frac{1}{2}\frac{\|d_k\|^4}{\langle Ad_k, d_k\rangle}. \tag{16.515}$$

Encore du calcul ... Vu que le produit $\langle Ad_k, d_k\rangle$ arrive tout le temps, nous allons étudier $\langle A^{-1}d_k, d_k\rangle$. Le truc malin est d'essayer d'exprimer ça en termes de $\bar{x}$ et $\bar{f} = f(\bar{x})$. Pour cela nous calculons $f(\bar{x})$:

$$\bar{f} = f(\bar{x}) = f(-A^{-1}b) = -\frac{1}{2}\langle b, A^{-1}b\rangle. \tag{16.516}$$

Ayant cela en tête nous pouvons calculer :

$$\begin{aligned}
\langle A^{-1}d_k, d_k\rangle &= \langle A^{-1}(Ax_k + b), Ax_k + b\rangle && (16.517a)\\
&= \langle x_k, Ax_k\rangle + \langle A^{-1}b, Ax_k\rangle + \langle b, x_k\rangle + \underbrace{\langle A^{-1}b, b\rangle}_{-2\bar{f}} && (16.517b)\\
&= \langle x_k, Ax_k\rangle + 2\langle x_k, b\rangle - 2\bar{f} && (16.517c)\\
&= 2\big(f(x_k) - \bar{f}\big) && (16.517d)
\end{aligned}$$

où nous avons utilisé le fait que $\langle x, Ay\rangle = \langle Ax, y\rangle$ parce que A est symétrique.

Erreur sur la valeur du minimum Nous voulons à présent estimer la différence $f(x_{k+1}) - \bar{f}$. Pour cela nous mettons en facteur $f(x_k) - \bar{f}$ dans $f(x_{k+1} - \bar{f})$; et d'ailleurs c'est pour cela que nous avons calculé $\langle A^{-1}d_k, d_k\rangle$: parce que ça fait intervenir $f(x_k) - \bar{f}$.

$$\begin{aligned}
f(x_{k+1}) - \bar{f} &= f(x_k) - \frac{1}{2}\frac{\|d_k\|^4}{\langle Ad_k, d_k\rangle} - \bar{f} && (16.518a)\\
&= \big(f(x_k) - \bar{f}\big)\left(1 - \frac{1}{2}\frac{\|d_k\|^4}{\langle Ad_k, d_k\rangle\big(f(x_k) - \bar{f}\big)}\right) && (16.518b)\\
&= \big(f(x_k) - \bar{f}\big)\left(1 - \frac{\|d_k\|^4}{\langle Ad_k, d_k\rangle\langle A^{-1}d_k, d_k\rangle}\right). && (16.518c)
\end{aligned}$$

Nous traitons le dénominateur à l'aide de l'inégalité de Kantorovitch 16.102. Nous avons

$$\frac{\|d_k\|^4}{\langle Ad_k, d_k\rangle\langle A^{-1}d_k, d_k\rangle} \geqslant \frac{\|d_k\|^4}{\frac{1}{4}\left(\sqrt{c_2(A)} + \frac{1}{\sqrt{c_2(A)}}\right)^2\|d_k\|^4} = \frac{4c_2(A)}{(c_2(A)+1)^2}. \tag{16.519}$$

Mettre cela dans (16.518c) est un calcul d'addition de fractions :

$$f(x_{k+1}) - \bar{f} \leqslant \big(f(x_k) - \bar{f}\big)\left(\frac{c_2(A) - 1}{c_2(A) + 1}\right)^2. \tag{16.520}$$

Par récurrence nous avons alors

$$f(x_k) - \bar{f} \leqslant \big(f(x_0) - \bar{f}\big)\left(\frac{c_2(A) - 1}{c_2(A) + 1}\right)^{2k}. \tag{16.521}$$

Notons qu'il n'y a pas de valeurs absolues parce que $\bar{f}$ étant le minimum de f, les deux côtés de l'inégalité sont automatiquement positifs.

Erreur sur la position du minimum Nous voulons à présent étudier la norme de $x_k - \bar{x}$. Pour cela nous l'écrivons directement avec la définition de f en nous souvenant que $b = -A\bar{x}$:

$$\begin{aligned}
f(x_k) - \bar{f} &= \frac{1}{2}\langle Ax_k, x_k\rangle + \langle A\bar{x}, x_k\rangle + \frac{1}{2}\langle A\bar{x}, \bar{x}\rangle + \langle A\bar{x}, \bar{x}\rangle && (16.522a)\\
&= \frac{1}{2}\langle Ax_k, x_k\rangle - \langle A\bar{x}, x_k\rangle + \frac{1}{2}\langle A\bar{x}, \bar{x}\rangle && (16.522b)\\
&= \frac{1}{2}\langle Ax_k, x_k\rangle - \frac{1}{2}\langle A\bar{x}, x_k\rangle - \frac{1}{2}\langle A\bar{x}, x_k\rangle + \frac{1}{2}\langle A\bar{x}, \bar{x}\rangle && (16.522c)\\
&= \frac{1}{2}\Big(\langle A(x_k - \bar{x}), x_k\rangle + \langle A\bar{x}, \bar{x} - x_k\rangle\Big) && (16.522d)\\
&= \frac{1}{2}\Big(\langle A(x_k - \bar{x}), (x_k - \bar{x})\rangle\Big) && (16.522e)
\end{aligned}$$

où à la dernière ligne nous avons fait $\langle A\bar{x}, \bar{x} - x_k\rangle = \langle \bar{x}, A(\bar{x} - x_k)\rangle$ en vertu de la symétrie de A.

Les produits de la forme $\langle Ay, y\rangle$ sont majorés par $\lambda_{min}\|y\|^2$ parce que λ_{min} est la plus grande valeur propre de A. Dans notre cas,

$$f(x_k) - \bar{f} \geqslant \frac{1}{2}\lambda_{min}\|x_k - \bar{x}\|^2 \tag{16.523}$$

Conclusion En combinant les inéquations (16.523) et (16.521) nous trouvons

$$\frac{1}{2}\lambda_{min}\|x_k - \bar{x}\|^2 \leqslant f(x_k) - \bar{f} \leqslant \big(f(x_0) - \bar{f}\big)\left(\frac{c_2(A) - 1}{c_2(A) + 1}\right)^{2k}, \tag{16.524}$$

c'est-à-dire

$$\|x_k - \bar{x}\| \leqslant \sqrt{\frac{2\big(f(x_0) - \bar{f}\big)}{\lambda_{min} + 1}}^{2k}. \tag{16.525}$$

□

Notons que lorsque $c_2(A)$ est proche de 1 la méthode converge rapidement. Par contre si $c_2(A)$ est proche de zéro, la méthode converge lentement.

16.14 Formes quadratiques, signature, et lemme de Morse

16.110.
Soit $(E, \|.\|_E)$ un espace vectoriel réel normé de dimension finie n. L'ensemble des formes quadratiques réelles [51] sur E est vu comme l'ensemble des matrices symétriques $S_n(\mathbb{R})$; il sera noté $Q(E)$ et le sous-ensemble des formes quadratiques non dégénérées est $S_n(\mathbb{R}) \cap \mathrm{GL}(n, \mathbb{R})$ qui sera noté $\Omega(E)$. Nous rappelons que la correspondance est donnée de la façon suivante.

Si $A \in S_n(\mathbb{R})$, la forme quadratique associée est q_A donnée par le produit scalaire $q_A(x) = x \cdot Ax$.

Pour information, le lemme de Morse est le lemme 19.193.

16.111.
Nous noterons encore $Q^+(E)$ les formes quadratiques positives sur E et $Q^{++}(E)$ les formes quadratiques strictement définies positives sur E.

Sur $Q(E)$ nous mettons la norme

$$N(q) = \sup_{\|x\|_E = 1} |q(x)|, \tag{16.526}$$

51. Définition 9.160.

qui du point de vue de $S_n(\mathbb{R})$ est

$$N(A) = \sup_{\|x\|_E=1} |x^t A x|. \tag{16.527}$$

Notons que à droite, c'est la valeur absolue usuelle sur $\mathbb{R}$.

Nous savons par le théorème de Sylvester (théorème 9.186) que dans $\mathbb{M}(n,\mathbb{R})$, toute matrice symétrique de signature (p,q) est semblable à la matrice

$$\mathbb{1}_{p,q} = \begin{pmatrix} \mathbb{1}_p & & \\ & \mathbb{1}_p & \\ & & 0_{n-p-q} \end{pmatrix}. \tag{16.528}$$

Donc deux matrices de S_n sont semblables si et seulement si elles ont la même signature (même si elles ne sont pas de rang maximum, cela soit dit au passage). Si nous notons $S_n^{p,q}(\mathbb{R})$ l'ensemble des matrices réelles symétriques de signature (p,q), alors

$$S_n^{p,q}(\mathbb{R}) = \{P^t A P \text{ tel que } P \in \mathrm{GL}(n,\mathbb{R})\} \tag{16.529}$$

où A est une quelconque ce ces matrices.

Nous voudrions en savoir plus sur ces ensembles. En particulier nous aimerions savoir si la signature est une notion « stable » au sens où ces ensembles seraient ouverts dans S_n. Pour cela nous considérons l'action de $\mathrm{GL}(n,\mathbb{R})$ sur S_n définie par

$$\begin{aligned} \alpha\colon \mathrm{GL}(n,\mathbb{R}) \times S_n(\mathbb{R}) &\to S_n(\mathbb{R}) \\ (P,A) &\mapsto P^t A P \end{aligned} \tag{16.530}$$

faite exprès pour que les orbites de cette action soient les ensembles $S_n^{p,q}(\mathbb{R})$.

La proposition suivante montre que lorsque $p+q=n$, c'est-à-dire lorsqu'on parle de matrices de rang maximum, les ensembles $S_n^{p,q}(\mathbb{R})$ sont ouverts, c'est-à-dire que la signature d'une forme quadratique est une propriété « stable » par petite variations des éléments de matrice. Notons tout de suite que si le rang n'est pas maximum, le théorème de Sylvester dit qu'elle est semblable à une matrice diagonale avec des zéros sur la diagonale ; en modifiant un peu ces zéros, on peut modifier évidemment la signature.

Proposition 16.112 ([54]).
Soit $(E, \|.\|_E)$ un espace vectoriel normé de dimension finie. Alors

(1) les formes quadratiques non dégénérées forment un ouvert dans l'ensemble des formes quadratiques,

(2) les ensembles $S_n^{p,q}(\mathbb{R})$ avec $p+q=n$ sont ouverts dans $S_n(\mathbb{R})$,

(3) les composantes connexes de $\Omega(E)$ sont les $S_n^{p,q}(\mathbb{R})$ avec $p+q=n$,

(4) les $S_n^{p,q}(\mathbb{R})$ non dégénérés sont connexes par arc.

Démonstration. Cette preuve est donnée du point de vue des matrices. La différence entre le point (3) et (4) est que dans le premier nous prouvons la connexité de $S_n^{p,q}(\mathbb{R})$ à partir de la connexité de $\mathrm{GL}^+(n,\mathbb{R})$, tandis que dans le second nous prouvons la connexité par arc de $S_n^{p,q}(\mathbb{R})$ à partir de la connexité par arc de $\mathrm{GL}^+(n,\mathbb{R})$. Bien entendu le second implique le premier.

(1) Il s'agit simplement de remarquer que $Q(E) = S_n(\mathbb{R})$, que $\Omega(E) = S_n(\mathbb{R}) \cap \mathrm{GL}(n,\mathbb{R})$ et que le déterminant est une fonction continue sur $\mathbb{M}(n,\mathbb{R})$.

(2) Soit $A_0 \in S_n^{p,q}(\mathbb{R})$. Le théorème de Sylvester 9.186 nous donne une matrice inversible P telle que $P^t A_0 P = \mathbb{1}_{p,q}$. Nous allons montrer qu'il existe un voisinage $\mathcal{U}$ de $\mathbb{1}_{p,q}$ contenu dans $S_n^{p,q}(\mathbb{R})$. À partir de là, l'ensemble $(P^{-1})^t \mathcal{U} P^{-1}$ sera un voisinage de A_0 contenu dans $S_n^{p,q}(\mathbb{R})$.

Nous considérons les espaces vectoriels

$$F = \mathrm{Span}\{e_1, \dots, e_p\} \tag{16.531a}$$
$$G = \mathrm{Span}\{e_{p+1}, \dots, e_n\} \tag{16.531b}$$

La norme euclidienne $\|.\|_p$ sur F est équivalente à la norme $|.|_E$ par le théorème 7.224. Donc il existe une constante $k_1 > 0$ telle que pour tout $x \in F$,

$$\|x\|_p \geqslant k_1 \|x\|_E. \tag{16.532}$$

De la même façon sur G, il existe une constante $k_2 > 0$ telle que

$$\|x\|_q \geqslant k_2 \|x\|_E. \tag{16.533}$$

Si nous posons $k = \min\{k_1^2, k_2^2\}$, alors nous avons

$$\forall x \in F, \quad \|x\|_p^2 \geqslant k_1^2 \|x\|_E^2 \geqslant k\|x\|_E^2 \tag{16.534a}$$
$$\forall x \in G, \quad \|x\|_q^2 \geqslant k_2^2 \|x\|_E^2 \geqslant k\|x\|_E^2. \tag{16.534b}$$

Soit une matrice $A \in S_n(\mathbb{R})$ telle que $N(A - \mathbb{1}_{p,q}) < k$, c'est-à-dire que A est dans un voisinage de $\mathbb{1}_{p,q}$ pour la norme sur $S_n(\mathbb{R})$ donné par (16.527). Si x est non nul dans E, nous avons

$$\left|x^t(A - \mathbb{1}_{p,q})x\right| \leqslant N(\mathbb{1}_{p,q} - A)\|x\|^2 \leqslant k\|x\|^2. \tag{16.535}$$

En déballant la valeur absolue, cela signifie que

$$-k\|x\|_E^2 \leqslant x^t(A - \mathbb{1}_{p,q})x \leqslant k\|x\|^2. \tag{16.536}$$

Si $x \in F$, alors la première inéquation et (16.532) donnent

$$x^t A x \geqslant \|x\|_p^2 - k\|x\|_E^2 > 0 \tag{16.537}$$

Si $x \in G$, alors la seconde inéquation et (16.533) donnent

$$x^t A x \leqslant k\|x\|_E^2 - \|x\|_q^2 < 0. \tag{16.538}$$

Nous avons donc montré que $x \mapsto x^t A x$ est positive sur F et négative sur G, ce qui prouve que A est bien de signature (p, q) et appartient donc à $S_n^{p,q}(\mathbb{R})$. Autrement dit nous avons

$$B(\mathbb{1}_{p,q}, k) \subset S_n^{p,q}(\mathbb{R}). \tag{16.539}$$

(3) Cette partie de la preuve provient essentiellement de [339], et fonctionne pour tous les $S_n^{p,q}(\mathbb{R})$, même pour ceux qui ne sont pas de rang maximum.
Soit $A \in S_n^{p,q}(\mathbb{R})$. Nous savons que $\mathrm{GL}(n, \mathbb{R})$ contient exactement deux composantes connexes (proposition 12.18). Vu que l'application

$$\begin{aligned} \alpha \colon \mathrm{GL}(n, \mathbb{R}) &\to S_n \\ P &\mapsto P^t A P \end{aligned} \tag{16.540}$$

est continue, l'image d'un connexe de $\mathrm{GL}(n, \mathbb{R})$ par α est connexe (proposition 7.90). En particulier, $\alpha\big(\mathrm{GL}^{\pm}(n, \mathbb{R})\big)$ sont deux connexes et nous savons que $S_n^{p,q}(\mathbb{R})$ a au plus ces deux composantes connexes.
Notre but est maintenant de trouver une intersection entre les parties $\alpha\big(\mathrm{GL}^{+}(n, \mathbb{R})\big)$ et $\alpha\big(\mathrm{GL}^{-}(n, \mathbb{R})\big)$ [52]. Soit par le théorème de Sylvester, soit par le théorème de diagonalisation des matrices symétriques réelles 9.150, il existe une matrice $P \in \mathrm{GL}(n, \mathbb{R})$ diagonalisant A. En suivant la remarque 9.151, et en notant Q la matrice obtenue à partir de P en changeant le signe de sa première ligne, nous avons

$$\alpha(Q) = Q^t A Q = P^t A P = \alpha(P). \tag{16.541}$$

Or si $P \in \mathrm{GL}^{+}(n, \mathbb{R})$, alors $Q \in \mathrm{GL}^{-}(n, \mathbb{R})$ et inversement. Donc nous avons trouvé une intersection entre $\alpha\big(\mathrm{GL}^{+}(n, \mathbb{R})\big)$ et $\alpha\big(\mathrm{GL}^{-}(n, \mathbb{R})\big)$.

52. À ce point, il me semble que [339] fait erreur parce que la matrice $-\mathbb{1}_n$ est de déterminant 1 lorsque n est pair. L'argument donné ici provient de [54]

(4) Soient A et B dans $S_n^{p,q}(\mathbb{R}) \cap \mathrm{GL}(n,\mathbb{R})$. Par le théorème de Sylvester, il existe P et Q dans $\mathrm{GL}(n,\mathbb{R})$ telles que $A = P^t \mathbb{1}_{p,q} P$ et $B = Q^t \mathbb{1}_{p,q} Q$. Par la remarque 9.151 nous pouvons choisir P et Q dans $\mathrm{GL}^+(n,\mathbb{R})$. Ce dernier groupe étant connexe par arc, il existe un chemin

$$\gamma\colon [0,1] \to \mathrm{GL}^+(n,\mathbb{R}) \tag{16.542}$$

tel que $\gamma(0) = P$ et $\gamma(1) = Q$. Alors le chemin

$$s \mapsto \gamma(s)^t \mathbb{1}_{p,q} \gamma(s) \tag{16.543}$$

est un chemin continu dans $S_n^{p,q}(\mathbb{R})$ joignant A à B.

□

Nous savons déjà de la proposition 16.112 que les ensembles $S_n^{p,q}(\mathbb{R})$ (pas spécialement de rang maximum) sont ouverts dans $S_n(\mathbb{R})$. Le lemme suivant nous donne une précision à ce sujet, dans le cas des matrices de rang maximum, en disant que la matrice qui donne la similitude entre A_0 et A est localement un C^1-difféomorphisme de A.

Lemme 16.113.
Soit $A_0 \in \Omega(\mathbb{R}^n) = S_n \cap \mathrm{GL}(n,\mathbb{R})$, une matrice symétrique inversible. Alors il existe un voisinage V de A_0 dans S_n et une application $\phi\colon V \to \mathrm{GL}(n,\mathbb{R})$ qui

(1) est de classe C^1,

(2) est telle que pour tout $A \in V$, $\varphi(A)^t A_0 \phi(A) = A$.

Démonstration. Nous considérons l'application

$$\begin{aligned} \varphi\colon \mathbb{M}(n,\mathbb{R}) &\to S_n \\ M &\mapsto M^t A_0 M. \end{aligned} \tag{16.544}$$

Étant donné que les composantes de $\varphi(M)$ sont des polynômes en les entrées de M, cette application est de classe C^1 – et même plus. Soit maintenant $H \in \mathbb{M}(n,\mathbb{R})$ et calculons $d\varphi_{\mathbb{1}}(H)$ par la formule (11.485) :

$$\begin{aligned} d\varphi_{\mathbb{1}}(H) &= \frac{d}{dt}\Big[\varphi(\mathbb{1} + tH)\Big]_{t=0} && \text{(16.545a)} \\ &= \frac{d}{dt}\Big[(\mathbb{1} + tH^t)A_0(\mathbb{1} + tH)\Big]_{t=0} && \text{(16.545b)} \\ &= \frac{d}{dt}\Big[A_0 + tA_0H + tH^tA_0 + t^2H^tA_0H\Big]_{t=0} && \text{(16.545c)} \\ &= A_0H + H^tA_0. && \text{(16.545d)} \end{aligned}$$

Donc

$$d\varphi_{\mathbb{1}}(H) = (A_0H) + (A_0H)^t. \tag{16.546}$$

Par conséquent

$$\ker(d\varphi_{\mathbb{1}}) = \{H \in \mathbb{M}(n,\mathbb{R}) \text{ tel que } A_0H \text{ est antisymétrique}\}, \tag{16.547}$$

et si nous posons

$$F = \{H \in \mathbb{M}(n,\mathbb{R}) \text{ tel que } A_0H \text{ est symétrique}\} \tag{16.548}$$

nous avons

$$\mathbb{M}(n,\mathbb{R}) = F \oplus \ker(d\varphi_{\mathbb{1}}) \tag{16.549}$$

parce que toute matrice peur être décomposée de façon unique en partir symétrique et antisymétrique. De plus l'application

$$\begin{aligned} f\colon F &\to S_n \\ H &\mapsto A_0H \end{aligned} \tag{16.550}$$

est une bijection linéaire. D'abord $A_0H = 0$ implique $H = 0$ parce que A_0 est inversible, et ensuite si $X \in S_n$, alors $X = A_0A_0^{-1}X$, ce qui prouve que X est l'image par f de $A_0^{-1}X$ et donc que f est surjective.

Maintenant nous considérons la restriction $\psi = \varphi_{|_F}$, $\psi\colon F \to S_n$. Remarquons que $\mathbb{1} \in F$ parce que $A_0 \in S_n$. L'application $d\psi_{\mathbb{1}}$ est une bijection. En effet d'abord

$$d(\varphi_{|_F})_{\mathbb{1}} = (d\varphi_{\mathbb{1}})_{|_F}, \tag{16.551}$$

ce qui prouve que

$$\ker(d\psi_{\mathbb{1}}) = \ker(d\varphi_{\mathbb{1}}) \cap F = \{0\}, \tag{16.552}$$

ce qui prouve que $d\psi_{\mathbb{1}}$ est injective. Pour montrer que $d\psi_{\mathbb{1}}$ est surjective, il suffit de mentionner le fait que $\dim F = \dim S_n$ du fait que l'application (16.550) est une bijection linéaire.

Nous pouvons utiliser le théorème d'inversion locale (théorème 16.48) et conclure qu'il existe un voisinage ouvert U de $\mathbb{1}$ dans F tel que ψ soit un difféomorphisme C^1 entre U et $V = \psi(U)$. Vu que $\mathrm{GL}(n,\mathbb{R})$ est ouvert dans $\mathbb{M}(n,\mathbb{R})$, nous pouvons prendre $U \cap \mathrm{GL}(n,\mathbb{R})$ et donc supposer que $U \subset \mathrm{GL}(n,\mathbb{R})$.

Pour tout $A \in V$, il existe une unique $M \in U$ telle que $\psi(M) = A$, c'est-à-dire telle que $A = M^tA_0M$. Cette matrice M est $\psi^{-1}(A)$ et est une matrice inversible. Bref, nous posons

$$\begin{aligned} \phi\colon V &\to \mathrm{GL}(n,\mathbb{R}) \\ A &\mapsto \psi^{-1}(A), \end{aligned} \tag{16.553}$$

et ce ϕ est de classe C^1 sur V parce que c'est ce que dit le théorème d'inversion locale. Cette application répond à la question parce que V est un voisinage de $\varphi(\mathbb{1}) = A_0$ et pour tout $A \in V$ nous avons

$$\phi(A)^tA_0\phi(A) = \varphi^{-1}(A)^tA_0\varphi^{-1}(A) = A. \tag{16.554}$$

□

16.15 Ellipsoïde de John-Loewner

C'est le moment de relire les conventions de notations données en 16.110 et 16.111 ainsi que la définition 9.160 d'une forme quadratique.

Soit q une forme quadratique sur $\mathbb{R}^n$ ainsi que $\mathcal{B}$ une base orthonormée de $\mathbb{R}^n$ dans laquelle la matrice de q est diagonale. Dans cette base, la forme q est donnée par la proposition 9.257 :

$$q(x) = \sum_i \lambda_i x_i \tag{16.555}$$

où les λ_i sont les valeurs propres de q.

Plus généralement nous notons $mat_{\mathcal{B}}(q)$ la matrice de q dans la base $\mathcal{B}$ de $\mathbb{R}^n$.

Proposition 16.114.
Soit $\mathcal{B}$ une base orthonormée de $\mathbb{R}^n$ et l'application[53]

$$\begin{aligned} D\colon Q(\mathbb{R}^n) &\to \mathbb{R} \\ q &\mapsto \det\big(mat_{\mathcal{B}}(q)\big). \end{aligned} \tag{16.556}$$

Alors :

(1) La valeur et D ne dépend pas du choix de la base orthonormée $\mathcal{B}$.

(2) La fonction D est donnée par la formule $D(q) = \prod_i \lambda_i$ où les λ_i sont les valeurs propres de q.

(3) La fonction D est continue.

53. L'ensemble $Q(E)$ est l'ensemble des formes quadratiques sur E.

Démonstration. Soit q une forme quadratique sur $\mathbb{R}^n$. Nous considérons $\mathcal{B}$ une base de diagonalisation de q :

$$q(x) = \sum_i \lambda_i x_i \tag{16.557}$$

où les x_i sont les composantes de x dans la base $\mathcal{B}$. Par définition, la matrice $mat_{\mathcal{B}}(q)$ est la matrice diagonale contenant les valeurs propres de q.

Nous considérons aussi $\mathcal{B}_1$, une autre base orthonormées de $\mathbb{R}^n$. Nous notons $S = mat_{\mathcal{B}_1}(q)$; étant symétrique, cette matrice se diagonalise par une matrice orthogonale : il existe $P \in \mathrm{O}(n, \mathbb{R})$ telle que

$$S = P mat_{\mathcal{B}}(q) P^t; \tag{16.558}$$

donc $\det(S) = \det(PP^t) \det\big(\operatorname{diag}(\lambda_1, \dots, \lambda_n)\big) = \lambda_1 \dots \lambda_n$. Ceci prouve en même temps que D ne dépend pas du choix de la base et que sa valeur est le produit des valeurs propres.

Passons à la continuité. L'application déterminant $\det\colon S_n(\mathbb{R}^n) \to \mathbb{R}$ est continue car polynôme en les composantes. D'autre par l'application $mat_{\mathcal{B}}\colon Q(\mathbb{R}^n) \to S_n(\mathbb{R})$ est continue par la proposition 9.256. L'application D étant la composée de deux applications continues, elle est continue. □

Lemme 16.115 (Volume d'un ellipsoïde[54, 1])**.**
Soit une forme quadratique strictement définie positive q sur $\mathbb{R}^n$. Nous considérons l'ellipsoïde

$$\mathcal{E} = \{x \in \mathbb{R}^n \text{ tel que } q(x) \leqslant r^2\} \tag{16.559}$$

pour un certain $r > 0$.

Alors le volume de $\mathcal{E}$ est donné par

$$V_q = \frac{V_0 r^n}{\sqrt{D(q)}} \tag{16.560}$$

où V_0 est une constante[54] *et D est l'application de la proposition 16.114.*

Démonstration. Vu que la matrice de q est symétrique, elle est diagonalisable par une matrice orthogonale (théorème 9.150). Autrement dit, il existe une base orthonormée $\mathcal{B} = \{e_1, \dots, e_n\}$ de $\mathbb{R}^n$ telle que

$$q(x) = \sum_{i=1}^n a_i x_i^2 \tag{16.561}$$

où $x_i = \langle e_i, x \rangle$ et les a_i sont tous strictement positifs.

À priori nous devrions calculer

$$\int_{\mathcal{E}} dx, \tag{16.562}$$

mais nous effectuons le changement de variable[55] associé à la matrice qui diagonalise q et nous devons simplement calculer

$$V_q = \int_{\sum_i a_i x_i^2 < r^2} dx \tag{16.563}$$

parce que le jacobien de ce changement de variable est 1 (déterminant d'une matrice orthogonale).

Tout cela pour dire que nous nommons $\mathcal{E}_q$ l'ellipsoïde associée à la forme quadratique q et V_q son volume que nous allons maintenant calculer[56] :

$$V_q = \int_{\sum_i a_i x_i^2 < r^2} dx \tag{16.564}$$

54. C'est le volume de la boule unité dans $\mathbb{R}^n$ et ce n'est pas tout à fait évident à calculer [340].
55. Théorème 13.255.
56. Le volume ne change pas si nous écrivons l'inégalité stricte au lieu de large dans le domaine d'intégration ; nous le faisons pour avoir un domaine ouvert.

Cette intégrale est écrite de façon plus simple en utilisant le C^1-difféomorphisme

$$\begin{aligned}\varphi\colon \mathcal{E}_q &\to B(0,1)\\ x &\mapsto \frac{1}{r}\Big(x_1\sqrt{a_1},\dots,x_n\sqrt{a_n}\Big).\end{aligned} \tag{16.565}$$

Le fait que φ prenne bien ses valeurs dans $B(0,1)$ est un simple calcul : si $x\in\mathcal{E}_q$, alors

$$\sum_i \varphi(x)_i^2 = \frac{1}{r^2}\sum_i a_i x_i^2 = \frac{1}{r^2}q(x) < 1. \tag{16.566}$$

Cela nous permet d'utiliser le théorème de changement de variables 13.255 :

$$V_q = \int_{\sum_i a_i x_i^2 < r^2} dx = \frac{r^n}{\sqrt{a_1\dots a_n}}\int_{B(0,1)} dx. \tag{16.567}$$

La dernière intégrale est le volume de la boule unité dans $\mathbb{R}^n$ et nous la notons V_0. La proposition 16.114 nous permet d'écrire V_q sous la forme

$$V_q = \frac{V_0 r^n}{\sqrt{D(q)}}. \tag{16.568}$$

□

16.116.
Le théorème suivant dit en substance que si K est compact, alors il existe un unique ellipsoïde de volume minimal centré en l'origine et contenant K. Il faut se rendre compte que l'ellipsoïde n'est pas celui que l'on croirait intuitivement parce que la contrainte *centrée en l'origine* est forte. Si $K = \overline{B\big((4,0),1\big)}$, alors l'ellipsoïde donnée n'est pas du tout K lui-même comme on pourrait s'y attendre. Ce serait probablement quelque chose comme la boule centrée en $(0,0)$ et de rayon 5 [57].

Ce que l'on voudrait est un ellipsoïde qui soit centrée où il faut pour que le volume soit minimal. Nous verrons que c'est possible en la proposition 16.118, mais qu'alors l'unicité est moins évidente (voir la remarque dans [341]). Si vous voulez en entendre parler, vous pouvez lire [342, 343].

Proposition 16.117 (Ellipsoïde de John-Loewner[54])**.**
Soit K compact dans $\mathbb{R}^n$ et d'intérieur non vide. Il existe une unique ellipsoïde [58] *(pleine) centrée en l'origine de volume minimal contenant K.*

Démonstration. Nous subdivisons la preuve en plusieurs parties.

Volume de l'ellipsoïde Soit $\mathcal{E}$ un ellipsoïde centré en l'origine. La proposition 9.265 et son corolaire 9.266 nous indiquent que

$$\mathcal{E} = \{x\in\mathbb{R}^n \text{ tel que } q(x)\leqslant 1\} \tag{16.569}$$

pour une certaine forme quadratique strictement définie positive q. Le lemme 16.115 nous donne alors le volume de $\mathcal{E}$ par

$$V_q = \frac{V_0}{\sqrt{D(q)}} \tag{16.570}$$

où V_0 est une constante.

Existence de l'ellipsoïde Nous voulons trouver un ellipsoïde contenant K de volume minimal, c'est-à-dire une forme quadratique $q\in Q^{++}(\mathbb{R}^n)$ telle que

— $D(q)$ soit maximal

57. Je ne sais pas très bien si il y a moyen de faire mieux. Ce serait sans doute un bon exercice ; faites-moi savoir si vous avez la réponse.
58. Définition 9.263.

— $q(x) \leqslant 1$ pour tout $x \in K$.

Nous considérons l'ensemble des candidats semi-définis positifs.

$$A = \{q \in Q^+ \text{ tel que } q(x) \leqslant 1 \forall x \in K\}. \tag{16.571}$$

Nous allons montrer que A est convexe, compact et non vide dans $Q(\mathbb{R}^n)$; il aura ainsi un maximum de la fonction continue D définie sur $Q(\mathbb{R}^n)$. Nous montrerons ensuite que le maximum est dans Q^{++}. L'unicité sera prouvée à part.

Non vide L'ensemble K est compact et donc borné par $M > 0$. La forme quadratique $q\colon x \mapsto \|x\|^2/M^2$ est dans A parce que si $x \in K$ alors

$$q(x) = \frac{\|x\|^2}{M^2} \leqslant 1. \tag{16.572}$$

Convexe Soient $q, q' \in A$ et $\lambda \in [0,1]$. Nous avons encore $\lambda q + (1-\lambda)q' \in Q^+$ parce que

$$\lambda q(x) + (1-\lambda)q'(x) \geqslant 0 \tag{16.573}$$

dès que $q(x) \geqslant 0$ et $q'(x) \geqslant 0$. D'autre part si $x \in K$ nous avons

$$\lambda q(x) + (1-\lambda)q'(x) \leqslant \lambda + (1-\lambda) = 1. \tag{16.574}$$

Donc $\lambda q + (1-\lambda)q' \in A$.

Fermé Pour rappel, la topologie de $Q(\mathbb{R}^n)$ est celle de la norme (9.563). Nous considérons une suite (q_n) dans A convergeant vers $q \in Q(\mathbb{R}^n)$ et nous allons prouver que $q \in A$, de sorte que la caractérisation séquentielle de la fermeture (proposition 7.265) conclue que A est fermé. En nommant e_x le vecteur unitaire dans la direction x nous avons

$$\big|q(x)\big| = \big|\|x\|^2 q(e_x)\big| \leqslant \|x\|^2 N(q), \tag{16.575}$$

de sorte que notre histoire de suite convergente donne pour tout x :

$$\big|q_n(x) - q(x)\big| \leqslant \|x\|^2 N(q_n - q) \to 0. \tag{16.576}$$

Vu que $q_n(x) \geqslant 0$ pour tout n, nous devons aussi avoir $q(x) \geqslant 0$ et donc $q \in Q^+$ (semi-définie positive). De la même manière si $x \in K$ alors $q_n(x) \leqslant 1$ pour tout n et donc $q(x) \leqslant 1$. Par conséquent $q \in A$ et A est fermé.

Borné La partie K de $\mathbb{R}^n$ est borné et d'intérieur non vide, donc il existe $a \in K$ et $r > 0$ tel que $\overline{B(a,r)} \subset K$. Si par ailleurs $q \in A$ et $x \in \overline{B(0,r)}$ nous avons $a + x \in K$ et donc $q(a+x) \leqslant 1$. De plus $q(-a) = q(a) \leqslant 1$, donc

$$\sqrt{q(x)} = \sqrt{q\big(x+a-a\big)} \leqslant \sqrt{q(x+a)} + \sqrt{q(-a)} \leqslant 2 \tag{16.577}$$

par l'inégalité de Minkowski 9.261. Cela prouve que si $x \in \overline{B(0,r)}$ alors $q(x) \leqslant 4$. Si par contre $x \in \overline{B(0,1)}$ alors $rx \in \overline{B(0,r)}$ et

$$0 \leqslant q(x) = \frac{1}{r^2} q(rx) \leqslant \frac{4}{r^2}, \tag{16.578}$$

ce qui prouve que $N(q) \leqslant \frac{4}{r^2}$ et que A est borné.

L'ensemble A est compact parce que fermé et borné, théorème de Borel-Lebesgue 7.127. L'application continue $D\colon Q(\mathbb{R}^n) \to \mathbb{R}$ de la proposition 16.114 admet donc un maximum sur le compact A. Soit q_0 ce maximum.

Nous montrons que $q_0 \in Q^{++}(\mathbb{R}^d)$. Nous savons que l'application $f\colon x \mapsto \frac{\|x\|^2}{M^2}$ est dans A et que $D(f) > 0$. Vu que q_0 est maximale pour D, nous avons

$$D(q_0) \geqslant D(f) > 0. \tag{16.579}$$

Donc $q_0 \in Q^{++}$.

Unicité S'il existe une autre ellipsoïde de même volume que celle associée à la forme quadratique q_0, nous avons une forme quadratique $q \in Q^{++}$ telle que $q(x) \leqslant 1$ pour tout $x \in K$. C'est-à-dire que nous avons $q_0, q \in A$ tels que $D(q_0) = D(q)$.

Nous considérons la base canonique $\mathcal{B}_c$ de $\mathbb{R}^n$ et nous posons $S = mat_{\mathcal{B}_c}(q)$, $S_0 = mat_{\mathcal{B}_c}(q_0)$. Étant donné que A est convexe, $(q_0 + q)/2 \in A$ et nous allons prouver que cet élément de A contredit la maximalité de q_0. En effet

$$D\left(\frac{q+q_0}{2}\right) = \det\left(\frac{S+S_0}{2}\right) \tag{16.580}$$

Nous allons utiliser le lemme 16.90 qui dit que le déterminant est log-concave sous la forme de l'équation (16.359) avec $\alpha = \beta = \frac{1}{2}$:

$$D\left(\frac{q+q_0}{2}\right) = \det\left(\frac{S+S_0}{2}\right) > \sqrt{\det(S)}\sqrt{\det(S_0)} = \det(S_0) = D(q_0). \tag{16.581}$$

Nous avons utilisé le fait que $D(q_0) = D(q)$ qui signifie que $\det(S_0) = \det(S)$. L'inéquation (16.581) contredit la maximalité de $D(q_0)$ et donne donc l'unicité.

□

Dans la proposition suivante nous oublions l'unicité, mais nous démontrons qu'il existe un ellipsoïde de volume minimal parmi les ellipsoïdes centrées où l'on veut et non seulement en zéro. La source de cette proposition est [1], et comme toujours avec cette source, vous devez regarder à la fois l'énoncé et la preuve avec un oeil encore plus prudent que d'habitude.

Proposition 16.118 ([1]).
Soit un compact d'intérieur non vide K dans $\mathbb{R}^n$. Il existe un ellipsoïde de volume minimal contenant K.

Démonstration. Au lieu de considérer le compact K et de chercher où centrer l'ellipsoïde afin qu'elle puisse contenir K en un volume minimal, nous allons chercher comment translater K pour qu'un ellipsoïde centré en zéro puisse contenir l'image translétée de K en un volume minimal.

Pour $a \in \mathbb{R}^n$, nous notons $\mathcal{E}_a$ la plus petite ellipsoïde centrée en 0 contenant $K + a$ (translation de K par le vecteur a). Elle est bien définie par la proposition 16.117. Notre jeu est maintenant d'étudier la fonction [59]

$$\begin{aligned} f\colon \mathbb{R}^n &\to \mathbb{R}^+ \\ a &\mapsto \mathrm{vol}(\mathcal{E}_a). \end{aligned} \tag{16.582}$$

Nous allons montrer que f est continue et qu'elle peut être restreinte à un compact sans risque de rater un minimum.

À propos de forme quadratique Nous notons q la forme quadratique de l'ellipsoïde $\mathcal{E}_0$ et A sa matrice symétrique associée. Si $x \in K + h$ nous avons, pour un certain $k \in K$:

$$\begin{aligned} q(x) &= q(k+h) && (16.583a)\\ &= \langle k+h, A(k+h)\rangle && (16.583b)\\ &= \langle k, Ak,\rangle + \langle k, Ahx,\rangle + \langle h, Ak\rangle + \langle h, Ah\rangle. && (16.583c) \end{aligned}$$

En tenant compte du fait que $A = A^t$ nous avons aussi $\langle k, Ah\rangle = \langle h, Ak\rangle$ et donc

$$q(x) = 2\langle k, Ah\rangle + q(h). \tag{16.584}$$

En ce qui concerne la norme, pour tout $x \in K + h$ nous avons donc la majoration

$$\begin{aligned} |q(x)| &\leqslant 1 + 2\|k\|\|Ah\| + q(h) && (16.585a)\\ &\leqslant 1 + 2|K|\|Ah\| + q(h) && (16.585b) \end{aligned}$$

59. Notons que cette fonction est bien définie, même sans la partie unicité de 16.117 parce que même si $\mathcal{E}_a$ n'était pas bien définie, son volume, lui, est bien défini.

où nous avons noté $|K| = \sup_{k\in K} \|k\|$. Vu que $q(x)$ est en fait toujours positif nous pouvons oublier la valeur absolue à gauche et conclure que $K + h$ est contenu dans l'ellipsoïde

$$\mathcal{F}_1 = \{x \in \mathbb{R}^n \text{ tel que } q(x) \leqslant 1 + 2|K|\|Ah\| + q(h)\}. \tag{16.586}$$

Volumes Nous avons donc $\mathrm{vol}(\mathcal{E}_h) \leqslant \mathrm{vol}(\mathcal{F}_1)$. En reprenant la formule du lemme 16.115 pour le volume de l'ellipsoïde nous avons la majoration

$$\mathrm{vol}(\mathcal{E}_h) \leqslant \mathrm{vol}(\mathcal{E}_0)\big(1 + 2|K|\|Ah\| + q(h)\big)^n. \tag{16.587}$$

Autrement dit, en ce qui concerne notre fonction f :

$$f(h) \leqslant f(0)\big(1 + 2|K|\|Ah\| + q(h)\big)^n. \tag{16.588}$$

Nous pouvons faire le même raisonnement en partant de $K+h$ et en voyant comment modifier le rayon de $\mathcal{E}_h$ pour que $\mathcal{E}_0$ y rentre. Autrement dit, nous refaisons le raisonnement en posant $K' = K+h$ et en étudiant $K'-h$. Si q_1 est la forme quadratique de $\mathcal{E}_h$ et si A_1 est sa matrice symétrique,

$$\mathrm{vol}(\mathcal{E}_0) \leqslant \mathrm{vol}(\mathcal{E}_h)\big(1 + 2|K + h|\|A_1h\| + q_1(h)\big)^n \tag{16.589}$$

En termes de f :

$$f(0) \leqslant f(h)\big(1 + 2|K + h|\|A_1h\| + q_1(h)\big)^n \tag{16.590}$$

Encadrement Les majoration (16.588) et (16.590) nous permettent de créer un encadrement pour $f(h)$. En effet, en écrivant (16.590) et en continuant la majoration en remplaçant $f(h)$ par (16.588), nous obtenons

$$\begin{aligned} f(0) &\leqslant f(h)\big(1 + 2|K + h|\|A_1h\| + q_1(h)\big)^n \\ &\leqslant f(0)\big(1 + 2|K|\|Ah\| + q(h)\big)^n\big(1 + 2|K + h|\|A_1h\| + q_1(h)\big)^n. \end{aligned} \tag{16.591}$$

Vu que nous avons dans l'idée de prendre des h petits (en norme) et que les parenthèses tendent vers 1 lorsque h est petit, nous pouvons supposer qu'elles sont non nulles et allègrement les passer au dénominateur. Nous divisons donc tout par le coefficient de $f(h)$:

$$\frac{f(0)}{\big(1 + 2|K + h|\|A_1h\| + q_1(h)\big)^n} \leqslant f(h) \leqslant f(0)\big(1 + 2|K|\|Ah\| + q(h)\big)^n \tag{16.592}$$

Nous utilisons la règle de l'étau [60] pour conclure que

$$\lim_{h\to 0} \mathrm{vol}(\mathcal{E}_h) = \mathrm{vol}(\mathcal{E}_0). \tag{16.593}$$

Continuité Si $a \in \mathbb{R}^n$ nous pouvons appliquer la limite (16.593) pour l'ellipsoïde $\mathcal{E}_a$ au lieu de $\mathcal{E}_0$, c'est-à-dire en partant du compact $K + a$ au lieu de K. Cela donne

$$\lim_{h\to 0} \mathrm{vol}(\mathcal{E}_{a+h}) = \mathrm{vol}(\mathcal{E}_a). \tag{16.594}$$

Donc f est continue sur $\mathbb{R}^n$.

Compact Nous avons en hypothèse que K est d'intérieur non vide. Il existe donc $a \in K$ et $r > 0$ tel que $B(a, r) \subset K$. Soit l'hyperplan affine H passant par a et perpendiculaire à a. Nous notons $S = H \cap B(a, r)$. Nous avons évidemment $S \subset K$ et donc $S \subset \mathcal{E}_0$. Vu que l'ellipsoïde $\mathcal{E}_0$ est convexe et contient K, tout le cône de base S et de hauteur a est dans $\mathcal{E}_0$.

Sans rentrer dans le détails du calcul du volume de ce cône, son volume tend vers l'infini si a tend vers l'infini [61]

60. Théorème 11.158.

61. Topologie du complété en un point, j'imagine que vous voyez de quoi il en retourne. Sinon c'est la définition 7.66, et il faut sans doute adapter le lemme 11.93 au cas de $\mathbb{R}^n$ pour tout faire dans les règles.

Nous avons donc une majoration

$$\mathrm{vol}(\mathcal{E}_h) \geqslant \text{volume du cône de base } S \text{ et de hauteur } a+h. \tag{16.595}$$

Dès que ce volume est plus grand que $\mathrm{vol}(\mathcal{E}_0)$, il est impossible que $\mathrm{vol}(\mathcal{E}_h) \leqslant \mathrm{vol}(\mathcal{E}_0)$. Soit $R > 0$ tel que pour tout $\|h\| \geqslant R$, le volume du cône soit plus grand que $\mathrm{vol}(\mathcal{E}_0)$. Le minimum de f est certainement dans $B(0,R)$ parce que au moins

$$\mathrm{vol}(\mathcal{E}_0) \in f\big(B(0,R)\big). \tag{16.596}$$

Minimum Nous considérons la fonction

$$\begin{aligned} f\colon \overline{B(0,R)} &\to \mathbb{R}^+ \\ h &\mapsto \mathrm{vol}(\mathcal{E}_h). \end{aligned} \tag{16.597}$$

Elle est continue sur un compact, et le théorème de Weierstrass 7.105 nous dit alors qu'elle atteint son minimum. Il n'y a cependant pas unicité de la valeurs de h pour laquelle $f(h)$ est minimum.

Envoi Soit $s \in \mathbb{R}^n$ tel que $f(s)$ soit minimal. Cela signifie que parmi tous les ellipsoïdes centrés en zéro et contenant des ensembles de la forme $K+h$, le plus petit est celui qui contient $K+s$. Soit $\mathcal{E}_s$ cet ellipsoïde.

Je prétend que $\mathcal{E}_s - s$ est le plus petit ellipsoïde contenant K, tout centres confondus. En effet, si $\mathcal{F}$ est un ellipsoïde centré en a et contenant K, alors $\mathcal{F} - a$ est un ellipsoïde centré en zéro et contenant $K - a$. Nous avons alors

$$\mathrm{vol}(\mathcal{F} - a) \geqslant \mathrm{vol}(\mathcal{E}_s) = \mathrm{vol}(\mathcal{E}_s - s). \tag{16.598}$$

L'invariance de la mesure par translation est le théorème 13.222.

□

16.16 Prolongement de fonctions

Lemme 16.119.
Soit E, un espace vectoriel normé complet et (A_n) une suite emboîtée de fermés non vides dont le diamètre tend vers zéro. Alors l'intersection $\bigcap_{n\in\mathbb{N}} A_n$ contient exactement un point.

Démonstration. Si l'intersection contenait deux points distincts a et b, alors nous aurions pour tout n la majoration $\mathrm{diam}(A_n) \geqslant \|a-b\|$ qui ne dépend pas de n. Cela contredirait la limite.

Soit une suite (x_n) avec $x_k \in A_k$ pour tout $k \in \mathbb{N}$. C'est une suite de Cauchy. En effet si $\epsilon > 0$, considérons N tel que $\mathrm{diam}(A_N) < \epsilon$. Dans ce cas dès que $n, m > N$ nous avons $x_n, x_m \in A_N$ et donc $\|x_n - x_m\| \leqslant \epsilon$. La suite x_n converge donc vers un élément dans E.

Nous devons montrer que $x \in A_k$ pour tout k. La queue de suite $(x_n)_{n\geqslant k}$ est une suite de Cauchy dans A_k qui converge donc vers un élément de A_k (ici nous utilisons le fait que A_k est fermé). Par unicité de la limite, cette dernière doit être x. Par conséquent $x \in \bigcap_{n\in\mathbb{N}} A_n$. □

Théorème 16.120 ([344]).
Soient X et Y des espaces vectoriels normés. Pour une application linéaire $f\colon X \to Y$, les assertions suivantes sont équivalentes :

(1) f est continue sur X,

(2) f est continue en un point de X,

(3) f est bornée.

Proposition 16.121 ([345]).
Soit un espace normé X, un espace de Banach F et une partie dense A de X. Si l'application linéaire

$$f\colon \big(A, \|.\|_X\big) \to F \tag{16.599}$$

est continue[62]*, alors il existe une unique application linéaire continue $\tilde{f}\colon X \to F$ prolongeant f. De plus $\|\tilde{f}\| = \|f\|$.*

Démonstration. Soient $x \in X$ et $(A_n)_{n\geqslant 0}$ la suite d'ensembles définie par

$$A_n = \{y \in A \text{ tel que } \|x - y\| \leqslant 2^{-n}\}. \tag{16.600}$$

Étant donné que A est dense, ces ensembles sont tous non vides. De plus $\operatorname{diam} A_n \to 0$ parce que si $y, y' \in A_n$ alors

$$\|y - y'\| \leqslant \|y - x\| + \|x - y'\| \leqslant 2^{-n+1}. \tag{16.601}$$

Vu que f est bornée, la suite d'ensembles $f(A_n)$ est une suite emboitée d'ensembles non vides de F. De plus leur diamètre tend vers zéro. En effet si $z, z' \in f(A_n)$, nous posons $z = f(y)$, $z' = f(y')$ et nous avons

$$\|z - z'\| \leqslant \|f(y) - f(x)\| + \|f(x) - f(y')\| \leqslant \|f\|\big(\|y - x\| + \|x - y'\|\big), \tag{16.602}$$

ce qui montre que $\operatorname{diam} f(A_n) \leqslant \|f\|2^{-n+1}$. Notons que nous avons utilisé la linéarité de f. Par le lemme 16.119, l'intersection $\bigcap_{n\in\mathbb{N}} \overline{f(A_n)}$ contient exactement un point. Nous posons

$$S(x) = \bigcap_{n\in\mathbb{N}} \overline{f(A_n)}. \tag{16.603}$$

Nous allons montrer que l'application $x \mapsto S(x)$ ainsi définie est l'application que nous cherchons.

Nous commençons par montrer que pour toute suite $y_k \to x$ avec $y_k \in A$ nous avons

$$f(y_k) \to S(x). \tag{16.604}$$

Pour cela nous considérons $n_0 \in \mathbb{N}$ et k_0 tel que $y_{k_0} \in A_{n_0}$. Avec cela nous avons

$$\|f(y_k) - S(x)\| \leqslant \operatorname{diam}(A_{n_0}) \leqslant \|f\|2^{-n_0+1}. \tag{16.605}$$

Pour montrer que S est linéaire, nous considérons deux suites dans A : $y_k \to x$ et $y'_k \to x'$ ainsi que la somme $y_k + y'k \to x + x'$. Nous écrivons la relation (16.604) pour ces trois suites :

$$f(y_k) \to S(x) \tag{16.606a}$$
$$f(y'_k) \to S(x') \tag{16.606b}$$
$$f(y_k + y'_x) \to S(x + x'). \tag{16.606c}$$

Cependant, étant donné que f est linéaire, pour tout k nous avons $f(y_k + y'_k) = f(y_k) + f(y'_k)$ et par conséquent

$$f(y_k + y'_k) \to S(x) + S(x'). \tag{16.607}$$

Par unicité de la limite, $S(x + x') = S(x) + S(x')$. Le même genre de raisonnement montre que $S(\lambda x) = \lambda S(x)$. L'application S est donc linéaire.

En ce qui concerna la continuité, nous avons

$$\|S(x)\| = \lim \|f(y_k)\| \leqslant \|f\|\|\lim y_k\| = \|f\|\|x\|, \tag{16.608}$$

donc $\|S\| \leqslant \|f\|$, c'est-à-dire que S est borné et donc continue parce que linéaire (théorème 16.120).

Nous montrons maintenant que S prolonge f. Si $x \in A$, alors nous avons $\bigcap_{n\in\mathbb{N}} f(A_n) = f(x)$, et donc $S(x) = f(x)$. Cela montre du même coup que $\|f\| \leqslant \|S\|$ et que par conséquent $\|f\| = \|S\|$.

Passons à la partie sur l'unicité. Soient donc S et T deux prolongements continus de f sur X. Soient $x \in X$ et $x_n \to x$ une suite dans A. Par continuité nous avons $T(x_n) \to T(x)$ et $S(x_n) \to S(x)$. Étant donné que par ailleurs pour tout n nous avons $S(x_n) = T(x_n)$, l'unicité de la limite montre que $T(x) = S(x)$. □

62. Nous avons bien mis sur A la topologie induite de X. Notons que ce n'est pas toujours celle qui est la plus naturelle sur A.

16.16.1 Encore du prolongement

Dans la même veine que la proposition 16.121 nous avons ce résultat.

Théorème 16.122 ([346]).
Soient E et F, deux espaces métriques complets ainsi que A dense dans E. Si $u\colon A \to F$ est uniformément continue, alors elle se prolonge de façon unique en une fonction continue $\tilde{u}\colon E \to F$. De plus ce prolongement est uniformément continu.

Démonstration. Soit $x \in E\backslash A$ et une suite (x_n) contenue dans A et convergente vers x. Nous voulons définir

$$\tilde{u}(x) = \lim_{n\to\infty} u(x_n) \tag{16.609}$$

mais pour ce faire nous devons prouver que la suite $\big(u(x_n)\big)$ converge dans F et que la limite ne dépend pas de la suite choisie parmi les suites de A qui convergent (dans E) vers x.

Commençons par montrer que $\big(u(x_n)\big)$ est de Cauchy dans F. Pour cela nous prenons $\epsilon > 0$ et $\eta > 0$ telle que $d_E(a,b) < \eta$ implique $d_F\big(u(a),u(b)\big) < \epsilon$ (uniforme continuité de u). Après, il suffit de choisir N tel que pour tout $n,m > N$ nous ayons $d(x_m,x_n) < \eta$ (parce que u_n est de Cauchy). Avec tout ça nous avons

$$d_F\big(u(x_m),u(x_n)\big) < \epsilon, \tag{16.610}$$

ce qui signifie que $\big(u(x_n)\big)$ est de Cauchy et donc convergente dans F.

Nous voulons montrer maintenant que si (x_n) et (y_n) sont deux suites dans A convergentes vers x alors $\lim_{n\to\infty} u(x_n) = \lim_{n\to\infty} u(y_n)$. Pour cela nous considérons la suite $z = (x_1,y_1,x_2,y_2,\ldots)$. Nous avons évidemment $z_n \to x$, et donc $u(z_n)$ converge dans F par ce qui a été dit plus haut. Mais $u(x_n)$ et $u(y_n)$ en sont deux sous-suites convergentes. Donc leurs limites sont égales.

Il reste à montrer que ce $\tilde{u}$ est continue et uniformément continue. Pour cela nous utilisons le module de continuité et le lemme 10.166. Étant donné que $\tilde{u}$ prolonge u nous avons

$$\omega_{\tilde{u}}(h) \geqslant \omega_u(h). \tag{16.611}$$

Soient $h > 0$ et $\epsilon > 0$; soient aussi $x,y \in E$ tels que $d(x,y) < h$. Nous prenons des suites $(a_n) \to x$ et $(y_n) \to y$ tout en choisissant n assez grand pour avoir $d_E(a_n,b_n) < h$. Nous avons

$$d_F\big(\tilde{u}(x),\tilde{u}(y)\big) \leqslant d_F\big(\tilde{u}(x),u(a_n)\big) + d\big(u(a_n),u(b_n)\big) + d_F\big(u(b_n),\tilde{u}(y)\big). \tag{16.612}$$

Si n est assez grand, par construction de $\tilde{u}$, le premier et le dernier terme sont plus petits que ϵ. Par définition du module de continuité nous avons d'autre part $d_F\big(u(a_n),u(b_n)\big) \leqslant \omega_u(h)$. Du coup

$$d_F\big(\tilde{u}(x),\tilde{u}(y)\big) \leqslant \omega_u(h) + 2\epsilon. \tag{16.613}$$

Si nous prenons le supremum sur les x et y vérifiant $d_E(x,y) < h$, à gauche nous obtenons $\omega_{\tilde{u}}(h)$ tandis que le membre de droite ne dépend pas de x et$y$. Donc pour tout ϵ, nous avons

$$\omega_{\tilde{u}}(h) \leqslant \omega_u(h) + 2\epsilon. \tag{16.614}$$

En comparaison avec (16.611), nous trouvons

$$\omega_{\tilde{u}(h)} \leqslant \omega_u(h). \tag{16.615}$$

Les fonctions u et $\tilde{u}$ ayant le même module de continuité, le lemme 10.166 nous enseigne que l'une est uniformément continue si et seulement si l'autre l'est. Vu que u est uniformément continue par hypothèse, le prolongement $\tilde{u}$ est uniformément continu. □

Définition 16.123.
*Un **plongement** de l'espace topologique X dans Y est une application $f\colon X \to Y$ telle que $f\colon X \to f(X)$ soit un homéomorphisme.*

Théorème 16.124 (Extensiton des isométries)**.**
Soit $\tilde{M}$ un espace métrique complet et une application isométrique

$$f\colon A \to \tilde{M} \tag{16.616}$$

où A est une partie dense d'un espace métrique M (pas spécialement complet). Alors f accepte une une unique extension isométrique

$$\tilde{f}\colon M \to \tilde{M} \tag{16.617}$$

Supposons de plus que M soit complet[63]*. Alors $\tilde{f}\colon M \to \tilde{M}$ est une bijection si et seulement si $f(A)$ est dense dans $\tilde{M}$.*

Démonstration. Nous commençons par prouver l'unicité. Soient $\tilde{f}_1$ et $\tilde{f}_2$ deux extensions de f et $x \in M$. Si (a_n) est une suite dans A convergeant vers x (possible parce que A est dense dans M), alors nous avons

$$\tilde{f}_1(a_n) = \tilde{f}_2(a_n) \tag{16.618}$$

et donc $\tilde{f}_1(x) = \tilde{f}_2(x)$ par continuité (une application isométrique est continue (proposition 7.260)).

Nous démontrons à présent l'existence.

Construction de $\tilde{f}$ Soient $x \in M$ et (a_n) une suite dans A qui converge vers x. Nous définissons

$$\tilde{f}(x) = \lim_{k\to\infty} f(a_k). \tag{16.619}$$

Note : nous pouvons prouver que cette définition ne dépend pas du choix de la suite (a_n) convergeant vers x, mais ce serait superflu parce que nous avons déjà prouvé l'unicité de $\tilde{f}$. Par contre nous devons expliquer pourquoi la limite du membre de droite de (16.619) existe dans $\tilde{M}$. D'abord la suite (a_n) est de Cauchy parce qu'elle est convergente (attention : M n'étant pas complet le fait d'être de Cauchy n'implique pas la convergence). Donc, étant donné que f est une isométrie, la suite $\big(f(a_n)\big)$ est de Cauchy dans $\tilde{M}$. Or ce dernier étant complet, la suite des images converge.

Montrons que cette application $\tilde{f}\colon M \to \tilde{M}$ répond à la question.

$\tilde{f}$ est isométrique Soient $a, b \in M$ et des suites dans A convergeant vers eux : $a_n \to a$, $b_n \to b$. Nous avons, par continuité de l'application distance,

$$\begin{aligned}
d\big(\tilde{f}(a), \tilde{f}(b)\big) &= \lim_{k\to\infty} d\big(\tilde{f}(a_k), \tilde{f}(b)\big) && (16.620a)\\
&= \lim_{k\to\infty}\lim_{l\to\infty} d\big(\tilde{f}(a_k), \tilde{f}(b_l)\big) && (16.620b)\\
&= \lim_{k\to\infty}\lim_{l\to\infty} d\big(f(a_k), f(b_l)\big) && (16.620c)\\
&= \lim_{k\to\infty}\lim_{l\to\infty} d\big(a_k), b_l\big) && (16.620d)\\
&= d(a, b). && (16.620e)
\end{aligned}$$

Cela prouve que $\tilde{f}$ est une isométrie.

Pour la suite nous supposons que M est complet. Notons tout de suite que $\tilde{f}$ est injective parce qu'elle est isométrique.

Bijection (premier sens) Nous supposons que $\tilde{f}\colon M \to \tilde{M}$ est une bijection. Par l'absurde nous supposons que $f(A)$ n'est pas dense dans $\tilde{M}$, c'est-à-dire que nous avons un point $x \in \tilde{M}$ et une boule n'intersectant par $f(A)$:

$$B(x, r) \cap f(A) = \varnothing. \tag{16.621}$$

Étant donné que $\tilde{f}$ a pour image des limites de suites dans $f(A)$, l'image de $\tilde{f}$ est contenue dans $\overline{f(A)}$. Donc si $\tilde{f}$ est surjective, c'est que $\tilde{M} \subset \overline{f(A)}$ et donc que $\overline{f(A)} = \tilde{M}$. Cela prouve que si $\tilde{f}$ est bijective, alors $f(A)$ est dense dans $\tilde{M}$.

63. Il me semble que cette hypothèse manque dans [347].

Bijection (l'autre sens) Nous supposons que $\overline{f(A)} = \tilde{M}$ et nous devons prouver que $\tilde{f}$ est surjective. Soient $x \in \tilde{M}$ et $f(a_n)$ une suite dans $f(A)$ qui converge vers x ; une telle suite existe parce que $f(A)$ est dense dans $\tilde{M}$. Cette suite est de Cauchy dans $\tilde{M}$ parce que dans un espace métrique, une suite convergente est de Cauchy. La suite (a_n) est elle-même également de Cauchy parce que

$$d(a_n, a_m) = d\big(f(a_n), f(a_m)\big). \tag{16.622}$$

Étant donné que (a_n) est de Cauchy dans M, elle converge vers un élément que nous nommons $a \in M$. Par continuité de f nous avons alors

$$\tilde{f}(a) = \lim_{k\to\infty} f(a_k) = x. \tag{16.623}$$

Cela prouve que x est bien dans l'image de $\tilde{f}$ et donc que $\tilde{f}$ est surjective.

□

16.17 Complétion d'un espace métrique

Une conséquence du théorème de prolongement est le théorème suivant qui permet de compléter un espace métrique.

Théorème 16.125 (Complétion d'un espace métrique[348, 347])**.**
Tout espace métrique se plonge par une isométrie à image dense dans un espace métrique complet. De plus ce dernier est unique à isométrie près.

Plus précisément, soit (M, d) un espace métrique. Il existe un espace métrique complet $\tilde{M}$ muni d'un plongement isométrique $\varphi\colon M \to \tilde{M}$ tel que $\varphi(M)$ soit dense dans $\tilde{M}$.

Ce complété de M est unique au sens suivant. Si $\tilde{M}_1$ et $\tilde{M}_2$ sont deux espaces métriques complets munis de plongements isométriques $f_i\colon M \to \tilde{M}_1$ dont les images sont denses, alors il existe une bijection isométrique $\phi\colon \tilde{M}_1 \to \tilde{M}_2$ telle que $\phi \circ f_1 = f_2$.

Démonstration. Nous ne prouvons que l'existence.

Soit C_M l'ensemble des suites de Cauchy de M. Nous définissons

$$\begin{aligned} f\colon C_M \times C_M &\to \mathbb{R} \\ u, v &\mapsto \lim_{n\to\infty} d(u_n, v_n). \end{aligned} \tag{16.624}$$

Notre première tâche est de nous assurer que cela est bien défini, c'est-à-dire que la limite existe toujours. En effet, si u et v sont des suites de Cauchy dans M, nous avons

$$|d(u_n, v_n) - d(u_m, v_m)| \leqslant d(u_n, v_n) + d(u_m, v_m) \leqslant 2\epsilon \tag{16.625}$$

dès que m et n sont assez grand. Cela prouve que la suite $n \mapsto d(u_n, v_n)$ est de Cauchy dans $\mathbb{R}$. Par complétude de $\mathbb{R}$, elle converge[64].

Nous considérons la relation d'équivalence $u \sim v$ si et seulement si $f(u, v) = 0$. Nous posons $\tilde{M} = C_M/\sim$ et nous y mettons la distance

$$d([u], [v]) = f(u, v) \tag{16.626}$$

et nous devons encore vérifier que cela est bien défini. Prenons $u' \sim u$ et $v' \sim v$. Alors nous avons

$$d(u'_n, v'_n) \leqslant d(u'_n, u_n) + d(u_n, v_n) + d(v_n, v'_n), \tag{16.627}$$

et donc

$$d(u', v') = \lim_{n\to\infty} d(u'_n, v'n) \leqslant \lim_{n\to\infty} d(u_n, v_n) = d(u, v). \tag{16.628}$$

64. Ici nous utilisons la complétude de $\mathbb{R}$. Cette dernière doit donc être démontrée indépendamment, ce qui est fait dans le théorème 7.294. De plus nous ne pouvons pas définir $\mathbb{R}$ comme étant le complété de $\mathbb{Q}$ en utilisant ce théorème.

Le même argument en inversant les primes et les non primes montre l'inégalité inverse. Donc $d(u,v) = d(u',v')$ dans C_M, et donc la distance (16.626) est bien définie sur $\tilde{M}$.

Afin de s'assurer que $\tilde{M}$ répond bien à la question du théorème, il faut encore démontrer les points suivants :

— M se plonge isométriquement dans $\tilde{M}$.

— l'image de M par le plongement est dense dans $\tilde{M}$.

— $\tilde{M}$ est complet.

Nous allons maintenant considérer l'application

$$\begin{aligned} \varphi\colon M &\to \tilde{M} \\ x &\mapsto \text{la classe de la suite constante } x. \end{aligned} \tag{16.629}$$

Plongement isométrique Nous allons montrer que cela est une isométrie bijective et que $\varphi(M)$ est dense dans $\tilde{M}$. Le fait que φ soit bijective entre M et $\varphi(M)$ est évident. C'est une isométrie parce que

$$d\big(\varphi(x),\varphi(y)\big) = \lim_{n\to\infty} d\big(\varphi(x)_n,\varphi(y)_n\big) = d(x,y). \tag{16.630}$$

Densité Soit $[u] \in \tilde{M}$. Tous les termes u_n sont des éléments de M. Nous considérons la suite dans $\varphi(M)$ donnée par

$$a_n = \varphi(u_n) \tag{16.631}$$

Chaque a_n est un élément[65] de $\tilde{M}$. Montrons que (a_n) converge dans $\tilde{M}$ vers u. Nous avons

$$\begin{aligned} d(a_n,u) &= \lim_{k\to\infty} d\big((a_n)_k,u_k\big) && \text{(16.632a)}\\ &= \lim_{k\to\infty} d(u_n,u_k) && \text{(16.632b)}\\ &= d(u_n,\ell) && \text{(16.632c)} \end{aligned}$$

en notant ℓ la limite de la suite (u_n). Ici nous avons utilisé le fait que la fonction distance était continue pour l'inverser avec la limite, par le théorème 11.156. Nous avons alors

$$\lim_{n\to\infty} d(a_n,[u]) = \lim_{n\to\infty} d(u_n,\ell) = 0. \tag{16.633}$$

Complétude Nous passons maintenant à la preuve du fait que $\tilde{M}$ est complet. Soit (y_n) une suite de Cauchy dans $\tilde{M}$. Soit $\epsilon > 0$; nous définissons $K(n)$ par

$$d\big((y_n)_k,(y_n)_l\big) < \epsilon \tag{16.634}$$

dès que $k,l \geqslant K(n)$. Cette définition fonctionne parce que pour chaque n, y_n est une suite de Cauchy dans M. Nous posons

$$x_n = (y_n)_{K(n)} \in M \tag{16.635}$$

et nous allons montrer que (x_n) est de Cauchy dans M –donc est un élément de $\tilde{M}$– et que $y_k \to (x_n)$ dans $\tilde{M}$.

Nous commençons par montrer que (x_n) est de Cauchy dans M. Nous avons

$$\begin{aligned} d(x_n,x_m) &= d\big((y_n)_{K(n)},(y_m)_{K(m)}\big) && \text{(16.636a)}\\ &\leqslant d\big((y_n)_{K(n)},(y_n)_l\big) + d\big((y_n)_l,(y_m)_l\big) + d\big((y_m)_l,(y_m)_{K(m)}\big) && \text{(16.636b)} \end{aligned}$$

Nous choisissons n,m tels que $d(y_n,y_m) < \epsilon$, ce qui nous permet de choisir l de telle façon à avoir $d\big((y_n)_k,(y_m)_k\big) < \epsilon$ pour tout $k \geqslant l$. De plus, quitte à encore augmenter l, nous

65. À partir de maintenant nous n'écrivons plus explicitement la classe d'équivalence.

supposons que $l > K(m)$ et $l > K(m)$. Avec ces choix nous voyons que $d(x_n, x_m) < 3\epsilon$, ce qui signifie que la suite (x_n) est de Cauchy dans M.
En ce qui concerne la convergence $y_n \to (x)$, on a

$$d\big(y_n, (x)\big) = \lim_{k\to\infty} d\big((y_n)_k, (y_k)_{K(k)}\big) \tag{16.637a}$$

$$\leqslant \lim_{k\to\infty} d\big((y_n)_k, (y_n)_l\big) + \lim_{k\to\infty} d\big((y_n)_l, (y_k)_l\big) + \lim_{k\to\infty} d\big((y_k)_l, (y_k)_{K(k)}\big) \tag{16.637b}$$

Nous devons trouver un n tel que si k est suffisamment grand, le tout est majoré par ϵ. Voici nos choix :

— n tel que $d(y_n, y_m) < \epsilon$ dès que $m \geqslant n$,
— $k > n$,
— $k > K(n)$,
— $l > k$,
— $l > K(k)$,
— l suffisamment grand pour que $d\big((y_n)_l, (y_k)_l\big) < \epsilon$.

Avec tous ces choix, les trois termes de (16.637b) sont plus petits que ϵ.
Ceci prouve que $\tilde{M}$ est complet.

□

16.17.1 Principe des zéros isolés

Théorème 16.126 (Principe des zéros isolés [349]).
Soit un ouvert $\Omega \subset \mathbb{C}$. Soient une fonction analytique[66] *$f\colon \Omega \to \mathbb{C}$ et $a \in \Omega$, une zéro non isolé de f. Alors f est nulle sur un voisinage de a.*

Démonstration. Nous écrivons f sous la forme d'une série entière[67] autour de a :

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n \tag{16.638}$$

valable sur une boule $B(a, r)$. Soit c_m le premier coefficient non nul (si il n'existe pas c'est que f est nulle sur tout $B(a, r)$ et alors le théorème est prouvé). Nous avons alors

$$f(z) = c_m (z-a)^m \big(1 + \sum_{k=1}^{\infty} d_k (z-a)^k\big) \tag{16.639}$$

avec $d_k = c_{k+m}/c_m$. Le rayon de convergence de la série $\sum_k d_k (z-a)^k$ est le même que celui de (16.638) parce que la suite $\frac{d_k r^{m+k}}{c_m}$ reste bornée (critère d'Abel, lemme 14.10). Si nous posons

$$g(z) = 1 + \sum_{k=1}^{\infty} d_k (z-a)^k, \tag{16.640}$$

alors g est une fonction continue et $g(a) = 1$. De plus

$$f(z) = c_m (z-a)^m g(z). \tag{16.641}$$

Soit une suite (z_n) de zéros de f qui converge vers a. Étant donné que g est continue, nous devrions avoir $\lim_{k\to\infty} g(z_k) = g(a) = 1$, mais si $f(z_k) = 0$ avec $z_k \neq a$, alors $g(z_k) = 0$. Cela est un paradoxe qui nous permet de conclure que si la suite z_n existe bien, alors f est identiquement nulle sur un voisinage, c'est-à-dire que tous les c_n sont nuls. □

66. Définition 11.368.
67. Définition de fonction analytique 11.368.

Corolaire 16.127.
Soit f une fonction analytique sur un ouvert connexe $\Omega \subset \mathbb{C}$. Si f s'annule sur un ouvert (non vide) de Ω, alors f s'annule sur tout Ω.

Démonstration. soit

$$N = \{z \in \Omega \text{ tel que } f = 0 \text{ sur un ouvert autour de } z\}. \tag{16.642}$$

Le fait que N soit ouvert est évident à partir de sa définition. Nous allons montrer que N est également fermé dans Ω, et donc conclure que $N = \Omega$. Soit (z_n) une suite dans N convergente vers $z \in \Omega$. Étant donné que $f(z_n) = 0$ et que f est continue, nous avons

$$f(z) = \lim_{n\to\infty} f(z_n) = 0, \tag{16.643}$$

ce qui fait de z un zéro non isolé de f. Par conséquent le principe des zéros isolés (théorème 16.126) nous enseigne que f s'annule dans un voisinage autour de z, c'est-à-dire que $z \in N$. L'ensemble N est donc fermé. □

Théorème 16.128 (Principe du prolongement analytique[1])**.**
Soit Ω un ouvert connexe de $\mathbb{C}$. Si deux fonctions analytiques coïncident sur un sous-ensemble D de Ω contenant un point d'accumulation dans Ω, alors elles sont égales sur Ω.

Démonstration. Soient deux fonctions analytique $f, g\colon \Omega \to \mathbb{C}$ égales sur D. Nous considérons la fonction $h = f - g$. Il existe dans D une suite (z_n) qui converge vers $a \in \Omega$ et telle que $f(z_n) = g(z_n)$, c'est-à-dire $h(z_n) = 0$ pour tout n.

Vu que h est continue nous avons aussi $h(a) = 0$. Le point a est donc un zéro non isolé de h. Le théorème 16.126 conclu que $h = 0$ sur Ω. □

16.18 Un petit extra

Soit f une fonction de $\mathbb{R}$ dans $\mathbb{R}$. Supposons que

(1) $f(1) = 1$,

(2) $f(x + y) = f(x) + f(y)$ pour tout réels x et y.

Nous pouvons montrer[68] que la seule fonction *continue* qui possède ces propriétés est la fonction identité $f(x) = x$ pour tout $x \in \mathbb{R}$.

De la même manière, il est aisé de voir que les seules applications linéaires de $\mathbb{R}^n$ dans $\mathbb{R}^n$ sont de la forme

$$f(x) = Ax \tag{16.644}$$

pour une constante réelle A. Une question naturelle qu'on peut alors se poser est la suivante :

Est-il possible de définir une fonction non continue ayant les propriétés (1) et (2) ?

En fait, il est possible de démontrer que si E est un espace vectoriel de dimension finie, alors toute application linéaire $f : E \to F$ (où F est un espace vectoriel) sera continue. Ceci ne reste plus vrai si l'espace vectoriel E est de dimension infinie. Donc une manière de trouver une réponse positive à la question posée plus haut, serait de voir $\mathbb{R}^n$ comme espace vectoriel de dimension infinie. Après un peu de réflexion, la réponse est venue à nous (merci à Nicolas et à Samuel).

Si nous admettons l'axiome du choix, alors nous pouvons appliquer le théorème de Zorn et nous savons que tout espace vectoriel admet une base. En particulier, l'ensemble des réels vu comme espace vectoriel sur $\mathbb{Q}$ admet une base, i.e. $\exists (e_i)_{i\in I}$ des éléments de $\mathbb{R}^n$ tels que tout réel s'écrit comme combinaison linéaire à coefficients rationnels de ces e_i, i.e.

$$\forall r \in \mathbb{R}^n, \exists (\lambda_i)_{i\in I} \text{ des éléments de } \mathbb{Q} \text{ tels que } r = \sum_{i\in I} \lambda_i e_i. \tag{16.645}$$

68. et toi, tu le peux ?

Utilisons cette base pour définir une fonction h de la manière suivante.

$$\forall i \in I, \text{ on définit } h(e_i) = \alpha_i \tag{16.646}$$

où les α_i doivent être bien choisis dans $\mathbb{R}^n$. Pour satisfaire la propriété (1), choisissons sans perte de généralité $e_1 = 1$ et $h(e_1) = 1$. Ajoutons à cette propriété la linéarité en imposant que

$$h(\sum \lambda_i e_i) = \sum \lambda_i \alpha_i. \tag{16.647}$$

Les équations (1) et (2) nous permettent de voir que, moyennant le choix des α_i, la fonction h est bien définie sur $\mathbb{R}^n$ et linéaire. Il est clair que si nous prenons par exemple

$$\alpha_i = e_i \; \forall i \in I$$

nous obtenons que la fonction h est en fait la fonction identité sur $\mathbb{R}^n$. Par contre, si nous définissons la fonction h comme satisfaisant la propriété (2) et si nous choisissons les α_i dans (1) de la manière suivante

$$\begin{aligned} h(e_1) &= e_2 \\ h(e_2) &= e_1 \\ h(e_i) &= e_i \quad \forall i \in I \backslash \{1, 2\} \end{aligned} \tag{16.648}$$

alors la fonction ainsi obtenue est linéaire et bien définie mais n'est plus l'identité. Donc nous avons trouvé une application linéaire de $\mathbb{R}^n$ dans $\mathbb{R}^n$ qui n'est pas continue.

Chapitre 17

Trigonométrie, isométries

17.1 Trigonométrie

17.1.1 Définitions, périodicité et quelques valeurs remarquables

Proposition-définition 17.1 (Défintion du cosinus et du sinus)**.**
La série

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n} \tag{17.1}$$

définit une fonction $\cos\colon \mathbb{R} \to \mathbb{R}$ *de classe* C^∞*. Nous l'appelons* ***cosinus****.*

La série

$$\sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^{2n+1} \tag{17.2}$$

définit une fonction $\sin\colon \mathbb{R} \to \mathbb{R}$ *de classe* C^∞*. Nous l'appelons* ***sinus****.*

Démonstration. La série entière définissant $\cos(x)$ a pour coefficients

$$a_n = \begin{cases} 0 & \text{si } n \text{ est impair} \\ \frac{(-1)^{n/2}}{n!} & \text{si } n \text{ est pair.} \end{cases} \tag{17.3}$$

Nous pouvons la majorer par la série entière donnée par les coefficients

$$b_n = \begin{cases} 1/n! & \text{si } n \text{ est impair} \\ \frac{(-1)^{n/2}}{n!} & \text{si } n \text{ est pair.} \end{cases} \tag{17.4}$$

Quelle que soit la parité de k nous avons toujours

$$|\frac{b_{k+1}}{b_k}| = \frac{1}{k+1}, \tag{17.5}$$

de telle sorte que la formule d'Hadamard (14.31) nous donne $R = \infty$ pour la série $\sum_{k=0}^{\infty} b_k x^k$. A fortiori [1] le rayon de convergence pour la série du cosinus est infini.

L'assertion concernant le sinus se démontre de même.

En ce qui concerne le fait que les fonctions sin et cos sont de classe C^∞ sur $\mathbb{R}$, il faut invoquer le corolaire 14.27. □

Par substitution directe dans les séries, nous avons immédiatement

$$\cos(0) = 1 \tag{17.6a}$$
$$\sin(0) = 1. \tag{17.6b}$$

1. Remarque 14.9.

Lemme 17.2.
En ce qui concerne la dérivation, nous avons

$$\sin' = \cos \tag{17.7a}$$
$$\cos' = -\sin . \tag{17.7b}$$

Démonstration. Il s'agit de se permettre de dériver terme à terme (proposition 14.25) les séries qui définissent le sinus et le cosinus. □

Lemme 17.3.
Les fonctions sinus et cosinus vérifient

$$\cos^2(x) + \sin^2(x) = 1 \tag{17.8}$$

pour tout $x \in \mathbb{R}$.

Démonstration. Posons $f(x) = \sin^2(x) + \cos^2(x)$ et dérivons :

$$f'(x) = 2\sin(x)\cos(x) + 2\cos(x)(-)\sin(x) = 0. \tag{17.9}$$

La fonction f est donc constante par le corolaire 11.134. Nous avons donc pour tout x :

$$f(x) = f(0) = \sin^2(0) + \cos^2(0) = 1. \tag{17.10}$$

Le dernier calcul s'obtient en substituant directement x par zéro dans les séries : $\sin(0) = 0$ et $\cos(0) = 1$. □

17.1.2 Fonction puissance (pour les complexes)

La fonction puissance a déjà fait l'objet de nombreuses définitions et extensions. Voir le thème 33. Nous allons maintenant définir a^z pour $a > 0$ et $z \in \mathbb{C}$.

Définition 17.4.
Pour le nombre $e \in \mathbb{R}$ *et le nombre imaginaire pur* iy *(*$y \in \mathbb{R}$*), nous définissons*

$$e^{iy} = \exp(iy) \tag{17.11}$$

où exp *est la série usuelle de la définition 14.37. Pour un nombre complexe général* $x + yi$ *nous définissons*

$$e^{x+iy} = e^x e^{iy}. \tag{17.12}$$

Et enfin, si $a > 0$ *et si* $z \in \mathbb{C}$ *nous définissons*

$$a^z = e^{z\ln(a)}, \tag{17.13}$$

la fonction logarithme ici étant celle $\ln\colon]0,\infty[\to \mathbb{R}$ *définie par la proposition 14.58.*

Si $z \in \mathbb{C}$ et si $n \in \mathbb{Z}$, la définition de z^n ne pose pas de problèmes, c'est la définition 1.95.

17.5.
Soit $z = x + iy \in \mathbb{C}$. L'exponentielle $\exp(x+yi)$ est déjà définie en 14.37 ; elle est la fonction donnée par

$$\begin{aligned} \exp\colon \mathbb{C} &\to \mathbb{C} \\ z &\mapsto \sum_{n=0}^{\infty} \frac{z^n}{n!}. \end{aligned} \tag{17.14}$$

Proposition 17.6.
Le rayon de convergence[2] *de la série exponentielle est infini.*

2. Définition 14.8.

Démonstration. La formule de Hadamard de la proposition 11.354 est à utiliser avec $a_k = j!$. Nous avons

$$\frac{1}{R} = \lim_{k\to\infty} \left| \frac{(n+1)!}{n!} \right| = \lim_{k\to\infty} (n+1) = \infty. \tag{17.15}$$

Donc $R = \infty$. □

Proposition 17.7.
Pour tout $z \in \mathbb{C}$ nous avons

$$\exp(z) = e^z. \tag{17.16}$$

Proposition 17.8 ([350])**.**
Quelques propriétés de l'exponentielle.

(1) Le fonction exp *est continue.*

(2) Nous avons la formule $e^{z+w} = e^z e^w$ pour tout $z, w \in \mathbb{C}$.

(3) $(e^z)^{-1} = e^{-z}$

(4) $(\exp(z))^n = \exp(nz)$.

Démonstration. La proposition 17.6 nous enseigne que le rayon de convergence est infini. La fonction ainsi définie est alors continue par la proposition 11.305.

Les séries $\exp(z)$ et $\exp(w)$ ayant un rayon de convergence infini nous pouvons utiliser le produit de Cauchy (théorème 14.15) :

$$\begin{aligned}
e^z e^w &= \sum_{n=0}^{\infty} \left(\sum_{i+j=n} \frac{z^i w^j}{i!j!} \right) && (17.17a)\\
&= \sum_{n=0}^{\infty} \left(\sum_{i=0}^{n} \frac{z^i w^{n-i}}{i!(n-i)!} \right) && (17.17b)\\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{i=0}^{n} \binom{n}{i} z^i w^{n-i} && (17.17c)\\
&= \sum_{n=0}^{\infty} \frac{1}{n!} (z+w)^n && (17.17d)\\
&= \exp(z+w). && (17.17e)
\end{aligned}$$

Nous avons utilisé la formule du binôme (proposition 3.47).

Les autres propriétés énoncées sont des corolaires :

$$e^z e^{-z} = e^0 = 1. \tag{17.18}$$

□

D'autres propriétés de l'exponentielle sur $\mathbb{C}$, entre autres l'holomorphie, sont données dans le théorème 25.43.

Lemme 17.9 ([1])**.**
Soient $a > 0$, $z \in \mathbb{C}$ et $n \in \mathbb{Z}$. Alors

$$(a^z)^n = a^{nz}. \tag{17.19}$$

17.1.3 Formules de trigonométrie

Le lemme suivant est un premier pas pour le paramétrage du cercle dont nous parlerons dans la proposition 17.48.

Lemme 17.10.
Nous avons la formule

$$e^{ix} = \cos(x) + i\sin(x) \tag{17.20}$$

pour tout $x \in \mathbb{R}$.

En particulier pour tout x, nous avons $|e^{ix}| = 1$.

Démonstration. La définition de l'exponentielle sur $\mathbb{C}$ est la définition 14.37. Cette définition fonctionne parce que $\mathbb{C}$ est une algèbre normée, et que $\mathbb{C}$ est un $\mathbb{C}$-module vérifiant l'inégalité $|zz'| \leqslant |z||z'|$ (en l'occurrence, une égalité).

Nous remarquons que que i^k vaut 1, i, -1, $-i$. Donc un terme sur deux est imaginaire pur et parmi ceux-là, un sur deux est positif. À bien y regarder, les termes imaginaires purs forment la série du sinus et ceux réels la série du cosinus.

Si vous aimez les formules,

$$e^{iy} = \sum_{n=0}^{\infty} \frac{(iy)^n}{n!} = \sum_{n=0}^{\infty} (-1)^n \frac{y^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} (-1)^n \frac{y^{2n+1}}{(2n+1)!} = \cos(y) + i\sin(y). \tag{17.21}$$

Nous avons utilisé le fait que $i^{2n} = (-1)^n$ et $i^{2n+1} = i(-1)^n$. □

Lemme 17.11.
Nous avons les formules d'addition d'angles[3]

$$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b) \tag{17.22a}$$
$$\sin(a+b) = \cos(a)\sin(b) + \sin(a)\cos(b) \tag{17.22b}$$
$$\cos(a-b) = \cos(a)\cos(b) + \sin(a)\sin(b) \tag{17.22c}$$

pour tout a, b réels.

Démonstration. Nous utilisons la formule d'addition dans l'exponentielle, proposition (14.203) et la formule (17.20) avant de séparer les parties réelles et imaginaires :

$$e^{i(a+b)} = e^{ia}e^{ib} = \cos(a)\cos(b) - \sin(a)\sin(b) + i\big(\cos(a)\sin(b) + \sin(a)\cos(b)\big). \tag{17.23}$$

Cela est également égal à

$$\cos(a+b) + i\sin(a+b). \tag{17.24}$$

En identifiant les parties réelle et imaginaires, nous obtenons les formules (17.22a) et (17.22c) annoncées.

Pour la formule (17.22c), il suffit de se souvenir que $\sin(-b) = -\sin(b)$ et $\cos(-b) = \cos(b)$ (ces deux égalités sont immédiatement visibles sur les développements en série : l'un a uniquement des puissances paires et l'autre impaires) et d'écrire (17.22a) avec $-b$ au lieu de b. □

Corolaire 17.12.
Les formules suivantes pour les duplications d'angles s'ensuivent :

$$\cos(2a) = \cos^2(a) - \sin^2(a) \tag{17.25a}$$
$$\sin(2a) = 2\cos(a)\sin(a). \tag{17.25b}$$

Démonstration. Poser $b = a$ dans les relations du lemme 17.11. □

Lemme 17.13.
Un sous-groupe de $(\mathbb{R}, +)$ est soit dense dans $\mathbb{R}$ soit de la forme $p\mathbb{Z}$ pour un certain réel $p \neq 0$.

3. Rien ne nous empêche de donner ce nom à ces formules, mais seriez-vous capable de définir précisément le mot « angle » ?

Démonstration. Soit A, un sous-groupe de $(\mathbb{R},+)$ qui ne soit pas dense. Soit un intervalle $]a,b[$ qui n'intersecte pas A (si vous voulez frimer, vous noterez ici que nous utilisons le fait que les intervalles ouverts forment une base de la topologie de $\mathbb{R}$). Si $d=|b-a|$, l'ensemble A ne contient pas deux éléments séparés par strictement moins de d. Soit p, le plus petit élément strictement positif de A ; nous avons $p\geqslant d$ (parce que $0\in A$ de toutes façons).

Vu que A est un groupe nous avons $p\mathbb{Z}\subset A$.

Pour l'inclusion inverse, si $x\in A$ est hors de $p\mathbb{Z}$, il existe un $y\in p\mathbb{Z}$ avec $|x-y|<p$. Et donc le nombre $|x-y|$ est dans A tout en étant plus petit que p. Contradiction. □

Proposition-définition 17.14 (Périodicité, le nombre π[351])**.**
Plusieurs choses à propos de la périodicité de la fonction cos.

(1) La fonction cos *est périodique.*

(2) Un nombre $T>0$ est une période si et seulement si $\cos(T)=1$ *et* $\sin(T)=0$.

Nous définissons le nombre $\pi>0$ comme étant la moitié de la période de la fonction cos *:*

$$2\pi=\min\{T>0 \text{ tel que } \cos(x+T)=\cos(x)\,\forall x\}. \tag{17.26}$$

Démonstration. Plusieurs étapes.

La fonction cosinus n'est pas toujours positive Supposons d'abord que $\cos(x)>0$ pour tout $x\in\mathbb{R}$. Dans ce cas, la fonction sin est strictement croissante. Mais les deux fonctions sont bornées par 1 du fait de la formule $\cos^2(x)+\sin^2(x)=1$. La fonction sin étant croissante et bornée, elle est convergente vers un réel par la proposition 11.66 :

$$\lim_{x\to\infty}\sin(x)=\ell \tag{17.27}$$

pour un certain $\ell>0$. Avec ça nous avons aussi (pour cause de dérivée) $\lim_{x\to\infty}\sin'(x)=0$, c'est-à-dire $\lim_{x\to\infty}\cos(x)=0$. Mais vu que $\cos^2(x)+\sin^2(x)=1$ nous en déduisons que $\lim_{x\to\infty}\sin(x)=1$. Mézalor $\lim_{x\to\infty}\cos'(x)=-1$, ce qui donne que la fonction cos n'est pas bornée. Cela est impossible. Nous en déduisons que $\cos(x)$ n'est pas toujours positive.

Il existe $T>0$ tel que $\cos(T)=1$ et $\sin(T)=0$ Par ce que nous venons de faire, il existe $r>0$ tel que $\cos(r)=0$. Pour cette valeur, nous avons aussi obligatoirement $\sin(r)=\pm1$. Nous avons aussi, en utilisant les formules (17.22),

$$\cos(2r)=\cos^2(r)-\sin^2(r)=-1 \tag{17.28a}$$
$$\sin(2r)=2\cos(r)\sin(r)=0. \tag{17.28b}$$

et par conséquent

$$\cos(4r)=\cos^2(2r)-\sin^2(2r)=1 \tag{17.29a}$$
$$\sin(4r)=2\cos(2r)\sin(2r)=0. \tag{17.29b}$$

Donc $T=4r$ fonctionne.

Si T est une période Nous entrons dans le vif de la preuve. Soit un $T>0$ tel que $\cos(x+T)=\cos(x)$ pour tout $x\in\mathbb{R}$. Avec la formule d'addition d'angle dans le cosinus nous cherchons un T tel que

$$\cos(x+T)=\cos(x)\cos(T)-\sin(x)\sin(T)=\cos(x) \tag{17.30}$$

et donc tel que

$$\cos(x)\big(\cos(T)-1\big)=\sin(x)\sin(T). \tag{17.31}$$

Nous dérivons cette équation :

$$-\sin(x)\big(\cos(T)-1\big)=\cos(x)\sin(T). \tag{17.32}$$

Nous multiplions chacune des deux équations (17.31) et (17.32) par $\sin(x)$ et $\cos(x)$ pour obtenir les quatre relations suivantes :

$$\cos^2(x)\big(\cos(T)-1\big)-\sin(x)\cos(x)\sin(T)=0 \tag{17.33a}$$
$$-\sin(x)\cos(x)\big(\cos(T)-1\big)-\cos^2(x)\sin(T)=0 \tag{17.33b}$$
$$\sin(x)\cos(x)\big(\cos(T)-1\big)-\sin^2(x)\sin(T)=0 \tag{17.33c}$$
$$\sin^2(x)\big(\cos(T)-1\big)-\sin(x)\cos(x)\sin(T)=0 \tag{17.33d}$$

En faisant (17.33a) moins (17.33d) nous trouvons $\cos(T)=1$. Et en sommant (17.33b) avec (17.33c) nous avons $-\sin(T)=0$.

Si $T>0$ est tel que $\sin(T)=0$ et $\cos(T)=1$ Alors les formules d'addition d'angle du lemme 17.11 donnent tout de suite

$$\cos(x+T)=\cos(x). \tag{17.34}$$

À de niveau nous croyons avoir prouvé que cos était périodique et que la période est donnée par

$$\min\{T>0 \text{ tel que } \sin(T)=0,\cos(T)=1\}. \tag{17.35}$$

Or rien n'est moins sûr parce qu'il pourrait arriver que ce minimum n'existe pas, c'est-à-dire que l'infimum soit zéro. Autrement dit, il peut arriver que l'ensemble des périodes soit dense. Plus précisément, soit $P\subset\mathbb{R}$ l'ensemble des périodes de cos. C'est un sous-groupe de $(\mathbb{R},+)$ et le lemme 17.13 nous dit que P est soit dense dans $\mathbb{R}$ soit de la forme $p\mathbb{Z}$ pour un $p>0$.

Si P est dense, soit $t\in\mathbb{R}$ et une suite (t_n) dans P telle que $t_n\to t$. Pour tout x et tout n nous avons

$$\cos(x+t_n)=\cos(x), \tag{17.36}$$

Vu que la fonction cosinus est continue, nous pouvons passer à la limite et écrire $\cos(x+t)=\cos(x)$. Cela étant valable pour tout x et pour tout t, la fonction cosinus est constante. Or nous savons que ce n'est pas le cas, donc P n'est pas dense. Donc cosinus est périodique. □

Proposition 17.15.
La fonction sin *est périodique de période π et*

$$2\pi=\min\{T>0 \textit{ tel que } \sin(T)=0,\cos(T)=1\}. \tag{17.37}$$

17.16.
Notons que tout ceci ne nous donne pas la plus petite indication d'ordre de grandeur de la valeur de π. Cela peut encore être 0.1 autant que 500.

Proposition 17.17 ([351, 1])**.**
Des propriétés à la chaine à propos des sinus, cosinus et de leurs périodes.

(1) Le nombre 2π est le plus petit nombre strictement positif tel que

$$\begin{cases}\cos(2\pi)=1 & (17.38a)\\ \sin(2\pi)=0. & (17.38b)\end{cases}$$

(2) Les fonctions sin *et* cos *sont périodiques de période 2π.*

(3) Nous avons $\cos(\pi)=-1$ et $\sin(\pi)=0$.

(4) Pour tout $a\in\mathbb{R}$ nous avons

$$\cos(a+\pi)=-\cos(\pi) \tag{17.39a}$$
$$\sin(a+\pi)=-\sin(\pi). \tag{17.39b}$$

(5) Le nombre π est le plus petit $T>0$ tel que $\cos(T)=-1$ et $\sin(T)=0$.

(6) *Nous avons*

$$\begin{cases} \cos(\pi/2) = 0 & (17.40a) \\ \sin(\pi/2) = 1. & (17.40b) \end{cases}$$

(7) *Nous avons les formules*

$$\begin{cases} \cos(x + \pi/2) = -\sin(x) & (17.41a) \\ \sin(x + \pi/2) = \cos(x) & (17.41b) \end{cases}$$

pour tout $x \in \mathbb{R}$.

(8) *Le nombre* $\pi/2$ *est le plus petit* $T > 0$ *vérifiant* $\sin(T) = 1$, $\cos(T) = 0$.

(9) *Nous avons les valeurs*

$$\begin{cases} \cos(\frac{3\pi}{2}) = 0 & (17.42a) \\ \sin(\frac{3\pi}{2}) = -1. & (17.42b) \end{cases}$$

(10) *Le nombre* $\pi/2$ *est le plus petit* $T > 0$ *tel que* $\cos(T) = 0$.

(11) *Pour tout* $x \in]0, \frac{\pi}{2}[$, *nous avons* $\cos(x) > 0$ *et* $\sin(x) > 0$.

Démonstration. C'est parti.

(1) Le fond de la proposition 17.14 est que toutes les périodes $T > 0$ vérifient $\cos(T) = 1$ et $\sin(T) = 0$. La définition de π est que c'est la plus petite période.

(2) En utilisant le fait que l'une est la dérivée de l'autre, si T est une période de cos nous avons

$$\begin{aligned} \sin(x + T) &= -\cos'(x + T) && (17.43a) \\ &= -\lim_{\epsilon \to 0} \frac{\cos(x + T + \epsilon) - \cos(x + T)}{\epsilon} && (17.43b) \\ &= -\lim_{\epsilon \to 0} \frac{\cos(x + \epsilon) - \cos(x)}{\epsilon} && (17.43c) \\ &= -\cos'(x) && (17.43d) \\ &= \sin(x). && (17.43e) \end{aligned}$$

Nous déduisons que toute période de cos est une période de sin. De la même façon, nous pouvons prouver le contraire : toute période de sin est une période de cos.

(3) D'un côté nous avons

$$\cos(2\pi) = \cos^2(\pi) - \sin^2(\pi) = 1 \tag{17.44}$$

parce que $\cos(2\pi) = \cos(0) = 1$. Vu que $\cos(\pi)$ et $\sin(\pi)$ sont bornés par -1 et 1, nous devons avoir $\sin(\pi) = 0$ et $\cos(\pi) = \pm 1$.

Mais d'un autre côté, le nombre 2π est le plus petit T vérifiant $\cos(T) = 1$, $\sin(T) = 0$. Donc avoir $\cos(\pi) = 1$ n'est pas possible. Nous concluons

$$\begin{cases} \cos(\pi) = -1 & (17.45a) \\ \sin(\pi) = 0. & (17.45b) \end{cases}$$

(4) Il s'agit d'utiliser les formules d'addition d'angles du lemme 17.11 pour calculer $\cos(a + \pi)$ et $\sin(a + \pi)$ en tenant compte du fait que $\cos(\pi) = -1$ et $\sin(\pi) = 0$.

(5) Soit $a \in]0, \pi[$ tel que $\cos(a) = -1$ et $\sin(a) = 0$. Alors nous avons

$$\begin{aligned} \cos(a + \pi) &= -\cos(\pi) = 1 && (17.46a) \\ \sin(a + \pi) &= -\sin(\pi) = 0, && (17.46b) \end{aligned}$$

ce qui donnerait $a + \pi \in]\pi, 2\pi[$ dont le cosinus est 1 et le sinus est zéro. Mais nous savons déjà que 2π est le minimum pour cette propriété.

(6) Nous avons

$$-1 = \cos(\pi) = \cos^2(\pi/2) - \sin^2(\pi/2), \tag{17.47}$$

donc $\cos(\pi/2) = 0$ et $\sin^2(\pi/2) = 1$, ce qui donne $\sin(\pi/2) = \pm 1$.

Nous devons départager le $\pm$. Pour cela nous savons que $\sin'(0) = \cos(0) = 1$, donc il existe $\epsilon > \epsilon$ tel que pour tout $x \in]0, \epsilon[$ nous avons $0 < \cos(x) < 1$ et $0 \sin(x) < 1$. Nous choisissons ϵ plus petit que $\pi/2$.

Supposons que $\sin(\pi/2) = -1$. Le théorème des valeurs intermédiaires 11.50 dit qu'il existe $x_0 \in]\epsilon, \pi/2[$ tel que $\sin(x_0) = 0$. Pour cette valeur de x_0 nous devons aussi avoir $\cos(x_0) = \pm 1$. Mais vu que 2π est minium pour avoir $\cos = 1$ et $\sin = 0$ nous devons avoir $\cos(x_0) = -1$. Alors nous avons aussi

$$\cos(x_0 + \pi) = \cos(x_0)\cos(\pi) - \sin(x_0)\sin(\pi) = -\cos(x_0) = 1 \tag{17.48a}$$

$$\sin(x_0 + \pi) = \cos(x_0)\sin(\pi) + \sin(x_0)\cos(\pi) = \sin(x_0) = 0. \tag{17.48b}$$

Encore une fois par minimalité de 2π, cela ne va pas. Conclusion : $\sin(\pi/2) = 1$.

(7) Il s'agit encore d'utiliser les formules d'addition d'angle en tenant compte des valeurs remarquables $\cos(\pi/2) = 0$ et $\sin(\pi/2) = 1$.

(8) Supposons $x_0 \in]0, \pi/2[$ tel que $\sin(x_0) = 1$ et $\cos(x_0) = 0$. En utilisant les formules (17.41) nous avons

$$\cos(x_0 + \pi/2) = -1 \tag{17.49a}$$

$$\sin(x_0 + \pi/2) = 0, \tag{17.49b}$$

avec $x_0 + \pi/2 < \pi$. Cela contredirait la minimalité de π.

(9) Il s'agit d'utiliser les formules (17.41) :

$$\cos(\frac{3\pi}{2}) = \cos(\pi + \pi/2) = -\sin(\pi) = 0 \tag{17.50a}$$

$$\sin(\frac{3\pi}{2}) = \sin(\pi + \pi/2) = \cos(\pi) = -1. \tag{17.50b}$$

(10) Si $\cos(x_0) = 0$ alors $\sin(x_0) = -1$ (parce que $\sin(x_0) = 1$ est déjà exclu). Alors $\cos(x_0 + \pi/2) = 1$ et $\sin(x_0 + \pi/2) = 0$, ce qui est également impossible.

(11) La fonction cosinus est continue (proposition 17.1) et $\cos(0) = 1$. Le théorème des valeurs intermédiaires implique que si $\cos(x) \leqslant 0$, alors il existe $t \in]0, x]$ avec $\cos(x) = 0$. Cela n'est pas possible pour $x < \pi/2$ par par le point (8).

Le cosinus est positif sur l'intervalle considéré et $\sin'(x) = \cos(x)$. Donc $\sin(0) = 0$ et la dérivée est positive. La proposition 11.125 conclu que sin est strictement croissante et donc strictement positive.

□

Corolaire 17.18.
Des nombres $x, y \in \mathbb{R}$ vérifient $e^{ix} = e^{iy}$ si et seulement si il existe $k \in \mathbb{Z}$ tel que $y = x + 2k\pi$.

Tout cela nous permet de calculer quelques valeurs remarquables de cosinus et sinus ainsi que d'écrire le tableau de variations de sinus et cosinus.

Lemme 17.19.
Nous avons les valeurs remarquables

$$\sin(\frac{\pi}{4}) = \cos(\frac{\pi}{4}) = \frac{\sqrt{2}}{2}. \tag{17.51}$$

Démonstration. La relation (17.25b) donne

$$0 = \cos(\pi/2) = \cos^2(\pi/4) - \sin^2(\pi/4). \tag{17.52}$$

Donc $\cos^2(\pi/4) = \sin^2(\pi/4)$. Mais vu que $\sin(\pi/4)$ et $\cos(\pi/4)$ sont positifs, ils sont égaux.

Mais $\sin^2(\pi/4) + \cos^2(\pi/4) = 1$. Donc le nombre $x = \cos(\pi/4) = \sin(\pi/4)$ vérifie l'équation $2x^2 = 1$, donc l'unique solution positive est $x = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$. □

Lemme 17.20.
Nous avons la valeur remarquable

$$\cos(\frac{\pi}{3}) = \frac{1}{2}. \tag{17.53}$$

Démonstration. Il faut utiliser la formule (17.22a) avec $\cos(\pi) = \cos(2\pi/3 + \pi/3)$ en sachant que $\cos(\pi) = -1$. Ensuite $\cos(2\pi/3) = \cos(\pi/3 + \pi/3)$. En décomposant ainsi, nous exprimons $-1 = \cos(\pi)$ en termes de $\cos(\pi/3)$ et de $\sin(\pi/3)$. En substituant $\sin^2(\pi/3) = 1 - \cos^2(\pi/3)$ nous trouvons que le nombre $\cos(\pi/3)$ vérifie l'équation

$$4x^3 - 3x + 1 = 0. \tag{17.54}$$

Croyez-le ou non, les solutions de cette équation sont $x = -1$ et $x = 1/2$. Allez. Faisons comme si nous le savions pas. En tout cas, ces deux nombres sont des solutions, et nous avons la factorisation [4]

$$4x^3 - 3x + 1 = (2x - 1)^2(x + 1). \tag{17.55}$$

Donc $1/2$ est de multiplicité 2 et -1 de multiplicité 1. Le théorème 3.194 nous dit qu'il n'y a alors pas d'autres racines que ces deux-là [5].

Nous en déduisons que la valeur de $\cos(\pi/3)$ est soit $1/2$ soit -1. La proposition 17.17(5) nous dit qu'il est impossible que $\cos(\pi/3)$ soit égal à -1 parce que $\pi/3 < \pi$. Donc $\cos(\pi/3) = 1/2$ comme annoncé. □

Remarque 17.21.
Vous avez déjà sans doute vu la démonstration de $\cos(30^\circ) = 1/2$ à partir de la figure 17.4. Il n'est pas possible de l'utiliser parce que cela n'est en réalité pas loin d'être la définition de l'angle entre deux droites.

Si vous voulez savoir la définition de l'angle entre deux droites, il faut passer par la définition 17.147, laquelle se base sur le lemme 17.131 qui, elle-même, se base sur la proposition 17.46.

Bref, à notre niveau, nous sommes encore loin de pouvoir faire des raisonnements trigonométriques sur base de géométrie dans les triangles.

Proposition 17.22.
Pour tout $x \in [0, \pi/4[$ *nous avons* $\cos(x) > \sin(x)$.

Démonstration. Nous posons $f(x) = \cos(x) - \sin(x)$. Elle vérifie $f(0) = 1$. En utilisant les dérivées du lemme 17.2, nous trouvons

$$f'(x) = -\big(\sin(x) + \cos(x)\big). \tag{17.56}$$

Mais sur $]0, \pi/2[$ nous avons $\cos(x) > 0$ et $\sin(x) > 0$ (proposition 17.17(11)). Donc f est strictement décroissante. Elle ne peut donc passer qu'une seule fois par zéro. Le lemme 17.19 nous indique que $f(\pi/4) = 0$. Donc $f(x) > 0$ sur $[0, \pi/4[$. □

<++>

Proposition 17.23.
Quelque valeurs trigonométriques.

4. Factorisation d'un polynôme en sachant des racines, proposition 3.188.
5. Nous attirons votre attention sur le fait que cela n'est en aucun cas une trivialité.

(1) *Pour le sinus :*
  (1a) $\sin(0) = 0$
  (1b) $\sin(\pi/6) = 1/2$
  (1c) $\sin(\pi/4) = \sqrt{2}/2$
  (1d) $\sin(\pi/3) = \sqrt{3}/2$
  (1e) $\sin(\pi/2) = 1$
(2) *Pour le cosinus :*
  (2a) $\cos(0) = 1$
  (2b) $\cos(\pi/6) = \sqrt{3}/2$
  (2c) $\cos(\pi/4) = \sqrt{2}/2$
  (2d) $\cos(\pi/3) = 1/2$
  (2e) $\cos(\pi/2) = 0$
(3) *Pour la tangente :*
  (3a) $\tan(0) = 0$
  (3b) $\tan(\pi/6) = \sqrt{3}/3$
  (3c) $\tan(\pi/4) = 1$
  (3d) $\tan(\pi/3) = \sqrt{3}$
  (3e) $\tan(\pi/2)$ *est non défini.*

Démonstration. Plusieurs ont déjà été faites. Les autres ne seront pas démontrées dans l'ordre énoncé.

$\underline{\sin(0) = 0}$ Substitution dans la définition (17.2).

$\underline{\sin(\pi/4) = \sqrt{2}/2}$ C'est le lemme 17.19.

$\underline{\sin(\pi/3) = 1/\sqrt{2}}$ Nous utilisons la formule $\sin^2(x) + \cos^2(x) = 1$ avec $x = \pi/3$. Cela donne $\sin^2(\pi/3) = 1/2$. Nous en déduisons que $\sin(\pi/3)$ vaut $\pm\frac{1}{\sqrt{2}}$.

La proposition 17.17(5) nous dit que sin est positive sur $[0, \pi]$. Donc c'est bien la possibilité $1/\sqrt{2}$ qui est la bonne.

$\underline{\sin(\pi/6) = 1/2}$ **et** $\underline{\cos(\pi/6) = \sqrt{3}/2}$ Nous partons de l'équation (17.25b) pour écrire

$$\sin(\pi/3) = 2\cos(\pi/6)\sin(\pi/6). \tag{17.57}$$

Nous avons déjà vu que $\sin(\pi/3) = \sqrt{3}/2$. En posant $x = \sin(\pi/6)$ nous avons également $\cos(\pi/6) = \sqrt{1 - x^2}$ parce que nous savons que la fonction cosinus est positive sur $[0, \pi/2]$ (proposition 17.17(11)). Nous avons donc l'équation

$$\frac{\sqrt{3}}{2} = 2x\sqrt{1 - x^2}. \tag{17.58}$$

Nous passons au carré et posons $y = x^2$. Après quelque manipulations,

$$16y^2 - 16y + 3 = 0. \tag{17.59}$$

Cela donne deux possibilités pour y : $\frac{3}{4}$ et $\frac{1}{4}$. Vu que $x > 0$, nous pouvons simplement passer à la racine carré : $x = \sqrt{3}/2$ ou $x = 1/2$.

Notez que si nous avion posé $x = \cos(\pi/6)$ au lieu de $x = \sin(\pi/6)$, nous aurions obtenu le même résultat. Donc $\sin(\pi/6)$ et $\cos(\pi/6)$ peuvent tout deux avoir les valeurs $\sqrt{3}/2$ ou $1/2$. Cela fait 4 possibilités.

Étant donné que $\sin^2(\pi/6) + \cos^2(\pi/6) = 1$, les deux possibilités avec $\sin(\pi/6) = \cos(\pi/6)$ sont exclues.

La proposition 17.22 nous dit aussi que $\cos(\pi/6) > \sin(\pi/6)$. Donc $\cos(\pi/6) = \sqrt{3}/2$ et $\sin(\pi/6) = 1/2$.

$\underline{\sin(\pi/2) = 1}$ C'est dans (17.40).

$\underline{\cos(0) = 1}$ Substitution dans la définition.

$\underline{\cos(\pi/6) = \sqrt{3}/2}$ Déjà fait avec le sinus de $\pi/6$.

$\underline{\cos(\pi/4) = \sqrt{2}/2}$ Lemme 17.19.

$\underline{\cos(\pi/3) = 1/2}$ Lemme 17.20.

$\underline{\cos(\pi/2) = 0}$ Dans (17.40).

Toutes les valeurs pour la tangente s'obtiennent maintenant par la définition en calculant $\tan(x) = \frac{\sin(x)}{\cos(x)}$. □

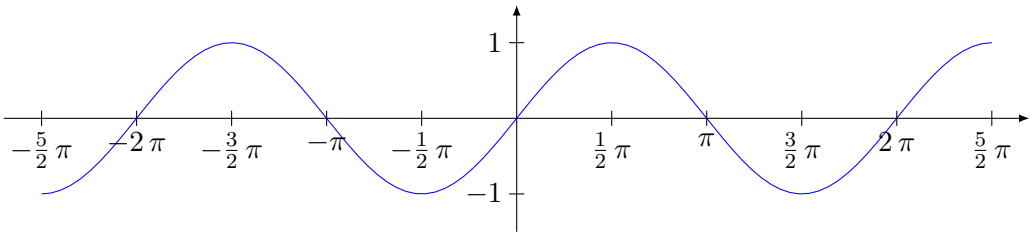
Voici un tableau qui rappelle les valeurs à retenir pour les fonctions sinus, cosinus et tangente.

x	$\sin(x)$	$\cos(x)$	$\tan(x)$
0	0	1	0
$\pi/6$	$1/2$	$\sqrt{3}/2$	$\sqrt{3}/3$
$\pi/6$	$1/2$	$\sqrt{3}/2$	$\sqrt{3}/3$
$\pi/4$	$\sqrt{2}/2$	$\sqrt{2}/2$	1
$\pi/3$	$\sqrt{3}/2$	$1/2$	$\sqrt{3}$
$\pi/2$	1	0	N.D.

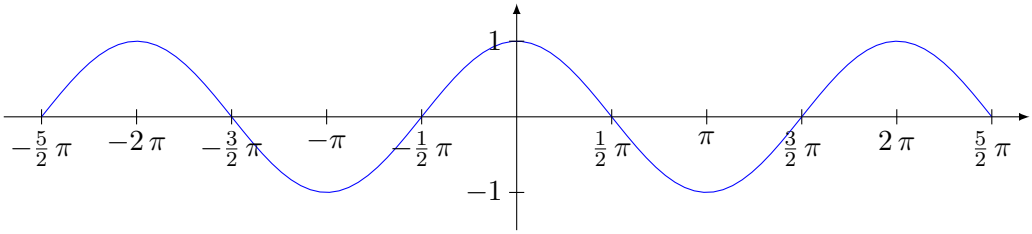
(17.60)

où « N.D. » signifie « non défini ».

Rappelons le graphe de la fonction sinus :



celui de la fonction cosinus :



Lemme 17.24.
Pour toute valeur de $x \in \mathbb{R}$ on a $|\sin(x)| \leqslant |x|$.

Démonstration. Nous séparons des cas en fonction des valeurs.

— Si $0 \leqslant x \leqslant \pi/2$ alors le sinus de x est la longueur du cathète verticale du triangle rectangle de sommets $O = (0,0)$, $A = (\cos(x), \sin(x))$ et $B = (\cos(x), 0)$. Le triangle de sommets A, B et $C = (1,0)$ est aussi rectangle et nous savons que chacun des cathètes ne peut pas être plus long que l'hypoténuse. Donc $\sin(x)$ est inférieur à la longueur du segment AC. À son tour le segment AC ne peut pas être plus long que l'arc de cercle $\widehat{A0C}$, car le chemin le plus court entre deux points du plan est toujours donné par un morceau de droite. La longueur de l'arc de cercle $\frown AC$ est *par définition* la mesure en radiants de l'angle $\widehat{AOC}$, qui est x et on a l'inégalité $\sin(x) \leqslant x$.

— Si $-\pi/2 \leqslant x \leqslant 0$ le même raisonnement que au point précédent permet de conclure que $\sin(x) \leqslant |x|$.

— Nous savons par ailleurs que la fonction sinus prend ses valeurs dans l'intervalle $[-1, 1]$ et donc pour tout x tel que $|x| \geqslant \pi/2 \equiv 1,57\ldots$ on a forcement $|\sin(x)| \leqslant |x|$.

□

$$\begin{cases} x = r\cos\theta & (17.61a) \\ y = r\sin\theta & (17.61b) \end{cases}$$

avec $r \in]0, \infty[$ et $\theta \in [0, 2\pi[$. Le jacobien vaut r.

Exemple 17.25.
Développer la fonction cos autour de $x = \frac{\pi}{3}$. Utiliser la valeur remarquable du lemme 17.20. Nous développons autour de $h = 0$ la fonction $\cos(\frac{\pi}{3} + h)$:

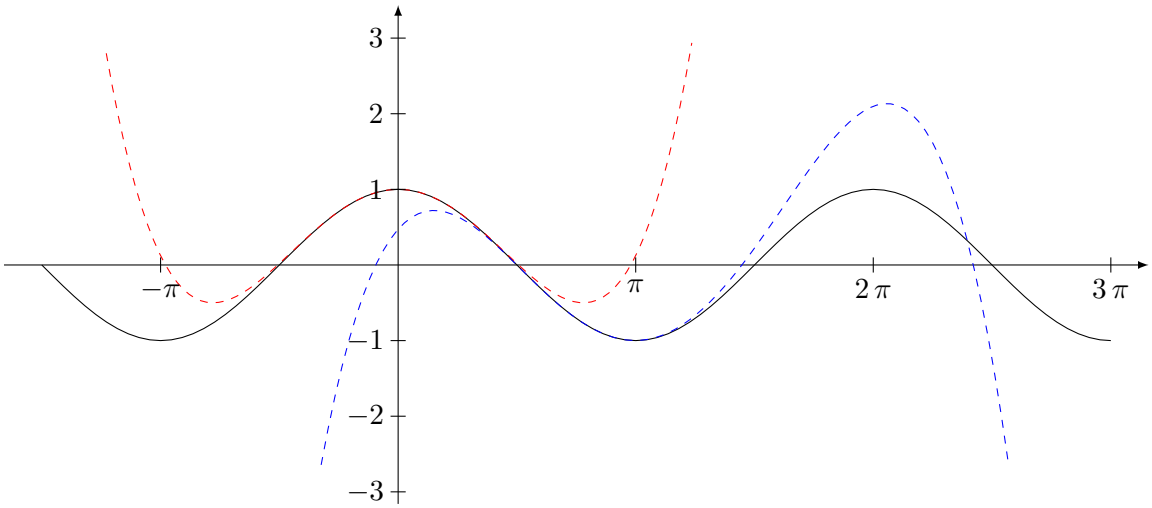
$$\cos\left(\frac{\pi}{3} + h\right) \sim \cos\left(\frac{\pi}{3}\right) + h\cos'(\frac{\pi}{3}) + \frac{h^2}{2}\cos''\left(\frac{\pi}{3}\right) = \frac{1}{2} - \frac{\sqrt{3}}{2}h - \frac{1}{4}h^2. \tag{17.62}$$

Il est aussi possible d'écrire cela en notant $x = x_0 + h$, c'est-à-dire en remplaçant h par $x - \frac{\pi}{3}$:

$$\cos(x) \sim \frac{1}{2} - \frac{\sqrt{3}}{2}(x - \frac{\pi}{3}) - \frac{1}{4}(x - \frac{\pi}{3})^2. \tag{17.63}$$

△

Pour donner une idée nous avons dessiné sur le graphe suivant la fonction sinus et ses développements d'ordre 4 autour de zéro et autour de $3\pi/4$.



17.2 Très modeste approximation de π

Nous sommes en droit de vouloir une valeur approchée de π.

Lemme 17.26.
Nous avons l'approximation numérique

$$2\sqrt{2} < \pi < 4. \tag{17.64}$$

Démonstration. Grace au lemme 17.19 nous savons que la fonction sin passe de 0 à $\sqrt{2}/2$ sur un intervalle de taille $\pi/4$ avec une dérivé majorée par 1. Par conséquent

$$\frac{\pi}{4} > \frac{\sqrt{2}}{2} \tag{17.65}$$

et donc [6]

$$\pi > 2\sqrt{2} \simeq 2.82 \tag{17.66}$$

De plus la fonction sin passe de 0 à $\sqrt{2}/2$ sur un intervalle de taille $\pi/4$ avec une dérivée majorée par $\sqrt{2}/2$, donc

$$\frac{\pi}{4} < \frac{\sqrt{2}/2}{\sqrt{2}/2}, \tag{17.67}$$

ce qui donne

$$\pi < 4. \tag{17.68}$$

□

6. Sérieusement, êtes vous capables de trouver une approximation de $\sqrt{2}$ en ne vous basant que sur des choses vues jusqu'ici ?

Pour avoir une meilleur approximation de π, nous pouvons remarquer que $\pi \in]2.82, 4[$, et que cet intervalle est suffisamment petit pour ne pas recouvrir l'intervalle correspondant pour 2π. L'équation $\cos(x) = -1$ possède donc une unique solution dans cet intervalle (et cette solution est π). Nous pouvons donc faire une dichotomie pour trouver la valeur de π, pourvu que nous ayons une façon d'évaluer des valeurs de $\cos(x)$ de façon pas trop ridicule.

17.3 Cercle trigonométriques

Proposition 17.27 ([352]).
Soient des fonctions $f, g\colon I \to \mathbb{R}$ de classe C^1 sur l'ouvert I de $\mathbb{R}$ telles que $f^2 + g^2 = 1$. Soit $t_0 \in I$ et θ_0 tel que $f(t_0) = \cos(\theta_0)$ et $g(t_0) = \sin(\theta_0)$.

Alors il existe une unique fonction continue $\theta\colon I \to \mathbb{R}$ telle que

$$\begin{cases} \theta(t_0) = \theta_0 & (17.69a) \\ f = \cos \circ \theta & (17.69b) \\ g = \sin \circ \theta. & (17.69c) \end{cases}$$

Démonstration. Nous commençons par l'existence, en passant par les nombres complexes. Soit $h\colon I \to \mathbb{C}$ définie par $h = f + ig$. Nous avons $h\bar{h} = 1$ et nous définissons

$$\theta(t) = \theta_0 - i \int_{t_0}^{t} h'(s)\overline{h(s)} ds. \tag{17.70}$$

Cette intégrale existe pour tout t parce que les fonctions f et g étant de classe C^∞, elles sont bornées sur le compact $[t_0, t]$. De plus θ est une fonction continue parce que c'est une primitive (proposition 13.235) [7].

La dérivée de θ est la fonction $s \mapsto -ih'(s)\overline{h(s)}$.

Utilisant la formule du lemme 17.10 sur la forme trigonométrique des nombres complexes, nous calculons :

$$\frac{d}{dt}\Big[he^{-i\theta}\Big]_{t=0} = e^{-i\theta}(h' - h\theta') = e^{-i\theta}(h' - ih(-i)h'\bar{h}) = 0. \tag{17.71}$$

Par conséquent il existe $c \in \mathbb{C}$ tel que $he^{-i\theta} = c$. Mais $h(t_0) = f(t_0) + ig(t_0) = \cos(\theta_0) + i\sin(\theta_0) = e^{i\theta_0}$, du coup

$$h(t_0)e^{-i\theta(t_0)} = c \tag{17.72}$$

donne immédiatement $c = 1$, ou encore $e^{i\theta(t)} = h(t)$, c'est-à-dire que

$$f + ig = \cos \circ \theta + i \sin \circ \theta, \tag{17.73}$$

ce qu'il fallait pour l'existence.

Pour l'unicité nous supposons avoir une autre fonction, α qui satisfait aux exigences. Pour tout $t \in I$ nous avons

$$e^{i\theta(t)} = e^{i\alpha(t)}. \tag{17.74}$$

Il existe donc une fonction $n\colon I \to \mathbb{N}$ telle que $\theta(t) = \alpha(t) + 2n(t)\pi$. Par continuité de θ et α, la fonction n doit être constante, mais vu que $\theta(t_0) = \alpha(t_0)$ nous avons $n = 1$. □

17.3.1 Les fonctions tangente et arc tangente

Définition 17.28.
*La fonction **tangente** est :*

$$\tan(x) = \frac{\sin(x)}{\cos(x)} \tag{17.75}$$

où sin *et* cos *sont de la définition 17.1.*

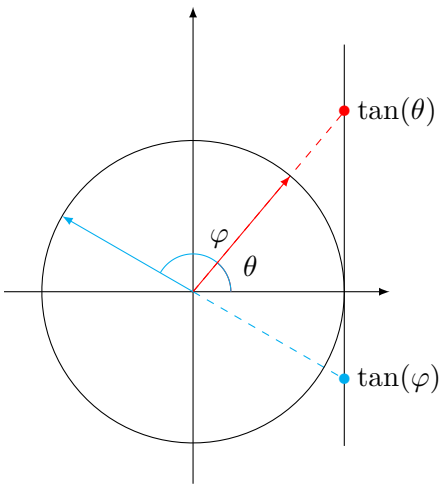


FIGURE 17.1 – Interprétation géométrique de la fonction tangente. La tangente de l'angle θ est positive (et un peu plus grande que 1) tandis que celle de la tangente de l'angle φ est négative.

La fonction tangente n'est pas définie sur les points de la forme $x = \frac{\pi}{2} + k\pi$, $k \in \mathbb{Z}$. Une interprétation géométrique, qui justifie le nom, est donnée sur la figure 17.1.

Proposition 17.29.
La fonction

$$\begin{aligned} \tan\colon \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[&\to \mathbb{R} \\ x &\mapsto \tan(x) \end{aligned} \tag{17.76}$$

est une bijection.

Démonstration. Le cosinus ne s'annulant pas sur l'intervalle donné, la fonction est bien définie. Nous avons

$$\lim_{x\to\pi/2^-} \tan(x) = +\infty \tag{17.77}$$

parce que la limite du sinus est 1 est celle du cosinus est zéro par les valeurs positives. Le même raisonnement donne la limite en $-\pi/2$ qui vaut $-\infty$. Le théorème des valeurs intermédiaires[8] dit que la fonction tangente est alors surjective sur $\mathbb{R}$.

Par ailleurs en utilisant les règles de calcul comme la dérivation du quotient 11.115(5) nous trouvons

$$\tan'(x) = \tan^2(x) + 1, \tag{17.78}$$

ce qui nous donne une dérivée partout strictement positive, et donc une fonction strictement croissante et donc injective. □

Le graphe de la fonction tangente est sur la figure 17.2.

En ce qui concerne la bijection réciproque nous avons le théorème suivant.

Théorème 17.30.
La fonction inverse de la tangente,

$$\begin{aligned} \arctan\colon \mathbb{R} &\to \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[\\ x &\mapsto \arctan(x) \end{aligned} \tag{17.79}$$

nommée ***arc tangente*** *est*

(1) impaire et strictement croissante sur $\mathbb{R}$.

7. En réalité nous appliquons la proposition 7.192 à chacune des parties réelles et imaginaires de la fonction $s \mapsto h'(s)\overline{h(s)}$.

8. Théorème 11.50.

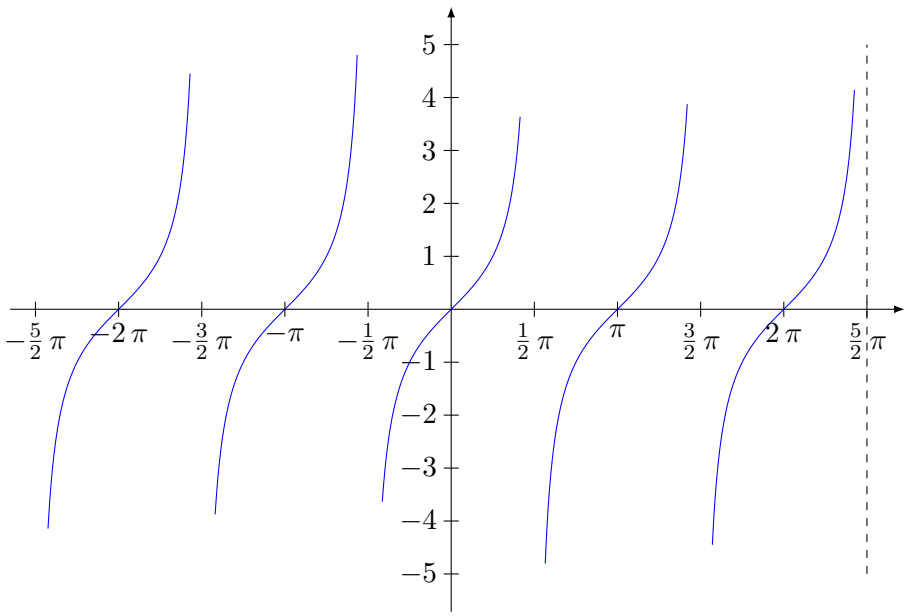


FIGURE 17.2 – Le graphe de la fonction tangente.

(2) dérivable sur $\mathbb{R}$ *de dérivée*

$$\arctan'(x) = \frac{1}{1+x^2}. \tag{17.80}$$

Démonstration. Il est immédiatement visible sur son développement de définition (17.2) que la fonction sinus est impaire. Une vérification similaire montre que la fonction cosinus est paire. La fonction tangente est alors impaire et sa réciproque l'est tout autant.

La fonction arc tangente est également dérivable (donc continue) par la proposition 11.120 parce que la fonction tangente l'est. Notons qu'ici nous nous sommes restreint à $]-\pi/2, \pi/2[$. Sinon, le résultat est faux.

La formule proposée pour la dérivée provient également de la proposition 11.120 et de la dérivée de la tangente : □

Lemme 17.31.
Nous avons les limites

(1) $\lim_{x\to\infty} \arctan(x) = \frac{\pi}{2}$,

(2) $\lim_{x\to-\infty} \arctan(x) = -\frac{\pi}{2}$.

Lemme 17.32.
Nous avons la valeur remarquable

$$\arctan(1/\sqrt{3}) = \frac{\pi}{6}. \tag{17.81}$$

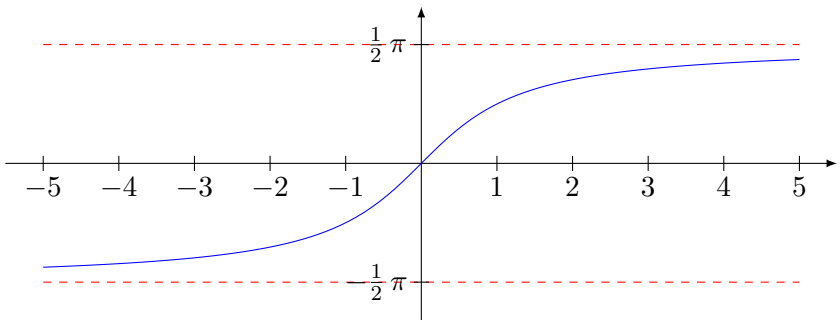
Le nombre $\arctan(x_0)$ se calcule en cherchant l'angle $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ dont la tangente vaut x_0. Nous obtenons le tableau de valeurs suivant :

Lemme 17.33.
Quelques valeurs remarquables de l'arc tangente :

x	0	$\frac{1}{\sqrt{3}}$	1	$\sqrt{3}$
$\arctan(x)$	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$

(17.82)

En ce qui concerne la représentation graphique de la fonction $x \mapsto \arctan(x)$, elle s'obtient « en retournant » la partie entre $-\frac{\pi}{2}$ et $\frac{\pi}{2}$ du graphique de la fonction tangente :



17.3.2 La fonction arc sinus

Nous voulons étudier la fonction

$$\begin{aligned} \sin\colon \mathbb{R} &\to [-1,1] \\ x &\mapsto \sin(x) \end{aligned} \tag{17.83}$$

et sa réciproque éventuelle.

La fonction sinus est continue sur $\mathbb{R}$ mais n'est pas bijective : elle prend une infinité de fois chaque valeur de $J = [-1,1]$. Pour définir une bijection réciproque de la fonction sinus en utilisant le théorème 11.61, nous devons donc choisir un intervalle à partir duquel la fonction sinus est monotone. Nous choisissons l'intervalle

$$I = [-\frac{\pi}{2}, \frac{\pi}{2}]. \tag{17.84}$$

La fonction

$$\begin{aligned} \sin\colon [-\frac{\pi}{2}, \frac{\pi}{2}] &\to [-1,1] \\ x &\mapsto \sin(x) \end{aligned} \tag{17.85}$$

est une bijection croissante et continue. Nous avons donc le résultat suivant.

Théorème 17.34 (Définition et propriétés de arc sinus)**.**
*Nous nommons **arc sinus** la bijection inverse de la fonction* $\sin\colon I \to J$*. La fonction*

$$\begin{aligned} \arcsin\colon [-1,1] &\to [-\frac{\pi}{2}, \frac{\pi}{2}] \\ x &\mapsto \arcsin(x) \end{aligned} \tag{17.86}$$

ainsi définie est

(1) continue et strictement croissante ;

(2) impaire : pour tout $x \in [-1,1]$ *nous avons* $\arcsin(-x) = -\arcsin(x)$.

Démonstration. Nous prouvons le fait que arcsin est impaire. Un élément de l'ensemble de définition de arcsin est de la forme $y = \sin(x)$ avec $x \in [-\pi/2, \pi/2]$. La relation (11.81) s'écrit dans notre cas

$$x = \arcsin\big(\sin(x)\big). \tag{17.87}$$

Nous écrivons d'une part cette équation avec $-x$ au lieu de x :

$$-x = \arcsin\big(\sin(-x)\big) = \arcsin\big(-\sin(x)\big) = \arcsin(-y); \tag{17.88}$$

et d'autre part nous multiplions (17.87) par -1 :

$$-x = -\arcsin\big(\sin(x)\big) = -\arcsin(y). \tag{17.89}$$

En égalisant les valeurs (17.88) et (17.89) nous trouvons

$$\arcsin(-y) = -\arcsin(y), \tag{17.90}$$

ce qui signifie que arcsin est une fonction impaire. □

Notons que cette preuve repose sur le fait que tout élément de l'ensemble de définition de la fonction arc sinus peut être écrit sous la forme $\sin(x)$ pour un certain x.

Si $x_0 \in [-1,1]$ est donné, calculer $\arcsin(x_0)$ revient à trouver un angle θ_0 dans $[-\frac{\pi}{2}, \frac{\pi}{2}]$ pour lequel $\sin(\theta_0) = x_0$. Un tel angle sera forcément unique.

Remarque 17.35.
La définition de arc sinus découle du choix de l'intervalle I, qui est une convention. Il aurait été possible de faire un choix différent : pourriez-vous trouver la réciproque de la fonction sinus sur l'intervalle $[\pi/2, 3\pi/2]$? Le mieux est de l'écrire comme une translatée de arc sinus, en utilisant le fait que sinus est une fonction périodique.

Exemple 17.36.
Pour calculer $\arcsin(1)$, il faut chercher un angle entre $-\frac{\pi}{2}$ et $\frac{\pi}{2}$ ayant 1 pour sinus : résoudre $\sin(\theta) = 1$. La solution est $\theta = \frac{\pi}{2}$ et nous avons donc $\arcsin(1) = \frac{\pi}{2}$. △

À l'aide des valeurs remarquables de la fonction sinus nous obtenons le tableau suivant de valeurs remarquables pour l'arc sinus.

x	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
$\arcsin(x)$	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$

Les autres valeurs remarquables peuvent être déduites du fait que l'arc sinus est une fonction impaire.

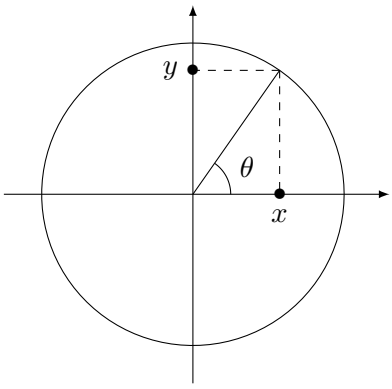
En ce qui concerne la dérivabilité de la fonction arc sinus, en application de la proposition 11.120 elle est dérivable en tout $y = \sin(x)$ tel que $\sin'(x) \neq 0$, c'est-à-dire tel que $\cos(x) \neq 0$. Or $\cos(x) = 0$ pour $x = \pm\frac{\pi}{2}$, ce qui correspond à $y = \sin(\pm\frac{\pi}{2}) = \pm 1$. La fonction arc sinus est donc dérivable sur $]-1,1[$. Nous avons donc la propriété suivante pour la dérivabilité.

Proposition 17.37.
La fonction arc sinus est continue sur $[-1,1]$ et dérivable sur $]-1,1[$. Pour tout $y \in]-1,1[$, la dérivée est donnée par la formule (11.278), qui dans ce cas s'écrit

$$\arcsin'(y) = \frac{1}{\cos\big(\arcsin(y)\big)} = \frac{1}{\sqrt{1-y^2}}. \tag{17.91}$$

La dernière égalité viens du fait que si $x = \arcsin(y)$ alors $y = \sin(x)$ et $\cos(x) = \sqrt{1-\sin^2(x)} = \sqrt{1-y^2}$.

Pour comprendre la dernière égalité, remarquer que dans le dessin suivant, $\theta = \arcsin(y)$, donc $y = \sin(\theta)$, et $x = \cos(\theta)$.



Notons enfin que le graphe de la fonction arc sinus est donné à la figure 17.3.

17.3.3 La fonction arc cosinus

Nous voulons étudier la fonction

$$\cos\colon \mathbb{R} \to [-1,1] \tag{17.92}$$

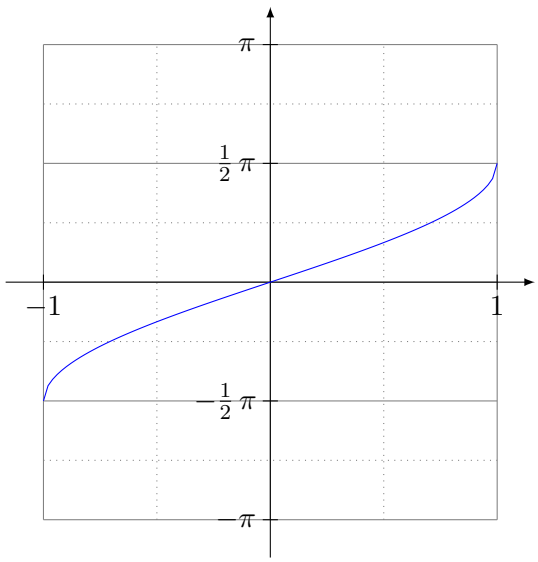


FIGURE 17.3 – Le graphe de la fonction $x \mapsto \arcsin(x)$

et son éventuelle réciproque. Encore une fois il n'est pas possible d'en prendre la réciproque globale parce que ce n'est pas une bijection ; ne fut-ce que parce qu'elle est périodique (proposition 17.14). Nous choisissons de considérer l'intervalle $[0, \pi]$ sur lequel la fonction cosinus est continue et strictement monotone décroissante.

Nous avons alors le résultat suivant :

Proposition-définition 17.38.
Pour définir la fonction arcsinus.

(1) La fonction
$$\cos\colon [0, \pi] \to [-1, 1] \tag{17.93}$$
est une bijection continue strictement décroissante.

(2) Sa bijection réciproque est la fonction
$$\arccos\colon [-1, 1] \to [0, \pi] \tag{17.94}$$
nommée ***arc cosinus****.*

(3) La fonction arc cosinus est continue, strictement décroissante.

(4) Elle est dérivable et pour tout $y \in]-1, 1[$, sa dérivée est donnée par
$$\arccos'(y) = \frac{1}{-\sin\big(\arccos(y)\big)} = \frac{-1}{\sqrt{1-y^2}}. \tag{17.95}$$

Démonstration. La fonction cosinus est continue et même de classe C^∞ par la proposition 17.1. Elle est strictement décroissant parce que sa dérivée $(-\sin)$ y est strictement positive (strictement à dans l'intérieur du domaine).

Le fait que arc cosinus soit une bijection continue strictement monotone est dans le théorème de la bijection 11.61. La dérivabilité et la formule sont de la proposition 11.120. □

Pour $y_0 \in [-1, 1]$, trouver la valeur de $\arccos(y_0)$ revient à résoudre l'équation $\cos(x_0) = y_0$. Cela nous permet de construire une tableau de valeurs :

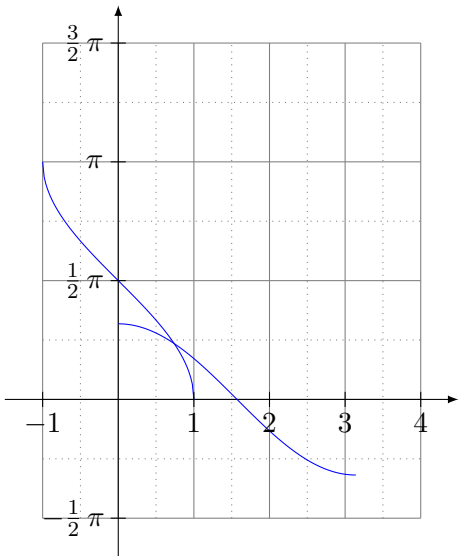
x	-1	$-\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{2}}{2}$	$-\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
$\arccos(x)$	π	$\frac{5\pi}{6}$	$\frac{3}{4}\pi$	$\frac{2}{3}\pi$	$\frac{1}{2}\pi$	$\frac{\pi}{3}$	$\frac{1}{4}\pi$	$\frac{1}{6}\pi$	0

Remarque 17.39.
Certes la fonction cosinus est paire (vue sur $\mathbb{R}$), mais la fonction arc cosinus ne l'est pas car elle est une bijection entre $[-1, 1]$ et $[0, \pi]$.

Exemple 17.40.
Cherchons $\arccos(\frac{1}{2})$. Il faut trouver un angle $\theta \in [0, \pi]$ tel que $\cos(\theta) = \frac{1}{2}$. La solution est $\theta = \frac{\pi}{3}$. Donc $\arccos(\frac{1}{2}) = \frac{\pi}{3}$.

Il n'est cependant pas immédiat d'en déduire la valeur de $\arccos(-\frac{1}{2})$. En effet $\theta = \arccos(-\frac{1}{2})$ si et seulement si $\cos(\theta) = -\frac{1}{2}$ avec $\theta \in [0, \pi]$. La solution est $\theta = \frac{2\pi}{3}$. △

En ce qui concerne la représentation graphique, il suffit de tracer la fonction cosinus entre 0 et π puis de prendre le symétrique par rapport à la droite $y = x$.



17.3.4 Une meilleure approximation de π

Nous avions laissé le nombre π avec l'approximation assez minable de $2\sqrt{2} < \pi < 4$ en le lemme 17.26. Nous pouvons maintenant faire nettement mieux.

Le lemme 17.32 donne

$$\arctan(1/\sqrt{3}) = \pi/6 \tag{17.96}$$

et l'idée est de donner un développement de arctan autour de zéro, de l'évaluer en $1/\sqrt{3}$ et d'égaliser le résultat à $\pi/6$. Tout cela donne lieu à des calcules peut-être fastidieux, mais comme un gars l'a fait dès l'an 1424[353] pour trouver 16 décimales correctes, nous faisons comme si c'était facile.

Pour trouver le développement en série de Taylor (théorème 11.375) de arc tangente autour de $x = 0$, il faut partir de la formule (17.80) et sans doute pas mal calculer et faire une récurrence [9]. Le résultat est :

$$\arctan(x) = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{2k+1}, \tag{17.97}$$

valable pour $x \in]-1, 1[$. Avec cela nous avons

$$\arctan(\frac{1}{\sqrt{3}}) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)3^k} \times \frac{1}{\sqrt{3}} = \frac{\pi}{6}, \tag{17.98}$$

et donc

$$\pi = \frac{6}{\sqrt{3}} \sum_{k=0}^{\infty} \frac{(-1)k}{(2k+1)3^k}. \tag{17.99}$$

Pour donner une idée du fait que ça fonctionne pas mal, voici le calcul pour quelques termes :

```
SageMath version 7.3, Release Date: 2016-08-04
```

9. Je n'ai pas fait le calcul, merci de me faire savoir si il y a une astuce.

```
 Type "notebook()" for the browser-based notebook interface.
 Type "help()" for help.

sage: n(taylor(arctan(x),x,0,5)(1/sqrt(3)))*6
3.15618147156995
sage: n(taylor(arctan(x),x,0,10)(1/sqrt(3)))*6
3.14260474566308
sage: n(taylor(arctan(x),x,0,20)(1/sqrt(3))*6-pi)
-2.14265171338823e-6
sage: n(taylor(arctan(x),x,0,58)(1/sqrt(3))*6-pi)
8.88178419700125e-16
```

tex/sage/sageSnip012.sage

Calculer 5 termes donne déjà 3.15. Et on est à 10^{-6} de la bonne réponse avec 20 termes. Et avec 58 termes, on n'est à 10^{-16}.

Problèmes et choses à faire

Pour bien faire, il faudrait étudier le reste et donner un encadrement.

17.3.5 Forme polaire ou trigonométrique des nombres complexes

Un nombre complexe étant représenté par deux nombres, on peut le représenter dans un plan appelé « plan de Gauss ». La plupart des opérations sur les nombres complexes ont leur interprétation géométrique dans ce plan.

Dans le plan de Gauss, le module d'un complexe z représente la distance entre 0 et z. On appelle *argument* de z (noté $\arg z$) l'angle (déterminé à 2π près) entre le demi-axe des réels positifs et la demi-droite qui part de 0 et passe par z. Le module et l'argument d'un complexe permettent de déterminer univoquement ce complexe puisqu'on a la formule

$$z = a + bi = |z|\left(\cos(\arg(z)) + i\sin(\arg(z))\right)$$

L'argument de z se détermine via les formules

$$\frac{a}{|z|} = \cos(\arg(z)) \quad \frac{b}{|z|} = \sin(\arg(z))$$

ou encore par la formule

$$\frac{b}{a} = \tan(\arg(z)) \quad \text{en vérifiant le quadrant.}$$

La vérification du quadrant vient de ce que la tangente ne détermine l'angle qu'à π près.

17.3.6 Angle entre deux vecteurs

Proposition-définition 17.41.
Soient des vecteurs $X, Y \in \mathbb{R}^2$. Il existe un unique $\theta \in [0, \pi]$ tel que

$$\cos(\theta) = \frac{X \cdot Y}{\|X\|\|Y\|}. \tag{17.100}$$

*Ce réel est appelé **angle** entre X et Y.*

Démonstration. Si a et b sont des réels, l'inégalité $|a| \leqslant b$ peut se développer en une double inégalité

$$-b \leqslant a \leqslant b. \tag{17.101}$$

L'inégalité de Cauchy-Schwarz (7.165) devient alors

$$-\|X\|\|Y\| \leqslant X \cdot Y \leqslant \|X\|\|Y\|. \tag{17.102}$$

Si $X \neq 0$ et $Y \neq 0$, nous en déduisons

$$-1 \leqslant \frac{X \cdot Y}{\|X\|\|Y\|} \leqslant 1. \tag{17.103}$$

Il existe donc par la proposition 17.38 un angle $\theta \in [0, \pi]$ tel que

$$\cos(\theta) = \frac{X \cdot Y}{\|X\|\|Y\|}. \tag{17.104}$$

□

17.42.
Certains n'hésitent pas à écrire la formule

$$X \cdot Y = \|X\|\|Y\| \cos(\theta). \tag{17.105}$$

comme une définition du produit scalaire. C'est ce qui arrive lorsqu'on défini les fonctions trigonométriques à partir de relations dans les triangles rectangles.

Notez que les angles entre deux vecteurs sont toujours plus petits ou égaux à 180°.

La longueur de la projection du point P sur la droite horizontale va naturellement être égale à $\cos(\theta)$. En effet, si nous notons X un vecteur horizontal de norme 1, cette projection est donné par $P \cdot X$. Mais en reprenant l'équation (17.105), nous voyons que

$$P \cdot X = \|P\|\|X\| \cos(\theta), \tag{17.106}$$

tandis qu'ici nous avons $\|P\| = \|X\| = 1$.

Nous appelons $\sin(\theta)$ la longueur de la projection sur l'axe vertical.

Quelques dessins nous convainquent que

$$\begin{aligned} \sin(\theta + 2\pi) &= \sin(\theta) & \cos(\theta + 2\pi) &= \sin(\theta), \\ \sin(\theta + \frac{\pi}{2}) &= \cos(\theta) & \cos(\theta + \frac{\pi}{2}) &= -\sin(\theta), \\ \sin(\pi - \theta) &= \sin(\theta) & \cos(\pi - \theta) &= -\cos(\theta). \end{aligned} \tag{17.107}$$

Le théorème de Pythagore nous montre aussi l'importante relation

$$\sin^2(\theta) + \cos^2(\theta) = 1. \tag{17.108}$$

Quelques valeurs remarquables pour les sinus et cosinus :

$$\begin{aligned} &\sin 0 = 0, \quad \sin\frac{\pi}{6} = \frac{1}{2}, \quad \sin\frac{\pi}{4} = \frac{\sqrt{2}}{2}, \quad \sin\frac{\pi}{3} = \frac{\sqrt{3}}{2}, \quad \sin\frac{\pi}{2} = 1, \quad \sin\pi = 0 \\ &\cos 0 = 1, \quad \cos\frac{\pi}{6} = \frac{\sqrt{3}}{2}, \quad \cos\frac{\pi}{4} = \frac{\sqrt{2}}{2}, \quad \cos\frac{\pi}{3} = \frac{1}{2}, \quad \cos\frac{\pi}{2} = 0, \quad \cos\pi = -1 \end{aligned} \tag{17.109}$$

Nous pouvons prouver simplement que $\sin(30°) = \frac{1}{2}$ et $\cos(30°) = \frac{\sqrt{3}}{2}$ en s'inspirant de la figure 17.4.

17.3.7 Aire du parallélogramme

Remarque 17.43.
Le nombre $\|a\|\|b\| \sin(\theta)$ est l'aire du parallélogramme formé par les vecteurs a et b, comme cela se voit sur la figure 17.5.

Proposition 17.44.
Nous avons

$$\|a \times b\| = \|a\|\|b\| \sin(\theta) \tag{17.110}$$

où $\theta \in [0.\pi]$ est l'angle formé par a et b.

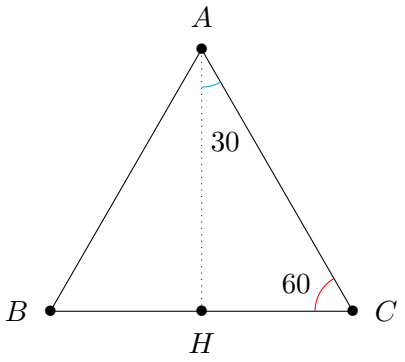


FIGURE 17.4 – Un triangle équilatéral de côté 1.

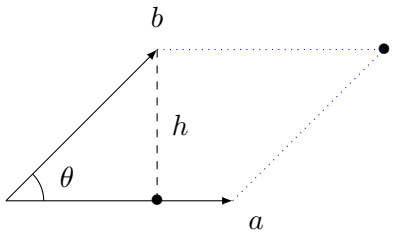


FIGURE 17.5 – Calculer l'aire d'un parallélogramme.

Démonstration. En utilisant la décomposition du produit vectoriel, nous avons

$$\begin{aligned}
\|a \times b\|^2 &= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}^2 + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}^2 \\
&= (a_2b_3 - b_2a_3)^2 + (a_1b_3 - a_3b_1)^2 + (a_1b_2 - a_2b_1)^2 \\
&= (a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2 \\
&= \|a\|^2\|b\|^2 - (a \cdot b)^2 \\
&= \|a\|^2\|b\|^2 - \|a\|^2\|b\|^2\cos^2(\theta) \\
&= \|a\|^2\|b\|^2\big(1 - \cos^2(\theta)\big) \\
&= \|a\|^2\|b\|^2\sin^2(\theta).
\end{aligned} \tag{17.111}$$

D'où le résultat. Nous avons utilisé la formule de la définition (17.41) donnant l'angle en fonction du produit scalaire. □

17.45.
Si les vecteurs a, b et c ne sont pas coplanaires, alors la valeur absolue du produit mixte (voir équation (7.230)) $a \cdot (b \times c)$ donne le volume du parallélépipède construit sur les vecteurs a, b et c.

En effet si φ est l'angle entre $b \times c$ et a, alors la hauteur du parallélépipède vaut $\|a\|\cos(\varphi)$ parce que la direction verticale est donnée par $b \times c$, et la hauteur est alors la « composante verticale » de a. Par conséquent, étant donné que $\|b \times c\|$ est l'aire de la base, le volume du parallélépipède vaut [10]

$$V = \|b \times c\|\|a\|\cos(\varphi). \tag{17.112}$$

Or cette formule est le produit scalaire de a par $b \times c$; ce dernier étant donné par le déterminant de la matrice formée des composantes de a, b et c grâce à la formule (7.230).

La valeur absolue du déterminant

$$\begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \tag{17.113}$$

10. Le calcul de ce volume mériterait une certaine réflexion, surtout à partir du moment où nous avons décidé de définir les fonctions trigonométriques à partir de son développement (définition 17.1).

est l'aire du parallélogramme déterminé par les vecteurs $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ et $\begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$. En effet, d'après la remarque 17.43, l'aire de ce parallélogramme est donnée par la norme du produit vectoriel

$$\begin{pmatrix} a_1 \\ a_2 \\ 0 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix} = \begin{vmatrix} e_x & e_y & e_z \\ a_1 & a_2 & 0 \\ b_1 & b_2 & 0 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} e_z, \tag{17.114}$$

donc la norme $\|a \times b\|$ est bien donnée par la valeur absolue du déterminant (17.113).

17.4 Paramétrisation du cercle

Nous allons parler de paramértisation du cercle. L'ensemble S^1 sera vu tantôt comme le cercle dans $\mathbb{R}^2$, tantôt comme le cercle dans $\mathbb{C}$. Nous n'allons pas pousser le vice jusqu'à écrire explicitement les isomorphismes lorsque nous passons d'une représentation à l'autre. Parmi les identifications que nous allons faire sans ménagement, il y a l'identification entre les applications

$$\begin{aligned} \gamma\colon [0,2\pi] &\to \mathbb{R}^2 \\ t &\mapsto \big(\cos(t), \sin(t)\big) \end{aligned} \tag{17.115}$$

et

$$\begin{aligned} \varphi\colon [0,2\pi[&\to \mathbb{C} \\ t &\mapsto e^{it}. \end{aligned} \tag{17.116}$$

C'est évidemment la formule $e^{ti} = \cos(t) + i\sin(t)$ (lemme 17.10) qui permet de transformer γ en φ et inversement. De plus $\mathbb{R}^2$ et $\mathbb{C}$ sont isomorphes en tant qu'espaces vectoriels normés (et aussi donc topologiques).

Nous allons voir deux choses à propos de cette application :

— Elle est continue, mais son inverse n'est pas continue. En considérant seulement la restriction $\varphi\colon]0,2\pi[\to S^2\backslash\{(1,0)\}$ nous avons un difféomorphisme, et donc une possibilité de changement de variables dans l'intégrale (théorème 13.255).

  Le fait qu'il manque un point est sans importante parce que nous n'allons considérer que la mesure de Lebesgue ou des variations simples autour de la mesure de Lebesgue.

— La fonction $\varphi\colon [0,2\pi[\to S^1$ est une bijection borélienne d'inverse borélien [11]. Donc nous pouvons transposer toute la théorie de la mesure de S^1 à $[0,2\pi[$ sans « triche ».

Tout cela pour dire que nous allons donner un tas de justifications pour écrire des égalités du type

$$\int_{S^1} f = \int_0^{2\pi} f \circ \varphi. \tag{17.117}$$

17.4.1 Bijection continue

Proposition 17.46.
L'application

$$\begin{aligned} \gamma\colon [0,2\pi[&\to S^1 \subset \mathbb{R}^2 \\ t &\mapsto \big(\cos(t), \sin(t)\big) \end{aligned} \tag{17.118}$$

est une bijection continue.

Démonstration. La continuité découle de la continuité des composantes. Le fait que l'image de γ soit dans S^1 découle immédiatement du fait que $\sin^2 + \cos^2 = 1$.

Pour la bijection, il faut injectif et surjectif.

11. Proposition 17.51.

Injectif Soient $x_1 < x_2$ tels que $\sin(x_1) = \sin(x_2)$ et $\cos(x_1) = \cos(x_2)$. Supposons pour fixer les idées que $\sin(x_1) > 0$ et $\cos(x_1) > 0$: si ce n'est pas le cas, il faut traiter séparément les 4 possibilités de combinaisons de signes.

Nous avons obligatoirement $x_1, x_2 \in [0, \frac{\pi}{2}[$. Vu que $\sin(x_1) = \sin(x_2)$, il existe par le théorème de Rolle 11.127 un élément $c \in]x_1, x_2[$ tel que $\sin'(c) = 0$, c'est-à-dire $\cos(c) = 0$. Cela contredirait la proposition 17.17(10) à moins que $x_1 = x_2$.

Surjectif Soient x, y tels que $x^2+y^2 = 1$. Supposons pour varier les plaisirs que $x < 0$ et $y > 0$. Vu que la fonction cos va de 0 à -1 lorsque x va de $\pi/2$ à π, le théorème des valeurs intermédiaires donne $t \in [\pi/2, \pi]$ tel que $\cos(t) = x$. Pour cette valeur de x nous avons

$$\cos^2(x) + \sin^2(x) = 1, \tag{17.119}$$

et donc $\sin^2(x) = y^2$, ce qui donne $\sin(x) = \pm y$. Mais pour $x \in [\pi/2, \pi]$ nous avons $\sin(t) > 0$. Par conséquent $\sin(t) = y$.

□

Exemple 17.47.
L'application

$$\begin{aligned} \varphi\colon]0, 2\pi[&\to S^1 \\ x &\mapsto \begin{pmatrix} \cos(x) \\ \sin(x) \end{pmatrix} \end{aligned} \tag{17.120}$$

est un continue par la proposition 17.46. Vu que $]0, 2\pi[$ est connexe (proposition 7.153) la proposition 7.90 implique que le cercle privé d'un point est connexe. △

Allez... Dans l'intro nous avions dit que nous n'allions pas faire explicitement les isomorphismes. Faisons-le quand même une fois, mais c'est bien parce que c'est vous hein.

Proposition 17.48.
L'application

$$\begin{aligned} f\colon [0, 2\pi[&\to S^1 \\ x &\mapsto e^{ix} \end{aligned} \tag{17.121}$$

est une bijection. Ici, S^1 est l'ensemble des nombres complexes de norme 1.

Démonstration. Nous savons que

$$\begin{aligned} \varphi\colon \mathbb{R}^2 &\to \mathbb{C} \\ (x, y) &\mapsto x + iy \end{aligned} \tag{17.122}$$

est une bijection isométrique. C'est pour cela que nous allons nous permettre de noter S^1 le cercle unité dans $\mathbb{R}^2$ aussi bien que l'ensemble des nombres complexes de norme 1.

Sur $\mathbb{R}^2$ nous avons l'application

$$\begin{aligned} \gamma\colon [0, 2\pi[&\to S^1 \subset \mathbb{R}^2 \\ t &\mapsto \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \end{aligned} \tag{17.123}$$

qui est une bijection continue (c'est la proposition 17.46). Et enfin le lemme 17.10 nous donne $e^{ix} = \cos(x) + i\sin(x)$.

Avec tout ça, l'application $\varphi^{-1} \circ f\colon [0, 2\pi[\to S^1$ est une bijection continue. Et comme φ l'est également, f est une bijection continue. □

La proposition suivante donne les coordonnées polaires sur $\mathbb{C}$. La régularité est l'objet du théorème 17.195 (à part le fait que ce dernier parle de $\mathbb{R}^2$ et non de $\mathbb{C}$).

Proposition 17.49.
Pour tout nombre complexe z, il existe un unique $\theta \in [0, 2\pi[$ tel que

$$z = |z|e^{i\theta}. \tag{17.124}$$

Démonstration. Soit $z \in \mathbb{C}$. Nous considérons $z' = z/|z|$ qui est de norme 1. Donc il existe un unique $\theta \in [0, 2\pi[$ tel que $z' = e^{i\theta}$ (proposition 17.48).

Pour ce θ nous avons $z = |z|e^{i\theta}$. □

Bien entendu, le θ est unique dans $[0, 2\pi[$, mais il n'est pas du tout unique dans $\mathbb{R}$.

17.4.2 Inverse

Nous pouvons écrire un inverse de la fonction φ grâce à la fonction arc tangente introduite au théorème 17.30. La fonction que nous écrivons à présent est la fonction $\arg_{0^-}$ définie par (25.273). Elle n'est pas exactement la fonction argument définie par (25.235).

Nous avons :

$$\begin{aligned} \varphi^{-1}\colon S^1 &\to [0, 2\pi[\\ x + iy &\mapsto \begin{cases} \operatorname{arctg}(y/x) & \text{si } x > 0, y \geqslant 0 \\ \frac{\pi}{2} & \text{si } (x, y) = (0, 1) \\ \pi - \operatorname{arctg}(-y/x) & \text{si } x < 0, y \geqslant 0 \\ \pi + \operatorname{arctg}(y/x) & \text{si } x < 0, y < 0 \\ 2\pi - \operatorname{arctg}(-y/x) & \text{si } x > 0, y < 0 \end{cases} \end{aligned} \tag{17.125}$$

Chacune des branche est continue parce que la fonction arc tangente l'est. Trois des raccords sont également continus grâce aux limites du lemme 17.31.

L'application φ^{-1} n'est cependant pas continue au point $(1, 0)$ [12]. C'est l'objet du lemme suivant.

Lemme 17.50.
L'application $\varphi^{-1}\colon S^1 \to [0, 2\pi[$ n'est pas continue en $(1, 0)$. Mais elle est continue ailleurs. Autrement dit,

$$\varphi^{-1}\colon S^1 \backslash \{(1, 0)\} \to]0, 2\pi[\tag{17.126}$$

est continue.

Démonstration. En effet, φ^{-1} serait continue si l'image de tout ouvert de $[0, 2\pi[$ par φ serait ouverte dans S^1 (topologie induite de $\mathbb{C}$). Prenons un petit ouvert $[0, \epsilon[$ (si vous êtes étonnés, c'est que vous n'avez pas bien la topologie induites en tête). Son image contient le point $(1, 0)$, mais aucun point (x, y) avec $y < 0$.

Montrons que tout voisinage de $(1, 0)$ dans $\mathbb{C}$ contient des points $x + iy$ de S^1 avec $y < 0$. Un point de S^1 est de la forme $\cos(t) + i\sin(t)$. Nous avons :

$$|\cos(t) + i\sin(t) - 1|^2 = \big(\cos(t) - 1\big)^2 + \sin^2(t) = 2\big(1 - \cos(t)\big). \tag{17.127}$$

Soit $\delta > 0$, et montrons que $B\big((1, 0), \delta\big) \cap S^1$ contient des points d'ordonnées négatives. D'abord il existe $\epsilon > 0$ tel que pour $t = 2\pi - \epsilon$,

$$2\big(1 - \cos(t)\big) < \delta. \tag{17.128}$$

Ensuite pour de tels t, nous avons $\sin(t) < 0$. Donc les points de S^1 correspondant à $2\pi - \epsilon$ sont dans $S^1 \cap B\big((1, 0), \delta\big)$.

Bref, l'image de $[0, \epsilon[$ n'est pas un ouvert de S^1. □

12. Vu que nous avons considéré $S^1 \subset \mathbb{C}$, nous aurions dû noter « 1 » ce point. Mais vous vous imaginez le clash de notation avec le $1 \in [0, 2\pi[\subset \mathbb{R}$?

17.4.3 Cercle trigonométrique

Le cercle trigonométrique est le cercle de rayon 1 représenté à la figure 17.6. Sa longueur est 2π.

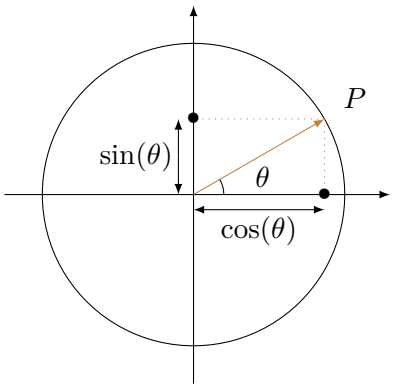


FIGURE 17.6 – Le cercle trigonométrique.

Nous verrons plus tard que la longueur de l'arc de cercle intercepté par un angle θ est égal à θ. Les radians sont donc l'unité d'angle les plus adaptés au calcul de longueurs sur le cercle.

17.4.4 Du point de vue de la tribu, mesure et co.

Nous avons considéré sur S^1 la topologie induite de $\mathbb{C}$. Nous allons y mettre la tribu induite de celle de Lebesgue de $\mathbb{C}$. Mais nous n'allons pas y mettre la *mesure* induite de $\mathbb{C}$; sinon tout serait toujours de mesure nulle.

Proposition 17.51 ([1]).
L'application φ est borélienne d'inverse borélien, c'est-à-dire

$$\mathcal{B}or(S^1) = \varphi\big(\mathcal{B}or([0,2\pi[)\big). \tag{17.129}$$

Démonstration. L'inclusion $\mathcal{B}\text{or}(S^1) \subset \varphi\big(\mathcal{B}\text{or}([0,2\pi[)\big)$ est la plus simple : si $A \in \mathcal{B}\text{or}(S^1)$, alors $\varphi^{-1}(A) \in \mathcal{B}\text{or}\big([0,2\pi[\big)$ parce que $\varphi\colon [0,2\pi[\to S^1$ est continue et donc borélienne (théorème 13.49).

Pour l'autre inclusion, il faudra faire par étapes.

Ouvert ne contenant pas zéro Si A est un ouvert de $[0,2\pi[$ ne contenant pas 0, il est un ouvert de $\mathbb{R}$ ou de $]0,2\pi[$. Le lemme 17.50 nous indique que son image par φ est ouverte dans S^1. En particulier, $\varphi(A) \in \mathcal{B}\text{or}(S^1)$.

Ouvert de la forme $[0,\epsilon[$ Nous supposons que ϵ est petit. Disons pour fixer les idées, plus petit que $\pi/2$. Nous avons :

$$\varphi\big([0,\epsilon[\big) = \varphi\big(]0,\epsilon[\big) \cup \varphi\big(\{0\}\big). \tag{17.130}$$

Le premier élément de l'union est un ouvert, et le second un unique point. L'union est un borélien.

Ouvert général Si un ouvert de $[0,2\pi[$ ne contient pas 0, son image est ouverte. Nous nous penchons sur le cas d'un ouvert contenant 0.

Si un ouvert de $[0,2\pi[$ contient 0, alors il contient un ouvert de la forme $[0,\epsilon[$, parce qu'un ouvert contient une boule autour de chacun de ses points (théorème 7.4 couplé au fait que nous sommes dans la topologie induite de $\mathbb{R}$).

Si A est un ouvert contenant zéro, alors

$$A = [0,\epsilon[\cup \big(A\backslash[0,\frac{\epsilon}{2}]\big). \tag{17.131}$$

Nous avons déjà vu que l'image du premier élément de l'union est un borélien. Étant donné que $A\backslash[0,\frac{\epsilon}{2}]$ est un ouvert ne contenant pas zéro, son image est un ouvert. Donc le l'image de A est un borélien.

Pause Nous avons déjà vu que l'image par φ de tout ouvert de $[0,2\pi[$ était un borélien de S^1. Nous devons en déduire que l'image de tout borélien de $[0,2\pi[$ est un borélien de S^1.

C'est ce que nous faisons maintenant

Boréliens Nous utilisons le lemme de transport 13.47 avec l'application φ^{-1} et l'ensemble des ouverts :

$$\varphi\big(\sigma(\mathcal{C})\big) = \sigma\big(\varphi(\mathcal{C})\big) \tag{17.132}$$

où $\mathcal{C}$ est la tribu des ouverts dans $[0,2\pi[$. L'ensemble $\sigma(\mathcal{C})$ est par définition l'ensemble $\mathcal{B}\mathrm{or}\big([0,2\pi[\big)$. D'autre part nous avons vu que l'image d'un ouvert est un borélien : $\varphi(\mathcal{C}) \subset \mathcal{B}\mathrm{or}(S^1)$. Nous avons donc

$$\varphi\big(\mathcal{B}\mathrm{or}([0,2\pi[)\big) = \sigma\big(\varphi(\mathcal{C})\big) \subset \sigma\big(\mathcal{B}\mathrm{or}(S^1)\big) \subset \mathcal{B}\mathrm{or}(S^1). \tag{17.133}$$

La preuve est terminée. □

Proposition 17.52 (Boréliens sur S^1[1])**.**
Soit la structure usuelle d'espace mesurable $(\mathbb{C}, \mathcal{B}or(\mathbb{C}))$. Nous considérons

- *la tribu $\mathcal{B}or(\mathbb{C})_{S^1}$ induite de la tribu des boréliens de $\mathbb{C}$ vers S^1,*
- *la tribu $\mathcal{B}or(S^1)$ des boréliens de S^1 construite à partir de la topologie induite de $\mathbb{C}$ vers S^1.*
- *la bijection $\varphi\colon [0,2\pi[\to S^1$,*
- *la mesure de Lebesgue sur $[0,2\pi[$ (induite de celle sur $\mathbb{R}$) et sur $\mathbb{C}$, que nous noterons toutes deux λ.*

Alors

(1) Nous avons les expressions

$$\mathcal{B}or(\mathbb{C})_{S^1} = \{A \in \mathcal{B}or(\mathbb{C}) \text{ tel que } A \subset S^1\} \tag{17.134a}$$
$$= \{A \cap S^1 \text{ tel que } A \in \mathcal{B}or(\mathbb{C})\} \tag{17.134b}$$

(2) Nous avons

$$\mathcal{B}or(S^1) = \mathcal{B}or(\mathbb{C})_{S^1} = \varphi\Big(\mathcal{B}or\big([0,2\pi[\big)\Big). \tag{17.135}$$

(3) En définissant $\mu\colon \mathcal{B}or(S^1) \to \mathbb{R}$ par

$$\mu(A) = \frac{\lambda\big(\varphi^{-1}(A)\big)}{2\pi}, \tag{17.136}$$

le triple $\big(S^1, \mathcal{B}or(S^1), \mu\big)$ est un espace mesuré.

(4) L'espace mesuré $\big(S^1, \mathcal{B}or(S^1), \mu\big)$ est fini et

$$\mu(S^1) = 1. \tag{17.137}$$

Démonstration. Point par point.

Pour (1) C'est la proposition 13.7.

Pour (2) La première égalité est le lemme 13.76. Le fait que $\mathcal{B}\mathrm{or}(S^1) = \varphi\Big(\mathcal{B}\mathrm{or}\big([0,2\pi[\big)\Big)$ est déjà la proposition 17.51.

Pour (3) Nous devons d'abord nous assurer que la formule ait un sens. Cela est chose aisée ; si $A \in \mathcal{B}\mathrm{or}(S^1)$, le point (2) nous indique que $\varphi^{-1}(A) \in \mathcal{B}\mathrm{or}\big([0,2\pi[\big)$. Ensuite, nous devons vérifier les deux conditions de la définition 13.20 pour avoir un espace mesuré.

En premier lieu,

$$\mu(\varnothing) = \frac{1}{2\pi}\lambda\big(\varphi^{-1}(\varnothing)\big) = \frac{1}{2\pi}(\varnothing) = 0. \tag{17.138}$$

En en second lieu, si les $A_i \in \mathcal{B}\mathrm{or}(S^1)$ sont disjoints, les $\varphi^{-1}(A_i)$ sont également disjoints parce que φ^{-1} est une bijection. Donc

$$\mu(\bigcup_i A_i) = \frac{1}{2\pi}\lambda(\bigcup_i \varphi^{-1}(A_i)) \tag{17.139a}$$
$$= \frac{1}{2\pi}\sum_i \lambda(\varphi^{-1}(A_i)) \tag{17.139b}$$
$$= \sum_i \frac{\lambda(\varphi^{-1}(A_i))}{2\pi} \tag{17.139c}$$
$$= \sum_i \mu(A_i). \tag{17.139d}$$

D'accord.

Pour (4) En ce qui concerne la mesure de S^1 pour μ nous avons simplement

$$\mu(S^1) = \frac{\lambda([0,2\pi[)}{2\pi} = 1. \tag{17.140}$$

□

Maintenant que $(S^1, \mathcal{B}\mathrm{or}(S^1), \mu)$ est un espace mesuré, nous pouvons compléter la tribu $\mathcal{B}\mathrm{or}(S^1)$ pour la mesure μ.

Définition 17.53.
*La **tribu de Lebesgue** sur S^1 est la mesure complétée pour*

$$(S^1, \mathcal{B}or(S^1), \mu) \tag{17.141}$$

où μ est la mesure définie par la proposition 17.52. Nous notons $\mathcal{L}eb(S^1)$ la tribu et encore μ la mesure.

Proposition 17.54 (Lebesgue sur S^1[1])**.**
Soit la structure d'espace mesuré complet $(S^1, \mathcal{L}eb(S^1), \mu)$. Nous considérons

- *la tribu $\mathcal{L}eb(\mathbb{C})_{S^1}$ induite de la tribu des boréliens de $\mathbb{C}$ vers S^1,*
- *la bijection $\varphi\colon [0,2\pi[\to S^1$,*

Alors

(1) La tribu $\mathcal{L}eb(\mathbb{C})_{S^1}$ est la tribu de toutes les parties de S^1.

(2) La tribu $\mathcal{L}eb(S^1)$ est donnée par

$$\mathcal{L}eb(S^1) = \varphi(\mathcal{L}eb(\mathbb{R})_{[0,2\pi[}) = \varphi(\mathcal{L}eb([0,2\pi[)) \tag{17.142}$$

où $\mathcal{L}eb([0,2\pi[)$ est la tribu sur $[0,2\pi[$ obtenue par completion de la tribu des boréliens de la topologie induite.

(3) Nous avons l'inclusion stricte

$$\mathcal{L}eb(S^1) \subsetneq \mathcal{L}eb(\mathbb{C})_{S^1}. \tag{17.143}$$

Démonstration. Point par point.

Pour (1) Si $A \subset S^1$, alors A est une partie de S^1 qui est mesurable et de mesure nulle pour $\mathbb{C}$. Donc A est λ-négligeable et par conséquent mesurable.

Pour (2) Il s'agit de prouver que

$$\widehat{\mathcal{B}\mathrm{or}(S^1)} = \varphi(\widehat{\mathcal{B}\mathrm{or}([0,2\pi[)}). \tag{17.144}$$

Ce n'est rien d'autre que la proposition 13.71. La seconde partie de l'égalité est la proposition 13.66

Pour (3) Comme indiqué au point (1), la tribu $\mathcal{L}\text{eb}(\mathbb{C})_{S^1}$ est la tribu de toutes les parties de S^1 ; l'incusion est donc évidente. Le point pas tout à fait évident à prouver est l'existence de parties de S^1 à n'être pas dans $\mathcal{L}\text{eb}(S^1)$.

Soit V non mesurable dans $[0, 2\pi[$ (prenez quelque chose comme l'ensemble de Vitali de l'exemple 13.141). Vu que, par le point (2),

$$\mathcal{L}\text{eb}(S^1) = \varphi\big(\mathcal{L}\text{eb}(\mathbb{R})_{[0,2\pi[}\big), \tag{17.145}$$

la partie $\varphi^{-1}(V)$ ne peut pas être dans $\mathcal{L}\text{eb}(S^1)$.

□

Si vous en voulez plus à propos de S^1 et la façon dont on passe la structure depuis $[0, 2\pi[$, vous pouvez lire la proposition 26.75 qui donne la structure de

$$L^2\big(S^1, \mathcal{L}\text{eb}(S^1), \mu\big) \tag{17.146}$$

qui sera, sans surprises la même que celle de

$$L^2\big([0, 2\pi[, \mathcal{L}\text{eb}\big([0, 2\pi[\big), \lambda\big). \tag{17.147}$$

17.5 Exemples trigonométriques

Nous mettons ici quelques exemples concernant les fonctions trigonométriques, qui n'ont pas pu être mis dans les chapitres le plus adapté, parce que ces derniers sont plus haut dans la table des matière.

Exemple 17.55.
Prouvons que la fonction[13] $f(x) = x\sin(x)$ tend vers zéro lorsque x tend vers 0. D'abord, nous coinçons la fonction entre deux fonctions connues :

$$0 \leqslant |x\sin(x)| = |x||\sin(x)| \leqslant |x|. \tag{17.148}$$

Donc $|x\sin(x)|$ est coincé entre $g(x) = 0$ et $h(x) = |x|$. Ces deux fonctions tendent vers 0 lorsque $x \to 0$, et donc $f(x)$ tend vers zéro. △

17.5.1 Quelques équations trigonométriques

La proposition suivante se voit très facilement sur le cercle trigonométrique, mais il faut le démontrer.

Proposition 17.56 ([1]).
Si $\theta_0 \in [0, 2\pi[$ vérifie $\cos(\theta_0) = x_0$, alors l'ensemble de solutions de l'équation $\cos(\theta) = x_0$ (d'inconnue θ) est

$$\{\theta_0, 2\pi - \theta_0\}. \tag{17.149}$$

Cet ensemble est un singleton si et seulement si $x_0 = \pm 1$.

Démonstration. Commençons par prouver que θ_0 et $2\pi - \theta_0$ sont des solutions. Le nombre θ_0 est solution par hypothèse. En ce qui concerne $2\pi - \theta_0$, il est possible d'utiliser la formule d'addition d'angle (17.22c) :

$$\cos(2\pi - \theta_0) = \cos(2\pi)\cos(\theta_0) + \sin(2\pi)\sin(\theta_0). \tag{17.150}$$

La proposition 17.17(1) nous indique que $\cos(2\pi) = 1$ et $\sin(2\pi) = 0$. Donc l'égalité (17.150) se réduit à $\cos(2\pi - x_0) = \cos(2\pi)$.

13. La définition de la fonction sinus est 17.1.

Le lemme 17.3 dit que si $\cos(\theta) = x_0$, alors

$$\sin(\theta) = \pm\sqrt{1 - x_0^2}. \tag{17.151}$$

Nous avons donc soit

$$\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = \begin{pmatrix} x_0 \\ \sqrt{1 - x_0^2} \end{pmatrix}, \tag{17.152}$$

soit

$$\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = \begin{pmatrix} x_0 \\ -\sqrt{1 - x_0^2} \end{pmatrix}, \tag{17.153}$$

Vu que $\theta \mapsto \big(\cos(\theta), \sin(\theta)\big)$ est une bijection avec S^1 (proposition 17.46), chacune de ces deux possibilités possède une unique solution. L'ensemble des solutions de $\cos(\theta) = x_0$ possède donc au maximum deux éléments.

L'ensemble des solutions possède exactement une solution lorsque les points $\big(x_0, \sqrt{1 - x_0^2}\big)$ et $\big(x_0, -\sqrt{1 - x_0^2}\big)$ sont identiques. Cela est le cas si et seulement si $\sqrt{1 - x_0^2} = 0$, c'est-à-dire si et seulement si $x_0 = \pm 1$. □

17.5.2 Développements en série

Proposition 17.57 (Taylor pour cosinus)**.**
Le développement du cosinus est donné par

$$\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} \cdots \tag{17.154}$$

C'est-à-dire que pout tout $n \in \mathbb{N}$, il existe une fonction $\alpha\colon \mathbb{R} \to \mathbb{R}$ telle que $\lim_{t\to 0} \alpha(t) = 0$ et

$$\cos(x) = \sum_{k=0}^{n} \frac{(-1)^{2k}}{(2k)!} x^{2k} + \alpha(x) x^{2n+1}. \tag{17.155}$$

En ce qui concerne le sinus, pour tout n nous avons une fonction $\alpha\colon \mathbb{R} \to \mathbb{R}$ telle que $\lim_{t\to 0} \alpha(t) = 0$ et

$$\sin(x) = \sum_{k=0}^{n} \frac{(-1)^k x^{2k+1}}{(2k+1)!} + x^{2n+2} \alpha(x). \tag{17.156}$$

Démonstration. Il s'agit d'utiliser la proposition 14.32, en faisant attention à l'ordre. Le fait est que dans (17.155), nous avons écrit le polynôme de degré $2n+1$ (et non seulement $2n$), en sachant que le terme d'ordre $2n+1$ est nul.

C'est pour cela que nous avons pu écrire $\alpha(x)x^{2n+1}$ au lieu de $\alpha(x)x^{2n}$ qui aurait été attendu.

Même raisonnement pour le développement du sinus. □

Remarque 17.58.
Quelques remarques concernant l'ordre du polynôme.

(1) Notons que nous aurions aussi pu écrire le reste sous la forme $\alpha(x)x^{2n}$, mais ça aurait été avec une autre fonction α : celle correspondant au développement à l'ordre $2n$ au lieu de $2n+1$.
(2) Les développements de sinus et de cosinus ont un terme sur deux qui est nul. C'est pour cela qu'en ayant une polynôme de degré $2p$, nous avons le développement d'ordre $2p+1$.
(3) Nous aurions pu utiliser les dérivées données dans la proposition 17.2 et les valeurs spéciales (17.6).

Corolaire 17.59.
Il existe une fonction $\alpha\colon \mathbb{R} \to \mathbb{R}$ telle que $\lim_{t\to 0} \alpha(t)/t = 0$ et

$$\sin(x) = x + \alpha(x). \tag{17.157}$$

Nous avons la limite

$$\lim_{x\to 0} \frac{\sin(x)}{x} = 1. \tag{17.158}$$

Démonstration. Il s'agit de prendre la formule (17.156) avec $n = 0$. Cela donne tout de suite (17.157). Pour la limite, on divise par x, ce qui donne (pour tout $x \neq 0$)

$$\frac{\sin(x)}{x} = 1 + \frac{\alpha(x)}{x}. \tag{17.159}$$

Et justement la fonction α la la propriété que $\lim_{x\to 0} \alpha(x)/x = 0$. □

Exemple 17.60.
Cherchons le développement limité à l'ordre 5 de $\tan(x) = \frac{\sin(x)}{\cos(x)}$. Nous utilisons les développements de la proposition 17.57 :

$$\sin(x) = x - \frac{x^3}{6} + \frac{x^5}{120} + x^5\alpha_1(x) \tag{17.160a}$$

$$\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{24} + x^5\alpha_2(x). \tag{17.160b}$$

Nous calculons alors la division des deux polynômes, en classant les puissances dans l'ordre croissant (c'est le sens inverse de ce qui est fait pour la divisions euclidienne !) :

$$\begin{array}{l|l}
x - \frac{1}{6}x^3 + \frac{1}{120}x^5 & 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 \\ \cline{2-2}
-\left(x - \frac{1}{2}x^3 + \frac{1}{24}x^5 \right) & x + \frac{1}{3}x^3 + \frac{2}{15}x^5 \\ \cline{1-1}
\quad \frac{1}{3}x^3 - \frac{1}{30}x^5 & \\
\quad -\left(\frac{1}{3}x^3 - \frac{1}{6}x^5 + \frac{1}{72}x^7 \right) & \\ \cline{1-1}
\qquad \frac{2}{15}x^5 - \frac{1}{72}x^7 & \\
\qquad -\left(\frac{2}{15}x^5 - \frac{1}{15}x^7 + \frac{1}{180}x^9 \right) & \\ \cline{1-1}
\qquad\qquad \frac{29}{360}x^7 - \frac{1}{180}x^9 &
\end{array}$$

Nous avons continué la division jusqu'à obtenir un reste de degré plus grand que 5. Le développement à l'ordre 5 de la fonction tangente autour de zéro est alors (proposition 11.399)

$$\tan(x) = x + \frac{1}{3}x^3 + \frac{2}{15}x^5 + x^5\alpha(x). \tag{17.161}$$

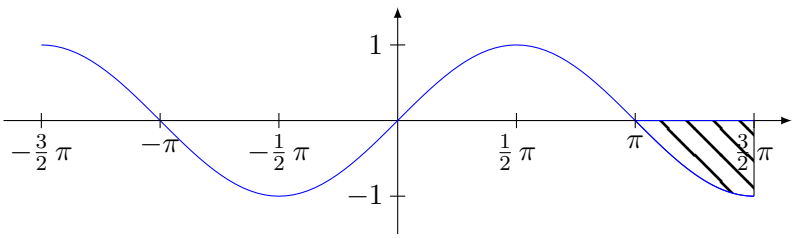
Notons que, vu que le reste ne nous intéresse pas vraiment, nous aurions pu ne pas calculer les coefficients des termes en x^7 et x^8. La dernière soustraction était également inutile. △

17.5.3 Intégration

Exemple 17.61.
Comme nous le voyons sur le dessin suivant,

$$\int_{-3\pi/2}^{3\pi/2} \sin(x)\,dx = 0 \tag{17.162}$$

parce que les deux parties bleues s'annulent avec les deux parties rouges (qui sont comptées comme des aires négatives).



△

17.5.4 Changement de variables dans une intégrale

Exemple 17.62.
Soit V la région trapézoïdale de sommets $(0,-1)$, $(1,0)$, $(2,0)$, $(0,-2)$, comme à la figure 17.7(a). Calculons ensemble l'intégrale double

$$\int_V e^{\frac{x+y}{x-y}}\,dV,$$

avec le changement de variable $\psi(x,y) = (x+y, x-y)$. C'est-à-dire que nous considérons les nouvelles variables

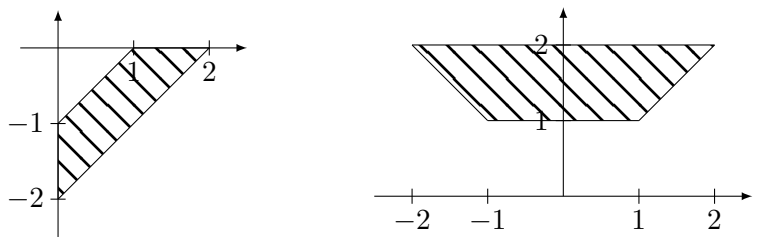
$$\begin{cases} u = x+y & (17.163a)\\ v = x-y. & (17.163b)\end{cases}$$

Il faut remarquer d'abord que le changement de variable proposé est dans le mauvais sens. On écrit alors $\phi(u,v) = \psi^{-1}(u,v) = \big((u+v)/2, (u-v)/2\big)$, c'est-à-dire

$$\begin{cases} x = \dfrac{u+v}{2} & (17.164a)\\ y = \dfrac{u-v}{2}. & (17.164b)\end{cases}$$

La région qui correspond à V est U, le trapèze de sommets $(-1,1)$, $(1,1)$, $(2,2)$ et $(-2,2)$, qu'on voit sur la figure 17.7(b) et qu'on décrit par

$$U = \{(u,v) \in \mathbb{R}^2 \,|\, 1 \leqslant v \leqslant 2,\ -v \leqslant u \leqslant v\}.$$



(a) La région V (b) La région $U = \phi^{-1}(V)$

FIGURE 17.7 – Avant et après le changement de variables

On observe que U est une région du premier type tandis que V n'est pas du premier ou du deuxième type. Le déterminant de la matrice jacobienne de ψ^{-1} est $J_{\psi^{-1}}$,

$$J_{\psi^{-1}}(u,v) = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}. \tag{17.165}$$

On a alors, en utilisant le fait que $F(x) = ae^{x/a}$ est une primitive de $f(x) = e^{x/a}$ (proposition 14.54) ainsi que le théorème fondamental de l'analyse (théorème 13.236),

$$\int_V e^{\frac{x+y}{x-y}}\,dV = \int_U e^{\frac{u}{v}}\,\frac{1}{2}\,dV = \int_1^2\int_{-v}^{v} e^{\frac{u}{v}}\,\frac{1}{2}\,du\,dv = \frac{3}{4}(e-e^{-1}).$$

△

17.6 Isométries de l'espace euclidien

Nous considérons l'espace affine euclidien $A = \mathcal{E}_n(\mathbb{R})$ modelé sur $\mathbb{R}^n$ avec sa métrique usuelle. Un premier grand résultat sera le théorème 9.249 qui dira que les isométries de cet espace sont des applications linéaires.

17.6.1 Structure du groupe $\mathrm{Isom}(\mathbb{R}^n)$

Si vous ne voulez pas savoir ce qu'est un produit semi-direct de groupes, vous pouvez lire seulement le point (1) du théorème suivant, et passer directement à la remarque 17.64.

Théorème 17.63.
Un peu de structure sur $\mathrm{Isom}(\mathbb{R}^n)$.

(1) L'application

$$\begin{aligned}\psi\colon T(n)\times \mathrm{O}(n)&\to \mathrm{Isom}(\mathbb{R}^n)\\ (v,\Lambda)&\mapsto \tau_v\circ\Lambda\end{aligned}\tag{17.166}$$

est une bijection. Ici, $T(n)$ *est le groupe des translations de* $\mathbb{R}^n$.

(2) Un couple $(v,\Lambda)\in T(n)\times \mathrm{SO}(n)$ *agit sur* $x\in\mathbb{R}^n$ *par*

$$(v,\Lambda)x=\Lambda x+v\tag{17.167}$$

au sens où $\psi(v,\Lambda)x=\Lambda x+v$.

(3) En tant que groupes,

$$\mathrm{Isom}(\mathbb{R}^n)\simeq T(n)\times_\rho \mathrm{O}(n)\tag{17.168}$$

où ρ *représente l'action adjointe de* $\mathrm{O}(n)$ *sur* $T(n)$ *et* $\times_\rho$ *dénotes le produit semi-direct de la définition 2.79.*

Démonstration. Point par point.

(1) Prouvons que l'application proposée est injective et surjective. Notons aussi que ce point ne parle pas de structure de groupe, mais seulement d'une bijection en tant qu'ensembles.

Injection Si $\psi(v,\Lambda)=\psi(w,\Lambda')$ alors en appliquant sur $x=0$ nous avons tout de suite $v=w$. Et ensuite $\Lambda=\Lambda'$ est immédiat.

Surjection Une isométrie $g\in\mathrm{Isom}(\mathbb{R}^n)$ est une application $g\colon\mathbb{R}^n\to\mathbb{R}^n$ telle que $d(x,y)=d\big(g(x),g(y)\big)$. Dans le cas de $\mathbb{R}^n$ cela se traduit par

$$\|x-y\|=\big\|g(x)-g(y)\big\|,\tag{17.169}$$

Vu que $x\mapsto\|x\|$ est une forme quadratique, elle tombe sous le coup du théorème 9.249, ce qui nous permet de dire que g est affine. Or par définition une application est affine lorsqu'elle est la composée d'une translation et d'une application linéaire.

Donc $g=\tau_v\circ\Lambda$ pour une certaine application linéaire isométrique $\Lambda\colon\mathbb{R}^n\to\mathbb{R}^n$. L'application Λ est donc dans $\mathrm{O}(n)$ par la proposition 9.42(3).

(2) C'est seulement le fait que $(\tau_v\circ\Lambda)x=\tau_v\big(\Lambda x\big)=\Lambda(x)+v$.

(3) Nous allons étudier l'application

$$\psi\colon T(n)\times_\rho O(n)\to\mathrm{Isom}(\mathbb{R}^n).\tag{17.170}$$

Le produit semi-direct est bien définit Il faut montrer que

$$\begin{aligned}\rho\colon O(n)&\to \mathrm{Aut}\big(T(n)\big)\\ \Lambda&\mapsto \mathbf{Ad}(\Lambda)\end{aligned}\tag{17.171}$$

est correcte.

D'abord pour $\Lambda\in O(n)$, nous avons bien $\rho_\Lambda(\tau_v)\in T(n)$ parce qu'en appliquant à $x\in\mathbb{R}^n$,

$$(\Lambda\tau_v\Lambda^{-1})(x)=\Lambda\big(\tau_v(\Lambda^{-1}x)\big)=\Lambda\big(\Lambda^{-1}x+v\big)=x+\Lambda(v)=\tau_{\Lambda(v)}(x).\tag{17.172}$$

Donc $\rho_\Lambda(\tau_v)=\tau_{\Lambda(v)}$.

De plus, $\rho_\Lambda\in\mathrm{Aut}\big(T(n)\big)$ parce que

$$\rho_\Lambda\big(\tau_v\circ\tau_w\big)=\rho_\Lambda(\tau_v)\circ\rho_\Lambda(\tau_v),\tag{17.173}$$

comme on peut aisément vérifier que les deux membres sont égaux à $\tau_{\Lambda(v+w)}$.

ψ est une bijection Cela est déjà vérifié.

ψ est un homomorphisme Nous avons d'une part

$$\psi\big((v,g)(w,h)\big) = \psi\big(v\rho_g(w), gh\big) = \tau_v \circ g \circ \tau_w \circ g^{-1} \circ g \circ h = \tau_v \circ g \circ \tau_w \circ h. \tag{17.174}$$

Et d'autre part,

$$\psi(v,g) \circ \psi(w,h) = \tau_v \circ g \circ \tau_w \circ h, \tag{17.175}$$

ce qui est la même chose.

□

Remarque 17.64.
Notons au passage la loi de groupe sur les couples qui est donnée, pour tout $v, v' \in \mathbb{R}^n$, $\Lambda, \Lambda' \in \mathrm{SO}(n)$, par

$$(v,\Lambda) \cdot (v',\Lambda') = (\Lambda v' + v, \Lambda\Lambda') \tag{17.176}$$

comme le montre le calcul suivant :

$$\begin{aligned}(v,\Lambda) \cdot (v',\Lambda')x &= (v,\Lambda)(\Lambda' x + v') && (17.177a)\\ &= \Lambda\Lambda' x + \Lambda v' + v && (17.177b)\\ &= (\Lambda v' + v, \Lambda\Lambda')x. && (17.177c)\end{aligned}$$

Proposition 17.65 ([354])**.**
Soient $n \geqslant 1$ et R un élément de $\mathrm{O}(n)$ de déterminant -1 tels que $R^2 = \mathrm{Id}$. En posant $C_2 = \{\mathrm{Id}, R\}$ nous avons

$$\mathrm{O}(n) = \mathrm{SO}(n) \times_\rho C_2 \tag{17.178}$$

Démonstration. Notons qu'un élément R comme décrit dans l'énoncé existe. Par exemple il y a l'application $(x_1, \dots, x_n) \mapsto (-x_1, x_2, \dots, x_n)$.

Cela étant dit, nous allons montrer que

$$\begin{aligned}\psi\colon \mathrm{SO}(n) \times C_2 &\to \mathrm{O}(n)\\ (A,h) &\mapsto Ah.\end{aligned} \tag{17.179}$$

est un isomorphisme.

Injectif Soient $A, B \in \mathrm{SO}(n)$ et $h, k \in C_2$ tels que $\psi(A,h) = \psi(B,k)$, c'est-à-dire tels que $Ah = Bk$. Vu que $\det(A) = \det(B) = 1$ nous avons $\det(h) = \det(k)$. Mais comme C_2 contient un élément de déterminant 1 et un élément de déterminant -1, nous avons $h = k$. De là $A = B$.

Surjectif Soit $X \in \mathrm{O}(n)$. Si $\det(X) = 1$ alors $X \in \mathrm{SO}(n)$ et $X = \psi(X, \mathbb{1})$. Si par contre $\det(X) = -1$ alors $XR \in \mathrm{SO}(n)$ parce que $\det(XR) = 1$ et nous avons

$$\psi(XR, R) = XR^2 = X. \tag{17.180}$$

Homomorphisme Nous avons

$$\psi\Big((A,h)(B,k)\Big) = \psi\big(A\rho_h(B), hk\big) = A(hBh^{-1})hk = AhBk, \tag{17.181}$$

tandis que

$$\psi(A,h)\psi(B,k) = AhBk, \tag{17.182}$$

qui est la même chose.

□

17.7 Isométries dans $\mathbb{R}^n$

Définition 17.66.
*Un **hyperplan** de $\mathbb{R}^n$ est un sous-espace affine de dimension $n-1$.*

Lemme-définition 17.67.
Si un hyperplan H de $\mathbb{R}^n$ est donné, et si $x \in \mathbb{R}^n$, il existe un unique point $y \in \mathbb{R}^n$ tel que

(1) $x - y \perp H$,

(2) Le segment $[x, y]$ coupe H en son milieu.

*La **réflexion** σ_H est l'application $\sigma_H \colon \mathbb{R}^n \to \mathbb{R}^n$ qui à x fait correspondre ce y.*

Démonstration. Il faut vérifier que les conditions données définissent effectivement un unique point de $\mathbb{R}^n$. Soit H_0 le sous-espace vectoriel parallèle à H et une base orthonormée $\{e_1, \dots, e_{n-1}\}$ de H_0. Nous complétons cela en une base orthonormée de $\mathbb{R}^n$ avec un vecteur e_n. Si $H = H_0 + v$, quitte à décomposer v en une partie parallèle et une partie perpendiculaire à H, nous avons

$$H = H_0 + \lambda e_n \tag{17.183}$$

pour un certain λ.

Une droite passant par x et perpendiculaire à H est de la forme $t \mapsto x + te_n$. Si $x = \sum_{i=1}^n x_i e_i$ alors l'unique point de cette droit à être dans H est le point tel que $x_n e_n + te_n = \lambda e_n$, c'est-à-dire $t = -x_n$. L'unique point y sur cette droite à être tel que $[x, y]$ coupe H en son milieu est celui qui correspond à $t = -2x_n$. □

Notons au passage que cette preuve donne une formule pour σ_H :

$$\sigma_H(x) = \sum_{i=1}^{n-1} x_i e_i - x_n e_n. \tag{17.184}$$

Il s'agit donc de changer le signe de la composante perpendiculaire à H.

Lemme 17.68.
Dans cette même base si H_0 est l'hyperplan parallèle à H et passant par l'origine, nous écrivons $H = H_0 + \lambda e_n$ pour un certain λ. Alors

$$\sigma_H = \sigma_{H_0} + 2\lambda e_n. \tag{17.185}$$

Démonstration. Un élément $x \in \mathbb{R}^n$ peut être décomposé dans la base adéquate en $x = x_H + x_n e_n$. Nous savons de la formule (17.184) que

$$\sigma_H(x) = x_H - x_n e_n. \tag{17.186}$$

Mais vu que $\sigma_{H_0}(x_H) = x_H - 2\lambda e_n$ nous avons

$$\sigma_{H_0}(x) + 2\lambda e_n = \sigma_{H_0}(x_H + x_n e_n) + 2\lambda e_N = x_H - 2\lambda e_n - x_n e_n + 2\lambda e_n = x_H - x_n e_n. \tag{17.187}$$

□

Le lemme suivant est une généralisation du fait que tous les points de la médiatrice d'un segment sont à égale distance des deux extrémités du segment (très utile lorsqu'on étudie les triangles isocèles).

Lemme 17.69 ([354]).
Soient deux points distincts $x_0, y_0 \in \mathbb{R}^n$ l'ensemble $H \subset \mathbb{R}^n$ donné par

$$H = \{x \in \mathbb{R}^n \text{ tel que } d(x, x_0) = d(x, y_0)\}. \tag{17.188}$$

Alors H est l'hyperplan orthogonal au vecteur $v = y_0 - x_0$ et H passe par le milieu du segment $[x_0, y_0]$.

Démonstration. Nous savons que

$$d(x,x_0)^2 = \langle x-x_0, x-x_0\rangle = \|x\|^2 + \|x_0\|^2 - 2\langle x, x_0\rangle, \tag{17.189}$$

ou encore

$$\|x_0\|^2 - \|y_0\|^2 = 2\langle x, x_0 - y_0\rangle. \tag{17.190}$$

En posant $v = y_0 - x_0$ et en considérant la forme linéaire

$$\begin{aligned}\beta\colon \mathbb{R}^n &\to \mathbb{R}\\ x &\mapsto \langle x, v\rangle,\end{aligned} \tag{17.191}$$

Nous avons $x \in H$ si et seulement si $\beta(x) = \frac{1}{2}\big(\|y_0\|^2 - \|x_0\|^2\big) = \lambda$. En d'autres termes, $H = \beta^{-1}(\lambda)$. Par la proposition 8.24 la partie H est un sous-espace affine. C'est même un translaté de $\ker(\beta)$, et comme $\ker(\beta)$ est l'espace vectoriel des vecteurs perpendiculaires à v, nous avons $\dim(H) = \dim\big(\ker(\beta)\big) = n-1$.

Le fait que H contienne le milieu du segment $[x_0, y_0]$ est par définition. □

Pour le lemme suivant, et pour que la récurrence se passe bien nous disons que l'ensemble vide est un espace vectoriel de dimension -1.

Lemme 17.70 ([355]).
Soit un espace euclidien E de dimension n.

(1) Si f est une isométrie de E satisfaisant

$$\dim\big(\mathrm{Fix}(f)\big) = n-k \tag{17.192}$$

alors f peut être écrit comme composition de k réflexions hyperplanes.

(2) Une isométrie de E peut être écrite sous la forme de $\mathrm{rang}(f-\mathrm{Id})$ réflexions, mais pas moins.

(3) Toute isométrie de $\mathbb{R}^n$ peut être écrite comme composition de $n+1$ réflexions.

Démonstration. Les deux parties importantes à démontrer sont les points (1) et la partie « pas moins » de (2). Le reste sont des reformulations.

Pour (1) Nous faisons une récurrence sur $k \geqslant 0$.

Pour l'initialisation, si $k = 0$ alors $\dim\big(\mathrm{Fix}(f)\big) = n$, c'est-à-dire que f fixe tout $\mathbb{R}^n$, autant dire que f est l'identité, une composition de zéro réflexions.

Pour la récurrence, nous supposons que le lemme est démontré jusqu'à $k \geqslant 0$. Soit donc $f \in \mathrm{Isom}(\mathbb{R}^n)$ tel que

$$\dim\big(\mathrm{Fix}(f)\big) = n-(k+1). \tag{17.193}$$

Vu que $k \geqslant 0$, la dimension de $\mathrm{Fix}(f)$ est strictement plus petite que n, donc il existe un $x_0 \in \mathbb{R}^n$ tel que $f(x_0) \neq x_0$. Nous posons

$$H = \{x \in E \text{ tel que } d(x,x_0) = d\big(x, f(x_0)\big)\}. \tag{17.194}$$

Par le lemme 17.69, ce H est l'hyperplan orthogonal à $v = f(x_0) - x_0$ et passant par le milieu du segment $[x_0, f(x_0)]$.

Nous posons $g = \sigma_H \circ f$. Vu que $g(x_0) = \sigma_H(f(x_0)) = x_0$, ce x_0 est un point fixe de g. Le fait que $\sigma_H\big(f(x_0)\big) = x_0$ est vraiment la définition de l'hyperplan H.

Nous avons donc

$$x_0 \in \mathrm{Fix}(g)\backslash\mathrm{Fix}(f). \tag{17.195}$$

Mais nous prouvons de plus que $\mathrm{Fix}(f) \subset \mathrm{Fix}(g)$. En effet si $y \in Fix(f)$ alors $y \in H$ parce que

$$d(y,x_0) = d\big(f(y), f(x_0)\big) = d\big(y, f(x_0)\big). \tag{17.196}$$

Vu que $y \in H$ nous avons $y \in \mathrm{Fix}(g)$ parce que

$$g(y) = \sigma_H\big(f(y)\big) = \sigma_H(y) = y. \tag{17.197}$$

Tout cela pour dire que l'ensemble $\mathrm{Fix}(g)$ est *strictement* plus grand que $\mathrm{Fix}(f)$. Et comme ce sont des espaces affines nous pouvons parler de dimension :

$$\dim\big(\mathrm{Fix}(g)\big) > \dim\big(\mathrm{Fix}(f)\big). \tag{17.198}$$

Par hypothèse de récurrence, l'application g peut être écrite comme composition de k réflexions. Donc l'application

$$f = \sigma_H \circ g \tag{17.199}$$

est une composition de $k+1$ réflexions.

Pour (2), existence Ce point est une reformulation du point (1). Le fait est que $\mathrm{Fix}(f) = \ker(f - \mathrm{Id})$ parce que $x \in \mathrm{Fix}(f)$ si et seulement si $f(x) = x$ si et seulement si $(f - \mathrm{Id})x = 0$. Nous utilisons le théorème du rang 4.39 à l'endomorphisme $f - \mathrm{Id}$:

$$\dim\big(\mathrm{Fix}(f)\big) = \dim\big(\ker(f - \mathrm{Id})\big) = \dim(E) - \mathrm{rang}(f - \mathrm{Id}). \tag{17.200}$$

En remplaçant par les valeurs :

$$n - k = n - \mathrm{rang}(f - \mathrm{Id}). \tag{17.201}$$

Or le point (1) donnait f comme composée de $n - k$ réflexions. Donc f est composée de $\mathrm{rang}(f - \mathrm{Id})$ réflexions.

Pour (2), « pas moins » Supposons que $f = \sigma_1 \circ \ldots \circ \sigma_r$ où σ_i est la réflexion de l'hyperplan H_i. Nous devons prouver que $r \geqslant \mathrm{rang}(f - \mathrm{Id})$. Nous avons

$$\bigcap_{i=1}^{r} H_i \subset \ker(f - \mathrm{Id}). \tag{17.202}$$

D'autre part, la proposition 9.239 nous donne $\dim \bigcap_i H_i \geqslant n - r$. Donc

$$n - r \leqslant \dim\big(\bigcap_{i=1}^{r}\big) \leqslant \dim\big(\ker(f - \mathrm{Id})\big) = n - \mathrm{rang}(f - \mathrm{Id}). \tag{17.203}$$

Donc $n - r \leqslant n - \mathrm{rang}(f - \mathrm{Id})$ ou encore

$$r \geqslant \mathrm{rang}(f - \mathrm{Id}). \tag{17.204}$$

Pour (3) La première partie de ce théorème n'est rien d'autre que le lemme 17.70 parce que le pire cas est celui où le fixateur de f est réduit à l'ensemble vide, et dans ce cas l'application f est une composition de $n+1$ réflexions.

□

Proposition 17.71.
Un élément de $\mathrm{SO}(3)$ *qui fixe deux vecteurs linéairement indépendants est l'identité.*

Démonstration. Soit un élément $A \in \mathrm{SO}(3)$ et deux vecteurs linéairement indépendants $v_1, v_3 \in \mathbb{R}^3$ tels que $Av_1 = v_1$ et $Av_2 = v_2$. Vu que v_1 et v_2 sont linéairement indépendants, le théorème de la base incomplète 4.11 nous permet de considérer $v_3 \in \mathbb{R}^3$ tel que $\{v_1, v_2, v_3\}$ soit une base. Dans cette base, la matrice de A est de la forme

$$A = \begin{pmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & c \end{pmatrix}. \tag{17.205}$$

Le déterminant de cette matrice est c. Or $\det(A) = 1$ parce qu'elle est dans $\mathrm{SO}(3)$. Donc $c = 1$. Le fait que A soit orthogonale implique que la troisième colonne doit être un vecteur de norme 1. Donc $a = b = 0$.

Donc $A = \mathrm{Id}$. □

Corolaire 17.72.
Tout élément de SO(3) *peut être écrit comme composée de deux réflexions.*

Démonstration. Un élément de SO(3) est une isométrie de $\mathbb{R}^3$ parce que si $A \in \mathrm{SO}(3)$ alors[14]

$$\langle Ax, Ay\rangle = \langle A^*Ax, y\rangle = \langle x, y\rangle. \tag{17.206}$$

Donc si le rang de A est k, alors A est la composée de $3-k$ réflexions par le lemme 17.70.

Si $A = \mathrm{Id}$, c'est bon parce que l'identité est la composée de deux réflexions égales. Nous supposons que A n'est pas l'identité.

Comme discuté dans l'exemple 11.101, l'opérateur A possède trois valeurs propres dans $\mathbb{C}$ dont une réelle, et deux complexes conjuguées. Nous les notons $\lambda \in \mathbb{R}$ et $\alpha, \bar{\alpha} \in \mathbb{C}$. Le déterminant de A, qui vaut 1, est le produit de ces trois valeurs propres, c'est-à-dire $\lambda|\alpha|^2$. En particulier $\lambda > 0$.

Si v est un vecteur propre correspondant à la valeur propre λ, nous avons $\|v\| = \|Av\| = |\lambda|\|v\|$ parce que A est une isométrie. Donc $\lambda = \pm 1$.

Au final, $\lambda = 1$. Cela signifie que A laisse au moins un vecteur invariant. Vu que A n'est pas l'identité, la proposition 17.71 nous indique qu'il n'y a pas d'autres vecteurs de $\mathbb{R}^3$ à être fixé par A. Donc $\dim\big(\mathrm{Fix}(A)\big) = 1$ et le lemme 17.70(1) nous s'écrit avec $n = 3$, $k = 2$ et implique que A est la composée de deux réflexions. □

Lemme 17.73.
Soit un hyperplan H et un vecteur v de $\mathbb{R}^n$. Nous avons

$$\tau_v \circ \sigma_H \circ \tau_v^{-1} = \sigma_{\tau_v(H)}. \tag{17.207}$$

Démonstration. Pour ce faire nous considérons une base adaptée. Les vecteurs $\{e_1, \dots, e_{n-1}\}$ forment une base orthonormée de H_0 et e_n complète en une base orthonormée de $\mathbb{R}^n$. Soit H_0 l'hyperplan parallèle à H et passant par l'origine ; nous avons, pour un certain $\lambda \in \mathbb{R}$,

$$H = H_0 + \lambda e_n \tag{17.208}$$

D'un autre côté, le vecteur v peut être décomposé en $v = v_1 + v_2$ où $v_1 \perp H$ et $v_2 \parallel H$. Alors

$$\tau_v(H) = H + v = H + v_2 = H_0 + \lambda e_n + v_2. \tag{17.209}$$

Nous pouvons maintenant utiliser le lemme 17.68 pour exprimer la transformation $\sigma_{\tau_v(H)}$:

$$\sigma_{\tau_v(H)}(x) = \sigma_{H_0}(x) + 2\lambda e_n + 2v_2 \tag{17.210}$$

Mais d'autre part,

$$(\tau_v \circ \sigma_H \circ \tau_v^{-1})(x) = v + \sigma_H(x - v) = v + \sigma_{H_0}(x - v) + 2\lambda e_n. \tag{17.211}$$

Vue la décomposition de $v = v_1 + v_2$ nous avons $\sigma_{H_0}(v) = -v_1 + v_2$ et donc

$$(\tau_v \circ \sigma_H \circ \tau_v^{-1})(x) = v + \sigma_{H_0}(x) + v_1 - v_2 + 2\lambda e_n = \sigma_{H_0} + 2v_1 + 2\lambda e_n. \tag{17.212}$$

Les expressions (17.210) et (17.212) coïncident, d'où l'égalité recherchée. □

Théorème 17.74 ([354]).
Une isométrie de $(\mathbb{R}^n, d)$ préserve l'orientation si et seulement si est elle composition d'un nombre pair de réflexions.

Démonstration. Nous définissons

$$\begin{aligned} \epsilon\colon \mathrm{Isom}(\mathbb{R}^n) &\to \{\pm 1\} \\ \tau_v \circ \alpha &\mapsto \det(\alpha) \end{aligned} \tag{17.213}$$

14. Opérateur orthogonal, définition 9.39.

où nous nous référons à la décomposition unique d'un élément de $\mathrm{Isom}(\mathbb{R}^n)$ sous la forme $\tau_v \circ \alpha$ avec $\alpha \in O(n)$ donnée par le théorème 17.63(3).

Le noyau de ϵ est alors la partie

$$\ker(\epsilon) = \mathbb{R}^n \times_{\mathbf{Ad}} \mathrm{SO}(n). \tag{17.214}$$

Une isométrie f préserve l'orientation si et seulement si $\epsilon(f) = 1$. Vu que toutes les isométries sont des compositions de réflexions (première partie), il nous suffit de montrer que $\epsilon(\epsilon_H) = -1$ pour qu'une isométrie préserve l'orientation si et seulement si elle est composition d'un nombre pair de réflexions.

Nous commençons par prouver que pour tout vecteur v, $\epsilon\big(\sigma_H\big) = \epsilon\big(\sigma_{\tau_v(H)}\big)$. Pour cela nous utilisons le lemme 17.73 et le fait que ϵ est un homomorphisme :

$$\epsilon(\sigma_{\tau_v(H)}) = \epsilon(\tau_v)\epsilon(\sigma_H)\epsilon(\tau_v^{-1}) = \epsilon(\sigma_H) \tag{17.215}$$

parce que la partie linéaire d'une translation est l'identité (et donc $\epsilon(\tau_v) = 1$ pour tout v).

Nous avons donc $\epsilon(\sigma_H) = \epsilon(\sigma_{H_0})$. En ce qui concerne σ_{H_0}, dans la base adaptée la matrice est

$$\sigma_{H_0} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & -1 \end{pmatrix}, \tag{17.216}$$

dont le déterminant est -1. □

Pour en savoir plus sur le groupe des isométries, il faut lire le théorème de Cartan-Dieudonné dans [356].

17.8 Groupes finis d'isométries

Définition 17.75.
*Si X est une partie finie de $\mathbb{R}^n$, le **barycentre** de X est le point*

$$B_X = \frac{1}{|X|} \sum_{x \in X} x \tag{17.217}$$

où $|X|$ est le cardinal de X.

Cela est à mettre en relation avec la définition dans le cadre affine 8.28.

Lemme 17.76 ([354])**.**
Les applications affines de $\mathbb{R}^n$ préservent le barycentre[15] *des parties finies.*

Démonstration. Soit une partie finie X de $\mathbb{R}^n$ et une application affine $f \in \mathrm{Aff}(\mathbb{R}^n)$. Nous devons prouver que

$$f(B_X) = B_{f(X)}. \tag{17.218}$$

Nous savons que toute application affine est une composée de translation et d'une application linéaire : $f = \tau_v \circ g$ avec $v \in \mathbb{R}^n$ et $g \in \mathrm{GL}(n, \mathbb{R})$. Nous vérifions le résultat séparément pour τ_v et pour g.

D'une part,

$$B_{\tau_v(X)} = \frac{1}{|\tau_v(X)|} \sum_{y \in \tau_v(X)} y = \frac{1}{|X|} \sum_{x \in X} (x + v) = B_x + \frac{1}{|X|} \sum_{x \in X} v = B_x + v = \tau_v(B_X). \tag{17.219}$$

Nous avons utilisé le fait que X et $\tau_v(X)$ possèdent le même nombre d'éléments, ainsi que le fait d'avoir une somme de $|X|$ termes tous égaux à v.

15. Définition 17.75.

D'autre part,

$$B_{g(X)} = \frac{1}{|X|}\sum_{x\in X} g(x) = g\big(\frac{1}{|X|}\sum_{x\in X} x\big) = g(B_X) \tag{17.220}$$

où nous avons utilisé la linéarité de g dans tous ses retranchements. □

Proposition 17.77.
Points fixes d'un sous-groupe.

(1) *Soit H un sous-groupe finie des isométries de $(\mathbb{R}^n, d)$. Alors il existe $v \in \mathbb{R}^n$ tel que $f(v) = v$ pour tout $f \in H$.*

(2) *Si H est un sous-groupe de $\mathrm{Isom}(\mathbb{R}^n, d)$ n'acceptant pas de points fixes, alors il est infini.*

Démonstration. Le groupe H agit sur $\mathbb{R}^n$, et si $x \in \mathbb{R}^n$ nous pouvons considérer son orbite

$$Hx = \{f(x) \text{ tel que } f \in H\}, \tag{17.221}$$

qui est une partie finie de $\mathbb{R}^n$. Considérons son barycentre $v = B_{Hx}$. Soit $f \in H$. Alors

$$\begin{aligned} f(v) &= f(B_{Hx}) && (17.222a)\\ &= B_{f(Hx)} && (17.222b)\\ &= B_{Hx} && (17.222c)\\ &= v, && (17.222d) \end{aligned}$$

Justifications :

— Pour (17.222b), c'est le lemme 17.76.

— Pour (17.222c), c'est le fait que, $f \in H$ étant donné, l'application $g \mapsto fg$ est une bijection de H, donc

$$f(Hx) = \{(fg)(x) \text{ tel que } h \in H\} = \{g(x) \text{ tel que } g \in H\} = Hx. \tag{17.223}$$

Bref, v est fixé par H.

La seconde affirmation n'est rien d'autre que la contraposée de la première. □

Proposition 17.78.
À propos de groupes finis d'isométries.

(1) *Tout sous-groupe fini de $\mathrm{Isom}(\mathbb{R}^n)$ est isomorphe à un sous-groupe fini de $\mathrm{O}(n)$.*

(2) *Tout sous-groupe fini de $\mathrm{Isom}^+(\mathbb{R}^n)$ est isomorphe à un sous-groupe fini de $\mathrm{SO}(n)$.*

Démonstration. Soit H un sous-groupe fini de $\mathrm{Isom}(\mathbb{R}^n)$ et $v \in \mathbb{R}^n$ un élément fixé par H (comme garantit par la proposition 17.77). Nous posons

$$\begin{aligned} \phi\colon H &\to \mathrm{Isom}(\mathbb{R}^n)\\ f &\mapsto \tau_v^{-1} \circ f \circ \tau_v. \end{aligned} \tag{17.224}$$

ϕ est un homomorphisme Les opération du type $\phi = \mathbf{Ad}(\tau_v)$ sont toujours des homomorphismes.

ϕ consiste à extraire la partie linéaire Si $f = \tau_w \circ g$ alors

$$\begin{aligned} \phi(f)(x) &= (\tau_{-v} \circ \tau_w \circ g \circ \tau_v)(x) && (17.225a)\\ &= \tau_{w-v}(g(x) + g(v)) && (17.225b)\\ &= g(x) + g(v) - v + w && (17.225c) \end{aligned}$$

Mais $g(v) + w = f(v)$ et nous savons que $f(v) = v$. Donc il ne reste que $\phi(f)(x) = g(x)$.

ϕ est injective Si $f = \tau_w \circ g$ vérifie $\phi(f) = \mathrm{Id}$, il faut en particulier que $g = \mathrm{Id}$. Mais H est fini et ne peut donc pas contenir de translations non triviales. Donc $w = 0$ et $f = \mathrm{Id}$.

Donc ϕ est une injection à valeur dans les transformations linéaires de $\mathrm{Isom}(\mathbb{R}^n)$. Autrement dit, ϕ est un isomorphisme entre H et son image, laquelle image est dans $\mathrm{O}(n)$.

En ce qui concerne la seconde partie, si $f \in \mathrm{Isom}^+(\mathbb{R}^n)$, alors $\phi(f)$ y est aussi, tout en étant linéaire. Donc $\phi(f) \in \mathrm{SO}(n)$. □

L'extraction de la partie linéaire est injective ? Certe c'est prouvé, mais on peut se demander ce qu'il se passe si H contient deux éléments qui ont la même partie linéaire. Cela n'est pas possible parce si $f_1 = \tau_{w_1} \circ g$ et $f_2 = \tau_{w_2} \circ g$ sont dans H alors $f_1 f_2^{-1} = \tau_{w_1+w_2}$ est également dans H, ce qui n'est pas possible si H est fini.

17.8.1 Groupe diédral

Proposition 17.79.
Les racines de l'unité dans $\mathbb{C}$, c'est-à-dire la partie

$$\{e^{2ik\pi/n}, k = 0, \dots, n-1\}, \tag{17.226}$$

forment un polynôme régulier.

17.8.1.1 Définition et générateurs : vue géométrique

Définition 17.80.
*Le **groupe diédral** D_n ($n \geqslant 3$) est le groupe des isométries de $(\mathbb{C}, d)$ laissant invariant l'ensemble*

$$\{e^{2ik\pi/n}, k = 0, \dots, n-1\} \tag{17.227}$$

des racines de l'unité.

17.81.
La proposition 17.79 nous permet de dire que le groupe diédral est le groupe des isométries de $\mathbb{R}^2$ laissant invariant un polygone régulier à n côtés. C'est un peu pour cela que nous n'avons défini D_n que pour $n \geqslant 3$; et un peu aussi pour une raison technique qui arrivera dans la preuve du lemme 17.82.

Lemme 17.82.
Nous avons

$$D_n \subset O(2, \mathbb{R}). \tag{17.228}$$

Démonstration. Si $f \in D_n$, alors $f(e^{2ik\pi/n})$ doit être l'un des $e^{2ik'\pi/n}$, et vu que f conserve les longueurs dans $\mathbb{C}$, nous devons avoir

$$1 = d(0, e^{2ik\pi/n}) = d\big(f(0), e^{2ik'\pi/n}\big). \tag{17.229}$$

Donc $f(0)$ est à l'intersection de tous les cercles de rayon 1 centrés en les $e^{2ik\pi/n}$, ce qui montre que $f(0) = 0$ (dès que $n \geqslant 3$). Par conséquent notre étude du groupe diédral ne doit prendre en compte que les isométries vectorielles de $\mathbb{R}^2$. En d'autres termes

$$D_n \subset O(2, \mathbb{R}). \tag{17.230}$$

□

Proposition 17.83 ([357]).
Le groupe D_n contient un sous-groupe cyclique d'ordre 2 et un sous-groupe cyclique d'ordre n.

Démonstration. Nous notons s la conjugaison complexe [16]. C'est un élément d'ordre 2 qui est dans D_n parce que

$$s\big(e^{2ki\pi/n}\big) = e^{2(n-k)i\pi/n}. \tag{17.231}$$

16. C'est une réflexion ; la réflexion d'axe $\mathbb{R}$ dans $\mathbb{C}$.

De la même façon, la rotations d'angle $2\pi/n$, que l'on note r, agit sur les racines de l'unité et engendre un le groupe d'ordre n des rotations d'angle $2k\pi/n$. □

Proposition 17.84 ([357]).
Si s est la conjugaison complexe et r la rotation d'angle $2\pi/n$, alors $(sr)^2 = \mathrm{Id}$.

Démonstration. Si $z^n = 1$, alors

$$(srsr)z = srse^{2i\pi/n}z = sr\big(e^{-2\pi i/n\bar{z}}\big) = s\bar{z} = z. \tag{17.232}$$

□

Proposition 17.85 ([357]).
Nous notons s la conjugaison complexe et r la rotation d'angle $2\pi/n$.

(1) Le groupe diédral D_n est engendré par s et r.

(2) Tous les éléments de D_n s'écrivent sous la forme r^m ou $s \circ r^m$.

Démonstration. Nous considérons les points $A_0 = 1$ et $A_k = e^{2ki\pi/n}$ avec $k \in \{1, \dots, n-1\}$. Par convention, $A_n = A_0$. L'action des éléments s et r sur ces points est

$$r(A_k) = A_{k+1} \tag{17.233a}$$
$$s(A_k) = A_{n-k}. \tag{17.233b}$$

Cette dernière est l'équation (17.231).

Soit $f \in D_n$. Étant donné que c'est une isométrie de $\mathbb{R}^2$ avec un point fixe (le point 0), f est soit une rotation soit une réflexion.

Supposons pour commencer que un des A_k est fixé par f. Dans ce cas f a deux points fixes : O et A_k et est donc la réflexion d'axe (OA_k). Dans ce cas, nous avons $f = s \circ r^{n-2k}$. En effet

$$s \circ r^{n-2k}(A_k) = s(A_{k+n-2k}) = s(A_{n-k}) = A_k. \tag{17.234}$$

Donc O et A_k sont deux points fixes de l'isométrie f ; donc f est bien la réflexion sur le bon axe.

Nous passons à présent au cas où f ne fixe aucun des A_k.

(1) Supposons que f soit une rotation. Si $f(A_k) = A_m$, alors l'angle de la rotation est

$$\frac{2(m-k)\pi}{n}, \tag{17.235}$$

et donc $f = r^{m-k}$, qui est de la forme demandée.

(2) Supposons à présent que f soit une réflexion d'axe Δ. Cette fois, Δ ne passe par aucun des points A_k, par contre Δ passe par 0. Nous commençons par montrer que Δ doit être la médiatrice d'un des côtés $[A_p, A_{p+1}]$ du polygone. Vu que Δ passe par O et n'est aucune des droites (OA_k), cette droite passe par l'intérieur d'un des triangles OA_pA_{p+1} et intersecte donc le côté correspondant.

Notre tâche est de montrer que Δ coupe $[A_p, A_{p+1}]$ en son milieu. Dans ce cas, Δ sera automatiquement perpendiculaire parce que le triangle OA_pA_{p+1} est isocèle en O. Nommons l la longueur des côtés du polygone, $P = \Delta \cap [A_p, A_{p+1}]$, $x = d(A_p, P)$ et $\delta = d(A_p, \Delta)$. Vu que f est la symétrie d'axe Δ, nous avons aussi $d\big(f(A_p), \Delta\big) = \delta$ et $d\big(A_p, f(A_p)\big) = 2\delta$. D'autre part, par la définition de la distance, $\delta < x$. Si $x < \frac{l}{2}$, alors $\delta < \frac{\delta}{2}$ et donc $d\big(A_p, f(A_p)\big) < l$. Or cela est impossible parce que le polygone ne possède aucun sommet à distance plus courte que l de A_p.

De la même manière si $x > \frac{l}{2}$, nous raisonnons avec A_{p+1} pour obtenir une contradiction. Nous en concluons que la seule possibilité est $x = \frac{l}{2}$, et donc $f(A_p) = A_{p+1}$. Montrons alors

que $f = s \circ r^{n-2p-1}$. Il faut montrer que c'est une réflexion qui envoie A_p sur A_{p+1}. D'abord c'est une réflexion parce que

$$\det(sr^{n-2p-1}) = \det(s)\det(r^{n-2p-1}) = -1 \tag{17.236}$$

parce que $\det(s) = -1$ alors que $\det(r^k) = 1$ parce que r est une rotation dans $\mathrm{SO}(2)$. Ensuite nous avons

$$s \circ r^{n-2p-1}(A_p) = s(A_{p+n-2p-1}) = s(A_{n-p-1}) = A_{n-(n-p-1)} = A_{p+1}. \tag{17.237}$$

Donc $s \circ r^{n-2p-1}$ est bien une réflexion qui envoie A_p sur A_{p+1}.

□

Corolaire 17.86.
La liste des éléments de D_n est

$$D_n = \{1, r, \dots, r^{n-1}, s, sr, \dots, sr^{n-1}\} \tag{17.238}$$

et $|D_n| = 2n$.

Démonstration. Nous savons par la proposition 17.85 que tous les élément de D_n s'écrivent sous la forme r^k ou sr^k. Vu que r est d'ordre n, il ne faut considérer que $k \in \{1, \dots, n-1\}$. Les éléments 1, r,..., r^{n-1} sont tous différents, et sont (pour des raisons de déterminant) tous différents des sr^k. Les isométries sr^k sont toutes différentes entre elles pour essentiellement la même raison :

$$sr^k(A_p) = s(A_{p+k}) = A_{n-p+k} \tag{17.239}$$

donc si $k \neq k'$, $sr^k(A_p) \neq sr^{k'}(A_p)$. La liste des éléments de D_n est donc

$$D_n = \{1, r, \dots, r^{n-1}, s, sr, \dots, sr^{n-1}\} \tag{17.240}$$

et donc $|D_n| = 2n$. □

Exemple 17.87.
Nous considérons le carré $ABCD$ dans $\mathbb{R}^2$ et nous cherchons les isométries de $\mathbb{R}^2$ qui laissent le carré invariant. Nous nommons les points comme sur la figure 17.8. La symétrie d'axe vertical est nommée s et la rotation de 90 degrés est notée r.

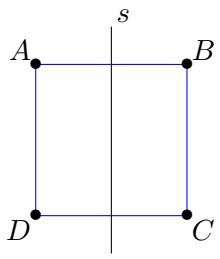


FIGURE 17.8 – Le carré dont nous étudions le groupe diédral.

Il est facile de vérifier que toutes les symétries axiales peuvent être écrites sous la forme $r^i s$. De plus le groupe engendré par s agit sur le groupe engendré par r parce que

$$(srs^{-1})(A, B, C, D) = sr(B, A, D, C) = s(A, D, C, B) = (B, C, D, A), \tag{17.241}$$

c'est-à-dire $srs^{-1} = r^{-1}$. Nous sommes alors dans le cadre du corolaire 2.81 et nous pouvons écrire que

$$D_4 = \mathrm{gr}(r) \times_\sigma \mathrm{gr}(s). \tag{17.242}$$

△

17.8.1.2 Table de multiplication

La proposition 17.85 nous indique que tous les éléments de D_n s'écrivent sous la forme $s^\epsilon \circ r^m$ avec $\epsilon \in \{0,1\}$. Nous allons maintenant écrire la table de multiplication pour de telles transformations de $\mathbb{C}$.

Lemme 17.88.
Si R est une rotation autour de 0 (dans $\mathbb{C}$), et si s est la conjugaison complexe, alors

$$rs = sr^{-1} \tag{17.243}$$

Démonstration. Il s'agit seulement d'un calcul en écrivant R comme la multiplication par $e^{i\alpha}$. Nous avons

$$(Rs)z = e^{i\alpha}\bar{z} = s\big(e^{-i\alpha}z\big) = sR^{-1}z. \tag{17.244}$$

□

Proposition 17.89.
Si $\epsilon_1, \epsilon_2 \in \{0,1\}$ et si $k, l \in \mathbb{Z}$ nous avons

$$(s^{\epsilon_1}r^k)(s^{\epsilon_2}r^l) = s^{\epsilon_1+\epsilon_2}r^{l+(-1)^{\epsilon_1}k}. \tag{17.245}$$

Démonstration. Si $\epsilon_2 = 1$ alors nous utilisons le lemme 17.88 pour trouver

$$(s^{\epsilon_1}r^k)(s^{\epsilon_2}r^l) = s^{\epsilon_1}(r^k s^{\epsilon_2})r^l = s^{\epsilon_1}s^{\epsilon_2}r^{-k}r^l. \tag{17.246}$$

La proposition est déjà prouvée dans ce cas.

Passons à $\epsilon_2 = 0$. Dans ce cas nous avons

$$(s^{\epsilon_1}r^k)(s^{\epsilon_2}r^l) = s^{\epsilon_1}r^{k+l}, \tag{17.247}$$

et c'est bon. □

17.8.1.3 Générateurs : vue abstraite

17.90.
Nous allons montrer que D_n peut être décrit de façon abstraite en ne parlant que de ses générateurs. Nous considérons un groupe G engendré par des éléments a et b tels que

(1) a est d'ordre 2,
(2) b est d'ordre n avec $n \geqslant 3$,
(3) $abab = e$.

Nous allons prouver que ce groupe doit avoir la même liste d'éléments que celle du corolaire 17.86.

Proposition 17.91 ([357]).
Le groupe G n'est pas abélien.

Démonstration. Nous savons que $abab = e$, donc $abab^{-1} = b^{-2}$, mais $b^{-2} \neq e$ parce que b est d'ordre $n > 2$. Donc $abab^{-1} \neq e$. En manipulant un peu :

$$e \neq abab^{-1} = (ab)(ba^{-1})^{-1} = (ab)(ba)^{-1} \tag{17.248}$$

parce que $a^{-1} = a$. Donc $ab \neq ba$. □

Lemme 17.92 ([357]).
Pour tout k entre 1 et $n-1$ nous avons

$$\boldsymbol{Ad}(a)b^k = ab^k a^{-1} = ab^k a = b^{-k}. \tag{17.249}$$

Démonstration. Nous faisons la démonstration par récurrence. D'abord pour $k = 1$, nous devons avoir $aba = b^{-1}$, ce qui est correct parce que par construction de G nous avons $abab = e$. Ensuite nous supposons que le lemme tient pour k et nous regardons ce qu'il se passe avec $k + 1$:

$$ab^{k+1}ba = ab^k ba = \underbrace{ab^k a}_{b^{-k}} \underbrace{aba}_{b^{-1}} = b^{-k}b^{-1} = b^{-(k+1)}. \tag{17.250}$$

□

Proposition 17.93.
L'élément a n'est pas une puissance de b.

Démonstration. Supposons le contraire : $a = b^k$. Dans ce cas nous aurions

$$e = (ab)(ab) = b^{k+1}b^{k+1} = b^{2k+2} = b^{2k}b^2 = a^2b^2 = b^2, \tag{17.251}$$

ce qui signifierait que b est d'ordre 2, ce qui est exclu par construction. □

Proposition 17.94 ([357])**.**
La liste des éléments de G est donnée par

$$G = \{1, b, \cdots, b^{n-1}, a, ab, \ldots, ab^{n-1}\} = \{a^\epsilon b^k\}_{\substack{\epsilon=0,1 \\ k=0,\ldots,n-1}} \tag{17.252}$$

Les éléments de ces listes sont distincts.

Démonstration. Étant donné que a n'est pas une puissance de b, les éléments 1, a, b,..., b^{n-1} sont distincts. De plus si k et $m = k + p$ sont deux éléments distincts de $\{1, \ldots, n-1\}$, nous avons $ab^k \neq ab^m$ parce que si $ab^k = ab^{k+p}$, alors $a = ab^p$ avec $p < n$, ce qui est impossible. Pour la même raison, $ab^k \neq e$, et $ab^k \neq b^m$.

Au final les éléments $1, a, b, \ldots, b^{n-1}, ab, \ldots, ab^{n-1}$ sont tous différents. Nous devons encore voir qu'il n'y en a pas d'autres.

Par définition le groupe G est engendré par a et b, donc tout élément $x \in G$ s'écrit $x = a^{m_1}b^{k_1} \ldots a^{m_r}b^{k_r}$ pour un certain r et avec pour tout i, $k_i \in \{1, \ldots, n-1\}$ (sauf k_r qui peut être égal à zéro) et $m_i = 1$, sauf m_1 qui peut être égal à zéro. Donc

$$x = a^m b^{k_1} ab^{k_2} a \ldots b^{k_{r-1}} ab^{k_r} \tag{17.253}$$

où m et k_r peuvent éventuellement être zéro. En utilisant le lemme 17.92 sous la forme $b^{k_i}a = ab^{-k_i}$, quitte à changer les valeurs des exposants, nous pouvons passer tous les a à gauche et tous les b à droite pour finir sous la forme $x = a^k b^m$.

Donc non, il n'existe pas d'autres éléments dans G que ceux déjà listés. □

Lemme 17.95 ([1])**.**
Tout élément de G s'écrit de façon unique sous la forme $a^\epsilon b^k$ ou $b^k a^\epsilon$ avec $\epsilon = 0, 1$ et $k = 0, \ldots, n-1$.

Démonstration. Nous commençons par la forme $a^\epsilon b^k$. L'existence est la proposition 17.94. Pour l'unicité nous supposons $a^\epsilon b^k = a^\sigma b^l$ et nous décomposons en 4.

$\epsilon = 0$, $\sigma = 0$ Alors $b^k = b^l$. Mais b étant d'ordre n et k, l étant égaux au maximum à $n - 1$, cette égalité implique $k = l$.

$\epsilon = 0$, $\sigma = 1$ Alors $b^k = ab^l$, ce qui donne $a = b^{k-l}$, ce qui est interdit par la proposition 17.93.

$\epsilon = 1$, $\sigma = 0$ Même problème.

$\epsilon = 1$, $\sigma = 1$ Encore une fois $b^k = b^l$ implique $k = l$.

En ce qui concerne la forme $b^k a^\epsilon$, l'existence est à montrer. Soit l'élément $g = a^\epsilon b^k$ et cherchons à le mettre sous la forme $b^l a^\sigma$. Si $\epsilon = 0$ c'est évident. Sinon $\epsilon = 1$ et nous avons par le lemme 17.92

$$ab^k = b^{-k}a^{-1} = b^{-k}b^n a = b^{-k}a. \tag{17.254}$$

En ce qui concerne l'unicité, nous refaisons 4 cas pour $b^k a^\epsilon = b^l a^\sigma$ comme précédemment et ils se traitement exactement comme précédemment. □

Théorème 17.96.
Les groupes G et D_n sont isomorphes.

Démonstration. Nous utilisons l'application

$$\begin{aligned} \psi\colon G &\to D_n \\ a^k b^m &\mapsto s^k r^m. \end{aligned} \tag{17.255}$$

C'est évidemment bien défini et bijectif, mais c'est également un homomorphisme parce que si nous calculons ψ sur un produit, nous devons comparer

$$\psi\big(a^{k_1}b^{m_1}a^{k_2}b^{m_2}\big) \tag{17.256}$$

avec

$$\psi\big(a^{k_1}b^{m_1}\big)\psi\big(a^{k_2}b^{m_2}\big) = s^{k_1}r^{m_1}s^{k_2}r^{m_2}. \tag{17.257}$$

Vu que D_n et G ont les mêmes propriétés qui permettent de permuter a et b ou s et r, l'expression à l'intérieur du ψ dans (17.256) se simplifie en $a^k b^m$ avec les même k et n que l'expression à droite dans (17.257) ne se simplifie en $s^k r^m$. □

Corolaire 17.97.
Toutes les propriétés démontrées pour G sont vraies pour D_n. En particulier, avec quelques redites :

(1) Le groupe D_n peut être défini comme étant le groupe engendré par un élément s d'ordre 2 et un élément r d'ordre $n-1$ assujettis à la relation $srsr = e$.
(2) Le groupe D_n n'est pas abélien.
(3) Pour tout $k \in \{1, \dots, n-1\}$ nous avons $sr^k s = r^{-k}$.
(4) L'élément s ne peut pas être obtenu comme une puissance de r.
(5) La liste des éléments de D_n est

$$D_n = \{1, r, \dots, r^{n-1}, s, sr, \dots, sr^{n-1}\} \tag{17.258}$$

(6) Le groupe diédral D_n est d'ordre $2n$.

Proposition 17.98.
En posant $C_n = \{r^k\}_{k=0,\dots,n-1}$ et $C_2 = \{a^\epsilon\}_{\epsilon=0,1}$, nous pouvons exprimer D_n comme le produit semi-direct

$$D_n = C_n \times_\rho C_2 \tag{17.259}$$

où ρ désigne l'action adjointe.

Démonstration. L'isomorphisme est :

$$\begin{aligned} \psi\colon C_n \times_\rho C_2 &\to D_n \\ (b^k, a^\epsilon) &\mapsto b^k a^\epsilon. \end{aligned} \tag{17.260}$$

<u>**Action adjointe**</u> L'application $\rho_{a^\epsilon} = \mathbf{Ad}(a^\epsilon)$ est toujours un homomorphisme. Vu que a^ϵ est soit e soit a, nous allons nous restreindre à a et oublier l'exposant ϵ. Il faut montrer que$\mathbf{Ad}(a) \in \mathrm{Aut}(C_n)$. En utilisant le lemme 17.92,

$$\mathbf{Ad}(a)b^k = ab^k a^{-1} = b^{-k} = b^{n-k}. \tag{17.261}$$

L'application $\mathbf{Ad}(a)\colon C_n \to C_n$ est donc bijective et homomorphique. Ergo isomorphisme.

Injectif Si $\psi(b^k, a^\epsilon) = \psi(b^l, a^\sigma)$, alors par unicité du lemme 17.95 nous avons $k = l$ et $\epsilon = \sigma$.

Surjectif Par la partie « existence » du lemme 17.95.

Homomorphisme Lorqu'on prend deux sous-groupes d'un même groupe (ici le groupe des isométrique de $\mathbb{R}^2$), et que l'on tente de faire un produit demi-direct en utilisant l'action adjointe, nous avons toujours un homomorphisme. Dans notre cas, le calcul est :

$$\psi\big((b^k, a^\epsilon)(b^l, a^\sigma)\big) = b^k \rho_{a^\epsilon}(b^l) a^{\epsilon+\sigma} = b^k a^\epsilon b^l a^{-\epsilon} a^{\epsilon+\sigma} = b^k a^\epsilon b^l a^\sigma = \psi(b^k, a^\epsilon)\psi(b^l, a^\sigma). \tag{17.262}$$

□

17.8.1.4 Classes de conjugaison

Pour les classes de conjugaison du groupe diédral nous suivons [358].

D'abord pour des raisons de déterminants[17], les classes des éléments de la forme r^k et de la forme sr^k ne se mélangent pas. Nous notons $C(x)$ la classe de conjugaison de x, et $y \cdot x = yxy^{-1}$.

Les relations que nous allons utiliser sont

$$sr^k s = r^{-k} \tag{17.263a}$$

$$rs = sr^{-1} = sr^{n-1}. \tag{17.263b}$$

La classe de conjugaison qui ne rate jamais est bien entendu $C(1) = 1$. Nous commençons les vraies festivités $C(r^m)$. D'abord $r^k \cdot r^m = r^m$, ensuite

$$(sr^k) \cdot r^m = sr^k r^m r^{-k} s^{-1} = sr^m s^{-1} = r^{-m}. \tag{17.264}$$

Donc

$$C(r^m) = \{r^m, r^{-m}\}. \tag{17.265}$$

À ce niveau il faut faire deux remarques. D'abord si $m > \frac{n}{2}$, alors $C(r^m)$ est la classe de C^{n-m} avec $n - m < \frac{n}{2}$. Donc les classes que nous avons trouvées sont uniquement à lister avec $m < \frac{n}{2}$. Ensuite si $m = \frac{n}{2}$ alors $r^m = r^{-m}$ et la classe est un singleton. Cela n'arrive que si n est pair.

Nous passons ensuite à $C(s)$. Nous avons

$$r^k \cdot s = r^k s r^{-k} = ssr^k sr^{-k} = sr^{-k} r^{-k} = sr^{n-2k}, \tag{17.266}$$

et

$$(sr^k) \cdot s = \underbrace{sr^k s}_{r^{-k}} r^{-k} s^{-1} = r^{-2k} s = r^{n-2k} s = sr^{(n-1)(n-2k)} = sr^{n^2-2kn-n+2k} = sr^{2k}. \tag{17.267}$$

donc

$$C(s) = \{sr^{n-2k}, sr^{2k}\}_{k=0,\dots,n-1}. \tag{17.268}$$

Ici aussi l'écriture n'est pas optimale : peut-être que pour certains k il y a des doublons. Nous reportons l'écriture exacte à la discussion plus bas qui distinguera n pair de n impair. Notons juste que si n est pair, l'élément sr n'est pas dans la classe $C(s)$.

Nous en faisons donc à présent le calcul en gardant en tête le fait qu'il n'a de sens que si n est pair. D'abord

$$s \cdot (sr) = ssrs = rs = sr^{n-1}. \tag{17.269}$$

Ensuite

$$(sr^k) \cdot (sr) = sr^k srr^{-k} s = r^{-2k+1} s = sr^{2k-1}. \tag{17.270}$$

Avec $k = \frac{n}{2}$, cela rend $s \cdot (sr)$, donc pas besoin de le recopier. Nous avons

$$C(sr) = \{sr^{2k-1}\}_{k=1,\dots,n-1}. \tag{17.271}$$

17. Vous notez qu'ici nous utilisons un argument qui utilise la définition de D_n comme isométries de $\mathbb{R}^2$. Si nous avions voulu à tout prix nous limiter à la définition « abstraite » en termes de générateurs, il aurait fallu trouver autre chose.

17.8.1.5 Le compte pour n pair

Si n est pair, nous avons les classes

$$C(1) = \{1\} \qquad 1 \text{ élément} \tag{17.272a}$$
$$C(r^m) = \{r^m, r^{m-1}\} \qquad \text{pour } 0 < m < \frac{n}{2} \qquad \frac{n}{2} - 1 \text{ fois } 2 \text{ éléments} \tag{17.272b}$$
$$C(r^{n/2}) = \{r^{n/2}\} \qquad 1 \text{ élément} \tag{17.272c}$$
$$C(s) = \{sr^{2k}\}_{k=0,\dots,\frac{n}{2}-1} \qquad \frac{n}{2} \text{ éléments} \tag{17.272d}$$
$$C(sr) = \{sr^{2k+1}\}_{k=0,\dots,\frac{n}{2}-1} \qquad \frac{n}{2} \text{ éléments.} \tag{17.272e}$$

Au total nous avons bien listé $2n$ éléments comme il se doit, dans $\frac{n}{2} + 3$ classes différentes.

17.8.1.6 Le compte pour n impair

Si n est impair, nous avons les classes

$$C(1) = \{1\} \qquad 1 \text{ élément} \tag{17.273a}$$
$$C(r^m) = \{r^m, r^{m-1}\} \qquad \text{pour } 0 < m < \frac{n-1}{2} \qquad \frac{n-1}{2} \text{ fois } 2 \text{ éléments} \tag{17.273b}$$
$$C(s) = \{sr^k\}_{k=0,\dots,n-1} \qquad n \text{ éléments} \tag{17.273c}$$

Au total nous avons bien listé $2n$ éléments comme il se doit, dans $\frac{n+3}{2}$ classes différentes.

17.8.2 Applications : du dénombrement

17.8.2.1 Le jeu de la roulette

Soit une roulette à n secteurs que nous voulons colorier en q couleurs[359]. Nous voulons savoir le nombre de possibilités à rotations près. Soit d'abord E l'ensemble des coloriages possibles sans contraintes ; il y a naturellement q^n possibilités. Sur l'ensemble E, le groupe cyclique G des rotations d'angle $2\pi/n$ agit. Deux coloriages étant identiques si ils sont reliés par une rotation, la réponse à notre problème est donné par le nombre d'orbites de l'action de G sur E qui sera donnée par la formule du théorème de Burnside 2.59.

Nous devons calculer $\operatorname{Card}\big(\operatorname{Fix}(g)\big)$ pour tout $g \in G$. Soit g, un élément d'ordre d dans G. Si g agit sur la roulette, chaque secteur a une orbite contenant d éléments. Autrement dit, g divise la roulette en n/d secteurs. Un élément de E appartenant à $\operatorname{Fix}(g)$ doit colorier ces n/d secteurs de façon uniforme ; il y a $q^{n/d}$ possibilités.

Il reste à déterminer le nombre d'éléments d'ordre d dans G. Un élément de G est donné par un nombre complexe de la forme $e^{2ik\pi/n}$. Les éléments d'ordre d sont les racines primitives[18] $d$ièmes de l'unité. Nous savons que –par définition– il y a $\varphi(d)$ telles racines primitives de l'unité. Bref il y a $\varphi(d)$ éléments d'ordre d dans G.

La formule de Burnside nous donne maintenant le nombre d'orbites :

$$\frac{1}{n}\sum_{d|n}\varphi(d)q^{n/d}. \tag{17.274}$$

Cela est le nombre de coloriage possibles de la roulette à n secteurs avec q couleurs.

18. Une racine non primitive 8ième de l'unité est par exemple i. Certes $i^8 = 1$, mais $i^4 = 1$ aussi. Le nombre i est d'ordre 4.

17.8.2.2 L'affaire du collier

Nous avons maintenant des perles de q couleurs différentes et nous voulons en faire un collier à n perles. Cette fois non seulement les rotations donnent des colliers équivalents, mais en outre les symétries axiales (il est possible de retourner un collier, mais pas une roulette). Le groupe agissant sur E est maintenant le groupe diédral [19] D_n conservant un polygone a n sommets.

Nous devons séparer le cas n impair du cas n pair.

Si n est impair, alors les axes de symétries passent par un sommet par le milieu du côté opposé. Le groupe D_n contient n symétries axiales. Nous avons donc maintenant

$$|G| = 2n. \tag{17.275}$$

Nous écrivons la formule de Burnside

$$\mathrm{Card}(\Omega) = \frac{1}{2n} \sum_{g \in G} \mathrm{Card}\big(\mathrm{Fix}(g)\big). \tag{17.276}$$

Si g est une rotation, le travail est déjà fait. Si g est une symétrie, nous avons le choix de la couleur du sommet par lequel passe l'axe et le choix de la couleur des $(n-1)/2$ paires de sommets. Cela fait

$$qq^{(n-1)/2} = q^{\frac{n+1}{2}} \tag{17.277}$$

possibilités. Nous avons donc

$$\mathrm{Card}(\Omega) = \frac{1}{2n} \left(\sum_{d|n} q^{n/d} \varphi(d) + nq^{\frac{n+1}{2}} \right). \tag{17.278}$$

Si n est pair, le choses se compliquent un tout petit peu. En plus de symétries axiales passant par un sommet et le milieu du côté opposé, il y a les axes passant par deux sommets opposés. Pour colorier un collier en tenant compte d'une telle symétrie, nous pouvons choisir la couleur des deux perles par lesquelles passe l'axe ainsi que la couleur des $(n-2)/2$ paires de perles. Cela fait en tout

$$q^2 q^{\frac{n-2}{2}} = q^{\frac{n+2}{2}}. \tag{17.279}$$

Le groupe G contient $n/2$ tels axes.

Notons que cette fois G ne contient plus que $n/2$ symétries passant par un sommet et un côté. L'ordre de G est donc encore $2n$. La formule de Burnside donne

$$\mathrm{Card}(\Omega) = \frac{1}{2n} \left(\sum_{d\,|\,n} \varphi(d) q^{n/d} + \frac{n}{2} q^{(n+2)/2} + \frac{n}{2} q^{n/2} \right). \tag{17.280}$$

17.8.3 Points fixés par une affinité

Lemme 17.99 ([356]).
Si $n \geqslant 3$, alors toute droite est intersection de deux plans non isotropes.

Proposition 17.100 ([354]).
Si une isométrie de $\mathbb{R}^n$ fixe un ensemble F de points, alors elle fixe l'espace affine engendrée par F.

Démonstration. Soit $f \in \mathrm{Isom}(\mathbb{R}^n)$ fixant F. Par le théorème 9.249, c'est une application affine et l'ensemble $\mathrm{Fix}(f)$ des points fixés par f est un sous-espace affine de $\mathbb{R}^n$, grâce à la proposition 8.56.

Donc $\mathrm{Fix}(f)$ est un espace affine contenant F. Vu que l'espace affine engendré par F est l'intersection de tous les espaces affines contenant F, il est en particulier contenu dans $\mathrm{Fix}(f)$. □

19. Définition 17.80.

Corolaire 17.101.
Si f et g sont des isométries de $\mathbb{R}^n$ qui coïncident sur F, alors elles coïncident sur l'espace affine engendré par F.

Démonstration. Nous considérons $h = g^{-1} \circ f$ qui est une isométrie de $\mathbb{R}^n$ fixant F. Elle fixe donc, par la proposition 17.100, l'espace affine engendré par F. Or tout point fixé par h est un point sur lequel g et f coïncident. □

17.9 Classification des isométries dans $\mathbb{R}^2$

17.9.1 Réflexions

Soit un espace vectoriel E de dimension 2 muni d'un produit scalaire[20]. Cela pourrait très bien être $\mathbb{R}^2$, mais nous allons nous efforcer de l'appeler E pour rester un peu général.

Lemme-définition 17.102 (Caractérisation des réflexions)**.**
Soit une droite ℓ de $\mathbb{R}^2$. Il existe une unique application $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ telle que

(1) $f(x) = x$ pour tout $x \in \ell$.

(2) f échange les côtés de ℓ.

(3) f laisse invariants les droites perpendiculaires à ℓ et les cercles dont le centre est sur ℓ.

*Cette application est la **réflexion** d'axe ℓ.*

Démonstration. Soit x hors de ℓ et p la droite perpendiculaire à ℓ et passant par x. Nous avons $f(x) \in p$. En nommant P l'intersection entre ℓ et p, nous considérons le cercle $S(P, \|Px\|)$ qui est un cercle dont le centre est sur ℓ. Il contient x et donc $f(x) \in S(P, \|Px\|)$.

Donc $f(x) \in p \cap S(P, \|Px\|)$. L'intersection entre un cercle et une droite contient de façon générique deux points. L'un est x, mais $f(x) = x$ n'est pas possible parce que x est hors de ℓ et f doit inverser les côtés de ℓ. Donc $f(x)$ est l'autre.

Cela prouve l'unicité. En ce qui concerne l'existence, il suffit de noter que la réflexion σ_ℓ satisfait les contraintes. □

Lemme 17.103.
Soit une droite ℓ est $A \in E$. Alors

$$\sigma_\ell(A) = 2\,\mathtt{proj}_\ell(A) - A \tag{17.281}$$

où $\mathtt{proj}_\ell$ est l'opération de projection orthogonale sur la droite ℓ.

Démonstration. Nous posons $f(X) = 2\,\mathtt{proj}_\ell(X) - X$ et nous allons montrer que $f = \sigma_\ell$ en vérifiant les conditions de la définition 17.102. Nous nous gardons bien de faire un raisonnement du type « nous allons montrer que f et σ_ℓ coïncident sur deux points, et sont donc égales par le corolaire 17.101 » parce que nous ne savons pas encore que σ_ℓ est une application affine, ni même que c'est une isométrie.

Si $X \in \ell$ alors $\mathtt{proj}_\ell(X) = X$ et nous avons $f(X) = 2X - X = X$. Donc ℓ est conservée.

En ce qui concerne les deux côtés de ℓ, il existe une application linéaire $s\colon E \to \mathbb{R}$ et une constante $c \in \mathbb{R}$ telles qu'en posant $\ell(X) = s(X) + c$, la droite ℓ soit le lieux des points X tels que $\ell(X) = 0$. Un côté de la droite est $\ell < 0$ et l'autre côté est $\ell > 0$. Nous avons :

$$\begin{aligned}
\ell\big(f(A)\big) &= \ell\big(2\,\mathtt{proj}_\ell(A) - A\big) && (17.282a)\\
&= s(2\,\mathtt{proj}_\ell(A) - A) + c && (17.282b)\\
&= 2s\big(\mathtt{proj}_\ell(A)\big) - s(A) + c && (17.282c)\\
&= s\big(\mathtt{proj}_\ell(A)\big) - s(A) && (17.282d)\\
&= -c - s(A) && (17.282e)\\
&= -\ell(A) && (17.282f)
\end{aligned}$$

20. Définition 7.186.

où nous avons utilisé le fait que, $\mathtt{proj}_\ell(A)$ étant sur ℓ, $s\big(\mathtt{proj}_\ell(A)\big) + c = 0$. Nous avons donc $\ell\big(f(A)\big) = -\ell(A)$, ce qui indique que A et $f(A)$ sont de part et d'autre de ℓ.

Si d est une droite perpendiculaire à ℓ et si $A \in d$ alors $f(A) = 2\,\mathtt{proj}_\ell(A) - A = \big(\mathtt{proj}_\ell(A) - A\big) + A \in d$ parce que $\mathtt{proj}_\ell(A) \in d$ du fait que d soit précisément perpendiculaire à ℓ. Nous avons aussi utilisé le fait que si $A, B, C \in d$ alors $(B - A) + C \in d$; pensez que $B - A$ est un vecteur directeur et que C est un point de d.

Enfin soit $K \in \ell$ et un cercle $S(K, r)$ centré en K. Soit $A \in S(K, r)$; nous devons vérifier que $f(A) = S(K, r)$. Le segment $[A, f(A)]$ est par définition perpendiculaire à ℓ. Soit M, le milieu, qui est sur la droite ℓ. Les triangles AMK et $f(A)MK$ sont rectangles en M, et $\|AM\| = \|Mf(A)\|$. Le théorème de Pythagore donne $\|AK\| = \|f(A)K\|$. Donc le cercle centré en K est donc préservé par f.

Nous en déduisons que $f = \sigma_\ell$. □

Proposition 17.104 ([360]).
Une réflexion est une isométrie de (E, d) où $d(A, B) = \|A - B\|$.

Démonstration. Soient $A, B \in E$; il faut vérifier que $\|A - B\| = \|\sigma_\ell(A) - \sigma_\ell(B)\|$. Pour cela nous écrivons

$$B - A = \underbrace{B - \mathtt{proj}_\ell(B)}_{=a} + \underbrace{\mathtt{proj}_\ell(B) - \mathtt{proj}_\ell(A)}_{=b} + \underbrace{\mathtt{proj}_\ell(A) - A}_{=c}. \tag{17.283}$$

Vu que $b \perp a$ et $b \perp c$ nous avons

$$\|B - A\| = \langle B - A, B - A\rangle = \|a\|^2 + 2\langle a, c\rangle + \|b\|^2 + \|c\|^2. \tag{17.284}$$

Nous pouvons faire le même jeu avec $\sigma_\ell(B) - \sigma_\ell(A)$ en tenant compte du fait que $\mathtt{proj}_\ell\big(\sigma_\ell(X)\big) = \mathtt{proj}_\ell(X)$ et que

$$\sigma_\ell(A) - \mathtt{proj}_\ell(A) = 2\,\mathtt{proj}_\ell(A) - A - \mathtt{proj}_\ell(A) = -\big(A - \mathtt{proj}_\ell(A)\big). \tag{17.285}$$

Là nous avons utilisé le lemme 17.103. Ce que nous trouvons est que

$$\sigma_\ell(B) - \sigma_\ell(A) = -a + b - c, \tag{17.286}$$

et donc encore une fois

$$\|\sigma_\ell(B) - \sigma_\ell(A)\| = \|a\|^2 - 2\langle a, c\rangle + \|b\|^2 + \|c\|^2. \tag{17.287}$$

□

Remarque 17.105.
Il faut bien comprendre que si l'axe de la réflexion ne passe par par 0 (le zéro de l'espace vectoriel normé $(E, \|.\|)$), la réflexion n'est pas une isométrie de $(E, \|.\|)$ au sens où nous n'avons pas $\|\sigma_\ell(x)\| = \|x\|$.

17.9.2 Segment, plan médiateur et équidistance

Lemme 17.106.
Un point M est sur la médiatrice du segment $[A, B]$ si et seulement si $\|M - A\| = \|M - B\|$.

Lemme 17.107.
Soient A et B de points de $\mathbb{R}^3$. Alors le plan médiateur du segment $[A, B]$ est le lieu des points de $\mathbb{R}^3$ à être équidistants de A et B.

Démonstration. Nous nommons σ ce plan.

Soit X un point équidistant de A et B. Alors dans le plan (A, B, X), le triangle ABX est isocèle en X, et la hauteur issue de X coupe perpendiculairement $[A, B]$ en son milieu. Cela prouve que X est dans le plan médiateur du segment $[A, B]$ (lemme 17.106).

Mettons au contraire que X est dans le plan médiateur de $[A, B]$. Nous avons $(X, M) \perp (A, B)$. Donc le triangle A, B, X est isocèle en X et donc X est équidistant de A et B. □

Lemme 17.108.
Si A' est l'image de A par σ_ℓ alors ℓ est la médiatrice du segment $[A, A']$.

Démonstration. Soit $M \in \ell$. Nous avons

$$\|A - M\|^2 = \|\mathtt{proj}_\ell(A) - A\|^2 + \|\mathtt{proj}_\ell(A) - M\|^2 \tag{17.288}$$

parce que $A - \mathtt{proj}_{ell}(A) \perp M - \mathtt{proj}_\ell(A)$. Par ailleurs, vu que $\sigma_\ell(A) = 2\,\mathtt{proj}_\ell(A) - A$ et que $\mathtt{proj}_\ell(A) = \mathtt{proj}_\ell(A')$,

$$\|\mathtt{proj}_\ell(A) - A\| = \|\mathtt{proj}_\ell(A') - A'\|. \tag{17.289}$$

Nous avons donc

$$\|\sigma_\ell(A) - M\|^2 = \|A - M\|^2, \tag{17.290}$$

ce qui prouve que M est sur la médiatrice de $[A', A]$ par le lemme 17.106. □

17.109.
Si l est une droite dans $\mathbb{R}^2$, nous avons la réflexion $\sigma_l \in \mathrm{Isom}(\mathbb{R}^2)$ d'axe l. Cela est une isométrie et donc une application affine par le théorème 9.249. Le lemme suivant détermine comment la réflexion σ_ℓ se décompose en une translation et une application linéaire.

Lemme 17.110.
Soit une droite ℓ. Alors

$$\sigma_\ell = \tau_{2w} \circ \sigma_{\ell_0} \tag{17.291}$$

où ℓ_0 est la droite parallèle à ℓ passant par l'origine, et w est le vecteur perpendiculaire à ℓ tel que $\ell_0 = \ell + w$.

Démonstration. Il faut trouver trois points non alignés sur lesquels les deux applications coïncident ; cela suffira par le corolaire 17.101.

Pour tous les points de ℓ_0, l'égalité fonctionne parce que si $x \in \ell_0$,

$$\sigma_\ell(x) = x + 2w, \tag{17.292}$$

tandis que

$$\sigma_{\ell_0}(x) + 2w = x + 2w \tag{17.293}$$

du fait que $\sigma_{\ell_0}(x) = x$.

Si $x \in \ell$, alors

$$\sigma_\ell(x) = x \tag{17.294}$$

tandis que

$$\sigma_{\ell_0}(x) + 2w = x - 2w + 2w = x. \tag{17.295}$$

Donc les applications affines σ_ℓ et $x \mapsto \sigma_{\ell_0}(x) + 2w$ coïncident sur ℓ et ℓ_0. Elles coïncident donc partout. □

17.9.3 Translations et réflexions

Lemme 17.111.
Si $A \in \mathbb{R}^2$, si ℓ est une droite de $\mathbb{R}^2$, alors nous avons

$$\mathtt{proj}_{\tau_A(\ell)} = \tau_A \circ \mathtt{proj}_\ell \circ \tau_A^{-1}. \tag{17.296}$$

Démonstration. Soit $x \in \mathbb{R}^2$. Soit v unitaire dans la direction[21] de ℓ. La condition $q = \mathtt{proj}_\ell(x)$ est le système

$$\begin{cases} q \in \ell & \text{(17.297a)} \\ (q - x) \cdot v = 0. & \text{(17.297b)} \end{cases}$$

21. Cela signifie qu'il existe $a \in \mathbb{R}^2$ tel que $\ell = a + \mathbb{R}v$.

Nous voulons prouver que $\tau_A(q) = \mathtt{proj}_{\tau_A(\ell)}\big(\tau_A(x)\big)$, c'est-à-dire que

$$\begin{cases} \tau_A(q) \in \tau_A(\ell) & (17.298a) \\ \big(\tau_A(q) - \tau_A(x)\big) \cdot v = 0. & (17.298b) \end{cases}$$

Nous avons utilisé le fait que v est un vecteur unitaire dans la direction de $\tau_A(\ell)$ aussi bien que de ℓ.

Vu que $q \in \ell$, bien entendu que $\tau_A(q) \in \tau_A(\ell)$. D'autre part, $\tau_A(q) - \tau_A(x) = q - x$, donc

$$\big(\tau_A(q) - \tau_A(x)\big) \cdot v = (q - x) \cdot v = 0. \tag{17.299}$$

□

Lemme 17.112.
Si ℓ est une droite de $\mathbb{R}^2$, si $A \in \mathbb{R}^2$, alors

$$\sigma_{\tau_A(\ell)} = \tau_A \sigma_\ell \tau_A^{-1}. \tag{17.300}$$

Démonstration. Il s'agit d'un calcul mettant en scène les lemmes 17.103 et 17.111 :

$$\begin{aligned} \big(\sigma_{\tau_A(\ell)}\tau_A\big)(x) &= 2\,\mathtt{proj}_{\tau_A(\ell)}\big(\tau_A(x)\big) && (17.301a) \\ &= 2\tau_A\big(\mathtt{proj}_\ell(x)\big) - \tau_A(x) && (17.301b) \\ &= 2\,\mathtt{proj}_\ell(x) + 2A - x - A && (17.301c) \\ &= 2\,\mathtt{proj}_\ell(x) - x + A && (17.301d) \\ &= \tau_A\big(\sigma_\ell(x)\big). && (17.301e) \end{aligned}$$

□

17.9.4 Rotations

Définition 17.113 (Rotation en dimension 2)**.**
*Une **rotation** d'un espace euclidien de dimension 2 est une composée de deux réflexions d'axes non parallèles. L'identité est une rotation.*

17.114.
Quelques remarques à propos de cette définition.

(1) Attention : nous ne parlons pas encore de rotations « vectorielles » : ici le centre de la rotation (que nous n'avons pas encore défini) peut ne pas être 0.

(2) Dans la même veine : plus tard, lorsque nous saurons que les rotations sont des isométries de (E, d) où $d(X, Y) = \|X - Y\|$, nous allons en réalité beaucoup plus souvent parler de rotations centrées en l'origine qu'en un point quelconque. C'est pourquoi à partir de 17.125 nous dirons le plus souvent « rotation » pour « rotation centrée en 0 ». D'où les énoncés comme « les rotations sont les matrice orthogonales » (corolaire 17.133), qui *stricto senus* de la définition 17.113 sont faux.

(3) Une rotation est composée de deux réflexions d'axes non parallèles. Il est cependant trop tôt pour décréter que l'intersection de ces axes est le centre de la rotation. Rien ne dit en effet pour l'instant que deux décompositions différentes de la même rotation, avec des axes différents donnent le même point d'intersection.

(4) Pourquoi ajouter l'identité ? Pour avoir un groupe. Dans le cas vectoriel, il est suffisant de demander d'être une composée de deux réflexions, parce que toutes les réflexions vectorielles ont des axes qui s'intersectent en 0. Le cas des axes parallèles est seulement le cas des axes confondus et revient à l'identité.

Si nous voulons avoir un groupe même pour les rotations centrées ailleurs qu'en zéro, nous devons ajouter « à la main » l'identité.

Toutes ces remarques se résument par : « tout devient compliqué du fait que nous voulons considérer également les rotations centrées ailleurs qu'en zéro ». En se contentent du cas vectoriel, de nombreuses choses sont plus simples.

Corolaire 17.115.
Si $A \neq B$ dans E alors il existe une unique réflexion envoyant A sur B.

Démonstration. En ce qui concerne l'existence, la réflexion dont l'axe est la médiatrice de $[A, B]$ fait l'affaire. En ce qui concerne l'unicité, le lemme 17.108 nous dit que si A est envoyé sur B, l'axe est forcément la médiatrice de $[A, B]$. □

Lemme-définition 17.116 ([204])**.**
Soit une rotation $r = \sigma_1 \circ \sigma_2$ différente de l'identité.

(1) Elle admet un unique point fixe.

(2) Ce point fixe est l'intersection des axes $\ell_1 \cap \ell_2$.

*Le **centre** d'une rotation (autre que l'identité) est cet unique point fixe.*

Démonstration. Nous nommons $O = \ell_1 \cap \ell_2$. Soit $A \in E$, et supposons que $r(A) = A$. Nous avons $\sigma_1 \circ r = s_2$ et donc

$$\sigma_1(A) = (\sigma_1 \circ r)(A) = s_2(A). \tag{17.302}$$

On pose $B = \sigma_1(A)$. Alors σ_1 et σ_2 envoient tout deux A sur B.

Si $A = B$ alors A est fixé par σ_1 et donc appartient à ℓ_1. Même chose pour A est fixé par σ_2 et donc $A \in \ell_2$. Cela donne $A = B = O$, et donc le point fixé par r est O.

Si $A \neq B$ alors il existe une unique réflexion envoyant A sur B (corolaire 17.115). L'unicité signifie que $\sigma_1 = \sigma_2$. Dans ce cas, $r = \sigma_1 \circ \sigma_2 = \mathrm{Id}$. □

17.117.
La rotation $\sigma_1 \circ \sigma_2$ laisse évidemment fixé le point $\ell_1 \cap \ell_2$. Si $\sigma_1 \circ \sigma_2 = \sigma_a \circ \sigma_b$ alors rien n'oblige les axes de σ_1 et σ_2 d'être identiques à ceux que σ_a et σ_b. Mais l'intersection $\ell_1 \cap \ell_2$ doit être la même que l'intersection $\ell_a \cap \ell_b$ parce que c'est l'unique point fixé par la composée. Cela nous permet de poser les définitions suivante.

Lemme 17.118.
Les rotation sont des isométries pour la distance : $\|X - Y\| = \|r(X) - r(Y)\|$.

Démonstration. Si $r = \sigma_1 \circ \sigma_2$, en utilisant le fait que σ_1 et σ_2 sont des isométries de (E, d) (17.104) nous avons :

$$d(X, Y) = d\big(\sigma_2(X), \sigma_2(Y)\big) = d\big(\sigma_1\sigma_2(X), \sigma_1\sigma_2(Y)\big) = d\big(r(X), r(Y)\big). \tag{17.303}$$

□

Ce lemme nous dit qu'une rotation de centre O vérifie $\|OX\| = \|Or(X)\|$ pour tout X.

Proposition 17.119 ([204])**.**
Soient $A, B, O \in E$ tels que $\|AO\| = \|BO\| \neq 0$. Alors il existe une unique rotation r centrée en O telle que $r(A) = B$.

Problèmes et choses à faire

Attention : la preuve qui suit contient de nombreuses galipettes et improvisations personnelles. Relisez-la attentivement avant de la prendre pour argent comptant.

La difficulté tient essentiellement à ce que cette preuve traite de façon vectorielle (tous les points sont des éléments de E) un énoncé qui est essentiellement affine : tous les points doivent être vus comme vecteurs partant de O.

Si vous comparez la preuve donnée ici avec celle de [204], vous remarquerez que dans ce dernier, seule la partie « A et O son alignés » est présente. C'est parce que lui, il se met directement dans le cas vectoriel et $O = 0$. Il a donc une preuve un tout petit peu moins générale, mais au moins ses isométries sont linéaires et non affines.

Démonstration. Existence et unicité séparément.

Existence Si $A = B$, l'identité fait l'affaire. Sinon, $\|A-O\| = \|B-O\|$ implique que la médiatrice de $[A, B]$ contient O. Soit σ_m la réflexion selon cette médiatrice. La rotation $\sigma_m \circ \sigma_{(AO)}$ convient.

Unicité Soit r une rotation de centre O et telle que $r(A) = B$. Si $A = B$ alors $r = \mathrm{Id}$ parce qu'une rotation autre que l'identité ne fixe que son centre par le lemme 17.116. Nous supposons que $A \neq B$.

Nous posons $g = \sigma_m \circ r$. Alors $g(A) = \sigma_m(B) = A$ parce que $\sigma_m(B) = A$ et $r(A) = B$. Cela signifie que g est une isométrie qui fixe A.

Si A et O ne sont pas alignés Attention : ici O est un point de E, pas le zéro de l'espace vectoriel E. Lorsqu'on dit que A et O ne sont pas alignés, nous parlons bien d'alignement avec le zéro de E.

Nous avons $g(A) = A$ et $g(O) = O$. Donc g coïncide avec $\sigma_{(AO)}$ en deux points non alignés, c'est-à-dire en deux points pour lesquels l'espace engendré est tout E. Nous en déduisons que $g = \sigma_{(AO)}$.

Si A et O sont alignés Soit maintenant un point C tel que $A - O \perp C - O$ et

$$\|OC\| = \|OA\| = \|OB\|. \tag{17.304}$$

Vu que g est une isométrie pour la distance sur E, pas pour la norme, nous ne pouvons pas écrire $g(C-O) \perp g(A-O)$ à partir de $C - O \perp A - O$. Nous décomposons $g(X) = s(X) + G$ où s est linéaire sur E. Il est vite vu que s est une isométrie de $(E, \|.\|)$:

$$\|X - Y\| = \|g(X) - g(Y)\| = \|s(X) + G - s(y) - G\| = \|s(X) - s(Y)\| = \|s(X - Y)\| \tag{17.305}$$

pour tout $X, Y \in E$. Nous avons de plus $g(A) = A$ et $g(O) = O$, ce qui donne $O = s(O) + G$ et $A = s(A) + G$. En égalisant les valeurs de G nous avons

$$O - s(O) = A - s(A). \tag{17.306}$$

Vu que s est une isométrie (une vraie) nous avons

$$s(A - O) \perp s(C - O), \tag{17.307}$$

mais $s(A - O) = s(A) - s(O) = A - O$ par (17.306). Donc

$$A - O \perp s(C - O). \tag{17.308}$$

Nous en concluons que $s(C - O) = \pm(C - O)$. Parce que les vecteurs $\pm(C - O)$ sont les deux seuls de norme $\|AO\| = \|CO\|$ à être perpendiculaire à $A - O$. Rappel : la définition de C et le fait que nous soyons en dimension 2.

Est-il possible d'avoir $s(C - O) = C - O$? Cela donnerait

$$g(A - O) = s(A) - s(O) + G = s(A) - O + O - s(O) + G = A - O + G \tag{17.309a}$$
$$g(C - O) = C - O + G, \tag{17.309b}$$

ce qui signifierait que g et τ_G coïncideraient sur les points $A - O$ et $C - O$, et donc seraient égaux par le corolaire 17.101. Cela est cependant impossible parce que g fixe au moins les points A et O alors que la translation ne fixe aucun point. Nous en déduisons $s(C - O) = -(C - O)$.

Nous avons aussi, parce que (AO) est une droite passant par l'origine que

$$\sigma_{(AO)}(A - O) = A - O \tag{17.310}$$

et parce que $C - O$ est perpendiculaire à cette droite :

$$\sigma_{(AO)}(C - O) = -(C - O). \tag{17.311}$$

Nous avons donc quand même que g et $\sigma_{(AO)}$ coïncident sur deux points non alignés : $A - O$ et $C - O$.

Dans tous le cas, $g = \sigma_{(AO)}$. Nous avons donc

$$\sigma_{(OA)} = \sigma_m \circ r, \tag{17.312}$$

et donc r est fixé à

$$r = \sigma_m \circ \sigma_{(OA)}. \tag{17.313}$$

□

17.120.
Anticipons un peu et faisons semblant de déjà connaitre les matrices et les fonctions trigonométriques. La proposition 17.119 nous dit qu'il existe une seule rotation amenant A sur B. Vous pourriez objecter que le point $(1,0)$ peut être amené sur $(0,-1)$ soit par la rotation d'angle $3\pi/2$, soit par celle d'angle $-\pi/2$. Il n'en est rien parce que ces deux rotations sont les mêmes ! Pensez-y. En tant qu'application $\mathbb{R}^2 \to \mathbb{R}^2$, la rotation $R_{3\pi/2}$ est égale à $R_{-\pi/2}$.

Une rotation donnée peut être écrite de beaucoup de façons comme composée de deux réflexions. En fait d'autant de façons qu'il y a de réflexions.

Proposition 17.121 ([204]).
Soit une rotation r de E centrée en O. Pour toute réflexion σ_ℓ telle que le centre de r soit sur ℓ, il existe une réflexion σ_1 tells que $r = \sigma_1 \circ \sigma_\ell$. Il existe aussi une réflexion σ_2 telle que $r = \sigma_\ell \circ s_2$.

Démonstration. Si $r = \mathrm{Id}$ c'est bon avec $s_1 = s_2 = \sigma_\ell$. Sinon nous considérons $A \neq O$ sur ℓ, et $B = r(A)$. Nous savons que $B \neq A$ parce que O est le seul point de E fixé par r (proposition 17.116). Il existe une réflexion (unique) σ_1 faisant $\sigma_1(A) = B$, et c'est le réflexion dont l'axe est la médiatrice de $[A, B]$. Le point O est sur cette médiatrice parce que les rotations sont des isométries de (E, d) (lemme 17.118).

La rotation $\sigma_1 \circ \sigma_\ell$ vérifie

$$(\sigma_1 \circ \sigma_\ell)(A) = \sigma_1(A) = B. \tag{17.314}$$

Or $\|OA\| = \|OB\|$, donc il y a unicité de la rotation centrée en O portant A sur B (proposition 17.119) ; nous avons donc $r = \sigma_1 \circ \sigma_\ell$.

En ce qui concerne $r = \sigma_\ell \circ \sigma_2$, il faut appliquer ce que nous venons de faire à la rotation r^{-1} : il existe σ_2 tel que $r^{-1} = \sigma_2 \circ \sigma_\ell$, ce qui donne

$$r = \sigma_\ell \circ \sigma_2. \tag{17.315}$$

□

17.9.5 Rotation d'un angle donné

Lemme-définition 17.122.
Soit $\theta \in \mathbb{R}$. Nous considérons l'application linéaire $R_0(\theta) \colon \mathbb{R}^2 \to \mathbb{R}^2$ dont la matrice dans la base canonique est

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}. \tag{17.316}$$

(1) Nous avons

$$R_0(\theta) = \sigma_\ell \circ s \tag{17.317}$$

où s est la réflexion d'axe horizontal et ℓ est la droite

$$\ell = \mathbb{R} \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix}. \tag{17.318}$$

(2) L'application $R_0(\theta)$ est une rotation autour de $(0,0)$.

(3) Si $A \in \mathbb{R}^2$, alors l'application

$$R_A(\theta) = \tau_A \circ R_0(\theta) \circ \tau_A^{-1} \tag{17.319}$$

est une rotation autour de A nommée ***rotation d'angle*** *θ.*

Démonstration. Nous allons prouver l'égalité (17.317) en calculant les deux membres sur les vecteurs $\begin{pmatrix}1\\0\end{pmatrix}$ et $\begin{pmatrix}1\\0\end{pmatrix}$ qui forment une base.

Pour $p = (1,0)$ D'abord la chose facile[22] : $s(1,0) = (1,0)$.

Pour calculer $\sigma_\ell(1,0)$, nous utilisons le lemme 17.103 ; nous commençons par chercher la projection orthogonale q de $p = (1,0)$ sur ℓ. Nous posons

$$q = \lambda \begin{pmatrix}\cos(\theta/2)\\ \sin(\theta/2)\end{pmatrix} \tag{17.320}$$

et nous cherchons λ satisfaisant $q \cdot (q-p) = 0$. Un peu de calculs passant par (17.8) nous donne

$$q \cdot (q-p) = \lambda\big(\lambda - \cos(\theta/2)\big). \tag{17.321}$$

Les deux solutions sont $\lambda = 0$ et $\lambda = \cos(\theta/2)$. Mais la solution $\lambda = 0$ revient à dire que la droite ℓ est verticale, c'est-à-dire $\cos(\theta/2) = 0$. Donc la solution est toujours donnée par

$$\lambda = \cos(\theta/2). \tag{17.322}$$

Nous introduisons cette valeur dans (17.320) pour fixer q, et nous utilisons la formule du lemme 17.103 :

$$\sigma_\ell \begin{pmatrix}1\\0\end{pmatrix} = 2q - p = \begin{pmatrix}2\cos^2(\theta/2) - 1\\ 2\cos(\theta/2)\sin(\theta/2)\end{pmatrix} = \begin{pmatrix}\cos(\theta)\\ \sin(\theta)\end{pmatrix}. \tag{17.323}$$

Nous avons utilisé le formules trigonométrique de duplication d'angle (corolaire 17.12).

Pour $p = (0,1)$ Cette fois $s(p) = (0,-1)$ et l'équation pour déterminer λ est

$$0 = q \cdot (q-p) = \begin{pmatrix}\lambda\cos(\theta/2)\\ \lambda\sin(\theta/2)\end{pmatrix} \cdot \begin{pmatrix}\lambda\cos(\theta/2)\\ \lambda\sin(\theta/2) - 1\end{pmatrix}. \tag{17.324}$$

Nous trouvons $\lambda\big(\lambda + \sin(\theta/2)\big) = 0$. Le cas $\lambda = 0$ signifie que la droite ℓ est horizontale et donc que $\sin(\theta/2) = 0$. Donc la solution est dans tous les cas

$$\lambda = -\sin(\theta/2). \tag{17.325}$$

Nous trouvons

$$\sigma_\ell \begin{pmatrix}0\\-1\end{pmatrix} = 2q - p = \begin{pmatrix}-\sin(\theta)\\ \cos(\theta)\end{pmatrix}. \tag{17.326}$$

Le calcul de $\sigma_\ell \circ s$ étant fait sur une base, il est facile de reconstituer la matrice

$$\begin{pmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{pmatrix}. \tag{17.327}$$

Cette matrice étant celle, par définition, de $R_0(\theta)$, nous avons montré que $R_0(\theta)$ était bien une rotation. Nous avons prouvé les points (1) et (2).

22. Ici comme partout dans le Frido nous ne faisons aucune différence entre (a,b) et $\begin{pmatrix}a\\b\end{pmatrix}$; ce sont seulement deux façons différentes d'écrire le même élément de $\mathbb{R}^2$. Nous ne faisons pas du semblant de croire que l'un ou l'autre serait un « covecteur » suivant que l'on tourne notre page dans un sens ou un autre.

Nous passons maintenant au point (3). Il est facile de voir que A est un point fixe de $R_A(\theta)$ parce que $\tau_A^{-1}(A) = (0,0)$.

Nommons ℓ' la droite horizontale $\mathbb{R}(1,0)$. Nous avons, par le point précédent $R_0(\theta) = \sigma_\ell \circ \sigma_{\ell'}$. En introduisant astucieusement $\tau_A^{-1}\tau_A$ dans l'expression définissant $R_A(\theta)$, nous avons

$$R_A(\theta) = \tau_A\sigma_\ell\sigma_{\ell'}\tau_A^{-1} = \underbrace{\tau_A\sigma_\ell\tau_A^{-1}}_{=\sigma_{\tau_A(\ell)}}\underbrace{\tau_A\sigma_{\ell'}\tau_A^{-1}}_{\sigma_{\tau_A(\ell')}} = \sigma_{\tau_A(\ell)}\sigma_{\tau_A(\ell')}. \tag{17.328}$$

Nous avons utilisé le lemme 17.112.

Nous voyons que $R_A(\theta)$ est une composée de deux réflexions se coupant en A. C'est donc une rotation centrée en A.

□

17.9.6 Rotations vectorielles

L'expression « rotation vectorielle » signifie rotation centrée en zéro. Elles sont « vectorielles » parce qu'elles sont linéaires comme nous le voyons à présent.

Proposition 17.123.
Quelques résultats à propos de rotations.

(1) Toutes les rotations de $\mathbb{R}^2$ centrées en 0 sont de la forme $R_0(\theta)$ pour un $\theta \in \mathbb{R}$.

(2) Les rotations de $\mathbb{R}^2$ centrées en 0 sont des applications linéaires.

(3) Si r est une rotation dans $\mathbb{R}^2$, il existe $A \in \mathbb{R}^2$ et $\theta \in \mathbb{R}$

$$r = \tau_A \circ R_0(\theta) \circ \tau_A^{-1}. \tag{17.329}$$

Démonstration. Soit r une rotation centrée en $(0,0)$. La proposition 17.121, il existe une droite ℓ passant par $(0,0)$ telle que telle que $r = \sigma_\ell \circ s$ où s est la réflexion d'axe horizontal : $s(x,y) = (x,-y)$.

Soit un vecteur unitaire $v \in \ell$. Vu que $v \in S^1$, la proposition 17.46 nous donne $t \in [0, 2\pi[$ tel que

$$v = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}. \tag{17.330}$$

En posant $\theta = 2t$ nous avons

$$\ell = \mathbb{R}\begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix}, \tag{17.331}$$

et donc $r = R_0(\theta)$ qui est l'application linéaire définie en 17.122.

Et enfin nous voyons le point (3). Soit A l'unique point fixe de la rotation r. Cette dernière s'écrit alors $r = \sigma_{\ell_1} \circ \sigma_{\ell_2}$ où ℓ_1 et ℓ_2 sont des droites telles que $\ell_1 \cap \ell_2 = \{A\}$.

En utilisant le lemme 17.112, nous avons $\sigma_{\ell_i} = \tau_A \circ \sigma_{\tau_A^{-1}(\ell_i)} \circ \tau_A^{-1}$. En substituant cela et en nous rendant compte que $\tau_A^{-1}\tau_A = \mathrm{Id}$ nous avons

$$r = \sigma_{\ell_1} \circ \sigma_{\ell_2} = \tau_A r_0 \tau_1^{-1} \tag{17.332}$$

où r_0 est la rotation $\sigma_{\tau_A^{-1}(\ell_1)} \circ \sigma_{\tau_A^{-1}}(\ell_2)$. Cette dernière est une rotation autour de $(0,0)$ parce que $\tau_A^{-1}(\ell_1) \cap \tau_A^{-1}(\ell_2) = \{0\}$. Elle est donc, par le point (1), de la forme $R_0(\theta)$ pour une certaine valeur de θ.

□

Proposition 17.124 ([204]).
Les rotations basées en O forment un groupe abélien.

Démonstration. L'identité est une rotation par définition. En ce qui concerne l'inverse, si $r = \sigma_1\sigma_2$ alors $r^{-1} = \sigma_2\sigma_1$. Nous commençons maintenant les choses pas tout à fait évidentes.

Composition Soient des rotations r, r' centrées en O. Soit également une réflexion σ dont l'axe contient O. Alors la proposition 17.121 nous donne l'existence de σ_1 et σ_2 tels que $r = \sigma_1\sigma$ et $r' = \sigma\sigma_2$. Avec ça, la composition donne

$$rr' = \sigma_1\sigma\sigma\sigma_2 = \sigma_1\sigma_2, \tag{17.333}$$

qui est encore une rotation.

Commutativité Soient deux rotations r et r' ainsi que des décompositions $r = \sigma_1\sigma$, $r' = \sigma\sigma_2$. Nous avons

$$rr' = \sigma_1\sigma_2 \tag{17.334a}$$
$$r'r = \sigma\sigma_2\sigma_2\sigma. \tag{17.334b}$$

Vu que $t = \sigma_2\sigma_1$ est une rotation nous pouvons encore appliquer la proposition 17.121 pour avoir $t = \sigma_2\sigma_1 = \sigma\sigma_3$. Avec ça,

$$r'r = \sigma\sigma\sigma_3\sigma = \sigma_3\sigma. \tag{17.335}$$

Mais aussi $rr' = \sigma_1\sigma_2 = t^{-1} = \sigma_3\sigma$. Nous avons donc bien $rr' = r'r$, et le groupe est commutatif.

□

17.125.
Jusqu'à présent nous avons parlé de rotations « affines ». Parmi elles, les rotations centrées en 0 (zéro, l'origine de E comme espace vectoriel) sont de particulière importance. Ce sont des applications linéaires, et même des isométries. Dans la suite, nous allons souvent dire simplement « rotation » pour dire « rotation centrée en 0 ».

Vu que nous allons maintenant prendre un point de vue plus vectoriel, nous allons noter les points de E avec des lettres comme x, y, u, v et plus avec des majuscules, comme quand on avait un point de vue afin. En même temps, nous allons noter les applications $E \to E$ par des lettres comme A et ne plus écrire les parenthèses. Bref, nous écrivons Au au lieu de $r(A)$.

Lemme 17.126.
En dimension 2*, les réflexions vectorielles (c'est-à-dire dont l'axe passe par* 0*) ont un déterminant* -1.

Démonstration. Soit une réflexion d'axe ℓ. Prenons une base orthonormale de E constituée de e_1 sur ℓ et de $e_2 \perp \ell$. Alors $\sigma_\ell(e_1) = e_1$ et $\sigma_\ell(e_2) = -e_2$. La formule du déterminant donne

$$\det(\sigma_\ell) = e_1^*\big(\sigma_\ell(e_1)\big)e_1^*\big(\sigma_\ell(e_2)\big) - e_2^*\big(\sigma_\ell(e_1)\big)e_1^*\big(\sigma_\ell(e_2)\big) = 1 \times (-1) - 0 \times 0 = -1. \tag{17.336}$$

Nous utilisons de façon cruciale le fait que le calcul du déterminant ne dépende pas de la base choisie, lemme 9.8. □

Proposition 17.127.
Les rotations[23] *sont*

(1) des applications linéaires orthogonales au sens de la définition 9.39,

(2) des applications de déterminant 1*,*

Démonstration. Le fait qu'elles soient linéaires est la proposition 17.123.

Nous avons, pour tout $u \in E$ l'égalité de la norme $\|u\|$ et $\|Au\|$ par le lemme 17.118 appliqué à $Y = 0$. En termes de produits scalaires nous avons alors $\langle Au, Au\rangle = \langle u, u\rangle$, et donc

$$\langle A^*Au, u\rangle = \|u\|^2. \tag{17.337}$$

23. Centrées en 0, nous ne le répéterons pas !

En particulier si $\{e_i\}_{i=1,\dots,n}$ est une base orthonormée de E nous avons

$$(A^*Ae_i)_i = \|e_i\|^2 = 1, \tag{17.338}$$

ce qui donne $\|A^*Ae_i\| \geqslant 1$, avec égalité si et seulement si $A^*Ae_i = e_i$. Ici nous avons utilisé le fait que $\langle x, e_i\rangle = x_i$, et le fait que pour tout i nous ayons $\|x\| \geqslant |x_i|$, avec égalité seulement si x est un multiple de e_i.

Par ailleurs l'inégalité de Cauchy-Schwarz 7.191 nous donne

$$\|u\|^2 = |\langle A^*Au, u\rangle| \leqslant \|A^*Au\|\|u\| \tag{17.339}$$

et donc

$$\|u\| \leqslant \|A^*Au\|. \tag{17.340}$$

Encore une fois, en appliquant cela à $u = e_i$ nous trouvons $1 \leqslant \|A^*Ae_i\|$. Vu que nous avions déjà l'inégalité dans l'autre sens, $\|A^*Ae_i\| = 1$. Et le cas d'égalité est uniquement possible avec $A^*Ae_i = e_i$.

Donc pour tout i de la base nous avons $A^*Ae_i = e_i$. Nous avons donc $A^*A = \mathrm{Id}$ et l'application A est orthogonale.

En ce qui concerne le déterminant, les réflexions sont de déterminant -1 par le lemme 17.126, donc $A = \sigma_1 \circ \sigma_2$ est de déterminant 1. Nous avons utilisé le fait que le déterminant était un morphisme : proposition 9.10(1). □

Remarque 17.128.
Nous ne savons pas encore que les rotations forment tout le groupe SO(2) des endomorphismes orthogonaux de déterminant 1. Il faudra attendre le corolaire 17.133 pour le savoir.

Lemme 17.129.
L'application $-\mathrm{Id}$ *est une rotation de* $\mathbb{R}^2$.

Démonstration. Soit une base orthonormée $\{e_1, e_2\}$ de E et la rotation $r = \sigma_1\sigma_2$ où σ_i est la réflexion le long de l'axe $\ell_i = \{te_i\}_{t\in\mathbb{R}}$. Faut-il vous prouver que $r = -\mathrm{Id}$? La réflexion σ_2 retourne la composante y d'un vecteur écrit dans la base $\{e_1, e_2\}$ sans toucher à la composante x. La réflexion σ_1 fait le contraire. □

17.9.7 Matrice des transformations orthogonales

Nous donnons maintenant une forme générale (trigonométrique) pour les matrices de SO(2). Nous ne pouvons cependant pas invoquer les lemmes 17.130 ou 17.131 pour prétendre avoir une matrice des rotations, parce que nous n'avons pas encore prouvé que les rotations étaient des transformations orthogonales. Ce sera pour la proposition 17.134.

Lemme 17.130.
Si $A \in \mathrm{O}(2)$ *alors il existe un unique* $\theta \in [0, 2\pi[$ *et un unique* $\epsilon = \pm 1$ *tels que*

$$A = \begin{pmatrix} \cos(\theta) & -\epsilon\sin(\theta) \\ \sin(\theta) & \epsilon\cos(\theta) \end{pmatrix} \tag{17.341}$$

Démonstration. Soit une matrice $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ et imposons qu'elle soit dans O(2). Le fait que A soit orthogonale impose

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2+c^2 & ab+cd \\ ab+cd & b^2+d^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{17.342}$$

Nous avons alors le système

$$\begin{cases} a^2 + b^2 = 1 & \text{(17.343a)} \\ b^2 + d^2 = 1 & \text{(17.343b)} \\ ab + cd = 0 & \text{(17.343c)} \end{cases}$$

La proposition 17.46 nous permet de déduire qu'il existe un unique $\theta \in [0, 2\pi[$ tel que $a = \cos(\theta)$, $c = \sin(\theta)$, ainsi que plusieurs $\alpha \in \mathbb{R}$ tel que $b = \cos(\alpha)$, $d = \sin(\alpha)$.

Note : si nous voulons $\alpha \in [0, 2\pi[$, alors il y a unicité. Ici nous ne nous attachons pas à cette contrainte ; nous savons qu'il en existe plusieurs, et nous allons en fixer un en fonction de θ. Le α ainsi fixé ne sera peut-être pas dans $[0, 2\pi[$, mais ce ne sera pas grave.

Les angles θ et α sont alors liés par la contrainte

$$\cos(\theta)\cos(\alpha) + \sin(\theta)\sin(\alpha) = 0. \tag{17.344}$$

Utilisant l'identité (17.22c) cela signifie que $\cos(\theta - \alpha) = 0$. Donc

$$\alpha \in \{\theta + \frac{\pi}{2} + k\pi\}_{k\in\mathbb{Z}}. \tag{17.345}$$

Si k est pair, ça donne

$$\cos(\alpha) = -\sin(\theta) \tag{17.346a}$$
$$\sin(\alpha) = \cos(\theta) \tag{17.346b}$$

et alors

$$A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}. \tag{17.347}$$

Si au contraire k est impair, alors

$$\cos(\alpha) = \sin(\theta) \tag{17.348a}$$
$$\sin(\alpha) = -\cos(\theta), \tag{17.348b}$$

et

$$A = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}. \tag{17.349}$$

Nous avons démontré qu'une matrice de $\mathrm{O}(2)$ était forcément d'une des deux formes (17.347) ou (17.349). Il est maintenant facile de vérifier que ces deux matrices sont effectivement dans $\mathrm{O}(2)$. □

Lemme 17.131.
Tout élément de $\mathrm{SO}(2)$ *s'écrit (dans la base canonique) de façon unique sous la forme*

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{17.350}$$

avec $\theta \in [0, 2\pi[$.

Démonstration. Vu que $\mathrm{SO}(2)$ est la partie de $\mathrm{O}(2)$ constitué des matrices de déterminant 1, nous pouvons reprendre la forme donnée par le lemme 17.130 et fixer ϵ par la contrainte sur le déterminant.

Nous avons, en utilisant la relation du lemme 17.3,

$$\det \begin{pmatrix} \cos(\theta) & -\epsilon\sin(\theta) \\ \sin(\theta) & \epsilon\cos(\theta) \end{pmatrix} = \epsilon, \tag{17.351}$$

et donc il faut et suffit de fixer $\epsilon = 1$. □

Corolaire 17.132 ([1])**.**
Nous avons une bijection

$$\begin{aligned} \psi\colon \mathrm{SO}(2) &\to [0, 2\pi[\\ \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} &\mapsto \theta, \end{aligned} \tag{17.352}$$

et un isomorphisme de groupe

$$\begin{aligned} \varphi\colon \mathrm{SO}(2) &\to \mathrm{U}(1) = \{e^{i\theta}\}_{\theta\in\mathbb{R}} \\ \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} &\mapsto e^{i\theta}. \end{aligned} \tag{17.353}$$

Démonstration. La première assertion est une paraphrase du lemme 17.131. Pour la seconde, il faut vérifier que c'est bien un morphisme et une bijection.

Pour le morphisme, ce sont les formules d'addition d'angle du lemme 17.11 qui jouent. En ce qui concerne la bijection...

Surjection Vu que $e^{i\theta+2ki\pi} = e^{i\theta}$, tout élément de $\mathrm{U}(1)$ est exponentielle de $i\theta$ pour un $\theta \in [0, 2\pi[$.

Injection Nous avons $\varphi(A) = \varphi(B)$ lorsque les formes matricielles de A et B sous forme trigonométrique sont avec des angles différents d'un multiple de 2π. Vu que les fonctions trigonométriques sont périodiques, nous avons $A = B$ parce que leurs matrices sont égales.

□

17.9.8 Rotations, SO(2) et matrice de rotation

Corolaire 17.133 ([1]).
Le groupe des rotations centrées en $(0, 0)$ *est le groupe* $\mathrm{SO}(2)$.

Démonstration. Nous devons prouver deux choses :

— Toutes les rotations sont des éléments de $\mathrm{SO}(2)$.

— Tous les éléments de $\mathrm{SO}(2)$ sont des rotations.

La proposition 17.123 nous indique que toute rotation de $\mathbb{R}^2$ centrée en $(0, 0)$ est de la forme $R_0(\theta)$, c'est-à-dire a une matrice de la forme

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}. \tag{17.354}$$

Donc toute rotation est dans $\mathrm{SO}(2)$.

D'autre part, le lemme 17.131 indique que tout élément de $\mathrm{SO}(2)$ a, dans la base canonique, une matrice de la forme (17.354). Le lemme 17.122 nous indique alors que c'est une rotation. □

Proposition 17.134.
Si r est une rotation de $\mathbb{R}^2$ centrée en $(0, 0)$, il existe un unique $\theta \in [0, 2\pi[$ tel que $r = R_0(\theta)$.

Démonstration. Soit une rotation r autour de $(0, 0)$. Le corolaire 17.133 nous dit qu'il existe $\theta \in \mathbb{R}$ tel que

$$r = R_0(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}. \tag{17.355}$$

Nous avons identifié l'application linéaire à sa matrice. L'élément $\big(\cos(\theta), \sin(\theta)\big)$ est dans S^1, et il existe donc, par la proposition 17.46, un unique $t \in [0, 2\pi[$ tel que $\cos(t) = \sin(\theta)$ et $\sin(t) = \sin(\theta)$. Pour ce t nous avons alors

$$r = R_0(\theta) = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix}. \tag{17.356}$$

□

Proposition 17.135.
Nous avons la formule suivante pour la composition :

$$R_0(\alpha) \circ R_0(\beta) = R_0(\alpha + \beta). \tag{17.357}$$

Démonstration. Par définition de $R_0(\theta)$, dans la base canonique de $\mathbb{R}^2$, la composition se calcule avec le produit suivant, en utilisant les formules du lemme 17.11 :

$$\begin{pmatrix}\cos(\alpha) & -\sin(\alpha)\\ \sin(\alpha) & \cos(\alpha)\end{pmatrix}\begin{pmatrix}\cos(\beta) & -\sin(\beta)\\ \sin(\beta) & \cos(\beta)\end{pmatrix} \tag{17.358a}$$

$$= \begin{pmatrix}\cos(\alpha)\cos(\beta)-\sin(\alpha)\sin(\beta) & -\cos(\alpha)\sin(\beta)-\sin(\alpha)\cos(\beta)\\ \sin(\alpha)\cos(\beta)+\cos(\alpha)\sin(\beta) & -\sin(\alpha)\sin(\beta)+\cos(\alpha)\cos(\beta)\end{pmatrix} \tag{17.358b}$$

$$= \begin{pmatrix}\cos(\alpha+\beta) & -\sin(\alpha+\beta)\\ \sin(\alpha+\beta) & \cos(\alpha+\beta)\end{pmatrix}. \tag{17.358c}$$

Donc dans la base canonique, la matrice de $R_0(\alpha)R_0(\beta)$ est celle de $R_0(\alpha+\beta)$. □

17.9.9 Rotation et application affine

Nous considérons à nouveau la définition 8.9 d'une application affine ainsi que sa décomposition en application linéaire et translation donnée par le lemme 8.10. Nous voyons maintenant comment ces choses se mettent dans le cas d'une rotation non centrée en l'origine.

Exemple 17.136.
Soit $A \in \mathbb{R}^2$ ainsi qu'une rotation f autour de A, c'est-à-dire une composition de deux symétries dont les axes se coupent en A [24]. Nous allons extraire de f la partie linéaire définie en 8.10.

Il existe des axes ℓ_1 et ℓ_2 tels que $\ell_1 \cap \ell_2 = \{A\}$ et tels que

$$f = s_{\ell_1} \circ s_{\ell_2}. \tag{17.359}$$

En utilisant le lemme 17.112,

$$f = t_A \circ s_{t_A^{-1}(\ell_1)} \circ t_A^{-1} \circ t_A \circ s_{t_A^{-1}(\ell_2)} \circ t_A^{-1} = t_A \circ s_{t_A^{-1}(\ell_1)} \circ s_{t_A^{-1}(\ell_2)} \circ t_A^{-1}. \tag{17.360}$$

Vu que $A \in \ell_i$, nous avons $O = (0,0) \in t_A^{-1}(\ell_i)$. Donc les axes $t_A^{-1}(\ell_1)$ et $t_A^{-1}(\ell_2)$ se coupent en O et nous pouvons écrire

$$f = t_A \circ R \circ t_A^{-1} \tag{17.361}$$

où R est une rotation centrée en O ; donc une application linéaire par la proposition 17.123.

Nous avons

$$f(x) = (t_A \circ R \circ t_A^{-1})(x) \tag{17.362a}$$
$$= (t_A \circ R)(x - A) \tag{17.362b}$$
$$= t_A\big(R(x) - R(A)\big) \tag{17.362c}$$
$$= R(x) + t_{R(A)+A} \tag{17.362d}$$
$$= (t_{R(A)+A} \circ R)(x). \tag{17.362e}$$

Donc

$$f = t_{R(A)+A} \circ R. \tag{17.363}$$

Il est maintenant aisé de montrer que R est la partie linéaire de f. Pour tout $M, x \in \mathbb{R}^2$ nous avons

$$f(M+x) = R(M+x) + R(A) + A \tag{17.364a}$$
$$= R(M) + R(x) + R(A) + A \tag{17.364b}$$
$$= R(x) + f(M). \tag{17.364c}$$

Donc ok pour la formule

$$f(M+x) = R(x) + f(M) \tag{17.365}$$

et R est la partie linéaire de f, voir la définition 8.10. △

24. Voir la définition 17.113.

Lemme 17.137 ([1]).
Tout sous-groupe fini de SO(2) *est cyclique.*

Démonstration. Soit uns sous-groupe fini G de SO(2). Nous savons que SO(2) est isomorphe à U(1) par le corolaire 17.132, et en bijection avec $[0, 2\pi[$. Vu que G est fini, l'ensemble $G\backslash\{e\}$ il possède, dans $[0, 2\pi[$ un élément minimum non nul. Soit g_0 ce minimum.

Soit un élément g_1 de G qui ne serait ni l'identité ni un multiple de g_0. En particulier tous les nombres du type $g_1 - kg_0$ sont dans G (l'image de G dans $[0, 2\pi[$ en fait). Si g_1 n'est pas un multiple de g_0, il n'en reste pas moins que $g_1 = \lambda g_0$; alors en prenant pour k la partie entière de λ, l'élément $g_1 - kg_0$ est plus petit que g_0. Contradiction. □

17.9.10 Angle entre deux droites

Avant d'aborder la classification des isométries, nous devons parler de l'angle entre deux droites. Si ℓ_1 et ℓ_2 sont deux droites, alors il est bien clair deux angles peuvent prétendre être « l'angle entre ℓ_1 et ℓ_2 ». De plus chacun de ces deux angles sont doubles parce que si α peut prétendre être l'angle entre ℓ_1 et ℓ_2, alors $-\alpha$ peut également prétendre.

Remarque 17.138.
Nous ne parlons pas de l'angle entre ℓ_1 et ℓ_2 mais bien de l'angle *de* ℓ_1 *à* ℓ_2. L'ordre des droites est important.

17.139.
Pour la suite, $R_O(\alpha)$ est la rotation d'angle α autour du point O tandis que $R(\alpha)$ est la rotation d'angle α autour de l'origine.

Proposition 17.140 ([361]).
Si u et v sont des vecteurs unitaires[25] *de $\mathbb{R}^2$ alors il existe une unique rotation*[26] *f telle que $f(u) = v$.*

Démonstration. C'est la proposition 17.119 appliquée à $O = (0, 0)$. □

Remarque 17.141.
Notons l'unicité. Nous ne faisons pas de différences entre R_θ et $R_{\theta+2\pi}$ et les autres $R_{\theta+2k\pi}$. En particulier si une rotation T est donnée, dire « $T = R_\theta$ » ne définit pas un nombre θ de façon univoque. Par contre ça définit une classe modulo 2π, c'est-à-dire un élément $\theta \in \mathbb{R}/2\pi$.

Nous avons déjà défini le groupe SO(2) en la définition 9.43 et nous avons déterminé ses matrices dans $\mathbb{R}^2$ en le lemme 17.131.

17.9.11 Angle orienté

La proposition 17.140 donne une application

$$T\colon S^1 \times S^1 \to \mathrm{SO}(2). \tag{17.366}$$

Et nous avons une relation d'équivalence sur $S^1 \times S^1$ donnée par $(u, v) \sim (u', v')$ si et seulement si il existe $g \in \mathrm{SO}(2)$ telle que $g(u) = u'$ et $g(v) = v'$.

Définition 17.142 (Angle orienté[361]).
Les classes de $S^1 \times S^1$ pour cette relation d'équivalence sont les ***angles orientés de vecteurs****. Nous notons $[u, v]$ la classe de (u, v).*

Proposition 17.143.
Nous avons $T(u, v) = T(u', v')$ si et seulement si $(u, v) \sim (u', v')$.

25. De norme 1.
26. Définition 17.113.

Démonstration. En utilisant la commutativité du groupe SO(2) nous avons équivalence entre les affirmations suivantes :

— $(u,v) \sim (u',v')$
— $T(u,u') = T(v',v')$
— $T(u,u') \circ T(u',v) = T(v,v') \circ T(u',v)$
— $T(u,v) = T(u',v')$.

□

Proposition 17.144.
Nous avons une bijection

$$\begin{aligned} S\colon \frac{S^1 \times S^1}{\sim} &\to \mathrm{SO}(2) \\ [u,v] &\mapsto T(u,v). \end{aligned} \tag{17.367}$$

Démonstration. En plusieurs points.

S est bien définie En effet si $[u,v] = [z,t]$ alors $T(u,v) = T(z,t)$.

Injectif Si $S[u,v] = S[z,t]$ alors $T(u,v) = T(z,t)$, qui implique $(u,v) \sim (z,t)$ par la proposition 17.143.

Surjectif Nous avons $R_\theta = T(u, R_\theta u)$.

□

Définition 17.145 (Somme d'angles orientés[361])**.**
Si $[u,v]$ et $[z,t]$ sont des angles orientés, nous définissons la somme par

$$[u,v] + [z,t] = S^{-1}\Big(S[u,v] \circ S[z,t]\Big). \tag{17.368}$$

Lemme 17.146.
Quelques propriétés des angles plats liées à la somme.

(1) $(S^1 \times S^1)/\sim$ *est un groupe commutatif.*
(2) Relations de Chasles :

$$[u,v] + [v,w] = [u,w]. \tag{17.369}$$

(3) $-[u,v] = [v,u]$.

Démonstration. Pour la relation de Chasles, ça se base sur la propriété correspondante sur T :

$$[u,v] + [v,w] = S^{-1}\Big(T(u,v) \circ T(v,w)\Big) \tag{17.370a}$$
$$= S^{-1}\big(T(u,w)\big) \tag{17.370b}$$
$$= [u,w]. \tag{17.370c}$$

Pour l'inverse, la vérification est que

$$[u,v] + [v,u] = [u,u] = 0. \tag{17.371}$$

□

Définition 17.147.
*La **mesure** de l'angle orienté $[u,v]$ est $[\theta]_{2\pi}$ si $T[u,v] = R_\theta$.*

Notons dans cette définition qu'écrire $T[u,v] = R_\theta$ dans SO(2) ne définit pas θ, mais seulement sa classe modulo 2π. C'est pour cela que la mesure de l'angle orienté n'est également définie que modulo 2π.

Pour la suite nous allons nous intéresser à des vecteurs qui ont, dans l'idée, un point de départ et un point d'arrivée. Si $A, B \in \mathbb{R}^2$ nous notons

$$\overrightarrow{AB} = \frac{B - A}{\|B - A\|}. \tag{17.372}$$

C'est le vecteur unitaire dans la direction « de B vers A ».

Théorème 17.148 (Théorème de l'angle inscrit[362])**.**
Soit un cercle Γ de centre O et trois points distincts $A, B, M \in \Gamma$. Alors

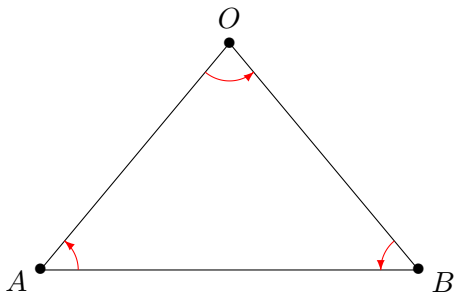
$$2(\overrightarrow{MA}, \overrightarrow{MB}) \in (\overrightarrow{OA}, \overrightarrow{OB})_{2\pi} \tag{17.373}$$

où l'indice 2π indique la classe modulo 2π.

Démonstration. Le triangle MOA est isocèle en O, donc les angles à la base sont égaux. Et de plus la somme des angles est dans $[\pi]_{2\pi}$. Bon, entre nous, nous savons que la somme des angles est exactement π, mais comme nous n'avons pas défini les angles autrement que modulo π, nous ne pouvons pas dire mieux. Donc

$$2(\overrightarrow{AB}, \overrightarrow{AO}) + (\overrightarrow{OB}, \overrightarrow{OA}) \in [\pi]_{2\pi}. \tag{17.374}$$

Il faut être sûr de l'orientation de tout cela. Le nombre $(\overrightarrow{AB}, \overrightarrow{AO})$ est l'angle qui sert à amener $\overrightarrow{AB}$ sur $\overrightarrow{AO}$. Vu que nous l'avons choisi dans le sens trigonométrique, il faut bien prendre les autres dans le sens trigonométrique et utiliser $(\overrightarrow{OA}, \overrightarrow{OB})$ et non $(\overrightarrow{OB}, \overrightarrow{OA})$.



De la même manière sur le triangle MOB nous écrivons

$$2(\overrightarrow{MB}, \overrightarrow{MO}) + (\overrightarrow{OM}, \overrightarrow{OB}) \in [\pi]_{2\pi}. \tag{17.375}$$

Nous faisons la différence entre les deux équations en remaquant que la différence de deux représentants de $[\pi]_{2\pi}$ est un représentant de $[0]_{2\pi}$ et en en nous souvenant que $-(\overrightarrow{MB}, \overrightarrow{MO}) = (\overrightarrow{MO}, \overrightarrow{MB})$ et les relations de Chasles du lemme 17.146(2) nous avons :

$$2(\overrightarrow{MA}, \overrightarrow{MB}) + (\overrightarrow{OB}, \overrightarrow{OA}) \in [0]_{2\pi}. \tag{17.376}$$

□

17.149.
Comment exprimer le fait qu'un angle orienté soit égal à θ modulo π alors que les angles orientés sont des classes modulo 2π ? Nous ne pouvons certainement pas écrire

$$(u, v) = [\theta]_\pi \tag{17.377}$$

parce que (u, v) est un élément de $S^1 \times S^1$ alors que $[\theta]_\pi$ est un ensemble de nombres. Nous pouvons écrire

$$[u, v] \subset [\theta]_\pi. \tag{17.378}$$

C'est cohérent parce que nous avons des deux côtés des ensembles de nombres. Les opérations permises sont l'égalité ou l'inclusion. L'égalité entre les deux ensembles n'est pas possible parce que la différence minimale ente deux éléments dans $[u, v]$ est 2π alors que celle dans $[\theta]_\pi$ est π.

Si u et v forment un angle droit, nous avons

$$[u,v] = \{\frac{\pi}{2} + 2k\pi\}_{k\in\mathbb{Z}}. \tag{17.379}$$

Et cela est bien un sous-ensemble de $[\pi/2]_\pi$.

Pour exprimer que la différence entre deux angles orientés diffèrent de π nous devrions écrire :

$$[u,v] \subset [a,b]_\pi \tag{17.380}$$

où le membre de droite signifie la classe modulo π d'un représentant de $[a,b]$.

Nous allons cependant nous permettre d'écrire

$$[u,v] = [a,b]_\pi \tag{17.381}$$

voire carrément

$$(u,v) = (a,b)_\pi. \tag{17.382}$$

Cette dernière égalité devant être comprise comme voulant dire que l'angle pour passer de u à v est soit le même que celui pour alle de a à b soit ce dernier plus π.

Théorème 17.150 ([362]).
Soient 4 points distincts du plan A, B, C, D. Ils sont alignés ou cocycliques[27] *si et seulement si*

$$(\overrightarrow{CA}, \overrightarrow{CB}) = (\overrightarrow{DA}, \overrightarrow{DB})_\pi. \tag{17.383}$$

Nous allons seulement démontrer l'implication directe.

Démonstration. Si les quatre points sont alignés nous avons $[\overrightarrow{CA}, \overrightarrow{CB}] = [0]_{2\pi}$ et $[\overrightarrow{DA}, \overrightarrow{DB}] = [0]_{2\pi}$. En particulier nous avons

$$[\overrightarrow{CA}, \overrightarrow{CB}] = [\overrightarrow{DA}, \overrightarrow{DB}] \tag{17.384}$$

et a fortiori l'égalité modulo π au lieu de 2π.

Nous nous relâchons en termes de notations. Si les quatre points sont cocycliques, nous pouvons utiliser le théorème de l'angle inscrit 17.148 dans les triangles ABC et ADB :

$$2(\overrightarrow{CA}, \overrightarrow{CB}) = (\overrightarrow{OA}, \overrightarrow{OB})_{2\pi} \tag{17.385a}$$
$$2(\overrightarrow{DA}, \overrightarrow{DB}) = (\overrightarrow{OA}, \overrightarrow{OB})_{2\pi}, \tag{17.385b}$$

ce qui donne $2(\overrightarrow{CA}, \overrightarrow{CB}) = 2(\overrightarrow{DA}, \overrightarrow{DB})_{2\pi}$ et donc

$$(\overrightarrow{CA}, \overrightarrow{CB}) = (\overrightarrow{DA}, \overrightarrow{DB})_\pi. \tag{17.386}$$

Comme annoncé, nous ne faisons pas la preuve dans l'autre sens ; elle peut être trouvée dans [362]. □

Exemple 17.151.
À propos de groupe engendré et de générateur[28]. Soit G le groupe des rotations d'angle[29] $k\pi/5$ (avec k entier). Ce groupe est constitué des « dixièmes de tour », puisque $\frac{k\pi}{5} = \frac{2k\pi}{10}$.

La rotation d'angle $2\pi/5$ n'est pas génératrice parce qu'elle n'engendre que des « cinquièmes de tour » : $4\pi/5$, $6\pi/5$,$8\pi/5$ et l'identité.

Par contre, la rotation d'angle $\pi/5$ est génératrice. △

27. C'est-à-dire sur un même cercle.
28. Définition 2.11 et 2.12
29. Voir la définition 17.147.

17.9.12 Angles et nombres complexes

Les nombres complexes peuvent être repérés par une norme et un angle, ce qui en fait un terrain propice à l'utilisation des angles orientés. Nous en ferons d'ailleurs usage dans $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ pour parler d'alignement, de cocyclicité et de birapport dans la proposition 22.85.

Soient deux éléments $z_1, z_2 \in \mathbb{C}$. Nous les écrivons sous la forme $z_1 = r_1 e^{i\theta_1}$ et $z_2 = r_2 e^{i\theta_2}$; remarquons que cela ne définit θ_i qu'à 2π près. Nous avons

$$[z_1, z_2] = [\theta_2 - \theta_1]_{2\pi}. \tag{17.387}$$

Soient maintenant $a, b, c, d \in \mathbb{C}$. Nous écrivons $\overrightarrow{ab}$ le vecteur unitaire dans le sens « de a vers b », c'est-à-dire un multiple positif bien choisi du nombre $b - a$. Nous notons θ_{ab} l'argument du nombre complexe $b - a$, et nous avons encore

$$[\overrightarrow{ab}, \overrightarrow{cd}] = [\theta_{ab} - \theta_{cd}]. \tag{17.388}$$

Avec toutes ces notations, ce qui est bien est que les produits et quotients de nombres complexes se comportent très bien par rapport aux angles : l'argument de a/b est $\theta_a - \theta_b$ et en particulier l'argument de

$$\frac{a-b}{c-d} \tag{17.389}$$

est dans la classe de l'angle orienté

$$[\overrightarrow{ba}, \overrightarrow{dc}]. \tag{17.390}$$

Définition 17.152.
Soient trois points $A, O, B \in \mathbb{R}^2$. Voici comment nous définissons l'angle $\widehat{AOB}$; informelement c'est l'angle de la rotation qui fait aller de A vers B.

- *Nous nous mettons en l'origine : $A' = A - O$ et $B' = B - O$.*
- *Nous normalisons : $A'' = A'/\|A'\|$ et $B'' = B'/\|B'\|$.*
- *Soit f, l'unique rotation telle que $f(A'') = B''$ (proposition 17.140).*
- *Soit θ l'unique élément de $[0, 2\pi[$ tel que la matrice de f dans la base canonique soit*

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{17.391}$$

  par la proposition 17.134.
- *L'angle $\widehat{AOB}$ est ce nombre.*

Nous voyons que l'angle est toujours un nombre entre 0 et 2π. Par abus de notation, nous admettrons de temps en temps de parler d'angle en-dehors de cet intervalle.

Lemme-définition 17.153.
Si ℓ_1 et ℓ_2 sont deux droites de $\mathbb{R}^2$ sécantes au point O et si $x \in \ell_1$ n'est pas O, alors il existe un unique $\alpha \in [0, \pi[$ tel que $R_O(\alpha)x \in \ell_2$. La valeur de α ne dépend pas du choix du point $x \in \ell_1$.

*Cet angle α est l'**angle** de ℓ_1 à ℓ_2.*

Proposition 17.154 ([1]).
Les angles sont invariants sous les translations.

Plus précisément, si $A, B, S, v \in \mathbb{R}^2$, alors

$$\widehat{T_v(A)T_v(S)T_v(B)} = \widehat{ASB} \tag{17.392}$$

où $T_v(X) = X + v$.

Démonstration. Nous notons $X_v = X + v$. Nous avons $A'_v = A_v - S_v = (A+v) - (S+v) = A - S = A''$. Donc les vecteurs A'' et B'' à partir desquels est calculé $\widehat{ASB}$ sont les mêmes que les vecteurs A''_v et B''_v qui servent à calculer $\widehat{T_v(A)T_v(S)T_v(B)}$. □

Proposition 17.155 ([1]).
Les angles sont invariants par rotations, c'est-à-dire que si $A, B, S \in \mathbb{R}^2$ et si R_θ est une rotation, alors

$$\widehat{ASB} = \widehat{R_\theta(A)R_\theta(S)R_\theta(B)}. \tag{17.393}$$

Démonstration. Pour être plus concis, nous écrivons A_θ for $R_\theta(A)$ et de même pour B et S. Afin de calculer l'angle $\widehat{R_\theta(A)R_\theta(S)R_\theta(B)}$, nous définissons

$$\begin{cases} A'_\theta = A_\theta - S_\theta & (17.394a) \\ S'_\theta = 0 & (17.394b) \\ B'_\theta = B_\theta - S_\theta. & (17.394c) \end{cases}$$

et

$$\begin{cases} A''_\theta = \dfrac{A_\theta - S_\theta}{\|A_\theta - S_\theta\|} & (17.395a) \\ B''_\theta = \dfrac{B_\theta - S_\theta}{\|B_\theta - S_\theta\|}. & (17.395b) \end{cases}$$

Par définition, l'angle est le α tel que $R_\alpha(A''_\theta) = B''_\theta$. Nous devons prouver que le même α vérifie $R_\alpha(A'') = B''$.

Le fait que R_θ soit une isométrie nous donne déjà

$$\|R_\theta(A) - R_\theta(B)\| = \|A - B\|. \tag{17.396}$$

Ensuite, la relation de définition de α s'écrit

$$\frac{R_\alpha R_\theta(A) - R_\alpha R_\theta(S)}{\|A_\theta - S_\theta\|} = \frac{R_\theta(B) - R_\theta(S)}{\|B_\theta - S_\theta\|}. \tag{17.397}$$

Vu que R_α et R_θ commutent, nous avons

$$R_\theta \frac{R_\alpha(A) - R_\alpha(B)}{\|A - S\|} = R_\theta \frac{B - S}{\|B - S\|}, \tag{17.398}$$

et comme R_θ est inversible, cela donne $R_\alpha(A'') = B''$. □

Lemme 17.156 ([1]).
Soit $A \in \mathbb{R}^2$ et une droite ℓ_1. Soit ℓ_2 une droite passant par A et intersectant ℓ_1 en O. Alors

$$\sigma_{\ell_1}(A) = R_O(-2\alpha)A \tag{17.399}$$

où α est l'angle de ℓ_1 à ℓ_2.

Démonstration. Nous allons utiliser des coordonnées autour de O. Il existe un vecteur v tel que

$$A = O + v \tag{17.400}$$

Par définition de l'angle α [30], la droite ℓ_2 s'obtient par rotation d'angle α depuis la droite ℓ_1. Donc le point

$$B = R_O(-\alpha)A \tag{17.401}$$

est sur ℓ_1.

Nous allons prouver que le point

$$D = R_O(-2\alpha)A \tag{17.402}$$

est $D = \sigma_{\ell_1} A$.

30. Définition 17.153.

Nous commençons par montrer que la droite (DA) est perpendiculaire à ℓ_1, c'est-à-dire que

$$(D-A)\cdot(B-O)=0. \tag{17.403}$$

En utilisant le fait que

$$R_O(\alpha)(O+X)=O+R(\alpha)X, \tag{17.404}$$

nous avons

$$D-A=R_O(-2\alpha)(O+v)-(O+v)=O+R(-2\alpha)v-O-v=R(-2\alpha)v-v \tag{17.405}$$

et de la même façon,

$$B-O=R(-\alpha)v. \tag{17.406}$$

Notons que tous les O se sont simplifiés et qu'il ne reste que des rotations usuelles. En utilisant le fait que $R(\alpha)$ est une isométrie, nous pouvons alors calculer

$$(D-A)\cdot(B-O)=\langle R(-2\alpha)v-v,R(-\alpha)v\rangle \tag{17.407a}$$
$$=\langle R(-\alpha)v-R(\alpha)v,v\rangle. \tag{17.407b}$$

En utilisant la matrice de rotation du lemme 17.131 nous trouvons

$$\big(R(-\alpha)-R(\alpha)\big)v=\begin{pmatrix}2\sin(\alpha)v_2\\-2\sin(\alpha)v_1\end{pmatrix} \tag{17.408}$$

et donc

$$\big\langle\big(R(-\alpha)-R(\alpha)\big)v,v\big\rangle=0. \tag{17.409}$$

Le point D est bien sûr la droite perpendiculaire à ℓ_1 et passant par A. Mais vu que D est obtenu à partir de A par une rotation, le point D est également sur le cercle de rayon $\|OA\|$ et centré en O. Ce cercle possède exactement deux intersections avec cette droite. Le premier est A et le second est $\sigma_{\ell_1}(A)$. Vu que D n'est pas A, nous avons $D=\sigma_\ell(A)$. □

17.9.13 Classification

Théorème 17.157 ([354]).
Toute isométrie du plan $(\mathbb{R}^2,d)$ est une composition d'au plus 3 réflexions.

Démonstration. Encore une fois nous décomposons la preuve en fonction du nombre de points fixes.

Si f n'a pas de points fixes Soit $x\in\mathbb{R}^2$. Nous considérons le segment $[f,f(x)]$ et nous nommons l sa médiatrice. Par construction, $f(x)=\sigma_l(x)$. Nous posons $g=\sigma_l\circ f$, et nous avons

$$g(x)=x. \tag{17.410}$$

Donc nous avons $f=\sigma_l\circ g$ avec $x\in\mathrm{Fix}(g)$.

Si f a un unique point fixe Soit x cet unique point fixe. Soit $y\neq x$ et l la médiatrice de $[y,f(y)]$. En posant $g=\sigma_l\circ f$ nous avons

$$g(y)=y \tag{17.411}$$

et $g(x)=x$ parce que

$$d\big(x,f(y)\big)=d\big(f(x),f(y)\big)=d(x,y), \tag{17.412}$$

ce qui donne que x est à égale distance de y et de $f(y)$, c'est-à-dire que $x\in l$ et par conséquent $g(x)=(\sigma_l\circ f)(x)=\sigma_l(x)=x$.

Donc g fixe x et y et donc toute la droite (xy).

Si f fixe une droite Soit l une droite fixée par f, et soient $x, y \in l$ et $z \notin l$ (avec $x \neq y$). Le fait que x et y soient des points fixes de f implique

$$\begin{cases} d\big(x, f(z)\big) = d(x, z) & (17.413a) \\ d\big(y, f(z)\big) = d(y, z) & (17.413b) \end{cases}$$

ce qui signifie que $f(z)$ est sur l'intersection des deux cercles[31] $S\big(x, d(x, z)\big)$ et $S\big(y, d(y, z)\big)$, et comme ce sont deux cercles centrés sur la droite l, les intersections sont liées par σ_l. Autrement dit, les intersections sont z et $\sigma_l(z)$.

Si $f(z) = z$ alors f fixe trois points non alignés et fixe dont $\mathbb{R}^2$, c'est-à-dire $f = \mathrm{Id}$.

Si par contre $f(z) = \sigma_l(z)$ alors les isométries f et σ_l coïncident sur trois points et coïncident donc partout par le corolaire 17.101 : $f = \sigma_l$.

Conclusion Nous avons montré que si $\mathrm{Fix}(f)$ a dimension m, alors il existe une droite pour laquelle $f = \sigma_l \circ g$ avec $\dim\big(\mathrm{Fix}(g)\big) > m$. Donc il faux au maximum trois pas pour avoir $\dim\big(\mathrm{Fix}(g)\big) = 2$ c'est-à-dire pour avoir $g = \mathrm{Id}$.

□

Définition 17.158.
*Une **réflexion glissée** est une transformation du plan de la forme $\tau_v \circ \sigma_\ell$ où le vecteur v est parallèle à la droite ℓ.*

Théorème 17.159 ([354]).
Les isométries du plan $(\mathbb{R}^2, d)$ sont exactement

(1) l'identité (composée de 0 réflexions),

(2) les réflexions,

(3) les translations (composées de 2 translations d'axes parallèles),

(4) les rotations (composées de 2 réflexions d'axes non parallèles),

(5) les réflexions glissées (composées de 3 réflexions)

Démonstration. Nous savons déjà que $f \in \mathrm{Isom}(\mathbb{R}^2)$ est une composée de 0, 1, 2 ou 3 réflexions.

Zéro réflexions Alors c'est l'identité. Ce n'est pas très profond.

Une réflexion Alors f est une réflexion. Toujours pas très profond.

Deux réflexions Soit $f = \sigma_{\ell_1} \circ \sigma_{\ell_2}$. Maintenant ça s'approfondit un bon coup.

Nous supposons d'abord que $\ell_1 \parallel \ell_2$. Dans ce cas nous allons prouver que $f = \tau_{2v}$ où v est le vecteur perpendiculaire à ℓ_1 tel que $\ell_1 + v = \ell_2$. Nous allons utiliser le lemme 17.110 pour montrer que $\sigma_{\ell_1} \circ \sigma_{\ell_2} = \tau_{2v}$. Nous avons

$$\ell_1 = \ell_0 + w \tag{17.414a}$$
$$\ell_2 = \ell_0 + w + v \tag{17.414b}$$

où w est un vecteur perpendiculaire à ℓ_1 et ℓ_0 est la droite passant par l'origine et parallèle à ℓ_1 et ℓ_2. Avec cela,

$$\begin{aligned} (\sigma_{\ell_1} \circ \sigma_{\ell_2})(x) &= \sigma_{\ell_1}\big(\sigma_{\ell_0}(x) + 2w\big) && (17.415a) \\ &= \sigma_{\ell_0}\big(\sigma_{\ell_0}(x) + 2w\big) + 2(v + w) && (17.415b) \\ &= x + \underbrace{\sigma_{\ell_0}(2w)}_{-2w} + 2v + 2w && (17.415c) \\ &= x + 2v. && (17.415d) \end{aligned}$$

Donc si f est composée de deux réflexions d'axes parallèles, alors f est une translation.

31. L'intersection existe pare que $d(x, z) + d(y, z) > d(x, y)$.

Toujours dans le cas où f est composée de deux réflexions, nous supposons que $f = \sigma_{\ell_2} \circ \sigma_{\ell_1}$ avec ℓ_1 et ℓ_2 non parallèles. Nous notons O le point d'intersection, et nous allons voir que $f = R_O(2\alpha)$ où α est l'angle de ℓ_1 à ℓ_2 donné par le lemme 17.153.

Soit $x \in \ell_1$. Alors

$$f(x) = \sigma_{\ell_2}(x), \tag{17.416}$$

et le lemme 17.156 nous donne un moyen de calculer $\sigma_{\ell_2}(x)$ parce que ℓ_1 est une droite passant par x et coupant ℓ_1 au point O. Le lemme dit que $\sigma_{\ell_2}(x) = R_O(2\alpha)$. Remarque : c'est bien 2α et non -2α parce qu'il s'agit de l'angle de ℓ_2 à ℓ_2 ; il y a inversion des numéros entre ici et l'énoncé du lemme.

Nous avons donc bien $f(x) = R_O(2\alpha)x$ pour $x \in \ell_1$.

Si $y \in \ell_2$ alors

$$f(y) = \sigma_{\ell_2}\big(R_O(-2\alpha)y\big) \tag{17.417}$$

Nous posons $z = \sigma_{\ell_1}(y) = R_O(-2\alpha)y$. Soit la droite ℓ_3 passant par O et z. Vu que $R_O(2\alpha)z = y \in \ell_2$, l'angle de ℓ_3 à ℓ_2 est 2α. Par conséquent

$$\sigma_{\ell_2}(z) = R_O\big(-2 \times (-2\alpha)\big)z = R_O(4\alpha)z = R_O(4\alpha)R_O(-2\alpha)y = R_O(2\alpha)y. \tag{17.418}$$

Donc les transformations f et $R_O(2\alpha)$ coïncident pour tous les points des droites ℓ_1 et ℓ_2, qui ne sont pas parallèles. Cela prouve que $f = R_O(2\alpha)$.

Trois réflexions Nous écrivons $f = \sigma_{\ell_3} \circ \sigma_{\ell_2} \circ \sigma_{\ell_1}$. Nous allons transformer cela progressivement en une symétrie glissée en passant par plusieurs étapes :

(1) $f = \sigma_\ell \circ \tau_v$,

(2) $f = \tau_v \circ \sigma_\ell$,

(3) $f = \tau_v \circ \sigma_\ell$ avec $v \parallel \ell$.

À chacune de ces étapes, v et ℓ vont changer. La dernière est une réflexion glissée.

Nous commençons par supposer $\ell_2 \parallel \ell_3$. Dans ce cas, $\sigma_{\ell_3} \circ \sigma_{\ell_2}$ est une translation, comme nous l'avons déjà vu. Alors $f = \tau_v \circ \sigma_{\ell_1}$ et nous sommes déjà dans le cas (2).

Nous supposons que ℓ_2 n'est pas parallèle à ℓ_3. Dans ce cas, si $O = \ell_2 \cap \ell_3$ nous avons

$$\sigma_{\ell_3} \circ \sigma_{\ell_2} = R_O(2\alpha) \tag{17.419}$$

où α est l'angle de ℓ_2 à ℓ_3. En réalité tant que l'angle de ℓ'_3 à ℓ'_2 est α nous avons

$$\sigma_{\ell'_3} \circ \sigma_{\ell'_2} = \sigma_{\ell_3} \circ \sigma_{\ell_2} = R_O(2\alpha). \tag{17.420}$$

Nous choisissons ℓ'_2 parallèle à ℓ_1, de telle sorte à ce que $\sigma_{\ell'_2} \circ \sigma_{\ell_1}$ soit une translation. Alors nous avons

$$f = \sigma_{\ell_3} \circ \sigma_{\ell_2} \circ \sigma_{\ell_1} = \sigma_{\ell_3} \circ \sigma_{\ell'_2} \circ \sigma_{\ell'_1} = \sigma_{\ell_3} \circ \tau_v. \tag{17.421}$$

où v est le vecteur de la translation en question.

Nous avons donc prouvé que toute composition de trois réflexions peut être écrite soit sous la forme (1) soit sous la forme (2).

Nous prouvons à présent que toute transformation de la forme (1) peut être écrite sous la forme (2). Plus précisément nous allons prouver que si ℓ est une droite, v un vecteur et ℓ_0 la droite parallèle à ℓ passant par l'origine, alors

$$\sigma_\ell \circ \tau_v = \tau_{\sigma_{\ell_0}(v)} \circ \sigma_l \tag{17.422}$$

D'abord nous savons que $\sigma_\ell(x) = \sigma_{\ell_0}(x) + 2w$ où w est le vecteur tel que $\ell = \ell_0 + w$. Ensuite c'est un simple calcul utilisant le fait que σ_{ℓ_0} est linéaire :

$$(\sigma_\ell \circ \tau_v)(x) = \sigma_l(x + v) = \sigma_{\ell_0}(x) + \sigma_{\ell_0}(v) + 2w, \tag{17.423}$$

et

$$(\tau_{\sigma_{\ell_0}(v)} \circ \sigma_\ell)(x) = \sigma_{\ell_0}(v) + \sigma_\ell(x) = \sigma_{\ell_0}(v) + \sigma_{\ell_0}(x) + 2w. \tag{17.424}$$

L'égalité est faite.

Nous montrons maintenant que toute transformation de la forme (2) peut être mise sous la forme (3). Soit donc $f = \tau_v \circ \sigma_\ell$ où v et ℓ ne sont pas spécialement parallèles.

Pour cela nous décomposons $v = v_1 + v_2$ avec $v_1 \perp \ell$ et $v_2 \parallel \ell$ et nous posons $\ell' = \ell + \frac{1}{2}v_1$. Nous montrons que

— $\tau_v \circ \sigma_\ell = \tau_{v_2} \circ \sigma_{\ell'}$

— $v_2 \parallel \ell'$.

Pour le deuxième point, $v_2 \parallel \ell$ et bien entendu $\ell' \parallel \ell$. Donc $v_2 \parallel \ell'$.

Soit ℓ_0 la droite parallèle à ℓ et ℓ' et passant par l'origine. Soit aussi le vecteur w tel que $\ell = \ell_0 + w$. Alors nous avons

$$\begin{cases} \sigma_\ell = \sigma_{\ell_0} + 2w & (17.425a) \\ \sigma_{\ell'} = \sigma_{\ell_0} + 2w + v_1 & (17.425b) \end{cases}$$

Nous avons

$$(\tau_v \circ \sigma_\ell)(x) = v + \sigma_{\ell_0}(x) + 2w \tag{17.426}$$

et

$$\begin{aligned} (\tau_{v_2} \circ \sigma_{\ell'})(x) &= v_2 + \sigma_{\ell_0}(x) + 2w + v_1 && (17.427a) \\ &= \sigma_{\ell_0}(x) + v + 2w && (17.427b) \end{aligned}$$

où dans la dernière ligne, nous avons regroupé $v_1 + v_2 = v$. Et voilà.

□

17.9.14 Classification des isométries de $\mathbb{R}$

Définition 17.160.
*Soit $x \in \mathbb{R}$; nous notons σ_x la **réflexion** par rapport à x, c'est-à-dire*

$$\sigma_x(y) = 2x - y. \tag{17.428}$$

Théorème 17.161 ([354])**.**
Toute isométrie de $\mathbb{R}$ est composée d'au plus 2 réflexions. Plus précisément toute isométrie de $\mathbb{R}$ est dans une des trois catégories suivantes :

— *l'identité (0 réflexions),*

— *les réflexions,*

— *les translations (2 réflexions)*

Démonstration. Nous divisions la preuve en fonction du nombre de points fixés par l'isométrie $f \in \mathrm{Isom}(\mathbb{R})$.

f fixe deux points distincts Alors elle fixe l'espace affine engendrée par ces deux points par la proposition 17.100. Donc f fixe tout $\mathbb{R}$ et est l'identité.

f fixe un unique point Soit x l'unique point fixé par f et considérons $y \neq x$. Vu que $x = f(x)$ et que f est une isométrie,

$$d\big(x, f(y)\big) = d\big(f(x), f(y)\big) = d(x, y). \tag{17.429}$$

Donc $f(y)$ est à égale distance de x que y. Autrement dit, $f(y)$ est soit y soit $\sigma_x(y)$. Mais comme x est unique point fixe, $f(y) = \sigma_x(y)$. Ce raisonnement étant valable pour tout $y \neq x$ nous avons $f = \sigma_x$.

f n'a pas de points fixes Soient $x \in \mathbb{R}$ et $y = \frac{x+f(x)}{2}$. Nous posons $g = \sigma_y \circ f$. Alors x est un point fixe de g parce que

$$g(x) = \sigma_y\big(f(x)\big) = 2y - f(x) = x. \tag{17.430}$$

Donc soit g est l'identité soit g est une réflexion (par les points précédents). La possibilité $g = \mathrm{Id}$ est exclue parce que cela ferait $f = \sigma_y$ alors que f n'a pas de points fixes. Donc g est une réflexion ; et comme x est un point fixe de g nous avons $g = \sigma_x$. Au final

$$f = \sigma_y \circ \sigma_x. \tag{17.431}$$

Montrons que cela implique que f est une translation :

$$\sigma_y\sigma_x(z) = \sigma_y(2x - z) = 2y - 2x + z = z + 2(y - x). \tag{17.432}$$

Donc $\sigma_y \circ \sigma_x$ est la translation de vecteur $2(y - x)$.

□

17.9.15 Isométries du tétraèdre régulier

Définition 17.162.
*Un polyèdre **régulier** est un polyèdre dont les faces sont des polygones réguliers identiques dont tous les sommets joignent le même nombre d'arrêtes.*

Définition 17.163.
*Le **tétraèdre** est une pyramide à base triangulaire dont toutes les faces sont des triangles équilatéraux.*

Proposition 17.164 (Isométries affines du tétraèdre régulier)**.**
Soient T un tétraèdre régulier et $\mathrm{Iso}(T)$ *son groupe d'isométries affines (définition 8.59). Alors*

$$\mathrm{Iso}(T) \simeq S_4 \tag{17.433}$$

où S_4 est le groupe des permutations de quatre objets.

Démonstration. Commençons par prouver qu'une isométrie préserve les sommets : l'image d'un sommet est un sommet. Pour cela nous considérons $g \in \mathrm{Iso}(T)$ et nous supposons que l'image d'un sommet x soit à l'intérieur d'une arrête. Soient $g(y)$ et $g(z)$ deux points distincts de cette arrête situés à égale distance de $g(x)$. Cela est possible parce que g est une bijection de $\mathbb{R}^3$. Aussi : $y \neq z$. Mais une application affine préserve l'alignement (vous ne le croyez pas ? regardez la forme donnée par le lemme (8.54)), donc x, y et z foment un triangle isocèle en x de points alignés et appartenant à T. Cela est impossible si x est un sommet.

Donc l'image d'un sommet est un sommet. Si nous numérotons les sommets $x_1, \dots, x_4$, nous obtenons un morphisme de groupe $\varphi\colon \mathrm{Isom}(T) \to S_4$ qui envoie g sur la permutation qui envoie 1 sur le numéro du sommet $g(x_1)$, 2 sur le numéro du sommet $g(x_2)$, etc.

Le morphisme φ est injectif Supposons $\varphi(g_1) = \varphi(g_2)$. Alors $g_1^{-1} \circ g_2$ est une isométrie de $(\mathbb{R}^3, d)$ qui fixe les quatre sommets. Une application affine $\mathbb{R}^4 \to \mathbb{R}^3$ fixant 4 point est l'identité par le lemme 8.51. Donc $g_1^{-1}g_2 = \mathrm{Id}$, ce qui prouve que $g_1 = g_2$. Vous noterez que nous utilisons l'unicité de l'inverse dans un groupe.

φ est surjectif Nous savons que S_4 est engendré par les transpositions (proposition 2.73). Or les transpositions sont dans l'image de φ. En effet, notons les sommets de notre tétraèdre par A, B, D et D et considérons la transposition $A \leftrightarrow B$. Elle est l'image par φ de la réflexion selon le plan σ, médiateur du segment $[A, B]$. Pour nous assurer de cela, nous devons nous assurer que C et D appartiennent à σ. Cela est le contenu du lemme 17.107.

Conclusion L'application φ est un morphisme bijectif, c'est-à-dire un isomorphisme.

□

17.9.16 Représentation de S_4 via les isométries du tétraèdre

17.165.
Lorsque le tétraèdre a son barycentre en l'origine de $\mathbb{R}^3$, l'isomorphisme $\varphi\colon \mathrm{Iso}(T) \to S_4$ donne une représentation de dimension 3 de S_4. Nous avons calculé les caractères de S_4 en la section 15.5 sans avoir besoin de savoir que l'une des représentations de dimension 3 est cella que nous venons de trouver via le groupe des isométries du tétraèdre. Nous allons cependant également y calculer les caractères de la représentation φ, pour le sport.

Une des représentations trouvées (la représentation ρ_s) peut être vue comme le groupe $\mathrm{Iso}(T)$ des isométries affine du tétraèdre grâce à la proposition 17.164 qui donne un isomorphisme de groupe $S_4 \simeq \mathrm{Iso}(T)$ lorsque T est un tétraèdre régulier de $\mathbb{R}^3$.

Si le barycentre de T est situé à l'origine de $\mathbb{R}^3$, alors les éléments de $\mathrm{Iso}(T)$ sont des applications linéaires parce que

— les affinités laissent invariantes les barycentres (proposition 8.41),

— les affinités qui laissent l'origine invariante sont linéaires (corolaire 8.55).

Nous allons à présent calculer la trace de cette représentation, en utilisant le fait que nous la connaissions explicitement. Nous savons que les caractères sont constants sur les classes de conjugaison ; nous allons donc écrire une matrice par classe de conjugaison (qui sont données dans l'exemple 2.71).

Pour tout cela nous allons considérer un tétraèdre dont le centre (isobary) est en $(0,0,0)$ et une base de $\mathbb{R}^3$ formée de trois sommets e_1, e_2 et e_3. Vu que l'isobarycentre des quatre sommes est en $(0,0,0)$, le quatrième somme est forcément le point de coordonnées $e_4(-1,-1,-1)$, de telle sorte que $e_1+e_2+e_3+e_4=0$.

Les transpositions Quelle isométrie de $\mathbb{R}^3$ permute deux sommets du tétraèdre sans bouger les autres ? Pour permuter les sommets e_1 et e_2 en laissant e_3 et e_4, c'est le symétrie par rapport au plan médiateur de $[e_1,e_2]$. Ce plan passe par les sommets e_3 et e_4, parce que le tétraèdre étant régulier, les points e_3 e_4 sont équidistants de e_1 et e_2. Le lemme 17.107 dit qu'alors ces points dont partie du plan médiateur.

Dans notre base, la matrice de la transposition précédemment nommée (12) est

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{17.434}$$

dont la trace est 1. Donc $\chi_s(12)=1$.

Les bitranspositions La bitransposition $(12)(34)$ est le produit des transpositions selon les plans médiateur de $[e_1,e_2]$ et $[e_3,e_4]$. Ces deux plans sont perpendiculaires, et l'intersection est la droite qui passe par les milieux. Cette droite est perpendiculaire aux deux segments en même temps. La matrice est :

$$\begin{pmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ 0 & 0 & -1 \end{pmatrix} \tag{17.435}$$

parce que $e_1 \mapsto e_2$, $e_2 \mapsto e_1$ et $e_3 \mapsto e_4$. Pour rappel, la matrice est formée des images des vecteurs de base. Cela donne

$$\chi_s\big((12)(34)\big) = -1. \tag{17.436}$$

Les 3-cycles La symétrie qui permute cycliquement les points e_1, e_2 et e_3 est la rotation d'angle[32] $2\pi/3$ dans le plan formé par les extrémités de ces trois vecteurs. Heureusement, la trace est invariante par changement de base ; donc nous pouvons calculer la trace d'une rotation d'angle

32. Angle d'une rotation, définition 17.122.

$2\pi/3$ dans n'importe quelle base. Par exemple :

$$\chi_s\big((12)(34)\big) = \mathrm{Tr}\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(2\pi/3) & \sin(2\pi/3) \\ 0 & -\sin(2\pi/3) & \cos(2\pi/3) \end{pmatrix} = 1 + 2\cos(2\pi/3) = 0. \tag{17.437}$$

Notons que, sans cette interprétation géométrique, nous y arrivons aussi facilement : dans notre base le 3-cycle est $e_1 \mapsto e_2 \mapsto e_3 \mapsto e_1$, donc la matrice est :

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \tag{17.438}$$

dont la trace est manifestement nulle : $\chi_s\big((123)\big) = 0$.

Le 4-cycle Il fait $e_1 \mapsto e_2 \mapsto e_3 \mapsto e_4 \mapsto e_1$, dont la matrice est

$$\begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}, \tag{17.439}$$

et la trace est $\chi_s\big((1,2,3,4)\big) = -1$.

Nous avons retrouvé les caractères de la représentation ρ_s, et nous pouvons vérifier qu'elle est irréductible.

17.10 Transformations de Lorentz

Nous considérons dans cette section un nombre réel $c > 0$ ainsi que l'espace $\mathbb{R}^2$ muni du produit pseudo-scalaire[33] donné par la matrice

$$\eta = \begin{pmatrix} c^2 & 0 \\ 0 & -1 \end{pmatrix}. \tag{17.440}$$

Et pour faire plus vrai, nous notons (x_0, x_1) les coordonnées sur $\mathbb{R}^2$. Ainsi

$$x \cdot y = c^2 x_0 y_0 - x_1 y_1. \tag{17.441}$$

Nous insistons sur le fait que cela n'est pas un produit scalaire.

Lemme 17.166 ([1]).
Soit $c > 0$. L'application

$$\begin{aligned} \varphi\colon]-c, c[&\to \mathbb{R} \\ v &\mapsto \frac{-v/c}{\sqrt{1 - \frac{v^2}{c^2}}} \end{aligned} \tag{17.442}$$

est une bijection.

Démonstration. Nous commençons par mentionner le fait que φ est continue du fait que le dénominateur ne s'annule pas. Une petite étude fonction montre que

$$\lim_{v \to -c} \varphi(v) = \infty, \tag{17.443}$$

et

$$\lim_{v \to c} \varphi(v) = -\infty, \tag{17.444}$$

33. Définition 7.187.

et

$$\varphi'(v) = -\frac{1}{c\sqrt{1-\frac{v^2}{c^2}}} - \frac{v^2/c^3}{\left(1-\frac{v^2}{c^2}\right)^{3/2}} < 0. \tag{17.445}$$

Tout cela fait que φ est bijective (entre autres par le théorème des valeurs intermédiaires 11.50 et la théorème dérivée et croissance 11.125). □

Lemme 17.167.
La forme bilinéaire

$$\begin{aligned} b\colon \mathbb{R}^2 \times \mathbb{R}^2 &\to \mathbb{R} \\ x, y &\mapsto x \cdot y \end{aligned} \tag{17.446}$$

est non dégénérée[34].

Démonstration. Soit $(x_0, x_1) \in \mathbb{R}^2$ tel que

$$b\big((x_0, x_1), (y_0, y_1)\big) = 0 \tag{17.447}$$

pour tout $(y_0, y_1) \in \mathbb{R}^2$. Nous avons

$$c^2 x_0 y_0 - x_1 y_1 = 0. \tag{17.448}$$

En écrivant cela avec $(y_0, y_1) = (1, 0)$ puis $(0, 1)$ nous obtenons immédiatement que $(x_0, x_1) = (0, 0)$. □

Théorème 17.168.
Soit une bijection[35] $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ *telle que*

$$f(x) \cdot f(y) = x \cdot y \tag{17.449}$$

pour tout $x, y \in \mathbb{R}^2$. *Alors :*

(1) f *est linéaire.*

(2) Il existe un unique choix de $(x, \sigma_1, \sigma_2) \in \mathbb{R} \times \{\pm 1\} \times \{\pm 1\}$ *tel que la matrice de* f *ait la forme*

$$f = \begin{pmatrix} \sigma_1 \cosh(\xi) & \frac{\sigma_1\sigma_2}{c} \sinh(\xi) \\ c \sinh(\xi) & \sigma_2 \cosh(\xi) \end{pmatrix}. \tag{17.450}$$

(3) Il existe un unique $v \in \,]-c, c[$ *tel que la matrice de* f *ait la forme*

$$f = \begin{pmatrix} \frac{\sigma_1}{\sqrt{1-\frac{v^2}{c^2}}} & -\frac{\sigma_1\sigma_2}{c^2} \frac{v}{\sqrt{1-\frac{v^2}{c^2}}} \\ \frac{-v}{\sqrt{1-\frac{v^2}{c^2}}} & \frac{\sigma_2}{\sqrt{1-\frac{v^2}{c^2}}} \end{pmatrix}. \tag{17.451}$$

Démonstration. Vu que notre produit pseudo-scalaire est non dégénéré (lemme 17.167), le fait que f soit linéaire est la proposition 9.249. Nous posons

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \tag{17.452}$$

et, conformément à la proposition 9.253 nous imposons $A^t \eta A = \eta$. Après un petit produit matriciel nous obtenons :

$$\begin{pmatrix} c^2\alpha^2 - \gamma^2 & c^2\alpha\beta - \gamma\delta \\ c^2\alpha\beta - \gamma\delta & c^2\beta^2 - \delta^2 \end{pmatrix} = \begin{pmatrix} c^2 & 0 \\ 0 & -1 \end{pmatrix}. \tag{17.453}$$

34. Définition 9.246.
35. À mon avis, il y a moyen d'affaiblir cette hypothèse. Écrivez-moi si vous avez une idée.

Voila quatre équations à résoudre pour les quatre inconnues $\alpha, \beta, \gamma, \delta$. Déjà les équations des termes anti-diagonaux sont les mêmes. Nous recopions le reste :

$$\begin{cases} c^2\alpha^2 - \gamma^2 = c^2 & (17.454a) \\ c^2\alpha\beta - \gamma\delta = 0 & (17.454b) \\ c^2\beta^2 - \delta^2 = 1. & (17.454c) \end{cases}$$

C'est le moment d'utiliser la proposition 14.93. La relation (17.454a) donne

$$\alpha^2 - \left(\frac{\gamma}{c}\right)^2 = 1, \tag{17.455}$$

ce qui implique l'existence (unique) de $\xi_1 \in \mathbb{R}$ et $\sigma_1 \in \{\pm 1\}$ tels que

$$\gamma = c\sinh(\xi_1) \tag{17.456a}$$
$$\alpha = \sigma_1 \cosh(\xi_1). \tag{17.456b}$$

La relation (17.454c) implique quant à elle l'existence de $\xi_2 \in \mathbb{R}$ et $\sigma_2 \in \{\pm 1\}$ tels que

$$\delta = \sigma_2 \cosh(\xi_2) \tag{17.457a}$$
$$\beta = \frac{1}{c}\sinh(\xi_2). \tag{17.457b}$$

Nous substituons maintenant toutes les valeurs (17.456) et (17.457) dans (17.454b). Cela donne

$$\sigma_1 \cosh(\xi_1)\sinh(\xi_2) = \sinh(\xi_1)\cosh(\xi_2). \tag{17.458}$$

Nous mettons cette relation au carré et nous substituons $\cosh(\xi_1)^2 = 1 + \sinh^2(\xi_1)$. Ce que nous trouvons est

$$\sinh(\xi_1)^2 = \sinh(\xi_2)^2, \tag{17.459}$$

qui implique que $\xi_1 = \pm\xi_2$. Nous posons donc $\xi_2 = \sigma_3\xi_1$ pour un certain $\sigma_3 \in \{\pm 1\}$. Cela nous permet d'alléger la notation et d'écrire ξ au lieu de ξ_1.

Nous remettons la valeur $\xi = \xi_1 = \sigma_3\xi_2$ dans l'équation (17.458) en tenant compte du fait que sinh est impaire et cosh est paire :

$$\sigma_1\sigma_3 \cosh(\xi)\sinh(\xi) = \sigma_2 \sinh(\xi)\cosh(\xi). \tag{17.460}$$

Et cela nous enseigne que $\sigma_3 = \sigma_1\sigma_2$.

Jusqu'à présent nous avons prouvé qu'il existe un unique $\xi \in \mathbb{R}$ et $\sigma_1, \sigma_2 \in \{\pm 1\}$ tels que

$$A = \begin{pmatrix} \sigma_1 \cosh(\xi) & \frac{\sigma_1\sigma_2}{c}\sinh(\xi) \\ c\sinh(\xi) & \sigma_2\cosh(\xi) \end{pmatrix}. \tag{17.461}$$

Nous utilisons à présent la bijection du lemme 17.166. Il existe un unique $v \in \left]-v, v\right[$ tel que $\sinh(\xi) = \varphi(v)$. En utilisant $\cosh(\xi)^2 = 1 + \varphi(v)^2$, nous trouvons

$$\cosh(\xi)^2 = \frac{1}{1 - \frac{v^2}{c^2}}. \tag{17.462}$$

Mais comme le cosinus hyperbolique est toujours strictement positif, nous pouvons prendre la racine carrée des deux côtés :

$$\cosh(\xi) = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{17.463}$$

En substituant dans (17.461), nous trouvons le résultat annoncé. □

17.10.1 Sous-groupe fini d'isométries du plan

Théorème 17.169 ([363]).
Soit un groupe fini G d'isométries de $(\mathbb{R}^2, d)$ contenant n éléments.

(1) Il existe un point $C \in \mathbb{R}^2$ fixé par tous les éléments de G.

(2) Si G ne contient pas de réflexions, alors il est cyclique[36] *et engendré par la rotation d'angle $2\pi/n$ autour de C.*

(3) Si G contient au moins une réflexion, et si C est un point fixe de G, alors

(3a) toutes les réflexions ont un axe qui passe par C,

(3b) n est pair,

(3c) Si σ est une réflexion dans G, alors nous avons $G = \operatorname{gr}\big(\sigma, R_C(4\pi/n)\big)$ où $R_C(\theta)$ est la rotation d'angle θ autour de C,

(3d) G est isomorphe au groupe diédral $D_{n/2}$.

Démonstration. Soit un groupe fini G constitué d'isométries de $(\mathbb{R}^2, d)$. Nous prouvons le théorème point par point.

Pour (1) C'est la proposition 17.77.

Questions de réflexions Le théorème 17.157 nous dit que les éléments de G sont des compositions d'au maximum 3 réflexions.

Exclure trois réflexions Il n'est pas possible que G contienne un élément composé de trois réflexions. En effet, les composées de trois réflexions, par le théorème 17.159 sont des réflexions glissées[37], c'est-à-dire des transformations de la forme $g = \tau_v \circ \sigma_\ell$ où v est un vecteur parallèle à la droite ℓ. Si $x \in \ell$, alors

$$g(x) = \tau_v(x) = x + v, \tag{17.464}$$

de telle sorte que $g^k(x) = x + kv$, qui signifie que tous les g^k sont différents. Le groupe G ne peut pas être fini si il contient une réflexion glissée.

G^+ et G^- Pour la même raison que celle qui exclu les réflexions glissées, G ne peut pas contenir de translations. Le théorème 17.159 nous donne la liste des possibilités. Après exclusion des translations et des réflexions glissées, il reste :

— l'identité

— les rotations,

— les réflexions.

Nous notons G^+ la partie de G contenant l'identité et les rotations et G^- celle contenant les réflexions. Notons que G^+ n'est pas vide parce qu'il contient au moins l'identité, tandis que G^- peut être vide, mais n'est certainement pas un groupe.

Même nombre d'éléments Nous prouvons à présent que si G^- est non vide, alors il a le même nombre d'éléments que G. Un élément de G^- est une réflexion. Soit $\sigma \in G^-$. Nous prouvons que

$$\begin{aligned} \varphi\colon G^+ &\to G^- \\ f &\mapsto \sigma \circ f \end{aligned} \tag{17.465}$$

est une bijection.

Surjective Soit $s \in G^-$. Posons $f = \sigma^{-1} \circ s$. Vu que σ^{-1} et s sont des réflexions, f est une rotation. Donc $f \in G^+$ et $\varphi(f) = s$.

Injective La condition $\varphi(f) = \varphi(g)$ dit que $\sigma \circ f = \sigma \circ g$. En composant par σ^{-1} nous obtenons $f = g$.

36. Définition 2.12.
37. Définition 17.158.

$G = \mathrm{gr}\big(R_C(2\pi/p)\big)$ Nous nommons p le nombre d'éléments de G^+. Si G^- est vide, $p = n$, et sinon $p = n/2$. Dans les deux cas, G^+ est un groupe de rotations à p éléments.

Le groupe G^+ contient seulement des rotations ; or le centre d'une rotation est l'unique point fixe. Donc tous les éléments de G^+ sont des rotations autour de C.

Le corolaire 2.35 au théorème de théorème de Lagrange nous indique que tous les éléments de G^+ vérifient $g^p = \mathrm{Id}$. Seules les rotations d'angle $2k\pi/p$ autour de C satisfont la condition $g^p = \mathrm{Id}$. Or il n'y a que p telles rotations. Donc elles sont toutes dans G^+. Nous en déduisons que

$$G^+ = \mathrm{gr}\big(R_C(2\pi/p)\big). \tag{17.466}$$

Pour (2) Dans le cas où G ne contient pas de réflexions, G^- est vide et G contient n éléments. La relation (17.466) devient

$$G = G^+ = \mathrm{gr}\big(R_C(2\pi/n)\big). \tag{17.467}$$

Pour (3) Nous supposons maintenant que G contienne au moins une réflexion. De la sorte $G^- \neq \emptyset$.

Pour (3a) Les seuls points fixes d'une réflexions sont ceux de l'axe. Donc C soit être sur tous les axes des réflexions contenues dans G^-.

Notons au passage que deux réflexions d'axes qui se coupent forment une rotation. Donc G^- ne forme pas un groupe, mais même pas en rêve.

Pour (3b) Vu que l'union $G = G^+ \cup G^-$ est disjointe et que G^+ et G^- ont le même nombre d'éléments par la bijection 17.465, si G^- est non vide, G possède un nombre pair d'éléments.

Pour (3c) Si $\sigma \in G$ est une réflexion, nous savons que G^+ possède $p = n/2$ éléments et que

$$G^+ = \{R_C(2k\pi/p)\} = \{R_C(4k\pi/n)\}_{k=1,\dots,n/2}. \tag{17.468}$$

L'élément $\sigma \in G^-$ étant fixé, la bijection (17.465) nous indique que tous les éléments de G^- sont de la forme $\sigma \circ f$ avec $f \in G^+$. Donc

$$G^- \subset \mathrm{gr}\big(\sigma, R_C(4\pi/n)\big). \tag{17.469}$$

Nous avons aussi

$$G^+ \subset \mathrm{gr}\big(\sigma, R_C(4\pi/n)\big). \tag{17.470}$$

Et comme σ et $R_C(4\pi/n)$ sont dans G nous avons $\mathrm{gr}\big(\sigma, R_C(4\pi/n)\big) \subset G$. Tout cela pour dire que

$$G = \mathrm{gr}\big(\sigma, R_C(4\pi/n)\big). \tag{17.471}$$

$R\sigma = \sigma R^{-1}$ Nous restons dans le cas où G^- n'est pas vide. Nous considérons R, la rotation d'angle θ autour de C. Si R_0 est la rotation d'angle θ autour de $(0,0)$, nous avons

$$R = \tau_C \circ R_0 \circ \tau_C^{-1}, \tag{17.472}$$

et si σ_0 est la symétrie d'axe parallèle à l'axe de σ, mais passant par $(0,0)$ nous avons :

$$\sigma = \tau_C \circ \sigma_0 \circ \tau_C^{-1}. \tag{17.473}$$

Si v est le vecteur directeur de la réflexion σ_0, nous considérons enfin α, la rotation qui fait $\alpha(v) = (1,0)$. Nous avons alors

$$\sigma_0 = \alpha^{-1} \circ s \circ \alpha \tag{17.474}$$

où s est la symétrie autour de l'axe horizontal. En n'ayant pas peur d'identifier $\mathbb{R}^2$ à $\mathbb{C}$, l'applicaiton s est la conjugaison complexe. Avec tout ça nous avons

$$R\sigma = \tau_C R_0 \tau_C^{-1} \tau_C \sigma_0 \tau_C^{-1} = \tau_C R_0 \alpha^{-1} s \alpha \tau_C^{-1} = \tau_C \alpha^{-1} R_0 s \alpha \tau_C^{-1} \tag{17.475}$$

où nous avons utilisé le fait que les rotations autour de $(0,0)$ forment un groupe abélien pour commuter α^{-1} avec R_0. Nous utilisaons à présent le lemme 17.88 pour commuter R avec s :

$$\begin{aligned}
R\sigma &= \tau_X \alpha^{-1} s R_0^{-1} \alpha \tau_C^{-1} && (17.476a)\\
&= \tau_C \underbrace{\alpha^{-1} s \alpha}_{\sigma_0} R_0^{-1} \tau_C^{-1} && (17.476b)\\
&= \tau_C \sigma_0 \tau_C^{-1} \tau_C R_0^{-1} \tau_C^{-1} && (17.476c)\\
&= \sigma R^{-1}. && (17.476d)
\end{aligned}$$

Nous avons utilisé le fat que $\tau_C R_0^{-1} \tau_C^{-1} = R^{-1}$ comme on peut s'en convaincre en calculant le produit.

Table de multiplication Nous considérons une réflexion $\sigma \in G$. Les éléments de G^+ sont des rotations autour de C et ceux de G^- de la forme σR où R est une rotation autour de C. Pour savoir la table de multiplication de G, nous devons écrire

$$(\sigma^{\epsilon_1} R^k)(\sigma^{\epsilon_2} R^l) = \sigma^{\epsilon} R^m \tag{17.477}$$

où $\epsilon_1, \epsilon_2 \in \{0,1\}$, R est la rotation d'angle $4\pi/n$ autour de C et α et m sont des constantes à exprimer en fonction de ϵ_1, ϵ_2, k et l.

Tous les éléments de G pouvant être écrits soit sous la forme R^m soit sous la forme σR^m, nous avons les possibilités suivantes :

(1) $R^m R^l = R^{m+l}$

(2) $(R^m)(\sigma R^l) = \sigma R^{-m} R^l = \sigma R^{l-m}$

(3) $(\sigma R^m) R^l = \sigma R^{m+l}$

(4) $(\sigma R^m)(\sigma R^l) = \sigma\sigma R^{l-m} = R^{l-m}$.

Pour (3d) Récoltons quelque faits.

— Nous venons de prouver que $R\sigma = \sigma R^{-1}$.

— Tout élément de G peut s'écrire soit sous la forme R^m soit sous la forme σR^m suivant que l'élément soit dans G^+ ou G^-.

— Tout élément du groupe diédral D_n s'écrit soit sous la forme r^m soit sous la forme sr^m (proposition 17.85(2)).

L'application $\varphi\colon G \to D_n$ suivante est donc une bijection :

$$\begin{cases} \varphi(R^m) = r^m & (17.478a)\\ \varphi(\sigma R^m) = sr^m. & (17.478b)\end{cases}$$

Il nous reste à prouver que c'est un morphisme. Cela se fait en utilisant la table de multiplication du groupe diédral donnée dans la proposition 17.89 et celle du groupe G que nous venons de faire.

□

Définition 17.170 (Groupe de symétrie d'une partie de $\mathbb{R}^n$[354]).
Si Y est une partie de $\mathbb{R}^n$, nous définissons le ***groupe des symétries*** *de Y par*

$$\mathrm{Sym}(Y) = \{f \in \mathrm{Isom}(\mathbb{R}^n) \text{ tel que } f(Y) = Y\}. \tag{17.479}$$

Nous définissons aussi le groupe des symétries propres de Y par

$$\mathrm{Sym}^+(Y) = \{f \in \mathrm{Isom}^+(\mathbb{R}^n) \text{ tel que } f(Y) = Y\}. \tag{17.480}$$

Théorème 17.171 ([354]).
Soit $Y \subset \mathbb{R}^2$ tel que le groupe $\mathrm{Sym}^+(Y)$ soit fini d'ordre n. Alors c'est un groupe cyclique d'ordre n.

Si $\mathrm{Sym}^+(Y)$ est fini, alors $\mathrm{Sym}(Y)$ est soit cyclique[38] *d'ordre n soit isomorphe au groupe diédral*[39] *d'ordre $2n$.*

Démonstration. Nous savons déjà par la proposition 17.78 que $\mathrm{Sym}^+(Y)$ est isomorphe à un sous-groupe H^+ d'ordre n de $\mathrm{SO}(2)$. Vérifions que ce groupe est cyclique. Si $n = 1$, c'est évident. Si $n \geqslant 2$ alors nous savons que H^+ est constitué de rotations d'angles dans $[0, 2\pi[$ et vu que c'est un ensemble fini, il possède une rotation d'angle minimal (à part zéro). Notons α_0 cet angle.

Nous montrons que H^+ est engendré par la rotation d'angle α_0. Soit une rotation d'angle α. Étant donné que $\alpha_0 < \alpha$ nous pouvons effectuer la division euclidienne[40] de α par α_0 et obtenir

$$\alpha = k\alpha_0 + \beta \tag{17.481}$$

avec $\beta < \alpha_0$. Mézalors $R(\beta) = R(\alpha)R(\alpha_0)^{-k}$ est également un élément du groupe. Cela contredit la minimalité dès que $\beta \neq 0$. Avoir $\beta = 0$ revient à dire que α est un multiple de α_0, ce qui signifie que le groupe H^+ est cyclique engendré par α_0.

Notons au passage que nous avons automatiquement $\alpha_0 = \frac{2\pi}{n}$ parce qu'il faut $R(\alpha_0)^n = \mathrm{Id}$. Nous avons prouvé que $\mathrm{Sym}^+(Y)$ est cyclique d'ordre n.

Nous étudions maintenant le groupe $\mathrm{Sym}(Y)$. Par la proposition 17.78 nous avons un homomorphisme injectif

$$\phi\colon \mathrm{Sym}(Y) \to \mathrm{O}(2), \tag{17.482}$$

et en posant $H = \phi\big(\mathrm{Sym}(Y)\big)$ nous avons un isomorphisme de groupes $\phi\colon \mathrm{Sym}(Y) \to H$. Nous savons aussi que ce ϕ se restreint en

$$\phi\colon \mathrm{Sym}^+(Y) \to H^+ \subset \mathrm{SO}(2) \tag{17.483}$$

où $H^+ = \phi\big(\mathrm{Sym}^+(Y)\big) = H \cap \mathrm{SO}(2)$. Le groupe H^+ est cyclique et est engendré par la rotation $R(2\pi/n)$.

Supposons un instant que $H \subset \mathrm{SO}(2)$. Alors nous avons $H = H^+$ et ϕ est un isomorphisme entre $\mathrm{Sym}(Y)$ et le groupe cyclique engendré par $R(2\pi/n)$.

Nous supposons à présent que H n'est pas un sous-ensemble de $\mathrm{SO}(2)$. Quelles sont les isométries de $\mathbb{R}^2$ qui ne sont pas de déterminant 1 ? Il faut regarder dans le théorème 17.159 quelles sont les isométries contenant un nombre impair de réflexions. Ce sont les réflexions et les réflexions glissées. Or il ne peut pas y avoir de réflexions glissées dans un groupe fini parce que si f est une réflexion glissée, tous les f^k sont différents.

Nous en déduisons que si H n'est pas inclus dans $\mathrm{SO}(2)$, il contient une réflexion que nous nommons σ. Nous allons en déduire que $H \simeq H^+ \times_{\mathbf{Ad}} C_2$ où $C_2 = \{\mathrm{Id}, \sigma\}$. Si $h \in H$ nous pouvons écrire

$$h = (h\sigma^\epsilon)\sigma^\epsilon \tag{17.484}$$

pour n'importe quelle valeur de ϵ, et en particulier pour $\epsilon = \pm 1$.

Si $h \in \mathrm{SO}(2)$ alors nous écrivons $h = h\epsilon^0$ et si $h \notin \mathrm{SO}(2)$ nous écrivons $h = (h\sigma)\sigma$. Vu que $h\sigma \in \mathrm{SO}(2)$, cette dernière écriture est encore de la forme $\mathrm{SO}(2) \times C_2$. Quoi qu'il en soit tout élément de H s'écrit comme un produit

$$H = H^+C_2. \tag{17.485}$$

Cette décomposition est unique parce que si $h_1c_1 = h_2c_2$ alors $h_2^{-1}h_1 = c_2c_1^{-1}$, et comme $h_2^{-1}h_1 \in H^+$ nous avons $c_2c_1^{-1} \in H^+$ et donc $c_1 = c_2$. Partant nous avons aussi $h_1 = h_2$. Pour avoir le produit

38. Définition 2.12.
39. Définition 17.80.
40. Théorème 3.6.

semi-direct il faut encore montrer que $\mathbf{Ad}(C_2)H^+ \subset H^+$. Le seul cas à vérifier est $\mathbf{Ad}(\sigma)H^+ \subset H^+$. Vu que les éléments de H^+ sont caractérisés par le fait d'avoir un déterminant positif, nous avons

$$\mathbf{Ad}(\sigma)R(\alpha) = \sigma R(\alpha)\sigma^{-1} \in H^+. \tag{17.486}$$

□

Remarque 17.172.
Tout ceci est cohérent avec le théorème de Burnside 9.242 parce que le sous-groupe fini de $\mathrm{SO}(n)$ engendré par la rotation $R(2\pi/n)$ est un groupe d'exposant fini, à savoir que si h est dans ce groupe, $h^n = \mathrm{Id}$.

17.10.2 Relations trigonométriques dans un triangle rectangle

Nous donnons maintenant quelques relations trigonométriques classiques dans un triangle rectangle. Le théorème de Pythagore est déjà le théorème 7.203 ; nous nous concentrons ici sur les angles.

Proposition 17.173.
Soient $A, B, S \in \mathbb{R}^2$ des points distincts et non alignés formant un triangle rectangle en A :

$$(A - S) \cdot (B - A) = 0. \tag{17.487}$$

En posant $\theta = \widehat{ASB}$ nous avons

$$\cos(\theta) = \frac{\|A - S\|}{\|B - S\|} \tag{17.488}$$

et

$$\sin(\theta) = \pm\frac{\|B - A\|}{\|B - S\|}. \tag{17.489}$$

Démonstration. Nous posons $C = A - S$ et $D = B - S$. Vu que $C \neq 0$, il existe une rotation R_α telle que

$$\begin{cases} (R_\alpha C)_x > 0 & \text{(17.490a)} \\ (R_\alpha C)_y = 0. & \text{(17.490b)} \end{cases}$$

Nous posons $X = R_\alpha C$ et $Y = R_\alpha D$.

Le triangle formé de O, X et Y est « posé » sur l'axe des abscisses et est rectangle en X, c'est-à-dire

$$X \cdot (Y - X) = 0. \tag{17.491}$$

De ce fait, le point Y satisfait à $Y_x = X_x$. Et enfin, grâce aux propositions 17.154 et 17.155 nous avons $\widehat{ASB} = \widehat{XOY}$.

Nous écrivons les relations qui définissent l'angle $\widehat{XOY}$. Pour cela nous posons $X' = X/\|X\|$ et $Y' = Y/\|Y\|$ et nous avons

$$\begin{cases} \cos(\theta) = X'_x Y'_x & \text{(17.492a)} \\ \sin(\theta) = X'_x Y'_y. & \text{(17.492b)} \end{cases}$$

Vu que $X = (X_x, 0)$ nous avons $X'_x = 1$. De plus

$$\|Y\| = \|R_\alpha(D)\| = \|D\| = \|B - S\|. \tag{17.493}$$

En substituant les valeurs dans (17.492),

$$\cos(\theta) = Y'_x = \frac{Y_x}{\|Y\|} = \frac{X_x}{\|B - S\|} = \frac{\|C\|}{\|B - S\|} = \frac{\|A - S\|}{\|B - S\|}. \tag{17.494}$$

Voila déjà une chose de prouvée.

Pour la seconde, nous avons $\sin(\theta) = Y'_y$. Selon le signe de Y_y nous avons $Y_y = \pm\|Y - X\|$ et donc

$$\sin(\theta) = Y'_y = \frac{Y_y}{\|Y\|} = \frac{\pm\|Y - X\|}{\|Y\|} = \pm\frac{\|D - C\|}{\|Y\|} = \pm\frac{\|B - A\|}{\|B - S\|}. \tag{17.495}$$

□

Le signe sur la formule du sinus revient au fait que la définition de l'angle $\widehat{AOB}$ est de considérer la rotation qui fait aller A vers B. Donc suivant la position relative de A, O et B, il se peut que l'angle mesuré soit l'angle *extérieur* au triangle.

La proposition suivante est parfois prise comme définition de l'angle.

Proposition 17.174.
Soient trois points non alignés $A, S, B \in \mathbb{R}^2$. Nous avons

$$\cos\left(\widehat{ASB}\right) = \frac{(A - S)\cdot(B - S)}{\|A - S\|\|B - S\|}. \tag{17.496}$$

Démonstration. Nous posons $C = A - S$, $D = B - S$, $X = C/\|C\|$ et $Y = D/\|D\|$. Avec cela, la définition 17.152 donne l'équation

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} X_x \\ X_u \end{pmatrix} = \begin{pmatrix} Y_x \\ Y_y \end{pmatrix} \tag{17.497}$$

que nous écrivons comme le système

$$\begin{cases} X_c\cos(\theta) - X_y\sin(\theta) = Y_x & (17.498a) \\ X_y\cos(\theta) + X_x\sin(\theta) = Y_y. & (17.498b) \end{cases}$$

Nous considérons maintenant cela comme un système pour $\big(\cos(\theta), \sin(\theta)\big)$:

$$\begin{pmatrix} X_x & -X_y \\ X_y & X_x \end{pmatrix} \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = \begin{pmatrix} Y_x \\ Y_y \end{pmatrix}. \tag{17.499}$$

Le déterminant de la dernière matrice est $X_x + X_y^2 = \|X\|^2 = 1$ parce que X est unitaire. Cette matrice est donc inversible et son inverse est vite calculée. Nous avons

$$\begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = \begin{pmatrix} X_x & X_y \\ -X_y & X_x \end{pmatrix} \begin{pmatrix} Y_x \\ Y_y \end{pmatrix}. \tag{17.500}$$

Cela donne ce que nous voulions :

$$\cos(\theta) = X\cdot Y = \frac{C\cdot D}{\|C\|\|D\|} = \frac{(A - S)\cdot(B - S)}{\|A - S\|\|C - S\|}. \tag{17.501}$$

□

Remarque 17.175.
Prendre la formule (17.496) comme définition de l'angle $\widehat{ASB}$ est cependant trompeur parce que ça ne permet de définir les angles que sur une partie de $[0, 2\pi[$ sur laquelle le cosinus est injectif. Pour réellement définir tous les angles, il faut alors un peu bricoler.

Sans vouloir être méchant, je crois que ceux qui prennent ça pour définition d'angle sont ceux qui donnent un cours sur le produit scalaire sans avoir l'intention de lier la définition d'une rotation comme composée de réflexions aux matrices de SO(2) et aux fonctions trigonométriques.

17.10.3 Pavages du plan

Définition 17.176.
*Une application affine $f\colon \mathbb{R}^n \to \mathbb{R}^n$ est un **déplacement** lorsqu'elle est une isométrie de $(\mathbb{R}^n, d)$ qui préserve l'orientation*[41].

Définition 17.177 ([268])**.**
*Un **pavage** de $\mathbb{R}^2$ est une paire (G, K) où G est un groupe de déplacements*[42] *de $\mathbb{R}^2$ et K un*

41. Définition 9.28.
42. Définition 17.176.

compact de $\mathbb{R}^2$ d'intérieur non vide telle que

(1) $G \cdot K = \mathbb{R}^2$,

(2) Si $g_1, g_2 \in G$ *satisfont* $g_1 \cdot \operatorname{Int}(K) \cap g_2 \cdot \operatorname{Int}(K) \neq 0$, *alors* $g_1 \cdot K = g_2 \cdot K$.

Nous disons qu'un groupe de déplacements de $\mathbb{R}^2$ est un ***groupe de pavage*** *de $\mathbb{R}^2$ si il existe un compact K tel que la paire (G, K) soit un pavage.*

En termes de notations,

$$G \cdot K = \bigcup_{g \in G} g(K) = \bigcup_{g \in G} \bigcup_{k \in K} g(k). \tag{17.502}$$

Lemme 17.178 ([1])**.**
Soient une bijection affine $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ ainsi qu'une droite d et un point A. Nous notons S la partie de $\mathbb{R}^2$ située du côté de d contenant A.

Alors $\varphi(S)$ est la partie de $\mathbb{R}^2$ située du côté de $\varphi(d)$ contenant $\varphi(A)$.

Démonstration. La droite d est donnée par une application affine $f\colon \mathbb{R}^2 \to \mathbb{R}$ et la définition

$$f = \{x \in \mathbb{R}^2 \text{ tel que } f(x) = 0\}. \tag{17.503}$$

Nous supposons que $f(A) > 0$; sinon, nous pouvons utiliser $-f$ au lieu de f. Donc

$$S = \{x \in \mathbb{R}^2 \text{ tel que } f(x) > 0\}. \tag{17.504}$$

La partie $\varphi(S)$ est alors donnée par

$$\varphi(S) = \{\varphi(x) \text{ tel que } f(x) > 0\}. \tag{17.505}$$

Comme φ est une bijection, cela s'écrit aussi bien

$$\varphi(S) = \{y \in \mathbb{R}^2 \text{ tel que } f\big(\varphi^{-1}(y)\big) > 0\}. \tag{17.506}$$

De même

$$\varphi(d) = \{s \in \mathbb{R}^2 \text{ tel que } f\big(\varphi^{-1}(s)\big) = 0\}. \tag{17.507}$$

Donc les deux côtés de la droite $\varphi(d)$ sont donnés par le signe de $f \circ \varphi^{-1}$. Nous avons

$$(f \circ \varphi^{-1})\big(\varphi(A)\big) = f(A) > 0. \tag{17.508}$$

Donc $\varphi(A) \in \varphi(S)$. □

Lemme 17.179.
Le groupe

$$G = \operatorname{gr}(\tau_{e_1}, \tau_{e_2}) \tag{17.509}$$

est un groupe de pavage du plan.

Démonstration. Il suffit de prendre le carré $K = [0,1] \times [0,1]$. En appliquant les translations, nous recouvrons tout le plan, sans intersections des intérieurs des carrés. Notons toutefois qu'il y a un recouvrement des bords. □

Lemme 17.180 ([1, 364])**.**
Le groupe

$$G = \operatorname{gr}(\tau_{e_1}, \tau_{e_2}, R_0(\pi)) \tag{17.510}$$

est un groupe de pavage du plan.

Démonstration. Le compact à considérer est $K = [0, \frac{1}{2}] \times [0, 1]$. Le compact K et son image par $R_0(\pi)$ sont représentés sur la figure 17.9.

En agissant sur K avec les translations verticales et horizontales, nous recouvrons des bandes verticales de largeur 1/2. En agissant de même sur $R_0(\pi)(K)$, nous recouvrons les autres bandes verticales.

Donc $G \cdot K$ recouvre bien $\mathbb{R}^2$. Il serait cependant un peu présomptueux de croire en avoir fini. Il faut vérifier la condition (2) de la définition 17.177 d'un pavage.

Supposons que $g_1 \cdot \operatorname{Int}(K) \cap g_2 \cdot \operatorname{Int}(K) \neq \emptyset$. Cela signifie qu'il existe $k_1, k_2 \in \operatorname{Int}(K)$ tels que $g_1(k_1) = g_2(k_2)$, ou encore que

$$(g_2^{-1} g_1) k_1 = k_2 \in \operatorname{Int}(K). \tag{17.511}$$

Quelle est la forme d'un élément général de G ? Le lemme 2.9 nous indique qu'un élément général de G est un produit fini de τ_{e_1}, τ_{e_2} et $R_0(\pi)$. Mais nous savons que si α est linéaire,

$$\alpha \circ \tau_u = \tau_{\alpha(u)} \circ \alpha. \tag{17.512}$$

Dans notre cas, dans un produit général, nous pouvons déplacer tous les facteurs $R_0(\pi)$ à droite en changeant des τ_{e_i} en $\tau_{R_0(\pi)e_i} = \tau_{-e_i}$. Les translations par contre commutent sans faire d'histoires. Donc un élément général de G est de la forme

$$g = \tau_{e_1}^k \tau_{e_2}^l R_0(\pi)^m \tag{17.513}$$

Nous pouvons évidemment restreindre m à $\{0, 1\}$. Supposons $k \in \operatorname{Int}(K)$ et $g(k) \in \operatorname{Int}(K)$. Nous avons $0 < k_x < 0.5$. Si $m = 1$, alors $g(k)_x \in]-1/2, 0[$ et aucun $\tau_{e_1}^k g(k)_x$ ne pourra plus être entre 0 et 1/2. Donc $m = 0$. À partir de là, pour avoir $g(k) \in \operatorname{Int}(K)$ nous devons avoir également $k = l = 0$. Donc $g = \mathrm{Id}$.

Deux éléments g_1 et g_2 vérifiant la condition (17.511) doivent donc vérifier $g_2^{-1} g_1 = \mathrm{Id}$, et donc $g_1 = g_2$. Par conséquent $g_1 \cdot K = g_2 \cdot K$. □

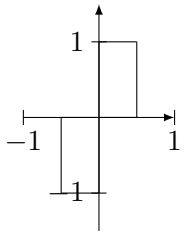


FIGURE 17.9 – Le compact K et son image par $R_0(\pi)$ pour le lemme 17.180.

Lemme 17.181 ([1, 364])**.**
Le groupe

$$\operatorname{gr}(\tau_{e_1}, \tau_{e_2}, R_0(\pi/2)) \tag{17.514}$$

est un groupe de pavage.

Démonstration. Le pavé à considérer est

$$K = [-\frac{1}{2}, 0] \times [0, \frac{1}{2}]. \tag{17.515}$$

En lui appliquant trois fois la rotation $R_0(\pi/2)$, nous reconstituons le carré $[-\frac{1}{2}, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]$. Ensuite, avec les translations, nous pavons tout le plan.

Pour la seconde condition, nous procédons comme dans la démonstration du lemme 17.180. D'abord $R_0(\pi/2)e_1 = e_2$ et $R_0(\pi/2)e_2 = -e_1$. Donc dans un produit général de τ_{e_1}, τ_{e_2} et $R_0(\pi/2)$

(et de leurs inverses), toutes les rotations peuvent être mises à droite ; nous avons donc un élément général de G sous la forme

$$g = \tau_a \circ R_0(\pi/2)^k \tag{17.516}$$

avec $a \in \mathbb{Z}e_1 + \mathbb{Z}e_2$ et $k \in \{0, 1, 2, 3\}$.

Vu que les translations se font par nombres entiers tandis que les différences de coordonnées entre les $R_0(\pi/2)^k K$ sont demi-entiers, si $k \in \mathrm{Int}(K)$, alors aucun τ_a ne permet d'avoir $(\tau_a \circ R_O(\pi/2))k \in \mathrm{Int}(K)$.

Bref, si $(g_2^{-1}g_2)k \in K$, alors encore une fois $g_2^{-1}g_1 = \mathrm{Id}$. □

Et c'est maintenant que les choses compliquées commencent.

Lemme 17.182.
Le groupe

$$\mathrm{gr}\left(\tau_{e_1}, \tau_{(-\frac{1}{2}, \frac{\sqrt{3}}{2})}, R_0(\pi/3)\right) \tag{17.517}$$

est un groupe de pavage.

Démonstration. Ici le compact K est le triangle de sommets $A = (0, 0)$, $B = (\frac{1}{2}, \frac{\sqrt{3}}{6})$ et $C = (1, 0)$. Plus précisément il s'agit de l'intersection des trois parties suivantes :

— le côté de la droite (AB) où est C,

— le côté de la droite (AC) où est B,

— le côté de la droite (BC) où est A.

Il est bon d'écrire ces trois conditions sous forme d'inéquations.

— La droite AB est donnée par l'équation $f_{AB}(x, y) = 0$ pour $f_{AB}(x, y) = x - \sqrt{3}y$. Vu que $f_{AB}(C) = 1$, la première inéquation pour K est

$$x - \sqrt{3}y \geqslant 0. \tag{17.518}$$

— Pour la droite (BC) nous avons $f_{BC}(x, y) = -x - \sqrt{3}y + 1$ et $f_{BC}(A) = 1$. Donc la seconde inéquation pour K est

$$-x - \sqrt{3}y + 1 \geqslant 0 \tag{17.519}$$

— La droite AC est donnée par l'application $f_{AC}(x, y) = y$. Vu que $f_{AC}(B) = \sqrt{3}/6$ nous avons la troisième inéquation pour K :

$$y \geqslant 0. \tag{17.520}$$

En résumé, la définition de K est le système

$$\begin{cases} x - \sqrt{3}y \geqslant 0 & \text{(17.521a)} \\ -x - \sqrt{3}y + 1 \geqslant 0 & \text{(17.521b)} \\ y \geqslant 0. & \text{(17.521c)} \end{cases}$$

La figure 17.10(a) nous montre ce triangle et l'action des puissances de $R_0(\pi/3)$ sur lui. Pour votre lanterne, la matrice de cette rotation est

$$R_0(\pi/3) = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix}. \tag{17.522}$$

La figure 17.10(c) vous montre une partie de ce pavage dans toute sa splendeur.

Vu que les translations sont $u_1 = e_1$ et $u_2 = (\frac{1}{2}, \frac{\sqrt{3}}{2})$, il est suffisant de montrer que notre pavage pave réellement le parallélogramme construit sur ces deux vecteurs. Nous vous avons très obligeamment dessiné ce parallélogramme pavé sur la figure 17.10(b).

Pour montrer que les triangles dessinés pavent effectivement le parallélogramme, nous allons faire une exhaustion de cas. Soit (x, y) dans le parallélogramme.

Nous commençons par couper le parallélogramme en deux parties suivant la diagonale allant de u_1 à u_2. Dans la vie mon p'ti gars, il y a deux types de points : ceux qui vérifient $x + \frac{1}{\sqrt{3}}y - 1 \leqslant 0$ et les autres.

Si $x + y/\sqrt{3} - 1 \leqslant 0$ Ceci est le côté de $(0,0)$. Nous subdivisons suivant la petite barre verticale (suivez le dessin), c'est-à-dire suivant les deux cas : $x \geqslant \frac{1}{2}$ et $x \leqslant \frac{1}{2}$.

Si $x \leqslant \frac{1}{2}$ Enfin nous coupons avec la droite diagonale partant de $(0,0)$, c'est-à-dire selon que $x - \sqrt{3}y \leqslant 0$ ou $x - \sqrt{3}y \geqslant 0$.

Si $x - \sqrt{3}y \leqslant 0$ Les points dont nous parlons sont les $(x,y) \in \mathbb{R}^2$ vérifiant

$$\begin{cases} x + \dfrac{1}{\sqrt{3}}y - 1 \leqslant 0 & (17.523a) \\ x \leqslant \dfrac{1}{2} & (17.523b) \\ x - \sqrt{3}y \leqslant 0. & (17.523c) \end{cases}$$

En suivant le dessin, vous remarquerez que l'élément de G à considérer est

$$g = \tau_{e_1} \circ \tau_{(-\frac{1}{2},\frac{\sqrt{3}}{2})} \circ R_0(\pi/3)^4. \tag{17.524}$$

Nous avons

$$g(A) = (\frac{1}{2}, \frac{\sqrt{3}}{2}) \tag{17.525a}$$
$$g(B) = (\frac{1}{2}, \frac{\sqrt{3}}{6}) \tag{17.525b}$$
$$g(C) = (0,0). \tag{17.525c}$$

Ce que les équations (17.523) décrivent est l'intersection des trois parties suivanteas[43] :

— le côté de la droite $\big(g(A)g(B)\big)$ contenant $g(C)$,

— le côté de la droite $\big(g(A)g(C)\big)$ contenant $g(B)$,

— le côté de la droite $\big(g(B)g(C)\big)$ contenant $g(A)$,

Le lemme 17.178 nous permet d'exprimer ces trois parties en termes de K :

— l'image par g du côté de la droite (AB) content C,

— l'image par g du côté de la droite (AC) content B,

— l'image par g du côté de la droite (BC) content A,

Comme g est une bijection, l'intersection des images par g et l'image par g de l'intersection. Bref, les équations (17.525) décrivent l'image par g de K.

Cette partie est donc pavée par (G,K).

Si $x - \sqrt{3}y \geqslant 0$ hop.

Si $x \geqslant \frac{1}{2}$ hop.

Si $x + y/\sqrt{3} - 1 \geqslant 0$ hop.

Les cas listés se traitement surement de la même façon. Nous tenons pour prouvé que (G,K) est bien surjectif sur $\mathbb{R}^2$.

Nous devons encore montrer la condition (2) de la définition 17.177. Commençons par déterminer la forme générale d'un élément de G sous la forme $\tau_a \circ r_0$ où r_0 est une application linéaire. Le lemme 2.9 nous indique qu'un élément général de $G = \mathrm{gr}(\tau_{e_1}, \tau_{(\frac{1}{2},\frac{\sqrt{3}}{2})}, R_0(\pi/3))$ est un produit arbitraire (mais fini) de τ_{e_1}, $\tau_{(\frac{1}{2},\frac{\sqrt{3}}{2})}$ et de $R_0(\pi/3)$ et de leurs inverses.

43. Personnellement, je n'ai pas vérifié, mais ça m'étonnerait que ce soit faux. Vérifiez et écrivez-moi.

Comme toujours nous avons $\alpha \circ \tau_v = \tau_{\alpha(v)} \circ \alpha$. Donc nous pouvons passer toutes les rotations $R_0(\pi/3)$ et $R_0(\pi/3)^{-1}$ à droite du produit quitte à produire des translations des formes suivantes :

$$R_0(\pi/3)^k e_1 \tag{17.526a}$$
$$R_0(\pi/3)^k (\frac{1}{2}, \frac{\sqrt{3}}{2}) \tag{17.526b}$$
$$R_0(\pi/3)^{-k} e_1 \tag{17.526c}$$
$$R_0(\pi/3)^{-k} (\frac{1}{2}, \frac{\sqrt{3}}{2}). \tag{17.526d}$$

En utilisant la matrice (17.522) nous trouvons assez vite que

$$R_0(\pi/3)e_1 = \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix} \tag{17.527a}$$
$$R_0(\pi/3)^2 e_1 = \begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix} \tag{17.527b}$$
$$R_0(\pi/3)^3 e_1 = -e_1. \tag{17.527c}$$

Les applications suivantes de $R_0(\pi/3)$ ne donnent rien de nouveau, si ce n'est le signe. Les puissances de $R_0(\pi/3)$ appliquées à $\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$ sont déjà parmi celles listées en (17.527). Quant aux inverses, $R_0(\pi/3)^{-1} = R_0(\pi/3)^5$; donc rien de nouveau non plus.

Un élément général de G est donc dans

$$\tau_v \circ R_0(\pi/3)^k \tag{17.528}$$

avec v dans

$$\mathbb{Z}e_1 + \mathbb{Z}\begin{pmatrix} 1/2 \\ \frac{\sqrt{3}}{2} \end{pmatrix} + \mathbb{Z}\begin{pmatrix} -1/2 \\ \frac{\sqrt{3}}{2} \end{pmatrix}. \tag{17.529}$$

Remarquons que

$$\begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix} = -e_1 + \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}. \tag{17.530}$$

Donc un élément général de G est

$$\tau_{e_1}^k \circ \tau_{(-\frac{1}{2}, \frac{\sqrt{3}}{2})}^l \circ R_0(\pi/3)^m. \tag{17.531}$$

Nous pouvons maintenant prouver notre point. Pour cela nous allons successivement considérer les 5 rotations de K présentées dans la sous-figure 17.10(a). Pour chacune nous allons montrer qu'aucune translation ne permet d'obtenir une intersection avec K.

Nous allons en faire un seul en détail.

Pour $L = R_0(\pi/3)K$

Pour $L = R_0(\pi/3)^2 K$ Nous allons prouver que si $(x, y) \in \mathrm{Int}(L)$, alors

$$\tau_{e_1}^k \circ \tau_{(-\frac{1}{2}, \frac{\sqrt{3}}{2})}^l (x, y) \tag{17.532}$$

ne peut pas être dans $\mathrm{Int}(K)$. Pour la simplicité des notations nous notons $r = R_0(\pi/3)^2$; nous avons

$$r = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}. \tag{17.533}$$

Nous considérons $g = \tau_{e_1}^k \circ \tau_{(-\frac{1}{2}, \frac{\sqrt{3}}{2})} \circ r$. Un calcul nous donne l'image de (x, y) par g :

$$g(x, y) = \tau_{e_1}^k \circ \tau_{(-\frac{1}{2}, \frac{\sqrt{3}}{2})}^l \begin{pmatrix} -\frac{1}{2}x - \frac{\sqrt{3}}{2}y \\ -\frac{\sqrt{3}}{2}x - \frac{1}{2}y \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}x - \frac{\sqrt{3}}{2}y + k - \frac{1}{2}l \\ \frac{\sqrt{3}}{2}x - \frac{1}{2}y + \frac{\sqrt{3}}{2}l \end{pmatrix}. \tag{17.534}$$

Nous considérons $(x,y) \in \mathrm{Int}(K)$ tel que $g(x,y) \in \mathrm{Int}(K)$, et nous allons trouver une contradiction. Le fait que $(x,y) \in \mathrm{Int}(K)$ signifie que (x,y) satisfait les inéquations (17.521) mais avec des inégalités strictes. Le fait que $g(x,y) \in \mathrm{Int}(K)$ nous donne trois inéquations de plus. Assez rapide calcul :

$$f_{AB}\big(g(x,y)\big) = -2x + k - 2l, \tag{17.535a}$$
$$f_{BC}\big(g(x,y)\big) = -x - \sqrt{3}y - k - l + 1, \tag{17.535b}$$
$$f_{AC}\big(g(x,y)\big) = \frac{\sqrt{3}}{2}x - \frac{1}{2}y + \frac{\sqrt{3}}{2}l. \tag{17.535c}$$

Au final, notre point (x,y) doit satisfaire le système suivant :

$$\begin{cases} x - \sqrt{3}y > 0 & (17.536a)\\ -x - \sqrt{3}y + 1 > 0 & (17.536b)\\ y > 0 & (17.536c)\\ -2x + k - 2l > 0 & (17.536d)\\ -x - \sqrt{3}y - k - l + 1 > 0 & (17.536e)\\ \frac{\sqrt{3}}{2}x - \frac{1}{2}y + \frac{\sqrt{3}}{2}l > 0. & (17.536f)\end{cases}$$

Les inéquations 17.536c et 17.536a donnent déjà $x > 0$. De même avec 17.536b nous trouvons $x < 1$. Voila déjà $x \in]0,1[$ qui est directement visible sur le dessin du triangle K.
Vu que $x > 0$, l'inéquation (17.536d) donne

$$k - 2l > -2x + k - 2l > 0. \tag{17.537}$$

Donc $k - 2l > 0$.
Vu que $x < 1$ et $y > 0$, l'inéquation (17.536f) donne

$$\frac{\sqrt{3}}{2} + \frac{\sqrt{3}}{2}l > \frac{\sqrt{3}}{2}x - \frac{1}{2}y + \frac{\sqrt{3}}{2}l > 0. \tag{17.538}$$

Donc $\frac{\sqrt{3}}{2}(1+l) > 0$. Comme l est entier, cela donne $l \geqslant 0$.
Enfin de (17.536e) nous tirons

$$-k + 1 > -x - \sqrt{3}y - k - l + 1 > 0, \tag{17.539}$$

Ce qui donne $k < 1$ et donc $k \leqslant 0$.
En résumé nous avons trouvé trois inéquations pour k et l :

$$\begin{cases} k \leqslant 0 & (17.540a)\\ l \geqslant 0 & (17.540b)\\ k - 2l > 0. & (17.540c)\end{cases}$$

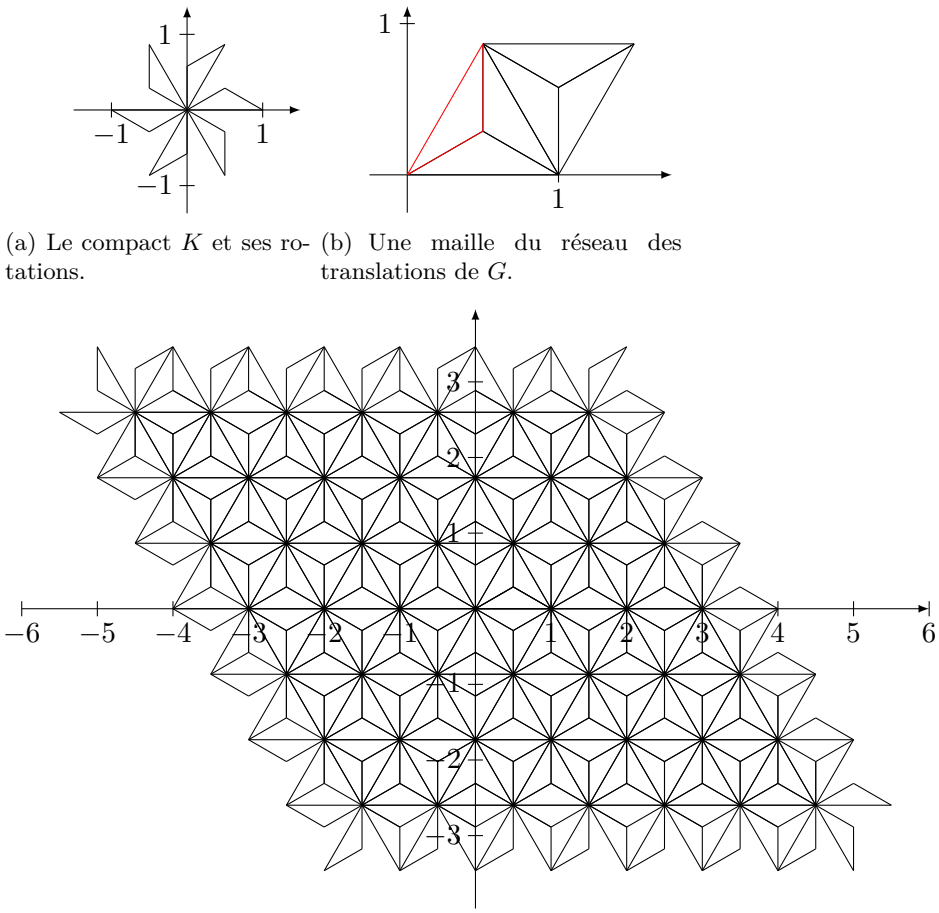
Ce système est impossible.
Il n'existe donc pas de translations qui, appliquée à $\mathrm{Int}(L)$, donne une intersection avec $\mathrm{Int}(K)$.

Pour $L = R_0(\pi/3)^3 K$ hop.

Pour $L = R_0(\pi/3)^4 K$ hop.

Pour $L = R_0(\pi/3)^5 K$ hop.

Voila. J'espère que toutes les idées sont en place, et que les parties manquantes sont seulement des vérifications qui se font mécaniquement de la même manière [44]. □



(a) Le compact K et ses rotations.

(b) Une maille du réseau des translations de G.

(c) Une partie du pavage complet.

FIGURE 17.10 – Illustrations pour le pavage du lemme 17.182.

Lemme 17.183.
Le groupe

$$\mathrm{gr}\left(\tau_1, \tau_{(-\frac{1}{2},\frac{\sqrt{3}}{2})}, R_0(\pi/3)\right) \tag{17.541}$$

est un groupe de pavage.

D'après [364], la démonstration du lemme 17.183 demande d'utiliser le losange de sommets $(0,0)$, $(1,0)$, $(\frac{1}{2},\frac{\sqrt{3}}{6})$ et $(\frac{1}{2},-\frac{\sqrt{3}}{6})$ [45].

Lemme 17.184 ([1])**.**
Soient $u_1, u_2 \in \mathbb{R}^2$ non colinéaires, ainsi que $\alpha, \beta \in [0,1]$. Nous considérons $v = \alpha u_1 + \beta u_2$.

Nous supposons pour fixer les idées que $\|u_1\| \geqslant \|u_2\|$. Alors

$$\min\{\|v\|, \|u_1 + u_2 - v\|\} \leqslant \|u_1\|. \tag{17.542}$$

Autrement dit, tout point intérieur d'un parallélogramme est plus proche d'un angle que la longueur du plus long côté.

Démonstration. Les points $\alpha u_1 + \beta u_2$ ($\alpha, \beta \in [0,1]$) se divisent en deux parties : ceux avec $0 \leqslant \alpha + \beta \leqslant 1$ et ceux avec $1 \leqslant \alpha + \beta \leqslant 2$.

Si $\alpha + \beta \leqslant 1$ alors

$$\|\alpha u_1 + \beta u_2\| < \|\alpha u_1\| + \beta\|u_2\| = \alpha\|u_1\| + \beta\|u_2\| \leqslant (\alpha+\beta)\|u_1\| \leqslant \|u_1\|. \tag{17.543}$$

44. Je n'ai pas vérifié. Faites-le et écrivez-moi pour me dire ce qui en est.
45. Je n'ai pas vérifié, mais à mon avis une preuve doit prendre les mêmes idées que celles du lemme 17.182.

L'inégalité est stricte parce que u_1 et u_2 ne sont pas colinéaires.

Si au contraire $\alpha + \beta \geqslant 1$ nous avons

$$\|u_1 + u_2 - \alpha u_1 - \beta u_2\| < \|(1-\alpha)u_1\| + \|(1-\beta)u_2\| < (2-\alpha-\beta)\|u_1\| \leqslant \|u_1\|. \tag{17.544}$$

□

Lemme 17.185.
Soient un ensemble E, et deux bijections $r, s\colon E \to E$ ayant chacune un unique point fixe. Si elles commutent, alors leurs points fixes sont égaux.

Démonstration. Nous nommons a le point fixe de r et b celui de s. Pour tout $x \in E$ nous avons $(rs)(x) = (sr)(x)$. En particulier pour $x = s^{-1}(a)$. D'une part

$$(rs)(x) = r(a) = a. \tag{17.545}$$

Et d'autre part,

$$(sr)(a) = (srs^{-1})(a) \tag{17.546}$$

Si nous imposons $(srs^{-1})(a) = a$, nous avons, en appliquant s^{-1} est deux côtés : $(rs^{-1})(a) = s^{-1}(a)$. Cela prouve que $s^{-1}(a)$ est un point fixe de r. Donc $s^{-1}(a) = a$.

Nous en déduisons que a est un point fixe de s et donc que $a = b$. □

Lemme 17.186.
Soit un sous-groupe fini T de $(\mathbb{R}^2, +)$ tel que $\|v\| > \delta$ pour tout $v \neq 0$ dans T.

Alors si u_1 et u_2 sont les plus petits éléments en norme de T, nous avons

$$T = \mathbb{Z}u_1 + \mathbb{Z}u_2. \tag{17.547}$$

Démonstration. Nous décomposons en plusieurs parties.

<u>$\mathbb{R}^+ v \cap T = \mathbb{N}v_m$</u> Soit $v \in T$. L'ensemble $\{\lambda \in \mathbb{R}^+ \text{ tel que } \lambda v \in T\}$ a un minimum parce que tous les éléments de T sont en norme plus grands que $\delta > 0$. Soit λ_m ce minimum et $v_m = \lambda_m v$.

Nous prétendons à présent que $\mathbb{R}^+ v \cap T = \mathbb{N}v_m$. Nous ne faisons d'ailleurs pas que prétendre ; nous *prouvons*. En effet, doit $\lambda v \in T$. Nous devons prouver que $\lambda = l\lambda_m$ pour un certain $l \in \mathbb{N}$. Soit $k \in \mathbb{N}$ tel que $k \leqslant \lambda < k+1$.

Nous avons

$$(k+1)\lambda_m - \lambda \leqslant (k+1)\lambda_m - k\lambda_m = \lambda_m. \tag{17.548}$$

Vu que T est un groupe pour l'addition, les faits que $\lambda_m v \in T$ et que $\lambda v \in T$ impliquent que $\big((k+1)\lambda_m - \lambda\big)v \in T$. Mais

$$|\big((k+1)\lambda_m - \lambda\big)v| \leqslant \lambda_m. \tag{17.549}$$

Vu la propriété de minimalité de λ_m nous avons forcément

$$(k+1)\lambda_m - \lambda = \lambda_m. \tag{17.550}$$

Cela prouve que $\lambda = k\lambda_m$.

Jusqu'ici nous avons prouvé que

$$\mathbb{R}^+ v \cap T = \mathbb{N}v_m \tag{17.551}$$

pour un certain multiple v_m de v.

<u>$\|u_1\| \leqslant \|v\|$ **pour tout** v</u> Nous montrons à présent qu'il existe $u_1 \in T$ tel que $\|u_1\| \leqslant \|v\|$ pour tout $v \in T$. Si tel n'était pas le cas, il existerait une suite $v_k \in T$ telle que $\|v_{k+1}\| < \|v_k\|$. Toute cette suite serait contenue dans le couronne (compacte) de rayons δ et $\|u_1\|$. Quitte à prendre une sous-suite, nous pouvons supposer que (v_k) converge[46]. Cette suite serait de Cauchy et pour tout ϵ (en particulier $\epsilon < \delta$), il existerait p, q tels que $\|v_p - v_q\| < \epsilon$. Vu que T est un groupe pour l'addition, nous aurions $v_p - v_q \in T$ avec $\|v_p - v_q\| < \epsilon \leqslant \delta$.

46. Proposition 7.138.

Première pause Si T est engendré seulement par u_1, nous avons fini. Autrement dit, si tout T est dans un sous-espace de dimension 1 de $\mathbb{R}^2$ nous avons terminé.

Dans la suite, nous supposons donc que T n'est pas contenu dans un sous-espace de dimension 1.

$T = \mathbb{Z}u_1 + \mathbb{Z}u_2$ Soient u_1 et u_2 les deux plus petits éléments de T en norme (peut-être ex-aequo). Ces deux éléments ne sont pas colinéaires, sinon leur différence serait plus petite.

Soit $v \in T$. Vu que $\{u_1, u_2\}$ est une base de $\mathbb{R}^2$, il existe $\alpha, \beta \in \mathbb{R}$ tels que

$$v = \alpha u_1 + \beta u_2. \tag{17.552}$$

Notre but est à présent de prouver que $\alpha, \beta \in \mathbb{Z}$.

Si ce n'était pas le cas, une simple translation nous mènerait dans les circonstances du lemme 17.184. Nous aurions alors que soit $\|v\|$ soit $\|u_1 + u_2 - v\|$ serait strictement plus petit que le plus grand entre $\|u_1\|$ et $\|u_2\|$. Cela contredirait le fait que $\|u_1\|$ et $\|u_2\|$ étaient les deux plus petits.

□

Proposition 17.187 ([1]).
Si G est un groupe de pavage[47] *de $\mathbb{R}^2$ et si $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ est une isométrie affine, alors le groupe*

$$G' = f \circ G \circ f^{-1} = \{f \circ g \circ f^{-1} \text{ tel que } g \in G\} \tag{17.553}$$

est un groupe de pavage.

Démonstration. Soit un compact K tel que (G, K) soit un pavage. Nous notons $K' = f(K)$. Nous devons prouver deux choses :

— $f \circ G \circ f^{-1}$ est un groupe de déplacements ;

— (G', K') est un pavage.

Ceci fait trois éléments à prouver.

G' est constitué de déplacements Les éléments de G sont des isométries, ainsi que f et f^{-1}. Donc les éléments de G' sont des isométries.

Soit $g \in G$. Vu que g est affine, il existe une décomposition $g = \tau_v \circ g_0$ où g_0 est linéaire. De même $f = \tau_w \circ f_0$. Les règles du produit et de l'inverse de la proposition 8.57(2)(3) nous indiquent que la partie linéaire de fgf^{-1} est $f_0 g_0 f_0^{-1}$.

En ce qui concerne le déterminant de $f_0 g_0 f_0^{-1}$, c'est la proposition 9.10 qui nous indique que

$$\det(f_0 g_0 f_0^{-1}) = \det(f_0)\det(g_0)\det(f_0^{-1}) = \det(f_0)\det(g_0)\det(f_0)^{-1} = \det(g_0) = 1. \tag{17.554}$$

$G' \cdot K' = \mathbb{R}^2$ Vu que f est affine, K' est encore compact[48] Nous avons :

$$\begin{aligned}
(f \circ G \circ f^{-1})\big(f(K)\big) &= \bigcup_{g\in G}\bigcup_{k\in K}(f \circ g \circ f^{-1})\big(\alpha(k)\big) && (17.555a)\\
&= \bigcup_{g\in G}\bigcup_{k\in K}(f \circ g)(k) && (17.555b)\\
&= \bigcup_{g\in G} f\big(\bigcup_{k\in K} g(k)\big) && (17.555c)\\
&= \bigcup_{g\in G}(f \circ g)(K) && (17.555d)\\
&= f\big(\bigcup_{g\in G} g(K)\big) && (17.555e)\\
&= f\big(G \cdot K\big) && (17.555f)\\
&= \mathbb{R}^2. && (17.555g)
\end{aligned}$$

47. Définition 17.177.
48. Il suffit de prouver que $f(K)$ est fermé et borné par le théorème de Borel-Lebesgue 7.127.

Pour la dernière égalité, nous avons utilisé le fait que f est bijective et que $G \cdot K = \mathbb{R}^2$.

L'autre condition Deux éléments de G' s'écrivent fg_1f^{-1} et fg_2f^{-1}. Nous les supposons tels que

$$(fg_1f^{-1}) \cdot \operatorname{Int}(K') \cap (fg_2f^{-1}) \cdot \operatorname{Int}(K') \neq \emptyset. \tag{17.556}$$

Nous avons :

$$\begin{aligned}(fg_1f^{-1}) \cdot \operatorname{Int}(K') \cap (fg_2f^{-1}) \cdot \operatorname{Int}(K') &= f\big(g_1 \cdot \operatorname{Int}(K)\big) \cap f\big(g_2 \cdot \operatorname{Int}(K)\big) && (17.557a)\\ &= f\big(\underbrace{g_1 \cdot \operatorname{Int}(K) \cap g_2 \cdot \operatorname{Int}(K)}_{\neq\emptyset}\big) && (17.557b)\\ &\neq \emptyset. && (17.557c)\end{aligned}$$

Justifications :

— Égalité (17.557a) parce qu'un peu de topologie nous enseigne que $f^{-1}\big(\operatorname{Int}(K')\big) = \operatorname{Int}\big(f^{-1}(K')\big) = \operatorname{Int}(K)$ parce que f est affine.

— Égalité (17.557b) parce que, f étant bijective, $f(A) \cap f(B) = f(A \cap B)$;

□

Théorème 17.188 ([268, 364]).
Nous notons τ_v la translation de vecteur v, $r_{A,\theta}$ la rotation de centre A et d'angle θ ainsi que $\tau_i = \tau_{e_i}$.

Un groupe G est un groupe de pavage de $\mathbb{R}^2$ si et seulement si il existe une bijection affine $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ telle que $f \circ G \circ f^{-1}$ est un groupe de la liste suivante :

(1) $\operatorname{gr}\big(\tau_1, \tau_2\big)$

(2) $\operatorname{gr}\big(\tau_1, \tau_2, R_0(\pi)\big)$

(3) $\operatorname{gr}\big(\tau_1, \tau_{(\frac{1}{2},\frac{\sqrt{3}}{2})}, R_0(2\pi/3)\big)$

(4) $\operatorname{gr}\big(\tau_1, \tau_2, R_0(\pi/2)\big)$

(5) $\operatorname{gr}\big(\tau_1, \tau_{(-\frac{1}{2},\frac{\sqrt{3}}{2})}, R_0(\pi/3)\big)$.

Démonstration. Les lemmes 17.179, 17.180, 17.181, 17.182, et 17.183 montrent que les groupes listés sont des groupes de pavage. La proposition 17.187 nous montre alors que si $H = fGf^{-1}$ est dans la liste, alors G est un groupe de pavage. Il nous reste à montrer que si G est un groupe de pavage, alors il existe une application affine α telle que $\alpha G \alpha^{-1}$ est un groupe de la liste.

Soit (G, K) un pavage de $\mathbb{R}^2$. Nous notons T l'ensemble des translations dans G, plus précisément,

$$T = \{v \in \mathbb{R}^2 \text{ tel que } \tau_v \in G\}. \tag{17.558}$$

Une borne pour T Nous prouvons qu'il existe $\delta > 0$ tel que $\|v\| > \delta$ pour tout $v \neq 0 \in T$. En effet, soit $m \in \operatorname{Int}(K)$ ainsi que r tel que $B(m, r) \subset \operatorname{Int}(K)$. Alors si $v \in T$ est tel que $\|v\| < r$ nous avons $\tau_v(m) \in \operatorname{Int}(K)$, ce qui donnerait

$$m \in \tau_v\big(\operatorname{Int}(K)\big) \cap \operatorname{Int}(K) \tag{17.559}$$

par hypothèse, cette intersection est non vide seulement si $v = 0$.

Donc il existe δ tel que $\|v\| \geqslant \delta$ pour tout $v \in T$.

Utilisation du lemme La partie T est donc dans la position du lemme 17.186 et nous avons

$$T = \mathbb{Z}u_1 + \mathbb{Z}u_2 \tag{17.560}$$

pour les vecteurs u_1 et u_2 les plus petits en norme de T.

En réalité, il se peut que T soit plus petit que ça, parce que G peut par exemple ne contenir aucune translations. Nous avons trois possibilités :

— $T = \{0\}$,

— $T = \mathbb{Z}u$ pour un certain $u \neq 0$ dans $\mathbb{R}^2$,

— $T = \mathbb{Z}u_1 + \mathbb{Z}u_2$ pour certains $u_1, u_2 \in \mathbb{R}^2 \backslash \{0\}$.

Translation Si $r, s \in G$, alors l'élément $rsr^{-1}s^{-1}$ est une translation. En effet, vu que les éléments de G sont des déplacements, ce sont des applications affines et donc il existe des applications linéaires A_r, A_s et des translations τ_r, τ_s telles que $r = A_r \circ \tau_r$ et $s = A_s \circ \tau_s$. Le lemme 8.57 nous donne les inverses. Nous avons

$$rsr^{-1}s^{-1} = (A_r \circ \tau_r)(A_s \circ \tau_s)(A_r^{-1} \circ \tau_{-A_r v_r})(A_s^{-1} \circ \tau_{-A_s v_s}). \tag{17.561}$$

La partie linéaire de cela est

$$A_r \circ A_s \circ A_r^{-1} \circ A_s^{-1}. \tag{17.562}$$

C'est donc une composée de rotations centrées en $(0,0)$. Mais ces rotations forment un groupe abélien (proposition 17.124). Donc nous pouvons écrire

$$A_r \circ A_s \circ A_r^{-1} \circ A_s^{-1} = A_r \circ A_r^{-1} \circ A_s \circ A_s^{-1} = \mathrm{Id}\,. \tag{17.563}$$

Tout ceci pour dire que dès que $r, s \in G$, l'élément $rsr^{-1}s^{-1}$ est une translation.

Les parties linéaires[1] Nous savons de l'exemple 8.14 que les éléments de G s'écrivent sous la forme $f = \tau_v \circ \alpha$ où $v \in \mathbb{R}^2$ et $\alpha\colon \mathbb{R}^2 \to \mathbb{R}^2$ est linéaire.

De plus, f étant une isométrie de $(\mathbb{R}^2, d)$, l'application α est une isométrie de $(\mathbb{R}^2, \|.\|)$. Vu que α est une isométrie, $\det(\alpha) = \pm 1$. Mais les déplacements conservent l'orientation ; donc α doit conserver l'orientation, et la proposition 9.27 nous dit que $\det(\alpha) > 0$. Donc

$$\det(\alpha) = 1. \tag{17.564}$$

Le théorème 17.74 dit que α est la composition d'un nombre pair de réflexions. Mais comme il y en a au plus trois (théorème 17.157), l'application α est composée de zéro ou deux réflexions.

Donc les parties linéaires des éléments de G sont des rotations.

Les autres Les parties linéaires des éléments de G sont des rotations. Mais les éléments de G eux-mêmes ne sont pas tellement mystérieux. Vu que ce sont des isométries de $(\mathbb{R}^2, d)$, elles sont composées de 0, 1, 2 ou 3 réflexions.

Mais ce sont des déplacements, donc ils préservent l'orientation et le théorème 17.74 dit qu'ils sont des composées de zéro ou deux réflexions (nombre pair). Ce sont donc des rotations.

Hein ? Les éléments linéaires de G sont des rotations. Et les autres aussi ? Les linéaires sont des rotations autour de $(0,0)$; les autres sont des rotations autour de points autres que $(0,0)$.

C'est pourquoi dans la suite, nous préciserons « rotation linéaire » pour une rotation autour de $(0,0)$ et nous dirons « rotation » pour une rotation en général. Dans le contexte affine, il faut toujours faire attention à ça : une rotation peut très bien n'être pas linéaire [49].

Les rotations linéaires stabilisent T Nous prouvons maintenant que les rotations linéaires de G stabilisent T, c'est-à-dire que si $v \in T$ et si α est une rotation linéaire de G, alors $\alpha(v) \in T$. La transformation $\alpha\tau_v\alpha^{-1}$ est dans G. Mais pour tout $x \in \mathbb{R}^2$ nous avons

$$(\alpha\tau_v\alpha^{-1})(x) = \alpha\big(\alpha^{-1}(x) + v\big) = x + \alpha(v) = \tau_{\alpha(v)}(x). \tag{17.565}$$

Donc $\alpha\tau\alpha^{-1} = \tau_{\alpha(v)}$ et $\alpha(v) \in T$.

49. Lorsque, ailleurs dans le Frido, nous disons « rotation », souvent nous pensons « rotation linéaire ». Gardez cependant à l'esprit qu'une rotation peut très bien être centrée ailleurs qu'en l'origine, et soyez toujours capable de préciser le cas échéant.

Exclusion de $T = \{0\}$ Le fait que $T = \{0\}$ ne signifie pas que tous les éléments de G sont des rotations ; il peut encore y avoir des composées de rotations et de translations $A \circ \tau$. Cela étant dit, si $T = \{0\}$, il n'en reste pas moins que $rsr^{-1}s^{-1}$ est une translation, c'est-à-dire est égal à Id. Mais $rsr^{-1}s^{-1} = e$ implique $rs = sr$.

Donc G est abélien. Les éléments de G sont donc des rotations qui commutent deux à deux. Vu qu'une rotation a son centre comme unique point fixe, le lemme 17.185 nous dit que tous les éléments de G sont des rotations de même centre.

Soit c le centre commun de tous les éléments de G. Vu que K est compact dans $\mathbb{R}^2$, il existe $r > 0$ tel que $K \subset B(c, r)$. Vu que G stabilise toutes les boules centrées en c, nous avons

$$G \cdot K \subset B(c, r). \tag{17.566}$$

Donc nous n'avons pas un recouvrement de $\mathbb{R}^2$. Le cas $T = \{0\}$ est exclu.

Exclusion de $T = \mathbb{Z}u$ Nous supposons à présent que $T = \mathbb{Z}u$ pour un certain $u \in \mathbb{R}^2$.

$r_0 = \pm \mathrm{Id}$ Nous savons que tous les éléments de G sont des rotations ; soit un élément r de G. La proposition 17.123(3) nous indique qu'il existe un point $a \in \mathbb{R}^2$ ainsi qu'une rotation linéaire r_0 telle que $r = \tau_a^{-1} r_0 \tau_a$. Nous allons prouver que r_0 est $\pm \mathrm{Id}$. D'abord,

$$r\tau_u r^{-1} = \tau_a^{-1} r_0 \tau_a \tau_u \tau_a^{-1} r_0^{-1} \tau_a = \tau_a^{-1} r_0 \tau_u r_0^{-1} \tau_a. \tag{17.567}$$

Ensuite, nous appliquons cela à $x \in \mathbb{R}^2$:

$$\begin{aligned}
(\tau_a^{-1} r_0 \tau_u r_0^{-1} \tau_a)(x) &= (\tau_a^{-1} r_0 \tau_u)\big(r_0^{-1}(x + a)\big) && \text{(17.568a)}\\
&= (\tau_a^{-1} r_0)\big(r_0^{-1}(x + a) + u\big) && \text{(17.568b)}\\
&= \tau_a^{-1}\big(x + a + r_0(u)\big) && \text{(17.568c)}\\
&= x + r_0(u). && \text{(17.568d)}
\end{aligned}$$

Donc $r \circ \tau_u \circ r^{-1} = \tau_{r_0(u)}$, ce qui prouve que $r_0(u) \in T$. Vu que $\|r_0(u)\| = \|u\|$ nous avons forcément $r_0(u) = \pm u$.

Si $r_0(u) = u$, alors $r_0 = \mathrm{Id}$ (parce que r_0 est une rotation fixant plus que un seul point). Dans ce cas, $r = \mathrm{Id}$.

Si au contraire $r_0(u) = -u$, alors $r_0 = -\mathrm{Id}$.

Forme générale Donc si r est un élément non trivial de G nous avons $r = \tau_a^{-1} \circ (-\mathrm{Id}) \circ \tau_a$, et alors

$$\big(\tau_a^{-1} \circ (-\mathrm{Id}) \circ \tau_a\big)(x) = \tau_a^{-1}(-\mathrm{Id})(x + a) = \tau_a^{-1}(-x - a) = -x - 2a. \tag{17.569}$$

Donc pour tout $r \in G$, il existe $a \in \mathbb{R}^2$ tel que

$$r(x) = -x - 2a. \tag{17.570}$$

Pour information, le centre de cette rotation est $-a$ (c'est le seul point fixe).

Les centres sont alignés Soit r une rotation de centre $-a$ et s de centre $-b$. Alors

$$(rs)(x) = r(-x - 2b) = x + 2b - 2a = x + 2(b - a). \tag{17.571}$$

Donc $rs = \tau_{2(b-a)}$.

Cela prouve que $2(b - a) \in \mathbb{Z}u$.

Une bande Soit la droite $D = \mathbb{R}u$. Nous considérons la bande

$$B_r = \{x \in \mathbb{R}^2 \text{ tel que } d(x, D) < r\}. \tag{17.572}$$

Une inclusion Nous prouvons à présent que pour tout r, nous avons

$$G \cdot B_r \subset B_r. \tag{17.573}$$

Nous savons qu'un élément général de G est une rotation centrée en un point de D, et que l'action d'une telle rotation est donnée par (17.570). Nous avons

$$\begin{aligned} d\big(r(x), D\big) &= d(x + 2a, D) && (17.574a)\\ &= d\big(\tau_{2a}^{-1}(x + 2a), \tau_{2a}^{-1}(D)\big) && (17.574b)\\ &= d(x, D). && (17.574c) \end{aligned}$$

Justifications :

— Pour (17.574a), nous avons $D = -D$ et $d(x, y) = d(-x, -y)$.

— Pour (17.574b), invariance par translation de la distance dans $\mathbb{R}^2$.

— Pour (17.574c), les éléments de D sont les multiples de a ; donc cette droite est invariante par cette translation.

Bref, si $d(x, D) = d\big(r(x), D\big)$. Donc pour tout $r > 0$ [50] et pour tout $g \in G$, si $x \in B_r$, alors $g(x) \in B_r$.

Exclusion Vu que K est compact et que la fonction $x \mapsto d(x, D)$ est continue, il existe $r > 0$ tel que $K \subset B_r$. Avec ça, $G \cdot K \subset G \cdot B_r \subset B_r$. Donc $G \cdot K$ ne recouvre par tout $\mathbb{R}^2$ et G n'est pas un groupe de pavage.

Et nous voila avec seulement $T = \mathbb{Z}u_1 + \mathbb{Z}u_2$ en lice.

Pause : quelques parties de G à ne pas confondre Il convient de ne pas se perdre entre différentes parties de G. Je vous laisse méditer quelque temps sur la liste suivante :

(1) G est le groupe que nous cherchons à déterminer ;

(2) T est le groupe des translations de G ;

(3) le groupe des rotations linéaires dans G ;

(4) l'ensemble des $\tau_A^{-1} \circ r \circ \tau_A$ où r est une rotation de G centrée en A.

(5) L est l'ensemble des parties linéaires des éléments de G.

En particulier les deux derniers ne sont pas les mêmes.

Dans le cas (4), il s'agit de dire que r est une rotation centrée en $A \in \mathbb{R}^2$ et écrire $r = \tau_A \circ r_0 \circ \tau_A^{-1}$ (proposition 17.123(3)) et considérer r_0. Dans ce cas, r_0 est une rotation, mais elle n'est pas ce que nous appelons la « partie linéaire » de r. Il n'y a pas de garanties que cela forme un groupe.

Si r est une rotation dans G, dans le cas (5) il s'agit de décomposer $r = \tau_v \circ \alpha$ (lemme 8.10) et considérer α. Dans ce cas, α est linéaire, mais il n'y a pas de garanties que α soit une rotation.

L est un groupe En trois conditions.

— Si $\alpha, \beta \in L$, il existe $v, w \in \mathbb{R}^2$ tels que $\tau_v \circ \alpha \in G$ et $\tau_w \circ \beta \in G$. La loi de produit 8.57(2) dit que $\tau_v \circ \alpha \circ \tau_w \circ \beta = \tau_{\alpha(w)+v} \circ \alpha\beta$. Donc $\alpha\beta$ est la partie linéaire d'un élément de G.

— De la même façon, en utilisant l'inverse 8.57(3), $(\tau_v \circ \alpha)^{-1} = \tau_{-\alpha^{-1}(v)} \circ \alpha^{-1}$. Donc α^{-1} est la partie linéaire d'un élément de G.

— Et enfin $\mathrm{Id} = \tau_0 \circ \mathrm{Id}$. Donc l'identité est dans L.

Ok pour que L soit un groupe.

Précision L'ensemble L est un groupe certes. Mais rien ne dit que L soit un sous-groupe de G.

50. Remarquez la notation malheureuse pour r qui est maintenant une distance alors que trois mots plus tôt, c'était un élément de G.

L préserve le réseau Soit $\alpha \in L$. Il existe $v \in \mathbb{R}^2$ tel que $g = \tau_v \circ \alpha \in G$. Soit $u \in T$. Nous allons montrer que $\alpha(u) \in T$. Vu que g et τ_u sont dans G, l'élément $g\tau_u g^{-1}$ est également dans G. Nous l'appliquons à $x \in \mathbb{R}^2$:

$$\begin{aligned}
(\tau_v\alpha\tau_u\alpha^{-1}\tau_v^{-1})(x) &= (\tau_v\alpha\tau_u\alpha^{-1})\alpha^{-1}(x-v) && (17.575a)\\
&= (\tau_v\alpha)\big(\alpha^{-1}(x-v)+u\big) && (17.575b)\\
&= \tau_v\big(x-v+\alpha(u)\big) && (17.575c)\\
&= x+\alpha(u) && (17.575d)\\
&= \tau_{\alpha(u)}(x). && (17.575e)
\end{aligned}$$

Donc $g\tau_u g^{-1} = \tau_{\alpha(u)} \in G$.

Question de trace Soit $\alpha \in L$; dans la base $\{u_1, u_2\}$ la matrice de α est

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{17.576}$$

avec $a, b, c, d \in \mathbb{Z}$. La trace de cette matrice est $a + d \in \mathbb{Z}$. Dans la base canonique de $\mathbb{R}^2$ par contre la proposition 17.134 nous dit qu'il existe $\theta \in [0, 2\pi[$ tel que la matrice de α soit

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}. \tag{17.577}$$

La trace est $2\cos(\theta)$. La trace est invariante par changement de base[51], donc $2\cos(\theta) = a + d \in \mathbb{Z}$. Les possibilités pour $\cos(\theta)$ sont donc -1, $-1/2$, 0, $1/2$ et 1.

Les angles possibles Nous savons que $\cos(\frac{\pi}{2}) = 0$ et $\cos(\pi/3) = 1/2$ (proposition 17.17(6) et lemme 17.20). La proposition 17.56 nous dit alors que, dans notre cas, les valeurs possibles pour θ dans $[0, 2\pi[$ sont

$$\{0, \frac{\pi}{3}, \frac{\pi}{2}, \frac{2\pi}{3}, \pi, \frac{4\pi}{3}, \frac{3\pi}{2}, \frac{5\pi}{3}\}. \tag{17.578}$$

Donc les rotations possibles dans L sont les rotations de ces angles.

Nous devons trouver quels sont les groupes qui peuvent être formés seulement avec ces éléments.

Quelques combinaisons impossibles Vu que L est un groupe, il y a des combinaisons impossibles. Par exemple si $R_0(\pi/3)$ et $R_0(\pi/2)$ sont dans L, alors la composée[52] $R_0(\pi/2)R_0(\pi/3) = R_0(5\pi/6)$ est également dans L. Mais comme $5\pi/6$ n'est pas dans la liste (17.578), $R_0(5\pi/6)$ n'est pas dans L.

En raisonnant de la sorte, nous voyons que si $R_0(\pi/2) \in L$, alors $L = \mathrm{gr}\big(R_0(\pi/2)\big)$.

La liste Plus généralement, les possibilités pour L sont

- $\{\mathrm{Id}\}$
- $\mathrm{gr}\big(R_0(\pi/2)\big)$
- $\mathrm{gr}\big(R_0(\pi/3)\big)$
- $\mathrm{gr}\big(R_0(\pi)\big)$
- $\mathrm{gr}\big(R_0(2\pi/3)\big)$

Une justification plus courte pour cette liste est d'invoquer le théorème 17.169(2) qui dit que L étant un sous-groupe fini des isométries de $(\mathbb{R}^2, d)$, il est cyclique et donc monogène. Notons pour cela que $\mathrm{gr}\big(R_0(4\pi/3)\big) = \mathrm{gr}\big(R_0(2\pi/3)\big)$ parce que $R_0(4\pi/3) = R_0(2\pi/3)^{-1}$. De la même manière $R_0(3\pi/2) = R_0(\pi/2)^{-1}$ et $R_0(5\pi/3) = R_0(\pi/3)^{-1}$.

51. Proposition 9.189.
52. Proposition 17.135 pour l'addition des angles.

Le cas un peu générique Nous supposons que $L = \mathrm{gr}\big(R_0(\theta)\big)$ pour un certain θ. Nous allons voir qu'à part dans les cas $\theta = 0$ et $\theta = \pi$, il est possible de trouver une application affine α telle que le groupe $\alpha G \alpha^{-1}$ est alors dans la liste.

Nous considérons un élément de G de la forme $\tau_{v_0} \circ R_0(\theta)$. Pour être bien clair, il n'est absolument pas garanti que v_0 soit dans T. Nous allons chercher un élément $w_0 \in \mathbb{R}^2$ tel que le groupe de pavage[53] $G' = \tau_{w_0} G \tau_{w_0}^{-1}$ contienne $R_0(\theta)$. Le groupe G' contient l'élément $g = \tau_{w_0}\tau_{v_0} R_0(\theta)\tau_{w_0}^{-1}$; nous l'appliquons à $x \in \mathbb{R}^2$:

$$g(x) = R_0(\theta)x - R_0(\theta)w_0 + w_0 + v_0. \tag{17.579}$$

Nous avons $g(x) = R_0(\theta)x$ lorsque

$$w_0 = \big(R_0(\theta) - \mathbb{1}\big)^{-1} v_0. \tag{17.580}$$

Voila pourquoi le cas $\theta = 0$ sera traité à part : dans le cas $\theta = 0$, l'opérateur $R_0(\theta) - \mathbb{1}$ n'est pas inversible. Dans les autres cas, nous avons un groupe $G' = \tau_{w_0} G \tau_{w_0}^{-1}$ qui contient $R_0(\theta)$. Vu qu'un élément général de G est de la forme $\tau_v \circ R_0(\theta)^k$, un élément général de G' est de la forme

$$\tau_{w_0} \circ \tau_v \circ R_0(\theta)^k \circ \tau_{w_0} = \tau_{w_0 + v - R_0(\theta)^k w} \circ R_0(\theta)^k. \tag{17.581}$$

Donc tous les éléments de G' sont encore de la forme

$$\tau_v \circ R_0(\theta)^k. \tag{17.582}$$

Mais dans G' nous avons une information capitale : $R_0(\theta)^k$ lui-même est dans G. Donc si $\tau_v \circ R_0(\theta)^k \in G'$, alors $\tau_v \in G'$.

Vu que G' est encore un groupe de pavage, tout ce qui a été dit précédemment tient et le groupe des translations dans G' est un réseau $T' = \mathbb{Z}u_1 + \mathbb{Z}u_2$ où u_1 et u_2 peuvent être choisis arbitrairement parmi les deux plus petits vecteurs non colinéaires de T.

Tous les éléments de G' sont de la forme $\tau_v \circ R_0(\theta)^k$ avec $v \in T'$. Donc

$$G' = \mathrm{gr}\big(\tau_{u_1}, \tau_{u_2}, R_0(\theta)\big). \tag{17.583}$$

Nous savons que $R_0(\theta)$ fixe T'. Or l'élément $R_0(\theta)u_1$ a la même norme que u_1 ; donc nous pouvons choisir $u_2 = R_0(\theta)$ pour peu que $R_0(\theta)u_1$ soit non colinéaire à u_1. Et c'est ici que nous laissons le cas $\theta = \pi$ de côté.

Nous écrivons donc sans vergogne que

$$G' = \mathrm{gr}\big(\tau_{u_1}, \tau_{R_0(\theta)u_1}, R_0(\theta)\big). \tag{17.584}$$

Nous allons maintenant nous occuper de u_1. Pour cela nous considérons une rotation suivie d'une homothétie α telle que $\alpha(u_1) = e_1$. Une telle opération α d'une part commute avec $R_0(\theta)$ et d'autre part fait $\alpha \circ \tau_v \circ \alpha^{-1} = \tau_{\alpha(v)}$. Donc le groupe $G'' = \alpha G' \alpha^{-1}$ est, par le lemme 2.10,

$$G'' = \mathrm{gr}\big(\tau_{e_1}, \tau_{\alpha R_0(\theta)u_1}, R_0(\theta)\big). \tag{17.585}$$

Cela s'écrit aussi bien

$$G'' = \mathrm{gr}\big(\tau_{e_1}, \tau_{R_0(\theta)e_1}, R_0(\theta)\big). \tag{17.586}$$

Et voila, ce groupe G'' est un de la liste.

Le cas $L = \{\mathrm{Id}\}$ Dans ce cas, tous les éléments de G sont de la forme τ_v avec $v \in T$. Nous considérons l'application linéaire $\alpha\colon \mathbb{R}^2 \to \mathbb{R}^2$ telle que $\alpha(u_1) = e_1$ et $\alpha(u_2) = e_2$. Nous avons

$$(\alpha \circ \tau_{u_i} \circ \alpha^{-1})(x) = (\alpha\tau_{u_i})\alpha^{-1}(x) = \alpha\big(\alpha^{-1}(x) + u_i\big) = x + \alpha(u_i) = \tau_i(x). \tag{17.587}$$

53. Proposition 17.187.

Donc $\alpha \circ \tau_{u_i} \circ \alpha^{-1} = \tau_i$. De la même façon, si $a \in \mathbb{Z}$ nous avons $\alpha \tau_{au_i} \alpha^{-1} = \tau_{ae_i}$. Et avec tout ça, si $v \in T$, alors $v = au_1 + bu_2$ et nous avons

$$\alpha \tau_v \alpha^{-1} = \alpha \tau_{au_1} \tau_{au_2} \alpha^{-1} = \alpha \tau_{au_1} \alpha^{-1} \alpha \tau_{bu_2} \alpha^{-1} = \tau_{ae1} \tau_{be_2}. \tag{17.588}$$

Nous avons donc

$$\alpha G \alpha^{-1} = \mathrm{gr}(\tau_1, \tau_2). \tag{17.589}$$

Cela est un des groupes de la liste.

Notez qu'à la place de ces calculs, nous pouvions aussi invoquer la proposition 2.10.

Le cas $L = \{-\mathrm{Id}\}$ Dans le cas $\theta = \pi$, nous pouvons aller jusqu'à (17.583) et écrire

$$G' = \mathrm{gr}\big(\tau_{u_1}, \tau_{u_2}, R_0(\pi)\big). \tag{17.590}$$

Vu que $R_0(\pi) = -\mathrm{Id}$ commute avec toutes les applications linéaires et que u_1 et u_2 ne sont pas colinéaires, nous pouvons considérer une application linéaire $\alpha\colon \mathbb{R}^2 \to \mathbb{R}^2$ telle que $\alpha(u_1) = e_1$, $\alpha(u_2) = e_2$. Nous avons alors

$$G'' = \alpha G' \alpha^{-1} = \mathrm{gr}\big(\tau_{e_1}, \tau_{e_2}, -\mathrm{Id}\big), \tag{17.591}$$

qui est encore dans la liste.

□

17.11 Un peu de structure de $\mathrm{O}(n)$

17.11.1 Valeurs propres dans $\mathrm{O}(n)$

Proposition 17.189 ([365]).
Soit une matrice $A \in O(n)$. Si $\lambda \in \mathbb{C}$ est une valeur propre de A, alors $\bar{\lambda}$ est également une valeur propre de A, et de plus $|\lambda| = 1$.

Démonstration. Dire que $\lambda \in \mathbb{C}$ est une valeur propre de A signifie qu'il existe $x \in \mathbb{C}^n$ (non nul) tel que $Ax = \lambda x$. Vu que les éléments de la matrice A sont réels,

$$A\bar{x} = \bar{A}\bar{x} = \overline{Ax} = \overline{\lambda x} = \bar{\lambda}\bar{x}. \tag{17.592}$$

Donc $\bar{\lambda}$ est une valeur propre de A pour le vecteur propre $\bar{x}$.

Soit λ une valeur propre de A de vecteur propre x. Alors nous avons d'une part

$$\langle \overline{Ax}, Ax \rangle = \langle A^t A\bar{x}, x \rangle = \langle x, \bar{x} \rangle, \tag{17.593}$$

et d'autre part

$$\langle \overline{Ax}, Ax \rangle = \langle \bar{\lambda}\bar{x}, \lambda x \rangle = |\lambda|^2 \langle \bar{x}, x \rangle. \tag{17.594}$$

Vu que $x \neq 0$ nous avons aussi $\langle \bar{x}, x \rangle \neq 0$. Par conséquent $|\lambda|^2 = 1$ et $|\lambda| = 1$. □

Lemme 17.190 ([365]).
Soit un espace vectoriel euclidien E de dimension finie et une isométrie f de E. Soit F un sous-espace de E stable par f. Alors $F^\perp$ est stable par f.

Démonstration. La restriction $f_F\colon F \to F$ est encore une isométrie ; elle est donc inversible : pour tout $y \in F$, il existe $x \in F$ tel que $y = f(x)$. Soit $a \in F^\perp$; nous montrons que $f(a) \in F^\perp$. Soit donc $y \in F$ et calculons :

$$\langle y, f(a) \rangle = \langle f(x), f(a) \rangle = \langle x, a, \rangle = 0 \tag{17.595}$$

parce que $x \in F^\perp$. □

Proposition 17.191 ([365]).
Soit une isométrie $f\colon \mathbb{R}^3 \to \mathbb{R}^3$.

(1) L'application linéaire f possède au moins une valeur propre réelle qui vaut ± 1.

(2) Il existe une base orthonormée de $\mathbb{R}^3$ dans laquelle la matrice de f est de la forme

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \cos(\theta) & -\epsilon\cos(\theta) \\ 0 & \sin(\theta) & \epsilon\cos(\theta) \end{pmatrix} \tag{17.596}$$

avec $\epsilon, \lambda = \pm 1$ et $\theta \in [0, 2\pi[$.

Démonstration. Le polynôme caractéristique de f, donné par $\det(f - \lambda\,\mathrm{Id})$, est à coefficients réels et de degré 3. Il possède dont au moins une solution réelle par le corolaire 11.100. Soit donc une valeur propre réelle λ de χ_f ; par le lemme 17.189 nous avons $\lambda = \pm 1$. Soit u_1 le vecteur propre correspondant. Nous notons F l'espace engendré par u_1.

Nous avons $f(F) = F$ et donc $f(F^\perp) = F^\perp$ par le lemme 17.190. Soit une base orthonormée $\{u_2, u_3\}$ de $F^\perp$ et la matrice B de la restriction f_p à $F^\perp$. Vu que l'application f_p est une isométrie de $F^\perp$, la matrice B est, par le lemme 17.130, de la forme

$$B = \begin{pmatrix} \cos(\theta) & -\epsilon\sin(\theta) \\ \sin(\theta) & \epsilon\cos(\theta) \end{pmatrix} \tag{17.597}$$

pour un certain $\theta \in [0, 2\pi[$ et $\epsilon = \pm 1$.

Dans la base $\{u_1, u_2, u_3\}$ de $\mathbb{R}^3$, la matrice de f est alors

$$\begin{pmatrix} \lambda & 0 \\ 0 & B \end{pmatrix}, \tag{17.598}$$

comme annoncé. □

Pour classifier les isométries de $\mathbb{R}^3$, nous pouvons nous baser sur les possibilités de la matrice donne dans le lemme 17.191. Il y a essentiellement quatre possibilités suivant les valeurs de $\lambda = \pm 1$ et $\epsilon = \pm 1$.

Si $\epsilon = \lambda = 1$ Alors la matrice est

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{17.599}$$

et l'isométrie correspondante est la rotation d'angle $-\theta$ autour de la droite de u_1.

Si $\epsilon = \lambda = -1$ Alors la matrice est

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & \sin(\theta) & -\cos(\theta) \end{pmatrix} \tag{17.600}$$

Cette application est plus subtile, parce que même dans le plan $\mathrm{Span}(u_2, u_3)$, ce n'est pas une rotation. Nous allons montrer qu'il s'agit d'une réflexion autour de la droite d'angle $\theta/2$ dans le plan $\mathrm{Span}(u_2, u_3)$. Nous nommons D cette droite. Dans la base $\{u_1, u_2, u_3\}$, les points de cette droite sont de la forme[54]

$$\big(0, \cos(\theta/2), \sin(\theta/2)\big). \tag{17.601}$$

L'image de u_1 par cette réflexion est $-u_1$, c'est clair.

Faisons en détail l'image de u_3. Nous devons démontrer que la droite D coupe le segment $[u_3, A(u_3)]$ perpendiculairement en son milieu.

54. Les plus acharnés remarqueront que $\{u_1, u_2, u_3\}$ est un ensemble, qui est une base. Mais un ensemble n'est pas ordonné, alors que pour écrire l'équation de droite qui suit, nous supposons un ordre. Je laisse au tel lecteur le soin de trouver une bonne notation.

Dans le plan $\mathrm{Span}(u_2, u_3)$ nous avons $u_3 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ et $A(u_3) = \begin{pmatrix} \sin(\theta) \\ \cos(\theta) \end{pmatrix}$. Le milieu du segment $[u_3, A(u_3)]$ est le point

$$M = \left(\frac{\sin(\theta)}{2}, \frac{1 - \cos(\theta)}{2} \right). \tag{17.602}$$

Les formules de duplication d'angle du corolaire 17.12 nous permettent d'écrire $\sin(\theta)$ et $\cos(\theta)$ en fonction de $\sin(\theta/2)$ et $\cos(\theta/2)$, et donc d'exprimer le point M de la façon suivante :

$$M = \left(\cos(\theta/2)\sin(\theta/2), \frac{1 - \big(\cos^2(\theta/2) - \sin^2(\theta/2)\big)}{2} \right) \tag{17.603a}$$

$$= \big(\cos(\theta/2)\sin(\theta/2), \sin^2(\theta/2)\big) \tag{17.603b}$$

$$= \sin(\theta/2)\big(\cos(\theta/2), \sin(\theta/2)\big). \tag{17.603c}$$

Ce point fait donc partie de la droite D. La droite D coupe le segment $[u_3, A(u_3)]$ en son milieu.

En ce qui concerne l'orthogonalité, nous calculons le produit scalaire

$$\big(A(u_3) - u_3\big) \cdot \begin{pmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} = \sin(\theta)\cos(\theta/2) - \big(1 + \cos(\theta)\big)\sin(\theta/2) = 0 \tag{17.604}$$

où nous avons encore utilisé les duplications d'angles et le fait que $1 = \cos^2(\theta/2) + \sin^2(\theta/2)$ (lemme 17.3).

Si $\epsilon = -1$ et $\lambda = 1$ Alors la matrice est

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & \sin(\theta) & -\cos(\theta) \end{pmatrix}. \tag{17.605}$$

C'est la symétrie orthogonale par le plan engendré par u_1 et $v = \cos(\theta/2)u_2 + \sin(\theta/2)u_3$. Le vecteur u_1 est bien évidemment préservé par A. En ce qui concerne le vecteur v,

$$A(v) = \cos(\theta/2) \begin{pmatrix} 0 \\ \cos(\theta) \\ \sin(\theta) \end{pmatrix} + \sin(\theta/2) \begin{pmatrix} 0 \\ -\sin(\theta) \\ \cos(\theta) \end{pmatrix} = \begin{pmatrix} 0 \\ \cos(\theta/2) \\ \sin(\theta/2) \end{pmatrix} = v. \tag{17.606}$$

Nous avons sauté quelques étapes de calcul mettant en scène les formules de duplication d'angle : exprimer $\cos(\theta) = \cos^2(\theta/2) - \sin^2(\theta/2)$ et $\sin(\theta) = 2\cos(\theta/2)\sin(\theta/2)$.

Pour achever, nous devons trouver un vecteur w perpendiculaire au plan, et montrer qu'il est envoyé par A sur $-w$. Un vecteur $w = xu_1 + yu_2 + zu_3$ est perpendiculaire au plan si les deux égalités suivantes sont satisfaites :

$$\big(\cos(\theta/2)u_2 + \sin(\theta/2)u_3\big) \cdot (xu_1 + yu_2 + zu_3) = 0 \tag{17.607a}$$

$$u_1 \cdot (xu_1 + yu_2 + zu_3) = 0. \tag{17.607b}$$

Nous avons immédiatement $x = 0$ et ensuite la relation

$$y\cos(\theta/2) + z\sin(\theta/2) = 0. \tag{17.608}$$

En ne regardant que les deux dernières composantes pour alléger l'écriture,

$$A(w) = y \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} + z \begin{pmatrix} \sin(\theta) \\ -\cos(\theta) \end{pmatrix}. \tag{17.609}$$

Le but est de montrer que cela est égal à $-y\cos(\theta/2) - z\sin(\theta/2)$.

Notons $c = \cos(\theta/2)$ et $s = \sin(\theta/2)$. Alors $A(w)_2 = y(c^2 - s^2) + 2zcs$. Évacuons tout de suite les deux cas limite : si $c = 0$ alors $A(w)_2 = -y$ (parce que $s = \pm 1$) et c'est bon. Si $s = 0$, alors $A(w)_2 = y$, mais la relation (17.608) donne $y = 0$, donc c'est bon aussi. Dans le cas générique, $z = -yc/2$ et

$$A(w)_2 = y(c^2 - s^2) - 2cs\frac{yc}{s} = -y(c^2 + s^2) = -y. \tag{17.610}$$

En ce qui concerne $A(w)_3$, c'est très similaire :

$$A(w)_3 = 2ysc - z(c^2 - s^2). \tag{17.611}$$

Avec $z = 0$ c'est $-z$, donc c'est bon. Avec $c = 0$ c'est z mais $z = 0$. Et pour le cas générique, la substitution $y = -zs/c$ donne le résultat.

Si $\epsilon = 1$ et $\lambda = -1$ Alors la matrice est

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}. \tag{17.612}$$

Cela est la composition entre la symétrie de plan $\mathrm{Span}(u_2, u_3)$ et la rotation d'angle θ dans ce plan.

17.11.2 Sous-groupes finis de SO(3)

Lemme 17.192 ([1]).
Points fixes pour SO(3).

(1) Tout élément de SO(3) *possède une droite de points fixes.*

(2) Tout élément non trivial de SO(3) *possède une seule droite de points fixes.*

Démonstration. Le polynôme caractéristique d'un élément de SO(3) est de degré trois et possède dont (en comptant les multiplicités) trois racines dont une réelle par le corolaire 11.100. Vu que nous sommes en dimension impaire, le coefficient du terme de degré 3 est -1 et le polynôme caractéristique de $g \in \mathrm{SO}(3)$ s'écrit

$$\chi_g(X) = -(X - \lambda_1)(X - \bar{\lambda}_1)(X - s) \tag{17.613}$$

avec $s = \pm 1$ que nous allons tout de suite fixer. Nous savons que $\det(g) = \chi_g(0)$ mais aussi que $\det(g) = 1$. Donc

$$1 = \det(g) = \chi_g(0) = \lambda_1\bar{\lambda}_1 s = s. \tag{17.614}$$

Tout cela pour dire que tout élément de SO(3) possède une valeur propre égale à 1, et donc une droite de points fixes.

Pour continuer, supposons que g possède deux droites distinctes de points fixes. En particulier g fixe un plan. Une base orthonormée de $\mathbb{R}^3$ peut être choisie en prenant deux vecteurs e_1, e_2 dans ce plan et un vecteur e_3 perpendiculaire au plan.

Vu que g est une isométrie, la base reste orthonormée sous l'action de g. Donc g a pour matrice

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \pm 1 \end{pmatrix}. \tag{17.615}$$

Pour que le déterminant soit 1, il faut que la matrice soit l'identité. □

Proposition 17.193 ([1, 366, 71, 367, 311]).
Les sous-groupes finis de SO(3) *sont :*

(1) les groupes cycliques $\mathbb{Z}/n\mathbb{Z}$,
(2) les groupes diédraux D_n,
(3) le groupe alterné A_4,
(4) le groupe alterné A_5
(5) le groupe symétrique S_4.

Démonstration. Soit G, un sous-groupe fini de SO(3). Par la proposition 9.42, les éléments de G sont des isométries de $\mathbb{R}^3$, et le lemme 17.192 dit que tout élément de G possède une droite de points fixes.

Un point de la sphère unité fixé par $g \in G$ est un **pôle** de g. Nous nommons Ω l'ensemble des pôles des éléments non triviaux de G.

Une action Le groupe G agit sur Ω. En effet si $x \in \Omega$, alors x est fixé par un élément g. Montrons que $h(x)$ est également fixé par un élément de G. Par dur : $(hgh^{-1})h(x) = h(x)$; donc $h(x)$ est un pôle de hgh^{-1}.

Les fixateurs sont cycliques Nous montrons à présent que pour tout $x \in \Omega$, le sous-groupe $\mathrm{Fix}(x)$ est cyclique. Soit donc $x \in \Omega$, le plan orthogonal $\sigma = \mathrm{Span}(x)^\perp$ et $h \in \mathrm{Fix}(x)$. Nous avons $h(\sigma) = \sigma$. En effet si $y \in \sigma$ nous avons

$$0 = y \cdot x = h(y) \cdot h(x) = h(y) \cdot x, \tag{17.616}$$

donc $h(y)$ est perpendiculaire à x. L'inclusion inverse se démontre de même : si $y \in \sigma$ alors $y = h\big(h^{-1}(y)\big)$ alors que $h^{-1}(y) \in \sigma$.

La restriction de h à σ est une isométrie de σ. Prenant une isométrie $f\colon \sigma \to \mathbb{R}^2$, l'application

$$\begin{aligned} \varphi\colon \mathrm{Fix}(x) &\to \mathrm{SO}(2) \\ h &\mapsto f \circ h \circ f^{-1}. \end{aligned} \tag{17.617}$$

est un morphisme injectif de groupes. En effet nous avons d'une part

$$\varphi(hh') = f \circ h \circ h' \circ f^{-1} = fhf^{-1}fh'f^{-1} = \varphi(h)\varphi(h'), \tag{17.618}$$

d'où le morphisme. Et d'autre part, si $\varphi(h) = \varphi(h')$ alors $f \circ h \circ f^{-1} = g \circ h' \circ f^{-1}$, qui donne immédiatement $h = h'$.

Nous en déduisons que $\mathrm{Fix}(x)$ est isomorphe à un sous-groupe de $SO(2)$ (l'image de φ). Le lemme 17.137 en fait un groupe cyclique.

Taille des fixateurs Soient Ω_i les orbites. Si $x, y \in \Omega_i$ alors nous montrons que $|\mathrm{Fix}(x)| = |\mathrm{Fix}(y)|$ avec la bijection

$$\begin{aligned} \varphi\colon \mathrm{Fix}(x) &\to \mathrm{Fix}(y) \\ h &\mapsto g^{-1}hg \end{aligned} \tag{17.619}$$

où g est choisi de façon à avoir $y = g(x)$ (possible parce que x et y sont dans la même orbite). Cela est surjectif parce que si $k \in \mathrm{Fix}(x)$ alors $k = \varphi(gkg^{-1})$ et l'on vérifie que $gkg^{-1} \in \mathrm{Fix}(y)$. L'application φ est également injective parce que si $ghg^{-1} = gh'g^{-1}$ alors $h = h'$.

Un peu de notations Vu que tous les fixateurs des éléments d'une orbite ont la même taille (finie), nous pouvons noter

$$n_i = |\mathrm{Fix}(x_i)| \tag{17.620}$$

pour $x_i \in \Omega_i$. Nous notons également r le nombre d'orbites de G.

La formule de Burnside du théorème 2.59, avec les notations d'ici, donne

$$r = \frac{1}{|G|} \sum_{g \in G} |\mathrm{Fix}(g)|. \tag{17.621}$$

Une belle formule Soit l'ensemble

$$A = \{(g, x) \text{ tel que } g \in G \backslash \{e\}, x \in \mathrm{Fix}(g)\} \tag{17.622}$$

où par $\mathrm{Fix}(g)$ nous entendons les pôles de G fixés par g.

Il y a $|G|-1$ possibilités pour la composante g, mais chaque élément $g \neq e$ possède exactement deux pôles, donc l'ensemble A contient exactement $2(|G|-1)$ éléments.

Nous pouvons calculer le nombre d'éléments dans A d'une autre façon : pour chaque $x \in \Omega$ nous avons $|\mathrm{Fix}(x)|-1$ éléments de $G\backslash\{e\}$ qui fixent x. Donc

$$|A| = \sum_{x\in\Omega} \big(|\mathrm{Fix}(x)|-1\big). \tag{17.623}$$

Mais $|\mathrm{Fix}(x)|$ est constant sur les orbites. Nous coupons donc la somme sur Ω en plusieurs sommes sur les orbites Ω_i :

$$|A| = \sum_{i=1}^{r} \sum_{x\in\Omega_i} \big(|\mathrm{Fix}(x)|-1\big) = \sum_i |\Omega_i|(n_i-1). \tag{17.624}$$

En égalisant les deux façons de calculer $|A|$, nous déduisons la formule

$$2\big(|G|-1\big) = \sum_i |\Omega_i|(n_i-1). \tag{17.625}$$

Nous utilisons ensuite la relation orbite-stabilisateur, proposition 2.52 : $|\mathrm{Fix}(x_i)||\Omega_i| = |G|$; la formule (17.625) devient

$$2\big(|G|-1\big) = \sum_i |G| - \sum_i \frac{|G|}{n_i} = r|G| + |G|\sum_i \frac{1}{n_i}, \tag{17.626}$$

ou encore, en simplifiant par $|G|$:

$$2 - \frac{2}{|G|} = r - \sum_i \frac{1}{n_i} = \sum_{i=1}^{r}\left(1-\frac{1}{n_i}\right). \tag{17.627}$$

Nous pouvons aussi repartir de (17.625) et sommer de façon plus simple $\sum_i |\Omega_i| = |\Omega|$ pour obtenir

$$2\big(|G|-1\big) = r|G| - |\Omega| \tag{17.628}$$

où Ω est l'ensemble des pôles de $G\backslash\{e\}$.

Quelles sont les possibilités ? Les nombres $|G|$, r et n_i sont des entiers. Nous allons voir qu'il n'y a pas des centaines de possibilités pour satisfaire la relation (17.627). D'abord, pour toute valeur de $|G|$ (strictement plus grande que 1),

$$1 \leqslant 2 - \frac{2}{|G|} < 2. \tag{17.629}$$

Ensuite, si g fixe x alors g^{-1} fixe également x, de sorte que $n_i = |\mathrm{Fix}(x_i)| \geqslant 2$ pour tout i. Donc tous les termes dans la somme à droite de (17.627) sont dans $[\frac{1}{2}, 1[$. Nous avons donc au minimum deux termes, et au maximum trois. Autrement dit : $r = 2$ ou $r = 3$.

Si $r = 2$ Le plus simple est de repartir de (17.628). En posant $r = 2$ nous trouvons tout de suite $|\Omega| = 2$. Il y a donc exactement deux pôles pour l'action de G sur la sphère unité.

Tous les éléments de G laissent donc le même axe invariant et G est un sous-groupe des isométries du plan qui lui est perpendiculaire. Autrement dit, G est un sous-groupe fini de SO(2) et donc cyclique par le lemme 17.137.

Les possibilités pour $r = 3$ Nous devons voir les solutions entières $(n_1, n_2, n_3, |G|)$ de

$$2 - \frac{2}{|G|} = 3 - \frac{1}{n_1} - \frac{1}{n_2} - \frac{1}{n_3} < 2. \tag{17.630}$$

Il faut en particulier que

$$\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}>1, \tag{17.631}$$

ce qui signifie qu'au moins un des n_i doit être 1 ou 2, mais qu'il n'est pas possible que tous les n_i soient plus grands ou égaux à 3. Vu que $n_i=|\operatorname{Fix}(x_i)|\geqslant 2$, nous en déduisons qu'au moins un des n_i doit valoir 2. Nous posons donc $n_1=2$.

De plus, nous savons que les n_i doivent diviser $|G|$. Donc $|G|$ est pair.

Si $n_2=2$ Nous sommes dans le cas $r=3$, $n_1=2$, $n_2=2$. Nous avons

$$\frac{1}{n_3}=\frac{2}{|G|}, \tag{17.632}$$

mais aussi $n_3=|G|/|\Omega_3|$ d'où nous déduisons que $|\Omega_3|=2$. Nous avons donc une orbite à deux éléments. Soit $\Omega_3=\{x,y\}$ avec $x\neq y$.

Le groupe $\operatorname{Fix}(x)$ est un groupe à $|G|/2$ éléments. Il est donc normal par le lemme 3.36. Si $g\in G$ est tel que $g(x)=y$ alors nous avons $\operatorname{Fix}(y)=g\operatorname{Fix}(x)g^{-1}$, mais comme $\operatorname{Fix}(x)$ est normal nous avons $\operatorname{Fix}(x)=\operatorname{Fix}(y)$. Donc tous les éléments de $\operatorname{Fix}(x)$ fixent x et y. Le groupe $\operatorname{Fix}(x)$ est donc un sous-groupe de $\mathrm{SO}(2)$ est est cyclique comme vu plus haut.

Mais de plus nous avons forcément $y=-x$ parce qu'un élément de G qui fixe un point fixe également l'opposé. Vu que $\Omega_3=\{x,-x\}$, il existe $s\in G$ tel que $s(x)=-x$. Évidemment, s n'est pas dans $\operatorname{Fix}(x)$ et les points fixes de s ne sont pas parmi x et $-x$. Donc l'élément s^2 a au moins 4 points fixes : les deux de s ainsi que x et $-x$. Il a donc au moins deux droites de points fixes et est donc l'identité : $s^2=e$.

De plus, vu que $s(y)$ doit être égal soit à x soit à y, et vu que $s(x)=y$, l'injectivité de s donne $s(y)=x$.

Soit a, un générateur de $\operatorname{Fix}(x)$. Nous allons montrer que $G=\operatorname{gr}(s,sa)$. Nous avons déjà

$$(sa)(x)=s(x)=y \tag{17.633a}$$
$$(sa)(y)=s(y)=x. \tag{17.633b}$$

Donc sa inverse x et y. Mais sa a ses propres deux points fixes (qui ne sont ni x ni y). L'élément $(sa)^2$ a deux quatre points fixes sur la sphère unité : x, y et les deux de sa. Nous en déduisons que $(sa)^2=e$.

Nous nous souvenons que a est un générateur $\operatorname{Fix}(x)$. Mais $a=s\cdot sa$, donc $a^k=(ssa)^k$. Nous en déduisons que $\operatorname{gr}(s,sa)$ contient au moins $\operatorname{Fix}(x)$.

D'autre part si h et h' sont des éléments distincts dans $\operatorname{Fix}(x)$, alors sh et sh' sont des éléments distincts de $\operatorname{gr}(s,sa)$ qui ne sont pas dans $\operatorname{Fix}(x)$. Autrement dit, la partie

$$A=\{sh \text{ tel que } h\in Fix(x)\} \tag{17.634}$$

est une partie de même cardinal que $\operatorname{Fix}(x)$ tout en n'ayant aucune intersection avec $\operatorname{Fix}(x)$ (note : l'identité n'est pas dans A). Mais $|\operatorname{Fix}(x)|=|G|/2$, donc $A\cup\operatorname{Fix}(x)=G$. Et justement $A\cup G\subset\operatorname{gr}(s,sa)$. Nous en déduisons que $\operatorname{gr}(s,sa)=G$.

Le théorème 17.96 nous assure que le groupe G est alors le groupe diédral parce que les éléments s et sa vérifient les relations données en 17.90.

Si $r=3$, les autres cas possibles Nous repartons de (17.627) en posant $r=3$. Nous obtenons ceci :

$$1+\frac{2}{|G|}=\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_3}. \tag{17.635}$$

Nous avons déjà vu que $n_1=2$ était obligatoire, et que tous les cas où deux des n_i sont égaux à 2 sont déjà couverts. Donc n_2 et n_3 valent 3 ou plus.

Nous trions les n_i dans l'ordre croissant. Si $n_2=4$ ou plus, alors n_3 vaut 4 ou plus. Mais

$$\frac{1}{2}+\frac{1}{4}+\frac{1}{4}=1<1+\frac{2}{|G|}. \tag{17.636}$$

Donc $n_3 = 3$ est obligatoire. Nous avons alors l'inégalité suivante qui restreint n_3 :

$$\frac{1}{n_3} = \frac{1}{6} + \frac{3}{|G|} > \frac{1}{6}. \tag{17.637}$$

Donc n_3 est 3, 4 ou 5.

Les derniers cas à couvrir sont :

— $(n_1, n_2, n_3) = (2, 3, 3)$. Dans ce cas, $\frac{7}{6} = 1 + \frac{2}{|G|}$, donc $|G| = 12$.

— $(n_1, n_2, n_3) = (2, 3, 4)$. Dans ce cas, $|G| = 24$.

— $(n_1, n_2, n_3) = (2, 3, 5)$. Dans ce cas, $|G| = 60$.

Le cas $(2, 3, 3)$ Nous utilisons les relations $n_i|\Omega_i| = |G|$ pour savoir la taille des orbites. Nous avons :

(1) $2|\Omega_1| = 12$, donc $|\Omega_1| = 6$, $3|\Omega_2| = 12$, donc $|\Omega_2| = 4$, $3|\Omega_3| = 12$, donc $|\Omega_3| = 4$.

Nous avons $G \cdot \Omega_2 = \Omega_2$. D'une part parce que, par définition d'une orbite, $G \cdot \Omega_2 \subset \Omega_2$, et d'autre part parce que si $x \in \Omega_2$, alors $g^{-1}(x) \in \Omega_2$ et $g\big(g^{-1}(x)\big) = x$; donc Ω_2 est bien dans l'image de Ω_2 par G. Nous avons donc un morphisme $s\colon G \to S_{\Omega_2}$ que nous allons immédiatement voir comme

$$s\colon G \to S_4 \tag{17.638}$$

où S_4 est le groupe des permutations de $\{1, 2, 3, 4\}$.

Voyons que s est injective. Si $s(g) = s(h)$, alors $s(gh^{-1}) = \mathrm{Id}$. Autrement dit, l'élément sh^{-1} de G est l'identité sur Ω_2 qui contient 4 éléments. Fixant 4 points (au moins), l'élément sh^{-1} est l'identité. Par conséquent

$$s\colon G \to s(G) \subset S_4 \tag{17.639}$$

est un isomorphisme entre G et un sous-groupe de S_4. Mais $|G| = 12$ et $|S_4| = 24$, donc G est d'indice deux dans S_4 et est donc le groupe alterné A_4 par la proposition 5.30(3).

Le cas $(2, 3, 4)$ Nous avons $|G = 24|$ et les orbites ont pour taille :

— $2|\Omega_1| = 24$, donc $|\Omega_1| = 12$,

— $3|\Omega_2| = 24$, donc $|\Omega_2| = 8$,

— $4|\Omega_3| = 24$, donc $|\Omega_3| = 6$.

Ω_2 vient par paires Soit $x \in \Omega$ tel que $|\operatorname{Fix}(x)| = 3$. Alors $x \in \Omega_2$ parce que x est forcément dans un des Ω_i et tout élément x_i de Ω_i vérifie $|\operatorname{Fix}(x_i)| = n_i$. Mais comme les éléments de SO(3) sont des applications linéaires, ceux qui fixent x fixent également $-x$. Cela pour dire que si $x \in \Omega_2$, alors $-x \in \Omega_2$. Nous avons donc quatre éléments distincts a_1, a_2, a_3 et a_4 tels que

$$\Omega_2 = \{\pm a_1, \pm a_2, \pm a_3, \pm a_4\}. \tag{17.640}$$

Action sur les couples Nous prétendons que G agit sur l'ensemble des couples $\{\pm a_i\}$. C'est encore la linéarité qui joue : l'élément $g(a_i)$ est forcément un des $\pm a_k$ (éventuellement $k = i$). Si $g(a_i) = a_k$, alors $g(-a_i) = -a_k$. Autrement dit, pour tout i, il existe un k tel que $g\big(\{a_i, -a_i\}\big) = \{a_k, -a_k\}$. Cette association $i \mapsto k$ est bijective (sinon g ne serait pas bijective), et fournit donc un morphisme de groupes

$$s\colon G \to S_4. \tag{17.641}$$

s est injective Nous prouvons à présent que $s(g) = \mathrm{Id}$ si et seulement si $g = e$. Dans un sens c'est évident : $g(e) = \mathrm{Id}$. Dans l'autre sens, nous devons prouver que si $g(a_i) \in \pm a_i$ pour tout i alors $g = e$.

Si $g(a_i) = a_i$ pour tout i, alors g stabilise 4 points et l'affaire est pliée. Nous supposons qu'au moins un des a_i n'est pas stabilisé par g. Pour fixer les idées nous disons que c'est a_1. Nous avons donc $g(a_1) = -a_1$. (oui : $g(a_1) = -a_1$ et non $\pm a_k$ pour un autre k parce que nous sommes sous l'hypothèse que g stabilise les couples)

L'élément g^2 fixe tout Ω_2 ; donc $g^2 = e$. Nommons $\pm b$ les points fixes de g. Si $b \in \Omega_2$ alors $|\operatorname{Fix}(b)| = 3$, c'est-à-dire que les éléments de G qui fixent b sont dans un groupe d'ordre 3, et le corolaire 2.35 nous indique que ces éléments ne peuvent être que d'ordre 1 ou 3, pas deux. Nous en déduisons que b n'est pas dans Ω_2 et donc que $g(a_i) = -a_i$ pour tout i.

Jusqu'à présent nous avons prouvé que si $g \in \ker(s)$ est non trivial, alors $g(a_i) = -a_i$ pour tout i.

Soit maintenant $h \in G$. Vu que Ω_2 est une orbite, $h(a_i) \in \Omega_2$ et nous notons $h(a_i) = \epsilon a_k$ avec $\epsilon = \pm 1$ et éventuellement $k = i$ ou éventuellement pas. Nous avons :

$$(h^{-1}gh)(a_i) = \epsilon(h^{-1}g)(a_k) = -\epsilon h^{-1}(a_k) = -\epsilon^2 a_i = -a_i. \tag{17.642}$$

Donc g et $h^{-1}gh$ ont même restriction à Ω_2. En particulier $h^{-1}ghg^{-1}$ est l'identité sur Ω_2 et est donc l'identité.

Pour tout h nous avons $g = h^{-1}gh$. Les points fixes de $h^{-1}gh$ sont $\pm h^{-1}(b)$, mais aussi $\pm b$. Nous avons donc égalité d'ensemble $\{h(b), -h(b)\} = \{b, -b\}$ pour tout $h \in G$ (notez le changement de notation $h \to h^{-1}$). Cela signifie que $\{b, -b\}$ est une orbite de G. Maizon'a pas d'orbites de cardinal deux ; contradition. Nous en déduisons que e est l'unique élément de $\ker(s)$.

Conclusion La partie $s(G)$ est un sous-groupe de S_4 isomorphe à G. Mais au niveau des cardinaux, $|G| = 24$ en même temps que $|S_4| = 24$. Donc $G \simeq s(G) \simeq S_4$.

Nous passons au cas $(2, 3, 5)$, et comme ça va être long et douloureux [55], nous sautons un niveau d'indentation.

Au niveau du cardinal de G,

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{5} = 1 + \frac{2}{|G|}, \tag{17.643}$$

donc $|G| = 60$. Et pour les orbites, $|\Omega_1| = 30$, $|\Omega_2| = 20$, $|\Omega_3| = 12$.

La proposition 5.40 nous indique que le seul groupe simple d'ordre 60 est le groupe A_5. Nous allons donc nous atteler à prouver que G est simple. Vous êtes prêts ?

Fixateurs et ordres Tous les éléments de G sont dans un fixateur de type $\operatorname{Fix}(x)$, et comme l'ordre d'un élément divise l'ordre du groupe (corolaire 2.35), tous les éléments de G ont un ordre 2, 3 ou 5. Nous sommes dans un cas très particulier parce que

— Les trois nombres 2, 3 et 5 sont des nombres premiers distincts. Donc « diviser n_i » signifie pratiquement « être égal à n_i », surtout lorsqu'on parle de l'ordre d'un élément, qui ne peut pas être 1.

— Il existe une seule orbite de chaque taille.

Nous notons $G(n_i)$ l'ensemble des éléments de G d'ordre n_i. Les parties $G(n_i)$ ne contiennent pas l'identité.

$g \in G(n_i)$ implique $\operatorname{Fix}(g) \subset \Omega_i$ Si $g \in G(n_i)$ et $x \in \operatorname{Fix}(g)$ alors $x \in \Omega_i$. En effet $x \in Fix(g)$ signifie $g(x) = x$ et donc aussi $g \in \operatorname{Fix}(x)$. Donc l'ordre de g divise $|Fix(x)|$, alors que l'ordre de g est n_i et que les possibilités pour $|\operatorname{Fix}(x)|$ sont exactement les n_i, lesquels sont premiers entre eux. Donc $|\operatorname{Fix}(x)| = n_i$ et $x \in \Omega_i$.

$|\operatorname{Fix}(x)| = n_i$ implique $x \in \Omega_i$ Comme plus haut, $g \in \operatorname{Fix}(x)$ implique que l'ordre de g divise n_i et est donc égal à n_i. Autrement dit, $g \in G(n_i)$. De plus $g \in \operatorname{Fix}(x)$ implique $x \in \operatorname{Fix}(g)$. Par le cas juste au-dessus nous déduisons $x \in \Omega_i$.

a et $-a$ dans la même orbite Nous avons évidemment $\operatorname{Fix}(a) = \operatorname{Fix}(-a)$ du fait que les éléments de G sont des applications linéaires. Si $|\operatorname{Fix}(a)| = n_i$ alors $a \in \Omega_i$ et aussi $|\operatorname{Fix}(-a)| = |\operatorname{Fix}(a)| = n_i$ et aussi $-a \in \Omega_i$.

55. Mais pas autant que le théorème 31.28, cependant.

Nombre de $\mathrm{Fix}(x_i)$ Soient $a, b \in \Omega_i$. Nous avons $|\mathrm{Fix}(a)| = |\mathrm{Fix}(b)| = n_i$ et $\mathrm{Fix}(a) = \mathrm{Fix}(b)$ si et seulement si $b = -a$ parce qu'un élément qui fixe a et b fixe automatiquement a, b, $-a$, et $-b$. Aucun élément non trivial ne peut fixer quatre points distincts. Autrement dit,

$$\mathrm{Fix}(a) \cap \mathrm{Fix}(b) = \begin{cases} \mathrm{Fix}(a) & \text{si } a = \pm b \\ \{e\} & \text{sinon.} \end{cases} \tag{17.644}$$

Chaque élément $x_i \in \Omega_i$ a son fixateur (il y en aurait $|\Omega_i| = 60/n_i$), mais ces fixateurs sont égaux deux à deux, donc il y a seulement $\frac{60}{2n_i}$ groupes distincts de la forme $|\mathrm{Fix}(x_i)|$ avec $x_i \in \Omega_i$.

Récapitulatif En reprenant ce que nous venons de dire avec $i = 1, 2, 3$ nous trouvons :

(1) $n_1 = 2$, avec $|\Omega_1| = 30$ et 15 groupes du type $\mathrm{Fix}(x_1)$ avec x_1 parcourant Ω_1.
(2) $n_2 = 3$, avec $|\Omega_2| = 20$ et 10 groupes du type $\mathrm{Fix}(x_2)$ avec x_2 parcourant Ω_2.
(3) $n_3 = 5$, avec $|\Omega_3| = 12$ et 6 groupes du type $\mathrm{Fix}(x_3)$ avec x_3 parcourant Ω_3.

Un élément non trivial de G se trouve forcément dans un et un seul de ces sous-groupes. Plus précisément, si $g \in G(n_i)$ alors g est dans un des $\mathrm{Fix}(x_i)$ avec $x_i \in \Omega_i$.

Comptons pour être sûr de ne pas s'être trompé. Chacune des lignes décrit 30 éléments de G ; par exemple pour la seconde ligne donne 10 groupes de taille $|\mathrm{Fix}(x_2)| = n_2 = 3$. Mais tous ces groupes ont pour intersection exactement $\{e\}$. Donc le comptage des éléments se fait comme suit :

$$3 \times 30 - 15 - 10 - 6 + 1. \tag{17.645}$$

Le dernier $+1$ est parce que nous aurions décompté l'identité une fois de trop. Bref, on a bien 60 éléments comme il se doit.

Un ensemble à calculer deux fois Soient les ensembles A_2, A_3 et A_5 définis par

$$A_i = \{(g, a) \in G(n_i) \times \Omega_i \text{ tel que } g(a) = a\} \tag{17.646}$$

où $G(n_i)$ est la partie de G des éléments d'ordre n_i.

Nous avons

$$|A_i| = \sum_{g \in G(n_i)} |\mathrm{Fix}(g) \cap \Omega_i|. \tag{17.647}$$

Mais les éléments de $G(n_i)$ sont d'ordre n_i, et par ce que nous avons dit plus haut, tous les éléments de $\mathrm{Fix}(g)$ sont dans Ω_i. Donc $\mathrm{Fix}(g) \cap \Omega_i = \mathrm{Fix}(g)$. Nous avons alors

$$|A_i| = \sum_{g \in G(n_i)} |\mathrm{Fix}(g)| = 2|G(n_i)| \tag{17.648}$$

parce que $|\mathrm{Fix}(g)| = 2$ pour tout g.

En compatant $|A_i|$ dans l'autre sens, nous avons

$$|A_i| = \sum_{x \in \Omega_i} |\mathrm{Fix}(x) \cap G(n_i)| \tag{17.649}$$

Vu que $x \in \Omega_i$, les éléments de $\mathrm{Fix}(x)$ sont d'ordre n_i [56] (sauf e), et comme $G(n_i)$ est justement l'ensemble des éléments d'ordre n_i dans G nous avons $\mathrm{Fix}(x) \cap G(n_i) : \mathrm{Fix}(x) \backslash \{e\}$. Cela pour dire que

$$\begin{aligned}
|A_i| &= \sum_{x \in \Omega_i} \Big(|\mathrm{Fix}(x)| - 1 \Big) && (17.650a)\\
&= \sum_{x \in \Omega_i} |\mathrm{Fix}(x)| - \sum_{x \in \Omega_i} 1 && (17.650b)\\
&= \sum_{x \in \Omega_i} n_i - |\Omega_i| \qquad\qquad |\mathrm{Fix}(x)| = n_i \text{ pcq } x \in \Omega_i && (17.650c)\\
&= |\Omega_i| n_i - |\Omega_i| = |G| - |\Omega_i|. && (17.650d)
\end{aligned}$$

56. Encore et toujours parce que les éléments de $\mathrm{Fix}(x)$ ont un ordre qui divise $|\mathrm{Fix}(x)| = n_i$ et que n_i est premier, et que nous avons exclu l'identité.

En égalisant cela à la valeur $2|G(n_i)|$ déjà trouvée, nous déduisons les valeurs des $|G(n_i)|$:

$$|G(n_i)| = \frac{|G| - |\Omega_i|}{2}. \tag{17.651}$$

Nous avons alors

(1) $|G(2)| = 15$

(2) $|G(3)| = 20$

(3) $|G(5)| = 24$

Les Sylow de G Les p-Sylow sont définis en 5.6, et le super théorème qui répond à toutes les questions est le théorème 5.11. Dans notre cas, les diviseurs premiers de $|G| = 60$ sont 2, 3 et 5. Il faut faire attention au 2 parce que sa plus haute puissance dans la décomposition de 60 est 4 et non 2. Nous avons :

(1) Un 2-Sylow est un sous-groupe d'ordre 4.

(2) Un 3-Sylow est un sous-groupe d'ordre 3.

(3) Un 5-Sylow est un sous-groupe d'ordre 5.

Entre autres :

(1) Les 10 sous-groupes $\mathrm{Fix}(x_2)$ avec $x_2 \in \Omega_2$ sont des 3-Sylow de G.

(2) Les 6 sous-groupes $\mathrm{Fix}(x_3)$ avec $x_3 \in \Omega_3$ sont des 5-Sylow de G.

(3) Les 15 sous-groupes $\mathrm{Fix}(x_1)$ avec $x_1 \in \Omega_1$ sont d'ordre 2 et ne sont donc pas des 2-Sylow de G.

Tous les 3-Sylow et les 5-Sylow Nous avons déjà trouvé 10 3-Sylow et 6 5-Sylow. Nous montrons à présent qu'il n'y en a pas d'autres. Le théorème de Sylow 5.11(4) nous indique que le nombre n_3 de 3-Sylow est :

— diviseur de 60,

— dans $[1]_3$

— au moins 10.

Les diviseurs de 60 sont :

$$1, 5, 3, 15, 2, 10, 6, 30, 4, 20, 12, 60. \tag{17.652}$$

Le seul qui vérifie toutes les conditions est 10. Donc G possède seulement 10 3-Sylow et ils sont tous de la forme $\mathrm{Fix}(x_2)$ avec $x_2 \in \Omega_2$.

Même raisonnement pour les 5-Sylow : il faut

— diviseur de 60,

— dans $[1]_5$

— au moins 6.

La seule possibilité est 6.

Sous-groupe normal Soit H, un sous-groupe normal de G. Notre but étant de prouver que G est simple, nous voulons prouver que H est soit $\{e\}$ soit G. Nous supposons que H est non trivial, et nous allons prouver que $H = G$.

Le théorème de Lagrange 2.34(1) nous dit que $|H|$ divise $|G|$. Le nombre $|H|$ ne peut donc avoir que 2, 3 et 5 comme facteurs premiers. Avec une mention spéciale pour le 2 : $|H|$ pourrait être divisible aussi par 4.

Diviseurs de $|H|$ Soit un sous-groupe normal H de G. Vu que c'est un sous-groupe sont ordre divise celui de G (encore et toujours le théorème de Lagrange 2.34), et donc les facteurs premiers de $|H|$ ne peuvent être que 2, 3 et 5.

Si $|H|$ est divisible en 3 Alors H contient au moins un 3-Sylow. Mais nous avons vu que les 3-Sylow de H sont les 3-Sylow de G. Donc H contient tous les 3-Sylow de G, parce que les 3-Syow sont conjugués et H est normal.

Soit E l'ensemble des sous-groupes de H. Vu qu'il est normal, H agit sur E par conjugaison, et les 3-Sylow forment une orbite. Si α est un 3-Sylow, la formule des classes (proposition 2.52(2)) nous donne

$$|H| = |\operatorname{Fix}(\alpha)||\mathcal{O}_\alpha|. \tag{17.653}$$

Mais l'orbite $\mathcal{O}_\alpha$ de α est l'ensemble des 3-Sylow, de sorte que $|\mathcal{O}_\alpha| = 10$. Donc $|H|$ est divisible en 10.

Mais il y a pire : H contient au moins les 10 sous-groupes $\operatorname{Fix}(x_2)$ pour $x_2 \in \Omega_2$. Ce sont 10 groupes de $|\operatorname{Fix}(x_2)| = 3$ éléments. En décomptant e qui est dans l'intersection, cela fait

$$10 \times |\operatorname{Fix}(x_2)| - 10 + 1 = 21 \tag{17.654}$$

éléments. Donc H contient au moins 21 éléments. Le nombre $|H|$ est donc :

— diviseur de 60

— multiple de 10

— au moins 21.

Donc c'est 30 ou 60.

Si $|H|$ est divisible en 5 Le même raisonnement tient et $|H|$ est 30 ou 60.

Nous restons avec les possibilités $|H|$ égal à 2, 4, 30 ou 60.

Si $|H| = 4$ Alors H contient au moins un 2-Sylow. Un 2-Sylow de H est un sous-groupe contenant 4 éléments qui sont d'ordre 2^m. Le seul m possible dans G est $m = 1$. Vu qu'un 2-Sylow de H contient 4 éléments, nous sommes dans le cas où H est un 2-Sylow. Il est donc le seul 2-Sylow de H parce que H est normal et que tous les 2-Sylow sont conjugués.

Mais tous les sous-groupes d'ordre 2 sont contenus dans un 2-Sylow. En particulier tous les 15 groupes $\operatorname{Fix}(x_1)$ sont dans l'unique 2-Sylow H qui est soit-disant d'ordre 4. IL y a là une belle impossibilité.

Donc le cas $|H| = 4$ est hors-concours.

Si $|H| = 2$ Alors $H = \{e, g\}$ avec $g^2 = e$. Si $h \in G$, l'élément hgh^{-1} ne peut être que e ou g (parce que H est normal). Le premier cas est $g = e$, et le second donne $gh = hg$. Donc g est dans le centre de G : il commute avec tous les éléments de G.

Vu que $g \in G(2)$, nous avons que les éléments $a \in \operatorname{Fix}(g)$ sont forcément dans Ω_1 parce que les points dont les fixateurs sont formés d'éléments d'ordre 2 sont dans Ω_1. Soit $h \in G$. Nous avons $g = hgh^{-1}$ et donc aussi

$$\big(hgh^{-1}\big)\big(h(a)\big) = hg(a) = h(a), \tag{17.655}$$

donc $h(a)$ et $-h(a)$ sont des points fixes de hgh^{-1}. Ce sont donc également de points fixes de g. Nous en déduisons que g a pour points fixes les points a, $-a$, $h(a)$ et $-h(a)$. Vu que g n'est pas e, ces quatre points ne peuvent pas être distincts. Vu que $h(a)$ ne peut pas être $-h(a)$, nous avons forcément $h(a) = \pm a$.

Donc l'orbite de a ne contiendrait que 2 éléments. Pas possible.

Si $|H| = 30$ À part $|H| = 60$, le dernier cas à traiter est $|H| = 30$. Nous rappelons obligeamment que

(1) $|G(2)| = 15$

(2) $|G(3)| = 20$

(3) $|G(5)| = 24$.

Si H possède 30 éléments, le théorème de Sylow dit que H contient au moins un 3-Sylow et un 5-Sylow, et donc tous. Vu que pour 3 et 5, les Sylow de H et de G sont les mêmes et bien identifiés, nous allons nous baser dessus. Le sous-groupe H contient tous les 3 et 5-Sylow, donc le comptage des éléments est :

$$10 \times |\operatorname{Fix}(x_2)| + 6 \times |\operatorname{Fix}(x_3)| - 15 = 45. \tag{17.656}$$

Nous aurions aussi pu ajouter $+4 - 1$ pour compter au moins un 2-Sylow.

Donc dès que H compte 30 éléments, il en compte 45 et donc 60 parce qu'il n'y a pas de diviseurs de 60 entre 45 et 60.

□

Problèmes et choses à faire

La démonstration des groupes finis de SO(3) est longue. Je me demande si il n'y a pas moyen de faire plus court. Par exemple [366] utilise le théorème de Cauchy 3.34 que je n'utilise pas. D'autre part, toutes les références me semblent utiliser plus ou moins implicitement le fait que si le sous-groupe normal H contient un élément de $G(n_i)$, alors il les contiennent tous. J'avoue ne pas trop comprendre pourquoi.

17.12 Systèmes de coordonnées

La trigonométrique nous offre de nouveaux systèmes de coordonnées qui peuvent se révéler pratiques de certains cas : les coordonnées polaires sur $\mathbb{R}^2$ ainsi que les coordonnées cylindriques et sphériques sur $\mathbb{R}^3$.

17.12.1 Coordonnées polaires

17.12.1.1 Ce que ça signifie intuitivement

On a vu qu'un point M dans $\mathbb{R}^2$ peut être représenté par ses abscisses x et ses ordonnées y. Nous pouvons également déterminer le même point M en donnant un angle et une distance comme montré sur la figure 17.11.

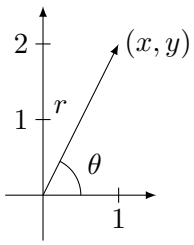


FIGURE 17.11 – Un point en coordonnées polaires est donné par sa distance à l'origine et par l'angle qu'il faut avec l'horizontale.

Le même point M peut être décrit indifféremment avec les coordonnées (x, y) ou bien avec (r, θ).

Remarque 17.194.
L'angle θ d'un point n'étant à priori défini qu'à un multiple de 2π près, nous convenons de toujours choisir un angle $0 \leqslant \theta < 2\pi$. Par ailleurs l'angle θ n'est pas défini si $(x, y) = (0, 0)$.

La coordonnée r est toujours positive.

Nous avons dans l'idée de définir r et θ par les formules

$$\begin{cases} x = r\cos(\theta) & \text{(17.657a)} \\ y = r\sin(\theta). & \text{(17.657b)} \end{cases}$$

17.12.1.2 Coordonnées polaires : le théorème

Théorème 17.195 (Coordonnées polaires[1]).
Soit l'application

$$\begin{aligned} T\colon [0,\infty[\times [0,2\pi[&\to \mathbb{R}^2 \\ (r,\theta) &\mapsto \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix}. \end{aligned} \tag{17.658}$$

(1) L'application T est surjective.

(2) L'application

$$T\colon]0,\infty[\times [0,2\pi[\to \mathbb{R}^2\backslash\{(0,0)\} \tag{17.659}$$

est bijective.

(3) En considérant la demi-droite $D = \{(x,0)\}_{x\geqslant 0}$, l'application

$$T\colon]0,\infty[\times]0,2\pi[\to \mathbb{R}^2\backslash D \tag{17.660}$$

est un C^∞-difféomorphisme[57].

Démonstration. Une bonne partie de ce théorème est une conséquence de 17.46. La vraie nouveauté de ce théorème sera la régularité. Nous démontrons point par point.

(1) Pour (1). Soit $a = (x,y) \in \mathbb{R}^2$. Nous avons $a/\|a\| \in S^1$. Par la proposition 17.46, il existe $\theta \in [0,2\pi]$ tel que

$$\frac{a}{\|a\|} = \big(\cos(\theta),\sin(\theta)\big). \tag{17.661}$$

Alors $a = \|a\|\big(\cos(\theta),\sin(\theta)\big) = T(\|a\|,\theta)$. Voila. L'application T est surjective.

(2) Pour (2). En ce qui concerne la surjectifivé,

$$T\big(0,[2,2\pi[\big) = \{(0,0)\}. \tag{17.662}$$

Donc le point (1) donne le surjectif lorsque nous enlevons d'un côté les points avec $r = 0$ et de l'autre le point $(0,0)$.

Pour l'injectivité, nous supposons $T(r_1,\theta_1) = T(r_2,\theta_2)$. Vu que $\|T(t,\theta)\| = r$, nous avons tout de suite $r_1 = r_2$. Nous restons donc avec l'égalité

$$\begin{pmatrix} \cos(\theta_1) \\ \sin(\theta_1) \end{pmatrix} = \begin{pmatrix} \cos(\theta_2) \\ \sin(\theta_2) \end{pmatrix}. \tag{17.663}$$

La proposition 17.46 dit alors que $\theta_1 = \theta_2$.

(3) Pour (3). L'application T est injective en tant que restriction d'une application injective. Pour le surjectif, soit $a \in \mathbb{R}^2\backslash D$. Vu que $a \notin D$, nous avons $\|a\| \neq 0$ et il est légitime de dire, comme plus haut, qu'il existe $\theta \in [0,2\pi[$ tel que

$$\frac{a}{\|a\|} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}. \tag{17.664}$$

Ce θ n'est pas zéro parce que $\theta = 0$ donne le point $(1,0)$ qui est sur D.

En ce qui concerne l'inverse, nous n'allons pas nous lancer dans une étude subtile de la fonction (17.125) ; nous avons déjà démontré la continuité dans le lemme 17.50, et monter dans les dérivées nous semble un peu compliqué. Au lieu de cela, nous allons faire en deux étapes :

— Prouver que T est de classe C^p pour tout p en invoquant seulement des théorèmes à proposition de différentielle,

57. L'application est de classe C^∞ et son inverse est également de classe C^∞. Le plus souvent, vous voulez seulement utiliser ce théorème dans le but de faire un changement de variables dans une intégrale ; vous n'avez donc besoin que d'un C^1-difféomorphisme.

— En déduire que T^{-1} est également C^p pour tout p en invoquant le théorème d'inversion locale 16.48.

Les applications $(r,\theta) \mapsto r$, $(r,\theta) \mapsto \sin(\theta)$ et $(r,\theta) \mapsto \cos(\theta)$ sont de classe C^∞ grâce au lemme 11.190. Le lemme 11.213 sur la différentiabilité du produit montre alors que les fonctions T_1 et T_2 données par

$$T_1(r,\theta) = r\cos(\theta) \tag{17.665a}$$
$$T_2(r,\theta) = r\sin(\theta) \tag{17.665b}$$

sont différentiables[58]. Mieux, la proposition 11.216 montre que ces fonctions T_1 et T_2 sont de classe C^p pour tout p, c'est-à-dire qu'elles sont de classe C^∞. Cela montre que les coordonnées polaires sont de classe C^∞, et il faut encore parler de l'inverse.

En ce qui concerne la différentielle,

$$dT_{(r,\theta)}(u,v) = \begin{pmatrix} u\cos(\theta) - rv\sin(\theta) \\ u\sin(\theta) + rv\cos(\theta) \end{pmatrix}. \tag{17.666}$$

Donc la matrice de la différentielle est

$$dT_{(r,\theta)} = \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta), \end{pmatrix} \tag{17.667}$$

dont le déterminant est r (lemme 17.3 utilisé). Donc la différentielle en (r,θ) est une application linéaire inversible parce que $r \neq 0$ aux points que nous considérons. L'application $dT_{(r,\theta)}$ est bicontinue parce que nous sommes en dimension finie. Tout cela pour dire que le théorème d'inversion local 16.48 fonctionne et T^{-1} est C^p dès que T est C^p.

Vu que T est de classe C^p pour tout p, l'inverse T^{-1} est également C^p pour tout p, c'est-à-dire que T^{-1} est de classe C^∞.

□

Définition 17.196.
Ce que nous appelons ***les coordonnées polaires*** *est l'application*

$$\begin{aligned} T\colon [0,\infty[\times [0,2\pi[&\to \mathbb{R}^2 \\ (r,\theta) &\mapsto \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix}. \end{aligned} \tag{17.668}$$

du théorème 17.195(3). Selon les circonstances, nous considérons l'une ou l'autre des restrictions pour avoir une bijection ou un difféomorphisme.

Exemple 17.197.
Soit à calculer

$$\lim_{(x,y)\to(0,0)} \frac{x^2+y^2}{x-y}. \tag{17.669}$$

Nous introduisons la fonction

$$\begin{aligned} f\colon \mathbb{R}^2\backslash\{x=y\} &\to \mathbb{R} \\ (x,y) &\mapsto \frac{x^2+y^2}{x-y}. \end{aligned} \tag{17.670}$$

Une idée souvent fructueuse pour traiter ce genre de limite est de passer aux coordonnées polaires. Attention, si on veut faire les choses très explicitement, c'est un peu lourd en notations. Il s'agit de poser

$$\begin{aligned} f\colon \big(]0,\infty[\times [0,2\pi[\big)\backslash\{\mathbb{R}\times\{\frac{\pi}{4}\} \cup \mathbb{R}\times\{\frac{5\pi}{4}\}\} &\to \mathbb{R} \\ (r,\theta) &\mapsto \frac{r^2}{r\big(\cos(\theta)-\sin(\theta)\big)}. \end{aligned} \tag{17.671}$$

58. Si vous voulez seulement avoir un C^1-difféomorphisme, calculez explicitement la différentielle et montrez que c'est continu. Vous n'avez pas à utiliser la proposition 11.216 ni rien des produits tensoriels.

Bon. À strictement parler, nous aurions pu dire que g est définie pour $r = 0$, mais vu que nous voulons seulement calculer la limite pour $r \to 0$, on n'a pas besoin de la valeur en zéro. De plus les coordonnées polaires ne sont pas bijectives en l'origine. Donc bon ... on s'en passe.

Quel est le lien entre f et g ? Du point de vue du calcul, le lien est qu'on a remplacé x par $r\cos(\theta)$ et y par $r\sin(\theta)$. Le vrai lien est l'égalité

$$g = f \circ T \tag{17.672}$$

où T est l'application de coordonnées polaires dont les principales propriétés sont données dans le théorème 17.195(2).

Soit un voisinage $B\big((0,0),R\big)$ de $(0,0)$ dans $\mathbb{R}^2$. Le but est de montrer que les valeurs $f(B)$ se regroupent autour d'une valeur ℓ lorsque $R \to 0$. Soyons plus précis et nommons ℓ le candidat limite. Soit $\epsilon > 0$; nous devons trouver $R > 0$ tel que $f\Big(B\big((0,0),R\big)\Big) \subset B(\ell,\epsilon)$.

Pour $R > 0$, nous avons

$$B\big((0,0),R\big) = T\big([0,R[\times [0,2\pi[\big), \tag{17.673}$$

donc

$$f(B) = g\big([0,R[\times [2,2\pi[\big). \tag{17.674}$$

Soit $r < R$. Nous avons

$$\lim_{\theta\to\pi/4} g(r,\theta) = \infty. \tag{17.675}$$

Donc $f(B)$ contient des valeurs arbitrairement grandes, quelle que soit la valeur de R. Il n'y a donc pas de limite possible.

Si vous voulez un argument un peu plus imagé, en voici un [59] basé sur une combinaison entre la méthode des coordonnées polaires et la méthode des chemins.

Certes *pour chaque* θ nous avons $\lim_{r\to 0} g(r,\theta) = 0$, mais il ne faut pas en déduire trop vite que la limite $\lim_{(x,y)\to(0,0)} g(x,y)$ vaut zéro parce que prendre la limite $r \to 0$ avec θ fixé revient à prendre la limite le long de la droite d'angle θ.

Il n'est pas possible de majorer $g(r,\theta)$ par une fonction ne dépendant pas de θ parce que cette fonction tend vers l'infini lorsque $\theta \to \pi/4$. Est-ce que cela veut dire que la limite n'existe pas ? Cela veut en tout cas dire que la méthode des coordonnées polaires ne parvient pas à résoudre l'exercice. Pour conclure, il faudra encore un peu travailler.

Nous pouvons essayer de calculer le long d'un chemin plus général $(r(t),\theta(t))$. Choisissons $r(t) = t$ puis cherchons $\theta(t)$ de telle sorte à avoir

$$\cos\theta(t) - \sin\theta(t) = t^2. \tag{17.676}$$

Le mieux serait de résoudre cette équation pour trouver $\theta(t)$. Mais en réalité il n'est pas nécessaire de résoudre : montrer qu'il existe une solution suffit. Nous pouvons supposer que $t^2 < 1$. Pour $\theta = \pi/4$ nous avons $\cos(\theta) - \sin(\theta) = 0$ et pour $\theta = 0$ nous avons $\cos(\theta) - \sin(\theta) = 1$. Le théorème des valeurs intermédiaires nous enseigne alors qu'il existe une valeur de θ qui résout l'équation (17.676).

Pour être rigoureux, nous devons aussi montrer que la fonction $\theta(t)$ est continue. Pour cela il faudrait utiliser le théorème de la fonction implicite 16.50. Nous verrons dans l'exemple 17.212 comment s'en sortir sans théorème de la fonction implicite, au prix de plus de calculs. △

Les coordonnées polaires sont données par le difféomorphisme

$$\begin{aligned} g\colon]0,\infty[\times]0,2\pi[&\to \mathbb{R}^2\backslash D \\ (r,\theta) &\mapsto \big(r\cos(\theta), r\sin(\theta)\big) \end{aligned} \tag{17.677}$$

59. Qui satisfera tous vos professeurs, pourvu que vous ayez compris que ce qui se cache est une histoire de valeurs de f prises sur un voisinage de $(0,0)$.

où D est la demi-droite $y = 0$, $x \geqslant 0$. Le fait que les coordonnées polaires ne soient pas un difféomorphisme sur tout $\mathbb{R}^2$ n'est pas un problème pour l'intégration parce que le manque de difféomorphisme est de mesure nulle dans $\mathbb{R}^2$. Le jacobien est donné par

$$Jg = \det \begin{pmatrix} \partial_r x & \partial_\theta x \\ \partial_r y & \partial_\theta y \end{pmatrix} = \det \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix} = r. \tag{17.678}$$

La fonction qui donne les coordonnées polaires est

$$\begin{aligned} \varphi\colon \mathbb{R}^+ \times]0, 2\pi[&\to \mathbb{R}^2 \\ (r, \theta) &\mapsto \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix}. \end{aligned} \tag{17.679}$$

Son Jacobien vaut

$$J_\varphi(r, \theta) = \det \begin{pmatrix} \frac{\partial x(r,\theta)}{\partial r} & \frac{\partial x(r,\theta)}{\partial \theta} \\ \frac{\partial y(r,\theta)}{\partial r} & \frac{\partial y(r,\theta)}{\partial \theta} \end{pmatrix} = \begin{vmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{vmatrix} = r. \tag{17.680}$$

Proposition 17.198.
Soit la fonction

$$\begin{aligned} T\colon]O, +\infty[\times \mathbb{R} &\to \mathbb{R}^2 \backslash \{(0,0)\} \\ (r, \theta) &\mapsto \big(r\cos(\theta), r\sin(\theta)\big). \end{aligned} \tag{17.681}$$

(1) Elle est surjective.
(2) Pour tout $a \in \mathbb{R}$, l'application T est bijective sur la bande $]0, +\infty[\times [a - \pi, a + \pi[$.
(3) Si $a = 0$, la fonction inverse est donnée par

$$T^{-1}(x, y) = \big(\sqrt{x^2 + y^2}, \operatorname{arctg}(y/x)\big). \tag{17.682}$$

Soit $P = (x, y)$ un élément dans $\mathbb{R}^2$, on dit que $r = \sqrt{x^2 + y^2}$ est le rayon de P et que $\theta = \operatorname{arctg}(y/x)$ est son argument principal. L'origine ne peut pas être décrite en coordonnées polaires parce que si son rayon est manifestement zéro, on ne peut pas lui associer une valeur univoque de l'angle θ.

Exemple 17.199.
L'équation du cercle de rayon a et centre $(0, 0)$ en coordonnées polaires est $r = a$. △

Exemple 17.200.
Une équation possible pour la demi-droite $x = y$, $x > 0$, est $\theta = \pi/4$. △

17.12.1.3 Transformation inverse : théorie

Voyons la question inverse : comment retrouver r et θ si on connait x et y ? Tout d'abord,

$$r = \sqrt{x^2 + y^2} \tag{17.683}$$

parce que la coordonnée r est la distance entre l'origine et (x, y). Comment trouver l'angle ? Nous supposons $(x, y) \neq (0, 0)$. Si $x = 0$, alors le point est sur l'axe vertical et nous avons

$$\theta = \begin{cases} \pi/2 & \text{si } y > 0 \\ 3\pi/2 & \text{si } y < 0 \end{cases} \tag{17.684}$$

Notez que si $y < 0$, conformément à notre convention $\theta \geqslant 0$, nous avons noté $\frac{3\pi}{2}$ et non $-\frac{\pi}{2}$.

Supposons maintenant le cas général avec $x \neq 0$. Les équations (17.657) montrent que

$$\tan(\theta) = \frac{y}{x}. \tag{17.685}$$

Nous avons donc

$$\theta = \tan^{-1}\left(\frac{y}{x}\right). \tag{17.686}$$

La fonction inverse de la fonction tangente est celle définie plus haut.

17.12.1.4 Transformation inverse : pratique

Le code suivant utilise Sage.

```
#! /usr/bin/sage -python
# -*- coding: utf8 -*-

from sage.all import *

def PointToPolaire(x,y):
  x=SR(x)
  y=SR(y)
  r = sqrt(x**2+y**2)
  if x == 0:
    if y > 0:
      alpha = pi/2
    if y < 0:
      alpha = 3*pi/2
    if y == 0 :
      raise ValueError,"Pas d'angle pour le point (0,0) !!"
  else :
    alpha = atan(y/x)
  if (x < 0) and (y == 0) :
    alpha = pi
  if (x < 0) and (y > 0) :
    alpha = alpha + pi
  if (x < 0) and (y < 0 ) :
    alpha = alpha + pi
  alpha=alpha.simplify_trig()
  return (r,alpha)

print PointToPolaire(1,1)
print PointToPolaire(-2,1)
print PointToPolaire(6*sqrt(3)/2,3)
```

tex/frido/calculAngle.py

Son exécution retourne :

```
(sqrt(2), 1/4*pi)
(sqrt(5), pi - arctan(1/2))
(6, 1/6*pi)
```

Notez que ce sont des valeurs *exactes*. Ce ne sont pas des approximations, Sage travaille de façon symbolique.

17.12.1.5 Coordonnées polaires : dérivées partielles

Le changement de coordonnées pour les polaires est la fonction

$$f\begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix}. \tag{17.687}$$

Considérons une fonction g sur $\mathbb{R}^2$, et définissons la fonction $\tilde{g}$ par

$$\tilde{g}(r,\theta) = g(r\cos\theta, r\sin\theta). \tag{17.688}$$

La formule (11.482) permet de trouver les dérivées partielles de g par rapport à r et θ en termes de celles par rapport à x et y de g.

Pour faire le lien avec les notations du point précédent, nous avons

$$\begin{aligned} f_1(r,\theta) &= r\cos(\theta) \\ f_2(r,\theta) &= r\sin(\theta) \\ (x_1,x_2) &\to (r,\theta) \\ (y_1,y_2) &\to (x,y). \end{aligned} \tag{17.689}$$

Nous avons donc

$$\begin{aligned} \frac{\partial \tilde{g}}{\partial r}(r,\theta) &= \sum_{i=1}^{2} \frac{\partial g}{\partial x_i}\big(f(r,\theta)\big)\frac{\partial f_i}{\partial r}(r,\theta) \\ &= \frac{\partial g}{\partial x}(r\cos\theta, r\sin\theta)\frac{\partial\big(r\cos\theta\big)}{\partial r}(r,\theta) \\ &\quad + \frac{\partial g}{\partial y}(r\cos\theta, r\sin\theta)\frac{\partial\big(r\sin\theta\big)}{\partial r}(r,\theta) \\ &= \cos\theta\frac{\partial g}{\partial x}(r\cos\theta, r\sin\theta) + \sin\theta\frac{\partial g}{\partial y}(r\cos\theta, r\sin\theta). \end{aligned} \tag{17.690}$$

Prenons par exemple $g(x,y) = \frac{1}{x^2+y^2}$. Étant donné que

$$\frac{\partial g}{\partial x} = \frac{-2x}{(x^2+y^2)^2}, \tag{17.691}$$

nous avons

$$\frac{\partial g}{\partial x}(r\cos\theta, r\sin\theta) = \frac{-2\cos\theta}{r^3}. \tag{17.692}$$

En utilisant la formule,

$$\frac{\partial \tilde{g}}{\partial r}(r,\theta) = \cos(\theta)\left(\frac{-2\cos\theta}{r^3}\right) + \sin(\theta)\left(\frac{-2\sin\theta}{r^3}\right) = -\frac{2}{r^3}. \tag{17.693}$$

Nous pouvons vérifier directement que cela est correct. En effet

$$\tilde{g}(r,\theta) = g(r\cos\theta, r\sin\theta) = \frac{1}{r^2}, \tag{17.694}$$

dont la dérivée par rapport à r vaut $-2/r^3$.

En ce qui concerne la dérivée par rapport à θ, nous avons

$$\begin{aligned} \frac{\partial \tilde{g}}{\partial \theta} &= \frac{\partial g}{\partial x}(r\cos\theta, r\sin\theta)\frac{\partial\big(r\cos(\theta)\big)}{\partial\theta} + \frac{\partial g}{\partial y}(r\cos\theta, r\sin\theta)\frac{\partial\big(r\sin(\theta)\big)}{\partial\theta} \\ &= \left(\frac{-2\cos\theta}{r^3}\right)(-r\sin\theta) + \left(\frac{-2\sin\theta}{r^3}\right)(r\cos\theta) \\ &= 0. \end{aligned} \tag{17.695}$$

En résumé et avec quelques abus de notation :

$$\begin{aligned} \frac{\partial \tilde{g}}{\partial r} &= \cos(\theta)\frac{\partial g}{\partial x} + \sin(\theta)\frac{\partial g}{\partial y} \\ \frac{\partial \tilde{g}}{\partial \theta} &= -r\sin(\theta)\frac{\partial g}{\partial x} + r\cos(\theta)\frac{\partial g}{\partial y} \end{aligned} \tag{17.696}$$

17.12.2 Coordonnées cylindriques

Les **coordonnées cylindriques** sont un perfectionnement des coordonnées polaires. Il s'agit simplement de donner le point (x, y, z) en faisant la conversion $(x, y) \mapsto (r, \theta)$ et en gardant le z. Les formules de passage sont

$$\begin{cases} x = r\cos(\theta) & (17.697a) \\ y = r\sin(\theta) & (17.697b) \\ z = z. & (17.697c) \end{cases}$$

Soit T la fonction de $]0, +\infty[\times\mathbb{R}^2$ dans $\mathbb{R}^3\backslash\{(0,0,0)\}$ définie par

$$\begin{array}{rccc} T: &]0,+\infty[\times\mathbb{R}\times\mathbb{R} & \to & \mathbb{R}^3\backslash\{(0,0,0)\} \\ & (r,\theta,z) & \mapsto & (r\cos\theta, r\sin\theta, z), \end{array} \tag{17.698}$$

Cette fonction est surjective. Elle est bijective sur chaque bande de la forme $]0, +\infty[\times[a-\pi, a+\pi[\times\mathbb{R}$, a dans $\mathbb{R}$. Il n'y a presque rien de nouveau par rapport aux coordonnées polaires. Les coordonnées cylindriques sont intéressantes si on décrit un objet invariant par rapport aux rotations autour de l'axe des z.

Exemple 17.201.
Il faut savoir ce que décrivent les équations les plus simples en coordonnées cylindriques,

— $r \leqslant a$, pour a constant dans $]0, +\infty[$, est le cylindre de hauteur infinie qui a pour axe l'axe des z et pour base le disque de rayon a centré à l'origine,

— $r = a$ est la surface du cylindre,

— $\theta = b$ est un demi-plan ouvert et sa fermeture contient l'axe des z,

— $z = c$ est un plan parallèle au plan x-y.

△

Exemple 17.202.
Un demi-cône qui a son sommet en l'origine et pour axe l'axe des z est décrit par $z = dr$. Si d est positif il s'agit de la moitié supérieure du cône, si $d < 0$ de la moitié inférieure. △

Exemple 17.203.
De même, la sphère de rayon a et centrée à l'origine est l'assemblage des calottes $z = \sqrt{a^2 - r^2}$ et $z = -\sqrt{a^2 - r^2}$. △

En ce qui concerne les coordonnées cylindriques, le Jacobien est donné par

$$J(r,\theta,z) = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial z} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta & 0 \\ \sin\theta & r\cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = r. \tag{17.699}$$

Nous avons donc $dx\,dy\,dz = r\,dr\,d\theta\,dz$.

$$\begin{cases} x = r\cos\theta & (17.700a) \\ y = r\sin\theta & (17.700b) \\ z = z & (17.700c) \end{cases}$$

avec $r \in]0, \infty[$, $\theta \in [0, 2\pi[$ et $z \in \mathbb{R}$. Le jacobien vaut r.

17.12.3 Coordonnées sphériques

Soit T la fonction de $]0,+\infty[\times\mathbb{R}^2$ dans $\mathbb{R}^3\backslash\{(0,0,0)\}$ définie par

$$\begin{array}{rccc} T: &]0,+\infty[\times\mathbb{R}\times\mathbb{R} & \to & \mathbb{R}^3\backslash\{(0,0,0)\} \\ & (\rho,\theta,\phi) & \mapsto & (\rho\cos\theta\sin\phi,\rho\sin\theta\sin\phi,\rho\cos\phi), \end{array} \tag{17.701}$$

Cette fonction est surjective. Elle est bijective sur chaque bande de la forme $]0,+\infty[\times[a-\pi,a+\pi[\times[b-\pi/2,b+\pi/2[$, a et b dans $\mathbb{R}$. Si $a=0$ et $b=-\pi/2$ la fonction inverse T^{-1} est donnée donnée

$$\begin{array}{rccc} T: & \mathbb{R}^3\backslash\{(0,0,0)\} & \to &]0,+\infty[\times[-\pi,\pi[\times[0,\pi[\\ & (x,y,z) & \mapsto & \left(\sqrt{x^2+y^2+z^2},\operatorname{arctg}\frac{y}{x},\arccos\left(\frac{z}{\sqrt{x^2+y^2+z^2}}\right)\right). \end{array} \tag{17.702}$$

Soit P un point dans $\mathbb{R}^3$. L'angle ϕ est l'angle entre le demi-axe positif des z et le vecteur $\overrightarrow{OP}$, ρ est la norme de $\overrightarrow{OP}$ et θ est l'argument en coordonnées polaires de la projection de $\overrightarrow{OP}$ sur le plan x-y.

Remarque 17.204.
Dans la littérature, les angles θ et ϕ sont parfois inversés (voire, changent de nom, par exemple φ au lieu de ϕ). Il faut donc être très prudent lorsqu'on veut utiliser dans un cours des formules données dans un autre cours.

Exemple 17.205.
Il faut connaitre le sens des équations plus simples,

— $\rho\leqslant a$, pour a constant dans $]0,+\infty[$, est la boule fermée de rayon a centrée à l'origine,

— $\rho=a$ est la sphère de rayon a centrée à l'origine,

— $\theta=b$ est un demi-plan ouvert et sa fermeture contient l'axe des z,

— $\phi=c$ est un demi-cône qui a son sommet à l'origine et pour axe l'axe des z. Si c est positif il s'agit de la moitié supérieure du cône, si $d<0$ de la moitié inférieure.

△

Les **coordonnées sphériques** sont ce qu'on appelle les « méridiens » et « longitudes » en géographie. Les formules de transformation sont

$$\begin{cases} x=\rho\sin(\theta)\cos(\varphi) & (17.703a) \\ y=\rho\sin(\theta)\sin(\varphi) & (17.703b) \\ z=\rho\cos(\theta) & (17.703c) \end{cases}$$

avec $0\leqslant\theta\leqslant\pi$ et $0\leqslant\varphi<2\pi$.

Remarque 17.206.
Attention : d'un livre à l'autre les conventions sur les noms des angles changent. N'essayez donc pas d'étudier par cœur des formules concernant les coordonnées sphériques trouvées autre part. Par exemple sur le premier dessin de wikipédia, l'angle φ est noté θ et l'angle θ est noté Φ. Mais vous noterez que sur cette même page, les conventions de noms de ces angles changent plusieurs fois.

Les coordonnées sphériques sont données par

$$\begin{cases} x=r\cos\theta\sin\varphi \\ y=r\sin\theta\sin\varphi \\ z=r\cos\varphi \end{cases} \quad \text{avec} \quad \begin{array}{l} r\in]0,\infty[\\ \theta\in]0,2\pi[\\ \phi\in]0,\pi[. \end{array} \tag{17.704}$$

Le jacobien associé est $Jg(r,\theta,\varphi) = -r^2 \sin\varphi$. Rappelons que ce qui rentre dans l'intégrale est la valeur absolue du jacobien.

$$\begin{cases} x = \rho\cos\theta\sin\phi & (17.705a) \\ y = \rho\sin\theta\sin\phi & (17.705b) \\ z = \rho\cos\phi & (17.705c) \end{cases}$$

avec $\rho \in]0,\infty[$, $\theta \in [0,2\pi[$ et $\phi \in [0,\pi[$. Le jacobien vaut $-\rho^2\sin(\phi)$.

N'oubliez pas que lorsqu'on effectue un changement de variables dans une intégrale, la *valeur absolue* du jacobien apparaît.

Cependant notre convention de coordonnées sphériques fait venir $\sin(\phi)$ avec $\phi \in [0,\pi[$; vu que le signe de $\sin(\phi)$ y est toujours positif, cette histoire de valeur absolue est sans grandes conséquent. Ce n'est pas le cas de toutes les conventions possibles.

17.12.3.1 Coordonnées sphériques : inverse

Trouvons le changement inverse, c'est-à-dire trouvons ρ, θ et φ en termes de x, y et z. D'abord nous avons

$$\rho = \sqrt{x^2+y^2+z^2}. \tag{17.706}$$

Ensuite nous savons que

$$\cos(\theta) = \frac{z}{\rho} \tag{17.707}$$

détermine de façon unique[60] un angle $\theta \in [0,\pi]$. Dès que ρ et θ sont connus, nous pouvons poser $r = \rho\sin\theta$ et alors nous nous trouvons avec les équations

$$\begin{cases} x = r\cos(\varphi) & (17.708a) \\ y = r\sin(\varphi), & (17.708b) \end{cases}$$

qui sont similaires à celles déjà étudiées dans le cas des coordonnées polaires.

17.13 Calcul de limites

Beaucoup de techniques de calcul de limites fonctionnent bien avec les fonctions trigonométriques, entre autres grâce à l'utilisation des coordonnées polaires de la proposition 17.198. De plus, le théorème de la fonction implicite Nous en voyons quelques exemples à présent.

Exemple 17.207 (Limite et prolongement par continuité)**.**
La fonction

$$f(x) = \frac{\cos(x)-1}{x} \tag{17.709}$$

n'est pas définie en $x = 0$.

Nous avons vu dans l'équation (17.6) que $\cos(0) = 1$, donc la limite

$$\lim_{x\to 0} \frac{\cos(x)-1}{x} \tag{17.710}$$

est la limite définissant la dérivée de cosinus en 0 (ici, le x joue le rôle de ϵ). Le lemme 17.2 nous donne la dérivée du cosinus comme étant le sinus. Nous avons donc :

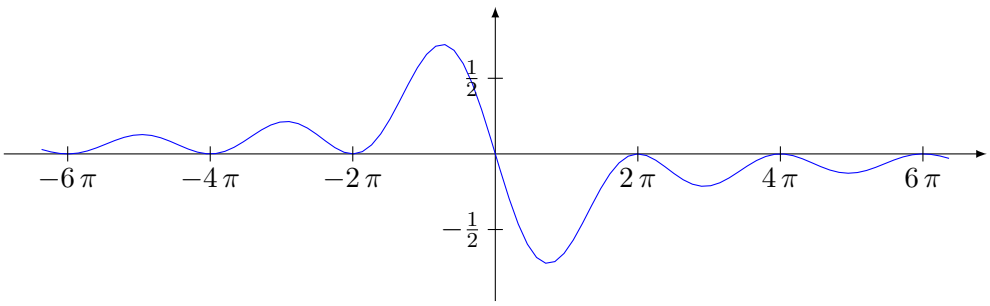
$$\lim_{x\to 0} \frac{\cos(x)-1}{x} = \sin(0) = 0, \tag{17.711}$$

et nous définissons le prolongement par continuité :

$$\tilde{f}(x) = \begin{cases} \frac{\cos(x)-1}{x} & \text{si } x \neq 0 \\ 0 & \text{sinon.} \end{cases} \tag{17.712}$$

Encore une fois, le graphe de la fonction $\tilde{f}$ ne présente aucune particularité autour de $x = 0$.

60. Le problème $\rho = 0$ ne se pose pas ; pourquoi ?



△

Exemple 17.208 (Un calcul heuristique de limite)**.**
Soit à calculer la limite suivante :

$$\lim_{x\to 0} \frac{e^{-2\cos(x)+2}\sin(x)}{\sqrt{e^{2\cos(x)+2}}-1}. \tag{17.713}$$

La stratégie que nous allons suivre pour calculer cette limite est de développer certaines parties de l'expression en série de Taylor, afin de simplifier l'expression. La première chose à faire est de remplacer $e^{y(x)}$ par $1+y(x)$ lorsque $y(x)\to 0$. La limite devient

$$\lim_{x\to 0} \frac{\big(-2\cos(x)+3\big)\sin(x)}{\sqrt{-2\cos(x)+2}}. \tag{17.714}$$

Nous allons maintenant remplacer $\cos(x)$ par 1 au numérateur et par $1-x^2/2$ au dénominateur. Pourquoi ? Parce que le cosinus du dénominateur est dans une racine, donc nous nous attendons à ce que le terme de degré deux du cosinus donne un degré un en dehors de la racine, alors que du degré un est exactement ce que nous avons au numérateur : le développement du sinus commence par x.

Nous calculons donc

$$\lim_{x\to 0} \frac{\sin(x)}{\sqrt{-2\left(1-\frac{x^2}{2}\right)+2}} = \lim_{x\to 0}\frac{\sin(x)}{x} = 1. \tag{17.715}$$

Tout ceci n'est évidemment pas très rigoureux, mais en principe vous avez tous les éléments en main pour justifier les étapes. △

17.13.1 Méthode des coordonnées polaires

La proposition suivante exprime la définition de la limite en d'autres termes, et va être pratique dans le calcul de certaines limites.

Proposition 17.209.
Soit $f\colon D\subset\mathbb{R}^m\to\mathbb{R}^n$, a un point d'accumulation de D et $\ell\in\mathbb{R}^n$. Nous définissons

$$E_r = \{f(x) \text{ tel que } x\in B(a,r)\cap D\}, \tag{17.716}$$

et

$$s_r = \sup\{\|v-\ell\| \text{ tel que } v\in E_r\}. \tag{17.717}$$

Alors nous avons $\lim_{x\to a} f(x)=\ell$ si et seulement si $\lim_{r\to 0} s_r = 0$.

Dans cette proposition, E_r représente l'ensemble des valeurs atteintes par f dans un rayon r autour de a. Le nombre s_r sélectionne, parmi toutes ces valeurs, celle qui est la plus éloignée de ℓ et donne la distance. En d'autres termes, s_r est la distance maximale entre $f(x)$ et ℓ lorsque x est à une distance au maximum r de a.

Lorsque nous avons affaire à une fonction $f\colon\mathbb{R}^2\to\mathbb{R}$, cette proposition nous permet de calculer facilement les limites en passant aux coordonnées polaires.

Exemple 17.210.
Reprenons la fonction de l'exemple 11.155 :

$$f(x,y) = \frac{xy}{x^2+y^2}. \tag{17.718}$$

Son domaine est $\mathbb{R}^2\backslash\{(0,0)\}$. Nous voulons calculer $\lim_{(x,y)\to(0,0)} f(x,y)$. Écrivons la définition de E_r :

$$E_r = \{f(x,y) \text{ tel que } (x,y) \in B\big((0,0),r\big)\}. \tag{17.719}$$

Les points de la boule sont, en coordonnées polaires, les points de la forme (ρ,θ) avec $\rho < r$. La chose intéressante est que $f(\rho,\theta)$ est relativement simple (plus simple que la fonction départ). En effet en remplaçant tous les x par $\rho\cos(\theta)$ et tous les y par $\rho\sin(\theta)$, et en utilisant le fait que $\cos^2(\theta)+\sin^2(\theta)=1$, nous trouvons

$$f(\rho,\theta) = \frac{\rho^2\cos(\theta)\sin(\theta)}{\rho^2} = \cos(\theta)\sin(\theta). \tag{17.720}$$

Cela signifie que

$$E_r = \{\cos(\theta)\sin(\theta) \text{ tel que } \theta \in [0,2\pi[\}. \tag{17.721}$$

Prenons ℓ quelconque. Le nombre s_r est le supremum des

$$\|\ell - \cos(\theta)\sin(\theta)\| \tag{17.722}$$

lorsque θ parcours $[0,2\pi]$. Nous ne sommes pas obligés calculer la valeur exacte de s_r. Ce qui compte ici est que s_r ne vaut certainement pas zéro, et ne dépend pas de r. Donc il est impossible d'avoir $\lim_{r\to 0} s_r = 0$, et la fonction donnée n'a pas de limite en $(0,0)$. △

Nous pouvons retenir cette règle pour calculer les limites lorsque $(x,y)\to(0,0)$ de fonctions $f\colon\mathbb{R}^2\to\mathbb{R}$:

(1) passer en coordonnées polaires, c'est-à-dire remplacer x par $\rho\cos(\theta)$ et y par $\rho\sin(\theta)$;
(2) nous obtenons une fonction g de ρ et θ. Si la limite $\lim_{r\to 0} g(r,\theta)$ n'existe pas ou dépend de θ, alors la fonction n'a pas de limite. Si on peut majorer g par une fonction ne dépendant pas de θ, et que cette fonction a une limite lorsque $r\to 0$, alors cette limite est la limite de la fonction.

La vraie difficulté de la technique des coordonnées polaires est de trouver le supremum de E_r, ou tout au moins de montrer qu'il est borné par une fonction qui a une limite qui ne dépend pas de θ. Une des situations classiques dans laquelle c'est facile est lorsque la fonction se présente comme une fonction de r multiplié par une fonction de θ.

Exemple 17.211.
Soit à calculer la limite

$$\lim_{(x,y)\to(0,0)} xy\left(\frac{x^2-y^2}{x^2+y^2}\right). \tag{17.723}$$

Le passage aux coordonnées polaires donne

$$f(r,\theta) = r^2\sin\theta\cos\theta(\cos^2\theta-\sin^2\theta). \tag{17.724}$$

Déterminer le supremum de cela est relativement difficile. Mais nous savons que de toutes façons, la quantité $\sin\theta\cos\theta(\cos^2\theta-\sin^2\theta)$ est bornée par 1. Donc

$$\|f(r,\theta)\| \leqslant r^2. \tag{17.725}$$

Maintenant la règle de l'étau montre que $\lim_{(x,y)\to(0,0)} f(x,y)$ est zéro.

La situation vraiment gênante serait celle avec une fonction de θ qui risque de s'annuler dans un dénominateur. △

L'exemple 17.197 donnera un cas où la méthode fonctionne plus difficilement. Entre autres parce qu'il utilisera en même temps la méthode des chemins et celle des coordonnées polaires.

Exemple 17.212.
Considérons fonction

$$f(x,y) = \frac{x^2+y^2}{x-y}. \tag{17.726}$$

Une mauvaise idée pour prouver que la limite n'existe pas pour $(x,y) \to (0,0)$ est de considérer le chemin (t,t). En effet, la fonction n'existe pas sur ce chemin. Or la méthode des chemins parle uniquement de chemins contenus dans le domaine de la fonction.

Nous prouvons que la limite n'existe pas en trouvant des chemins le long desquels les limites sont différentes. Si nous essayons le chemin (t,kt) avec k constant, nous trouvons

$$f(t,kt) = \frac{t(1+k^2)}{1-k}. \tag{17.727}$$

La limite $t \to 0$ est hélas toujours 0. Nous ne pouvons donc pas conclure.

Nous allons maintenant utiliser la même technique que celle utilisée en coordonnées polaires. Vous noterez que dans ce cas, travailler en cartésiennes donne lieu à des calculs plus longs. L'astuce consiste à prendre k non constant et à chercher par exemple $k(t)$ de façon à avoir

$$\frac{1+k(t)^2}{1-k(t)} = \frac{1}{t}. \tag{17.728}$$

Avec une telle fonction k, la fonction $t \mapsto f(t,tk(t))$ serait la constante 1. L'équation à résoudre pour k est

$$tk^2 + k + (t-1) = 0, \tag{17.729}$$

et les solutions sont

$$k(t) = \frac{-1 \pm \sqrt{1-4t(t-1)}}{2t}. \tag{17.730}$$

Nous proposons donc les chemins

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t \\ \frac{-1\pm\sqrt{1-4t(t-1)}}{2} \end{pmatrix} \tag{17.731}$$

Nous devons vérifier deux points. D'abord que ce chemin est bien défini, et ensuite que $tk(t)$ tend bien vers zéro lorsque $t \to 0$ (sinon $(t,k(t)t)$) n'est pas un chemin passant par $(0,0)$. Lorsque t est petit, ce qui se trouve sous la racine est proche de 1 et ne pose pas de problèmes. Ensuite,

$$\lim_{t\to 0} tk(t) = \frac{-1\pm 1}{2}. \tag{17.732}$$

En choisissant le signe $+$, nous trouvons un chemin qui nous convient.

Ce que nous avons prouvé est que

$$f\left(t, \frac{-1+\sqrt{1-4t(t-1)}}{2}\right) = 1 \tag{17.733}$$

pour tout t. Le long de ce chemin, la limite de f est donc 1. Cette limite est différente des limites obtenues le long de chemins avec k constant. La limite $\lim_{(x,y)\to(0,0)} f(x,y)$ n'existe donc pas. △

Exemple 17.213.
Considérons la fonction (figure 17.12)

$$f(x,y) = \begin{cases} \sqrt{x^2+y^2}\sin\frac{1}{x^2+y^2} & \text{si } (x,y) \neq (0,0) \\ 0 & \text{si } (x,y) = (0,0), \end{cases} \tag{17.734}$$

et cherchons la limite $(x, y) \to (0, 0)$. Le passage en coordonnées polaires[61] donne

$$f(\rho, \theta) = \rho \sin \frac{1}{\rho}. \tag{17.735}$$

Pour calculer la limite de cela lorsque $\rho \to 0$, nous remarquons que

$$0 \leqslant |\rho \sin \frac{1}{\rho}| \leqslant \rho \tag{17.736}$$

parce que $\sin(\frac{1}{\rho}) \leqslant 1$ quel que soit ρ. Or évidemment $\lim_{\rho \to 0} \rho = 0$, donc la limite de la fonction (17.735) est zéro et ne dépend pas de θ. Nous en concluons que $\lim_{(x,y)\to(0,0)} f(x, y) = 0$. △

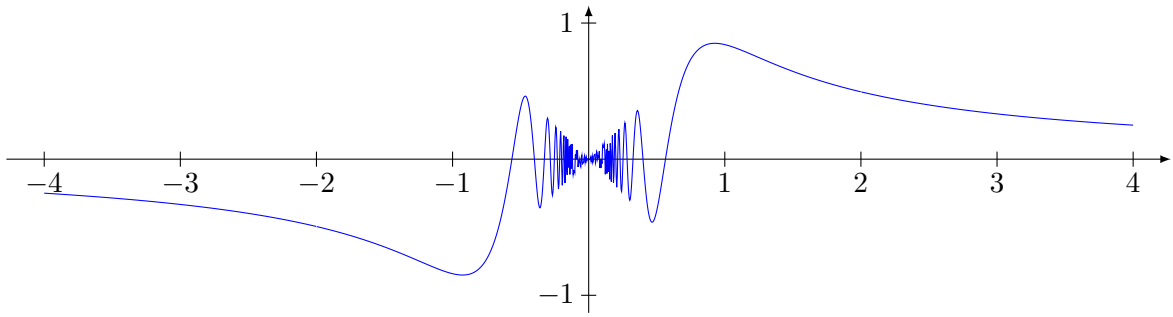


FIGURE 17.12 – La fonction de l'exemple 17.213.

17.13.2 Méthode du développement asymptotique

Nous savons que nous pouvons développer certaines fonctions en série grâce au développement de Taylor (théorème 11.375). Lorsque nous avons une limite à calculer, nous pouvons remplacer certaines parties de la fonction à traiter par la formule (11.1026b). Cela est très utile pour comparer des fonctions trigonométrique à des polynômes.

Lemme 17.214.
Nous avons la limite

$$\lim_{x \to 0} \frac{\sin(x)}{x} = 1. \tag{17.737}$$

Démonstration. Une manière de prouver cela est d'écrire

$$\sin(x) = x + h(x) \tag{17.738}$$

avec $h \in o(x)$, c'est-à-dire $\lim_{x\to 0} h(x)/x = 0$. Alors nous avons

$$\lim_{x \to 0} \frac{\sin(x)}{x} = \lim_{x \to 0} \frac{x + h(x)}{x} = \lim_{x \to 0} \frac{x}{x} + \lim_{x \to 0} \frac{h(x)}{x} = 1. \tag{17.739}$$

□

L'utilisation de la proposition 11.411 permet d'utiliser cette technique dans le cadre de limites à plusieurs variables. Reprenons le lemme 17.214 un tout petit peu modifié :

Lemme 17.215.
Pour tout $x > 0$ nous avons $\sin(x) < x$.

Démonstration. Nous posons $f(x) = x - \sin(x)$. Cette fonction vérifie $f(0) = 0$ et

$$f'(x) = 1 - \cos(x). \tag{17.740}$$

Vu que $|\cos(x)| \leqslant 1$, nous avons toujours $f'(x) \geqslant 0$ et même $f'(x) > 0$ pour $x \in]0, \delta]$. Donc f est au moins strictement croissante sur $]0, \delta]$ et ensuite strictement croissante presque partout. □

61. Proposition 17.198.

Exemple 17.216.
Soit à calculer $\lim_{(x,y)\to(0,0)} f(x,y)$ où

$$f(x,y) = \frac{\sin(xy)}{xy}. \tag{17.741}$$

La première chose à faire est de voir f comme la composée de fonctions $f = f_1 \circ f_2$ avec

$$\begin{aligned} f_1 \colon \mathbb{R} &\to \mathbb{R} \\ t &\mapsto \frac{\sin(t)}{t} \end{aligned} \tag{17.742}$$

et

$$\begin{aligned} f_2 \colon \mathbb{R}^2 &\to \mathbb{R} \\ (x,y) &\mapsto xy. \end{aligned} \tag{17.743}$$

Étant donné que $\lim_{(x,y)\to(0,0)} f_2(x,y) = 0$, nous avons $\lim_{(x,y)\to(0,0)} f(x,y) = \lim_{t\to 0} f_1(t) = 1$. △

Exemple 17.217.
Les dérivées partielles de la fonction $f(x,y) = xy^3 + \sin y$ au point $(0,\pi)$ sont

$$\partial_x f(0,\pi) = \frac{\partial f}{\partial x}(0,\pi) = \lim_{\substack{t\to 0\\ t\neq 0}} \frac{(t\pi^3 + \sin\pi) - (\sin\pi)}{t} = \pi^3,$$

$$\partial_y f(0,\pi) = \frac{\partial f}{\partial y}(0,\pi) = \lim_{\substack{t\to 0\\ t\neq 0}} \frac{0(\pi+t)^3 + \sin(t+\pi) - 0\cdot\pi^3}{t} = \cos\pi = -1,$$

△

17.14 Quelques intégrales avec de la trigonométrie

Le théorème 13.255 manque un peu d'exemples. Nous allons en voir quelques-uns maintenant.

17.14.1 Changement de variables

Le domaine $E = \{(x,y) \in \mathbb{R}^2 \text{ tel que } x^2+y^2 < 1\}$ s'écrit plus facilement $E = \{(r,\theta) \text{ tel que } r < 1\}$ en coordonnées polaires. Le passage aux coordonnées polaires permet de transformer une intégration sur un domaine rond à une intégration sur le domaine rectangulaire $]0,2\pi[\times]0,1[$. La question est évidemment de savoir si nous pouvons écrire

$$\int_E f = \int_0^{2\pi}\int_0^1 f(r\cos\theta, r\sin\theta)drd\theta. \tag{17.744}$$

Hélas, non ; la vie n'est pas aussi simple.

Définition 17.218.
*Un **difféomorphisme** est une application $g\colon A \to B$ telle que g et $g^{-1}\colon B \to A$ soient de classe C^1.*

Théorème 17.219.
Soit $g\colon A \to B$ un difféomorphisme. Soient $F \subset B$ un ensemble mesurable et borné et $f\colon F \to \mathbb{R}$ une fonction bornée et intégrable. Supposons que $g^{-1}(F)$ soit borné et que Jg soit borné sur $g^{-1}(F)$. Alors

$$\int_F f(x)dy = \int_{g^{-1}(F)f\left(g(x)\right)} |Jg(x)|dx \tag{17.745}$$

Pour rappel, Jg est le déterminant de la matrice jacobienne (aucun lien de parenté) donnée par

$$Jg = \det \begin{pmatrix} \partial_x g_1 & \partial_y g_1 \\ \partial_x g_2 & \partial_t g_2 \end{pmatrix}. \tag{17.746}$$

Comme dans les intégrales simples, il y a souvent moyen de trouver un changement de variables qui simplifie les expressions. Le domaine $E = \{(x, y) \in \mathbb{R}^2 \text{ tel que } x^2 + y^2 < 1\}$ par exemple s'écrit plus facilement $E = \{(r, \theta) \text{ tel que } r < 1\}$ en coordonnées polaires. Le passage aux coordonnées polaires permet de transformer une intégration sur un domaine rond à une intégration sur le domaine rectangulaire $]0, 2\pi[\times]0, 1[$. La question est évidemment de savoir si nous pouvons écrire

$$\int_E f = \int_0^{2\pi} \int_0^1 f(r\cos\theta, r\sin\theta) dr d\theta. \tag{17.747}$$

Hélas ce n'est pas le cas. Il faut tenir compte du fait que le changement de base dilate ou contracte certaines surfaces.

Soit $\varphi\colon D_1 \subset \mathbb{R}^2 \to D_2 \subset \mathbb{R}^2$ une fonction bijective de classe C^1 dont l'inverse est également de classe C^1. On désigne par x et y ses composantes, c'est-à-dire que

$$\varphi(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \end{pmatrix} \tag{17.748}$$

avec $(u, v) \in D_1$.

Théorème 17.220.
Soit une fonction continue $f\colon D_2 \to \mathbb{R}$. Alors

$$\int_{\varphi(D_1)} f(x, y) dx dy = \int_{D_1} f\big(x(u, v), y(u, v)\big) |J_\varphi(u, v)| du dv \tag{17.749}$$

où J_φ est le Jacobien de φ c'est-à-dire

$$J_\varphi(u, v) = \det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial u}{\partial v} \end{pmatrix}. \tag{17.750}$$

Ne pas oublier de prendre la valeur absolue lorsqu'on utilise le Jacobien dans un changement de variables.

17.14.2 Coordonnées polaires

Exemple 17.221.
Calculons la surface du disque D de rayon R. Nous devons calculer

$$\int_D dx dy. \tag{17.751}$$

Pour passer au polaires, nous savons que le disque est décrit par

$$D = \{(r, \theta) \text{ tel que } 0 \leqslant r \leqslant R, 0 \leqslant \theta \leqslant 2\pi\}. \tag{17.752}$$

Nous avons donc

$$\int_D dx dy = \int_D r\, dr d\theta = \int_0^{2\pi} \int_0^R r\, dr d\theta = 2\pi \int_0^R r\, dr = \pi R^2. \tag{17.753}$$

△

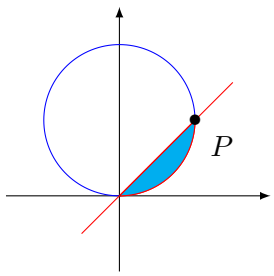


FIGURE 17.13 – Passage en polaire pour intégrer sur un morceau de cercle.

Exemple 17.222.
Montrons comment intégrer la fonction $f(x,y) = \sqrt{1-x^2-y^2}$ sur le domaine délimité par la droite $y = x$ et le cercle $x^2 + y^2 = y$, représenté sur la figure 17.13. Pour trouver le centre et le rayon du cercle $x^2+y^2 = y$, nous commençons par écrire $x^2+y^2-y = 0$, et ensuite nous reformons le carré : $y^2 - y = (y - \frac{1}{2})^2 - \frac{1}{4}$.

Le passage en polaire transforme les équations du bord du domaine en

$$\begin{aligned} \cos(\theta) &= \sin(\theta) \\ r^2 &= r\sin(\theta). \end{aligned} \tag{17.754}$$

L'angle θ parcours donc $]0, \pi/4[$, et le rayon, pour chacun de ces θ parcours $]0, \sin(\theta)[$. La fonction à intégrer se note maintenant $f(r,\theta) = \sqrt{1-r^2}$. Donc l'intégrale à calculer est

$$\int_0^{\pi/4} \left(\int_0^{\sin(\theta)} \sqrt{1-r^2} r\, rd \right). \tag{17.755}$$

Remarquez la présence d'un r supplémentaire pour le jacobien.

Notez que les coordonnées du point P sont $(1,1)$. △

En pratique, lors du passage en coordonnées polaires, le « $dxdy$ » devient « $r\,drd\theta$ ».

Exemple 17.223.
On veut évaluer l'intégrale de la fonction $f(x,y) = x^2 + y^2$ sur la région V suivante :

$$V = \{(x,y) \in \mathbb{R}^2 \,|\, x^2 + y^2 \leqslant 1,\ x > 0,\ y > 0\}.$$

On peut faire le calcul directement,

$$\int_V f(x,y)\, dV = \int_0^1 \int_0^{\sqrt{1-x^2}} x^2 + y^2\, dy\, dx = \int_0^1 \left(x^2\sqrt{1-x^2} + \frac{(1-x^2)^{3/2}}{3} \right) dx$$

mais c'est un peu ennuyeux. On peut simplifier beaucoup les calculs avec un changement de variables vers les coordonnées polaires. Dans ce cas, on sait bien que le difféomorphisme à utiliser est $\phi(r,\theta) = (r\cos\theta, r\sin\theta)$. Le jacobien J_ϕ est

$$J_\phi(r,\theta) = \begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = r, \tag{17.756}$$

qui est toujours positif. D'une part, la fonction f peut s'écrire sous la forme $f(\phi(r,\theta)) = r^2$ et d'autre part, $\phi^{-1}(V) =]0,1] \times]0, \pi/2[$. Par conséquent, la formule du changement de variables nous donne

$$\int_V f(x,y)\, dV = \int_0^{\pi/2} \int_0^1 r^3 dr\, d\theta = \int_0^{\pi/2} \frac{1}{4}\, d\theta = \frac{\pi}{8}.$$

△

Exemple 17.224.
Montrons comment intégrer la fonction $f(x,y) = \sqrt{1-x^2-y^2}$ sur le domaine délimité par la droite $y = x$ et le cercle $x^2 + y^2 = y$, représenté sur la figure 17.14. Pour trouver le centre et le rayon du cercle $x^2+y^2 = y$, nous commençons par écrire $x^2+y^2-y = 0$, et ensuite nous reformons le carré : $y^2 - y = (y - \frac{1}{2})^2 - \frac{1}{4}$.

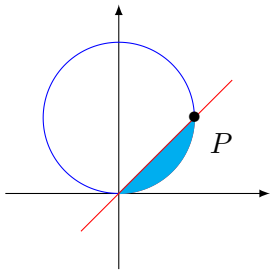


FIGURE 17.14 – Passage en polaire pour intégrer sur un morceau de cercle.

Le passage en polaire transforme les équations du bord du domaine en

$$\begin{aligned} \cos(\theta) &= \sin(\theta) \\ r^2 &= r\sin(\theta). \end{aligned} \tag{17.757}$$

L'angle θ parcours donc $]0, \pi/4[$, et le rayon, pour chacun de ces θ parcours $]0, \sin(\theta)[$. La fonction à intégrer se note maintenant $f(r,\theta) = \sqrt{1-r^2}$. Donc l'intégrale à calculer est

$$\int_0^{\pi/4} \left(\int_0^{\sin(\theta)} \sqrt{1-r^2} r\, rd \right). \tag{17.758}$$

Remarquez la présence d'un r supplémentaire pour le jacobien.

Notez que les coordonnées du point P sont $(1,1)$. △

17.14.3 Coordonnées cylindriques

Exemple 17.225.
On veut calculer le volume de la région A définie par l'intersection entre la boule unité et le cylindre qui a pour base un disque de rayon $1/2$ centré en $(0, 1/2)$

$$A = \{(x,y,z) \in \mathbb{R}^3 \,|\, x^2 + y^2 + z^1 \leqslant 1\} \cap \{(x,y,z) \in \mathbb{R}^3 \,|\, x^2 + (y-1/2)^2 \leqslant 1/4\}.$$

On peut décrire A en coordonnées cylindriques

$$\begin{aligned} A = \Big\{ &(r,\theta,z) \in]0,+\infty[\times[-\pi,\pi[\times\mathbb{R} \,| \\ &-\pi/2 < \theta < \pi,\ 0 < r \leqslant \sin\theta,\ -\sqrt{1-r^2} \leqslant z \leqslant \sqrt{1-r^2} \Big\}. \end{aligned} \tag{17.759}$$

Le jacobien de ce changement de variables, J_{cyl}, est

$$J_{cyl}(r,\theta), z = \begin{vmatrix} \cos\theta & \sin\theta & 0 \\ -r\sin\theta & r\cos\theta & 0 \\ 0 & 0 & \end{vmatrix} = r, \tag{17.760}$$

qui est toujours positif. Le volume de A est donc

$$\int_{\mathbb{R}^3} \chi_A(x,y,z)\, dV = \int_{-\pi/2}^{\pi/2} \int_0^{\sin\theta} \int_{-\sqrt{1-r^2}}^{\sqrt{1-r^2}} r dz\, dr\, d\theta = \frac{2\pi}{8} + \frac{8}{9}.$$

△

Exemple 17.226 (Volume d'un solide de révolution)**.**
Soit $g : [a, b] \to \mathbb{R}_+$ une fonction continue et positive. On dit que le solide A décrit par

$$A = \left\{(x, y, z) \in \mathbb{R}^3 \,|\, z \in [a, b],\ \sqrt{x^2 + y^2} \leqslant g^2(z)\right\}$$

est un solide de révolution. Afin de calculer son volume, on peut décrire A en coordonnées cylindriques,

$$A = \left\{(r, \theta, z) \in]0, +\infty[\times[-\pi, \pi[\times\mathbb{R} \,|\, a \leqslant z \leqslant b,\ 0 < r^2 \leqslant g^2(z)\right\}.$$

Le jacobien de ce changement de variables est $J_{cyl} = r$, comme dans l'exemple précédent. Le volume de A est donc

$$\int_{\mathbb{R}^3} \chi_A(x, y, z)\, dV = \int_a^b \int_{-\pi}^{\pi} \int_0^{g(z)} r\, dr\, d\theta\, dz = \int_a^b \pi g^2(z)\, dz.$$

Cette formule peut être utilisée pour tout solide de révolution. △

17.14.3.1 Coordonnées sphériques

Le calcul est un peu plus long :

$$\begin{aligned} J(\rho, \theta, \varphi) &= \begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \varphi} \\ \frac{\partial z}{\partial \rho} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \varphi} \end{vmatrix} \\ &= \begin{vmatrix} \sin\theta\cos\varphi & \rho\cos\theta\cos\varphi & -\rho\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & \rho\cos\theta\sin\varphi & -\rho\sin\theta\cos\varphi \\ \cos\theta & -\rho\sin\theta & 0 \end{vmatrix} \\ &= \rho^2 \sin\theta. \end{aligned} \tag{17.761}$$

Donc

$$dx\, dy\, dz = \rho^2 \sin(\theta)\, d\rho\, d\theta\, d\varphi. \tag{17.762}$$

17.14.4 Coordonnées sphériques

Exemple 17.227.
On veut calculer le volume du cornet de glace A

$$A = \left\{(x, y, z) \in \mathbb{R}^3 \,|\, (x, y) \in \mathbb{S}^2,\ \sqrt{x^2 + y^2} \leqslant z \leqslant \sqrt{1 - x^2 - y^2}\right\}.$$

On peut décrire A en coordonnées sphériques.

$$A = \{(\rho, \theta, \phi) \in]0, +\infty[\times[-\pi, \pi[\times[0, \pi[\,|\, 0 < \phi \leqslant \pi/4,\ 0 < \rho \leqslant 1\}.$$

Le jacobien de ce changement de variables J_{sph} est

$$J_{sph}(\rho, \theta, \phi) = \begin{vmatrix} \cos\theta\sin\phi & \sin\theta\sin\phi & \cos\phi \\ -\rho\sin\theta\sin\phi & \rho\cos\theta\sin\phi & 0 \\ \rho\cos\theta\cos\phi & \rho\sin\theta\cos\phi & -\rho\sin\phi \end{vmatrix} = \rho^2 \sin\phi, \tag{17.763}$$

Le volume de A est donc

$$\int_{\mathbb{R}^3} \chi_A(x, y, z)\, dV = \int_{-\pi}^{\pi} \int_0^{\pi/4} \int_0^1 \rho^2 \sin\phi\, d\rho\, d\phi\, d\theta = \frac{2\pi}{3}\left(1 - \frac{1}{\sqrt{2}}\right).$$

△

Exemple 17.228 (Une petite faute à ne pas faire)**.**
Si nous voulons calculer le volume de la sphère de rayon R, nous écrivons donc

$$\int_0^R dr \int_0^{2\pi} d\theta \int_0^{\pi} r^2 \sin(\phi) d\phi = 4\pi R = \frac{4}{3}\pi R^3. \tag{17.764}$$

Ici, la valeur absolue n'est pas importante parce que lorsque $\phi \in]0, \pi, [$, le sinus de ϕ est positif.

Des petits malins pourraient remarquer que le changement de variable (17.704) est encore un paramétrage de $\mathbb{R}^3$ si on intervertit le domaine des angles :

$$\begin{aligned} \theta &: 0 \to \pi \\ \phi &: 0 \to 2\pi, \end{aligned} \tag{17.765}$$

alors nous paramétrons encore parfaitement bien la sphère, mais hélas

$$\int_0^R dr \int_0^{\pi} d\theta \int_0^{2\pi} r^2 \sin(\phi) d\phi = 0. \tag{17.766}$$

Pourquoi ces « nouvelles » coordonnées sphériques sont-elles mauvaises ? Il y a que quand l'angle ϕ parcours $]0, 2\pi[$, son sinus n'est plus toujours positif, donc la *valeur absolue* du jacobien n'est plus $r^2 \sin(\phi)$, mais $r^2 \sin(\phi)$ pour les ϕ entre 0 et π, puis $-r^2 \sin(\phi)$ pour ϕ entre π et 2π. Donc l'intégrale (17.766) n'est pas correcte. Il faut la remplacer par

$$\int_0^R dr \int_0^{\pi} d\theta \int_0^{\pi} r^2 \sin(\phi) d\phi - \int_0^R dr \int_0^{\pi} d\theta \int_{\pi}^{2\pi} r^2 \sin(\phi) d\phi = \frac{4}{3}\pi R^3 \tag{17.767}$$

△

17.14.5 Un autre système utile

Un changement de variables que l'on voit assez souvent est

$$\begin{cases} u = x + y & \text{(17.768a)} \\ v = x - y. & \text{(17.768b)} \end{cases}$$

Afin de calculer son jacobien, il faut d'abord exprimer x et y en fonctions de u et v :

$$\begin{cases} x = (u + v)/2 & \text{(17.769a)} \\ y = (u - v)/2. & \text{(17.769b)} \end{cases}$$

La matrice jacobienne est

$$\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}. \tag{17.770}$$

Le déterminant vaut $-\frac{1}{2}$. Nous avons donc

$$dxdy = \frac{1}{2} dudv. \tag{17.771}$$

Nous insistons sur le fait que c'est $\frac{1}{2}$ et non $-\frac{1}{2}$ qui intervient parce que que la formule du changement de variable demande d'introduire la *valeur absolue* du jacobien.

Exemple 17.229.
Calculer l'intégrale de la fonction $f(x, y) = x^2 - y^2$ sur le domaine représenté sur la figure 17.15.

Les droites qui délimitent le domaine d'intégration sont

$$\begin{aligned} y &= -x + 2 \\ y &= x - 2 \\ y &= x \\ y &= -x \end{aligned} \tag{17.772}$$

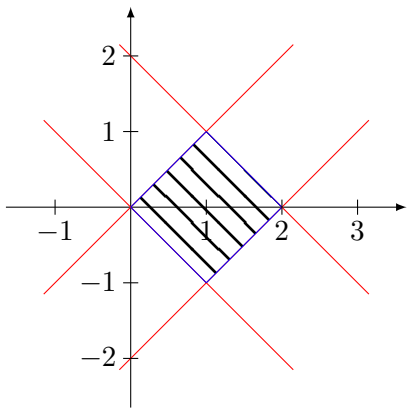


FIGURE 17.15 – Un domaine qui s'écrit étonnament bien avec un bon changement de coordonnées.

Le domaine est donc donné par les équations

$$\begin{cases} y + x < 2 & (17.773a) \\ y - x > -2 & (17.773b) \\ y - x < 0 & (17.773c) \\ y + x > 0. & (17.773d) \end{cases}$$

En utilisant le changement de variables $u = x + y$, $v = x - y$ nous trouvons le domaine $0 < u < 2$, $0 < v < 2$. En ce qui concerne la fonction, $f(x, y) = (x + y)(x - y)$ et par conséquent

$$f(u, v) = uv. \tag{17.774}$$

L'intégrale à calculer est simplement

$$\int_0^2 \int_0^2 uv\, dudv = \int_0^2 u\, du \left[\frac{v^2}{2}\right]_0^2 = 2\int_0^2 u\, du = 4. \tag{17.775}$$

△

17.15 Aire d'une surface de révolution

Soit γ une courbe dans le plan xy, paramétrée par

$$\gamma(u) = \begin{pmatrix} x(u) \\ y(u) \\ 0 \end{pmatrix} \tag{17.776}$$

avec $u \in [a, b]$. Nous supposons que la courbe est toujours positive, c'est-à-dire $y(u) > 0$ pour tout u.

Nous voulons considérer la surface obtenue en effectuant une rotation de cette ligne autour de l'axe X. Chaque point de la courbe va parcourir un cercle de rayon $y(u)$ dans le plan YX et centré en $(x(u), 0, 0)$. La surface est donc donnée par

$$\varphi(u, \theta) = \begin{pmatrix} x(u) \\ y(u)\cos\theta \\ y(u)\sin\theta \end{pmatrix} \tag{17.777}$$

avec $(u, \theta) \in [a, b] \times [0, 2\pi]$. Notez que la courbe de départ correspond à $\theta = 0$.

Les vecteurs tangents à la surface pour ce paramétrage sont

$$T_u = \frac{\partial\varphi}{\partial u} = \begin{pmatrix} x'(u) \\ y'(u)\cos\theta \\ y'(u)\sin\theta \end{pmatrix} \quad T_\theta = \frac{\partial\varphi}{\partial\theta} = \begin{pmatrix} 0 \\ -y(u)\sin\theta \\ y(u)\cos\theta \end{pmatrix}. \tag{17.778}$$

Le produit vectoriel de ces deux vecteurs vaut

$$\begin{aligned} T_u \times T_\theta &= \begin{vmatrix} e_x & e_y & e_z \\ x' & y'\cos\theta & y'\sin\theta \\ 0 & -y\sin\theta & y\cos\theta \end{vmatrix} \\ &= y'(u)y(u)\,e_x - x'(u)y(u)\cos\theta\,e_y + x'(u)y(u)\sin\theta\,e_z. \end{aligned} \tag{17.779}$$

En ce qui concerne la norme :

$$dS = \|T_u \times T_\theta\| = \sqrt{(y'y)^2 + (x'y)^2} = |y(u)|\sqrt{y'(u)^2 + x'(u)^2}. \tag{17.780}$$

Étant donné que nous avons supposé que $y(u) > 0$, nous pouvons supprimer les valeurs absolues, et l'aire de la surface de révolution devient :

$$\begin{aligned} Aire(S) &= \int_0^{2\pi} d\theta \int_a^b y(u)\sqrt{x'(u)^2 + y'(u)^2}du \\ &= 2\pi \int_a^b y(u)\sqrt{x'(u)^2 + y'(u)^2}du. \end{aligned} \tag{17.781}$$

Exemple 17.230.
Calculons la surface du cône de révolution de rayon (à la base) R et de hauteur h. La courbe de départ est le segment droite qui part de $(0,0)$ et qui termine en (R,h) de la figure 17.16.

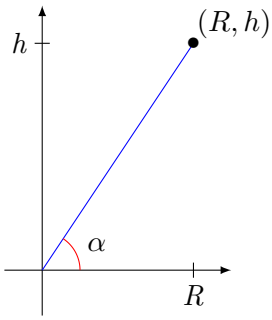


FIGURE 17.16 – En faisant tourner cette droite autour de l'axe X, nous obtenons un cône.

Ce segment peut être paramétré par

$$\gamma(u) = \begin{pmatrix} Ru \\ hu \\ 0 \end{pmatrix} \tag{17.782}$$

avec $u \in [0,1]$. Cela donne $x(u) = Ru$, $y(u) = hu$ et par conséquent

$$Aire = 2\pi \int_0^1 hu\sqrt{R^2 + h^2} = \pi h\sqrt{R^2 + h^2}. \tag{17.783}$$

Ce résultat peut aussi être exprimé en fonction de l'angle, grâce à la formule (17.489). En sachant que $h = \sqrt{h^2 + R^2}\sin(\alpha)$, nous trouvons

$$Aire = \pi(R^2 + h^2)\sin(\alpha). \tag{17.784}$$

△

Exemple 17.231.
Calculons la surface latérale du tore obtenu par révolution du cercle de la figure 17.17.

Le chemin qui détermine le cercle de départ est

$$\gamma(u) = \begin{pmatrix} R\cos(u) \\ a + R\sin(u) \\ 0 \end{pmatrix}, \tag{17.785}$$

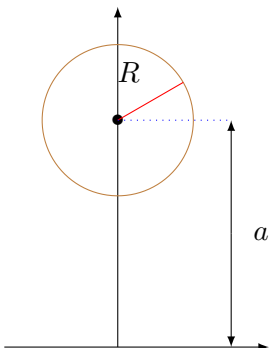


FIGURE 17.17 – Si nous tournons ce cercle autour de l'axe X, nous obtenons un tore de rayon « externe » a et de rayon « interne » R.

c'est-à-dire $x(u) = R\cos(u)$, $y(u) = a + R\sin(u)$ avec $u \in [0, 2\pi]$. Nous avons donc l'aire

$$\begin{aligned} Aire &= 2\pi \int_0^{2\pi} \big(a + R\sin(u)\big) R\, du \\ &= 2\pi R\big(2\pi a + R[-\cos(u)]_0^{2\pi}\big) \\ &= 4\pi^2 aR. \end{aligned} \tag{17.786}$$

△

17.16 Table de caractères du groupe diédral

Cette section vient de [54] ; nous avons comme but d'établir la table des caractères des représentations complexes du groupe diédral D_n.

17.16.1 Représentations de dimension un

Nous nous occupons des représentations de D_n sur $\mathbb{C}$. Les applications linéaires $\mathbb{C} \to \mathbb{C}$ sont seulement les multiplications par des nombres complexes. Nous cherchons donc $\psi\colon D_n \to \mathbb{C}^*$.

Nous savons que D_n est généré [62] par s et r. Vu que $s^2 = 1$, nous avons

$$\psi(s)^2 = \psi(s^2) = \psi(1) = 1, \tag{17.787}$$

donc $\psi(s) \in \{-1, 1\}$. Nous savons aussi que $srsr = 1$, donc

$$\psi(s)^2\psi(r)^2 = 1, \tag{17.788}$$

ce qui donne $\psi(r) \in \{-1, 1\}$.

Nous avons donc quatre représentations de dimension un données par

	$\psi(r) = 1$	$\psi(r) = -1$
$\psi(s) = 1$	ρ^{++}	ρ^{+-}
$\psi(s) = -1$	ρ^{-+}	ρ^{--}

Attention au fait que nous devons aussi avoir la relation $\psi(r)^n = \psi(r^n) = 1$. Donc $\psi(r)$ doit être une racine n^{e} de l'unité. Nous allons donc devoir avoir un compte différent selon la parité de n. Nous en reparlerons à la fin, au moment de faire les comptes. En ce qui concerne les caractères correspondants,

	r^k	sr^k
χ^{++}	1	1
χ^{+-}	$(-1)^k$	$(-1)^k$
χ^{-+}	1	-1
χ^{--}	$(-1)^k$	$(-1)^{k+1}$

62. Voir proposition 17.85 et tout ce qui suit.

Étant donné qu'ils sont tous différents, ce sont des représentations deux à deux non équivalentes, lemme 15.15.

17.16.2 Représentations de dimension deux

Nous cherchons maintenant les représentations $\rho\colon D_n \to \mathrm{End}(\mathbb{C}^2)$. Ici nous supposons connue la liste des éléments de D_n donnée par le corolaire 17.86. Soit $\omega = e^{2i\pi/n}$ et $h \in \mathbb{Z}$; nous considérons la représentation $\rho^{(h)}$ de D_n définie par

$$\rho^{(h)}(r^k) = \begin{pmatrix} \omega^{hk} & 0 \\ 0 & \omega^{-hk} \end{pmatrix} \tag{17.789a}$$

$$\rho^{(h)}(st^k) = \begin{pmatrix} 0 & \omega^{-hk} \\ \omega^{hk} & 0 \end{pmatrix}. \tag{17.789b}$$

Cela donne bien $\rho^{(h)}$ sur tous les éléments de D_n par la proposition 17.85. Nous pouvons restreindre le domaine de h en remarquant d'abord que $\rho^{(h)} = \rho^{(h+n)}$, et ensuite que les représentations $\rho^{(h)}$ et $\rho^{(-h)}$ sont équivalentes. Un opérateur d'entrelacement est donné par $T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, et il est facile de vérifier que $T\rho^{(h)}(x) = \rho^{-h}(x)T$ avec $x = r^k$ puis avec $x = sr^k$.

Donc $\rho^{(h)} \simeq \rho^{(-h)} \simeq \rho^{(n-h)}$ et nous pouvons restreindre notre étude à $0 \leqslant h \leqslant \frac{n}{2}$.

Nous allons séparer les cas $n = 0$, $h = n/2$ et les autres. En effet si nous notons par commodité $a = \omega^h$, alors un vecteur (x, y) est vecteur propre de $\rho^{(h)}(s)$ et de $\rho^{(h)}(r)$ si et seulement s'il vérifie les systèmes d'équations

$$\begin{cases} ax = \lambda x & (17.790a) \\ \frac{1}{a}y = \lambda y & (17.790b) \end{cases}$$

et

$$\begin{cases} \frac{1}{a}y = \mu x & (17.791a) \\ ax = \mu y & (17.791b) \end{cases}$$

avec λ et μ des nombres non nuls. Une représentation sera réductible si et seulement si ces deux systèmes acceptent une solution non nulle commune. Il est vite vu que si $x \neq 0$ et $y \neq 0$, alors $a^2 = 1$, ce qui signifie $h = 0$ ou $h = n/2$. Sinon, il n'y a pas de solutions, et la représentation associée est irréductible.

(1) $h = 0$. Nous avons

$$\rho^{(0)}(r^k) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \rho^{(0)}(sr^k) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{17.792}$$

donc le caractère de cette représentation est $\chi^{(0)}(r^k) = 2$ et $\chi^{(0)}(sr^k) = 0$. Donc nous avons

$$\chi^{(0)} = \chi^{++} + \chi^{-+}. \tag{17.793}$$

Il y a maintenant (au moins) quatre façons de voir que la représentation $\rho^{(0)}$ est réductible.

Première méthode Trouver un opérateur d'entrelacement. Pour cela nous calculons les matrices :

$$S(r) = (\rho^{++} \oplus \rho^{-+})(r^k) = \begin{pmatrix} \rho^{++}(r^k) & 0 \\ 0 & \rho^{-+}(r^k) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{17.794a}$$

$$S(sr^k) = (\rho^{++} \oplus \rho^{-+})(sr^k) = \begin{pmatrix} \rho^{++}(sr^k) & 0 \\ 0 & \rho^{-+}(sr^k) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{17.794b}$$

(17.794c)

Nous cherchons une matrice T telle que $TS(r^k) = \rho^{(0)}(r^k)T$ et $TS(sr^k) = \rho^{(0)}(sr^k)T$. Étant donné que $S(r^k) = \mathbb{1} = \rho^{(0)}(r^k)$, la première contrainte n'en est pas une. Nous pouvons vérifier qu'avec $T = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, nous avons bien

$$T \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{17.795}$$

Donc ce T entrelace $\rho^{++} \oplus \rho^{-+}$ avec $\rho^{(0)}$ qui sont donc deux représentations équivalentes. Donc $\rho^{(0)}$ est réductible et ça ne nous intéresse pas de la lister.

Seconde méthode Invoquer le théorème 15.22(1) et dire que les représentations sont équivalentes parce que les caractères sont égaux.

Troisième méthode Utiliser le théorème 15.22(2) et calculer $\langle \chi^{(0)}, \chi^{(0)} \rangle$:

$$\langle \chi^{(0)}, \chi^{(0)} \rangle = \frac{1}{|D_n|} \sum_{g \in D_n} |\chi^{(0)}(g)|^2 \tag{17.796a}$$

$$= \frac{1}{2n} \big(4 + 0 + 4(n-1)\big) \tag{17.796b}$$

$$= 2. \tag{17.796c}$$

Ici le 4 est pour le 1, le zéro est pour les termes sr^k et $4(n-1)$ est pour les $n-1$ termes r^k. Vu que le résultat n'est pas 1, la représentation $\rho^{(0)}$ n'est pas irréductible.

Quatrième méthode Regarder les solutions des systèmes (17.790) et (17.791) dont nous avons parlé plus haut.

La première méthode a l'avantage d'être simple et ne demander aucune théorie particulière à part les définitions. La seconde méthode est la plus rapide, mais demande un théorème très puissant. La troisième utilise également un théorème assez avancé, mais a l'avantage sur les deux autres méthodes de ne pas avoir besoin de savoir à priori un candidat décomposition de $\rho^{0)}$; cette méthode est applicable même sans faire la remarque que $\chi^{(0)} = \chi^{++} + \chi^{-+}$. Quoi qu'il en soit, nous ne listons pas $\chi^{(0)}$ dans notre table de caractères.

(2) $h = n/2$. Vu que $\omega^{n/2} = e^{i\pi} = -1$, nous avons

$$\rho^{(n/2)}(r^k) = \begin{pmatrix} (-1)^k & 0 \\ 0 & (-1)^k \end{pmatrix} \quad \rho^{(n/2)}(sr^k) = \begin{pmatrix} 0 & (-1)^k \\ (-1)^k & 0 \end{pmatrix} \quad , \tag{17.797}$$

et donc

$$\chi^{(n/2)}(r^k) = 2(-1)^k \tag{17.798a}$$

$$\chi^{(n/2)}(sr^k) = 0. \tag{17.798b}$$

Il est vite vu que $\chi^{(n/2)} = \chi^{+-} + \chi^{-+}$. Ergo la représentation $\rho^{(n/2)}$ n'est pas irréductible.

(3) $0 < h < \frac{n}{2}$. Dans ce cas nous avons $\omega^h \neq \omega^{-h}$, et en regardant les systèmes d'équations donnés plus haut, nous voyons que $\rho^{(h)}(s)$ et $\rho^{(h)}(r)$ n'ont pas de vecteurs propres communs. Donc ces représentations sont irréductibles.

Nous devons cependant encore vérifier si elles sont deux à deux non équivalentes. Supposons que pour $h \neq h'$ nous ayons une matrice $T \in \mathrm{GL}(2, \mathbb{C})$ telle que $T\rho^{(h)}(r)T^{-1} = \rho^{(h')}(r)$. Cela impliquerait en particulier que les matrices $\rho^{(h)}(r)$ et $\rho^{(h')}(r)$ aient même valeurs propres. Nous aurions donc $\{\omega^h, \omega^{-h}\} = \{\omega^{h'}, \omega^{-h'}\}$. Mais cela est impossible avec $0 < h < h' < \frac{n}{2}$. Donc toutes ces représentations sont distinctes.

Le caractère de la représentation $\rho^{(h)}$ est $\chi^{(h)}(r^k) = \omega^{hk} + \omega^{-hk} = 2\cos\left(\frac{2\pi hk}{n}\right)$.
Nous ajoutons donc la ligne suivante à notre liste :

	r^k	sr^k
$\chi^{(h)}$	$2\cos\left(\frac{2\pi hk}{n}\right)$	0

17.16.3 Le compte pour n pair

Nous avons 4 représentations de dimension 1 puis $\frac{n}{2}-1$ représentations de dimension 2. En tout nous avons

$$\frac{n}{2}+3 \tag{17.799}$$

représentations irréductibles modulo équivalence. Cela fait le compte en vertu des classes de conjugaisons listées en 17.8.1.5. Pour rappel, le nombre de représentations non équivalentes est égal au nombre de classes de conjugaison par le corolaire 15.25. Notons que c'est cela qui justifie le fait que nous ne devons pas chercher d'autres représentations. Nous sommes sûrs de les avoir toutes trouvées.

17.16.4 Le compte pour n impair

Nous avions fait mention plus haut du fait que si ψ est une représentation de dimension 1, le nombre $\psi(r)$ devait être une racine n^{e} de l'unité. Donc en dimension 1 nous avons seulement les représentations ρ^{++} et ρ^{-+}. Pour celles de dimension 2, nous en avons $\frac{n-1}{2}$. En tout nous avons donc

$$\frac{n+3}{2} \tag{17.800}$$

représentations irréductibles modulo équivalence. Cela fait le compte en vertu des classes de conjugaisons listées en 17.8.1.6.

Chapitre 18

Corps finis, racines de l'unité

18.1 Le groupe des racines de l'unité dans les nombres complexes

18.1.1 Le groupe

Définition 18.1.
*Une **racine** n^e **de l'unité** dans un anneau est une racine du polynôme $X^n - 1$.*

Lemme-définition 18.2.
L'ensemble

$$\mathrm{U}_n = \{e^{2i\pi k/n} \text{ tel que } k = 0, \dots, n-1\} \tag{18.1}$$

est un groupe pour la multiplication.

Ses éléments forment exactement l'ensemble des racines de l'unité dans $\mathbb{C}$.

Démonstration. Il est vite vu que tous les éléments de U_n sont des racines de l'unité parce que

$$\left(e^{2i\pi k/n}\right)^n = e^{2i\pi k} = 1, \tag{18.2}$$

entre autres à cause du lemme 17.10.

Cela nous donne déjà n racines pour $X^n - 1$ dans $\mathbb{C}$. Le théorème 6.101 nous indique qu'il ne peut pas y en avoir plus. □

Un des intérêts du groupe des racines est qu'il permet de factoriser $X^n - 1$, comme nous le verrons via les polynômes cyclotomiques dans le lemme 18.18.

Lemme 18.3.
L'ensemble U_n est un groupe cyclique[1] *d'ordre n généré par $\xi = e^{2i\pi/n}$.*

Démonstration. Il y a les trois propriétés à vérifier pour que ce soit un groupe.

Neutre Le nombre 1 est une racine de l'unité.

Inverse Si $\omega \in U_n$ alors $\omega^n = 1$ et donc $\omega\omega^{n-1} = 1$, ce qui signifie que ω^{n-1} est un inverse de ω. Il reste à voir que $\omega^{n-1} \in U_n$. En effet $\left(\omega^{n-1}\right)^n = (\omega^n)^{n-1} = 1^{n-1} = 1$.

Associativité Cas particulier de l'associativité dans $\mathbb{C}$.

Le fait que ce soit un groupe cyclique contenant n éléments est la définition. □

Le lemme suivant donne les autres générateurs.

Lemme 18.4.
Le nombre ξ^a est un générateur de U_n si et seulement si $\mathrm{pgcd}(a, n) = 1$.

1. Définition 2.12.

Démonstration. Si $\text{pgcd}(a, n) = 1$ alors le théorème de Bézout 3.13 nous fournit des entiers u et v tels que $ua + vn = 1$. Alors nous avons

$$e^{2i\pi/n} = e^{2(ua+vn)i\pi/n} = (e^{2ai\pi/n})^u, \tag{18.3}$$

ce qui signifie que ξ est dans le groupe engendré par ξ^a, et par conséquent tout U_n est engendré.

Pour l'implication inverse, nous utilisons Bézout dans le sens inverse. Soit ξ^a un générateur de U_n. Alors il existe u tel que $(\xi^a)^u = \xi$, donc $\xi^{au-1} = 1$, c'est-à-dire qu'il existe v tel que $au-1 = vn$. Cette dernière égalité implique que $\text{pgcd}(a, n) = 1$. □

Exemple 18.5.
Une conséquence tout à fait extraordinaire de ce lemme est que 7 est générateur de $\mathbb{Z}/12\mathbb{Z}$ (parce que $\text{pgcd}(7, 12) = 1$). Or en solfège, une quinte fait 7 demi-tons, et une gamme en fait 12. Le cycle des quintes est donc générateur de la gamme[368]. Ce fait est connu des pianistes[2] depuis des siècles. △

Proposition 18.6 (Intersection par deux)**.**
Les ensembles U_α et U_β ont une intersection réduite à $\{1\}$ si et seulement si α et β sont premiers entre eux.

Démonstration. Nous rappelons qu'une racine α^{e}de l'unité peut s'écrire sous la forme $e^{2i\pi k/\alpha}$ avec $0 \leqslant k < \alpha$.

Sens direct Par contraposée, nous supposons que α et β ne sont pas premiers entre eux, et nous notons d leur pgcd. Nous nommons $\alpha = d\alpha'$ et $\beta = d\beta'$. Pour trouver une intersection entre U_α et U_β nous devons trouver une valeur de $0 < k < \alpha$ telle que

$$(e^{2i\pi k/\alpha})^\beta = e^{2i\pi k\beta/\alpha} = 1, \tag{18.4}$$

c'est-à-dire une valeur de k telle que $k\beta/\alpha$ soit un entier. Mais $k\beta/\alpha = k\beta'/\alpha'$ et par conséquent prendre $k = \alpha'$ fonctionne. Surtout que par hypothèse $d > 1$ et donc $k = \alpha' < \alpha$.

Sens réciproque Supposons maintenant que α et β soient premiers entre eux. Soit $z \in U_\alpha \cap U_\beta$. Le fait que z soit une racine α^{e}de l'unité implique qu'il existe un $k < \alpha$ tel que $z = e^{2i\pi k/\alpha}$. Mais si z est également une racine β^{e}de l'unité, alors $z^\beta = 1$, c'est-à-dire que $k\beta/\alpha$ doit être un entier, soit l cet entier. Nous avons

$$k\beta = l\alpha. \tag{18.5}$$

Si $k > 0$, comme le nombre α divise $k\beta$, cela conduirait via le lemme de Gauss 3.26 à dire que α divise k. Mais α ne peut pas diviser k parce que nous avions supposé que k était strictement plus petit que α. Donc $k = 0$ et $z = 1$.

□

Proposition 18.7 (Intersection : le cas général[1])**.**
Soient des entiers positifs $\alpha_1, \ldots, \alpha_p$. Nous avons

$$\bigcup_{i=1}^{p} U_{\alpha_i} = \{1\} \tag{18.6}$$

si et seulement si $\text{pgcd}(\alpha_1, \ldots, \alpha_p) = 1$ *(c'est-à-dire que les α_i sont premiers dans leur ensemble).*

Démonstration. Nous le décomposons les α_i en facteurs premiers[3] de la façon suivante : $\alpha_i = \prod_{k\in\mathbb{N}} p_k^{\alpha_i^{(k)}}$ où les p_k sont les nombres premiers.

2. Même ceux qui ignorent le théorème de Bézout.
3. Théorème 3.28.

Caractérisation par une décomposition en facteurs premiers Les éléments z différents de 1 dans U_{α_1} s'écrivent sous la forme

$$z = e^{2i\pi k/\alpha_1} \tag{18.7}$$

avec $0 < k < \alpha_1$.

Pour tout $i \neq 1$, le fait que $z \in U_{\alpha_i} \cap U_{\alpha_1}$ se traduit par le fait que $\left(e^{2i\pi k/\alpha_1}\right)^{\alpha_i} = 1$, c'est-à-dire que $\alpha_i k/\alpha_1$ est entier, donc que α_1 divise $k\alpha_i$. Par conséquent il existera un élément différent de 1 dans l'intersection des U_{α_i} si et seulement s'il existe un entier k strictement compris entre 0 et α_1 pour lequel α_1 divise tous les $k\alpha_i$.

Un entier $0 < k < \alpha_1$ convient si et seulement si pour tout l, la puissance de p_l dans la décomposition de k est au moins égale à

$$\alpha_1^{(l)} - \alpha_i^{(l)} \tag{18.8}$$

pour tout l.

Sens direct L'hypothèse $\operatorname{pgcd}(\alpha_1, \dots, \alpha_p) \neq 1$ implique qu'il existe un l pour lequel tous les $\alpha_i^{(l)}$ sont non nuls. Nous construisons le k voulu en prenant pour tout p_i la même puissance que celle dans α_1, sauf pour p_l pour lequel nous prenons la puissance $\alpha_1^{(l)} - \min_i\{\alpha_i^{(l)}\}$. Le minimum en question est strictement positif, ce qui donne un k strictement inférieur à α_1.

Sens réciproque Si $\operatorname{pgcd}(\alpha_1, \dots, \alpha_p) = 1$ alors pour tout l, il existe un i tel que $\alpha_i^{(l)} = 0$. Donc pour tout l, la puissance de p_l dans la décomposition de k est au moins $\alpha_1^{(l)}$. Cela implique que $k \geqslant \alpha_1$, ce qui est impossible.

□

Définition 18.8.
Les générateurs de U_n *sont les* ***racines primitives***[4] *de l'unité dans* $\mathbb{C}$. *Nous nommons* Δ_n *leur ensemble :*

$$\Delta_n = \{e^{2ki\pi/n} \text{ tel que } 0 \leqslant k \leqslant n-1, \operatorname{pgcd}(k,n) = 1\}. \tag{18.9}$$

Nous avons par exemple

$$\Delta_1 = \{1\} \tag{18.10a}$$
$$\Delta_2 = \{e^{\pi i}\} \tag{18.10b}$$
$$\Delta_4 = \{e^{\pi i/2}, e^{3\pi i/2}\}. \tag{18.10c}$$

Notons que $1 \in \Delta_d$ seulement avec $d = 1$.

18.1.2 Fonction indicatrice d'Euler (première partie)

Nous introduisons ici la fonction indicatrice d'Euler et ses liens basiques avec les racines de l'unité. Pour les propriétés plus avancées, voir 18.1.5.

18.1.3 Introduction par les racines de l'unité

Définition 18.9.
La fonction φ *donnée par*

$$\varphi(n) = \operatorname{Card}(\Delta_n) \tag{18.11}$$

*est l'****indicatrice d'Euler****.*

Si p est un nombre premier, alors $\varphi(p) = p - 1$.

4. parce qu'en prenant les puissances successives de l'une d'entre elles, nous retrouvons toutes les racines de l'unité, voir aussi la définition 18.57.

Lemme 18.10.
Nous avons

$$\mathrm{U}_n = \bigcup_{d\,|\,n} \Delta_d \tag{18.12}$$

et l'union est disjointe. Nous avons aussi la formule

$$n = \sum_{d\,|\,n} \varphi(d). \tag{18.13}$$

Démonstration. À l'application $x \mapsto e^{2i\pi x}$ près, nous pouvons considérer

$$\Delta_d = \{\frac{k}{d} \text{ tel que } k = 0, \dots, d-1, \mathrm{pgcd}(k,d) = 1\}, \tag{18.14}$$

c'est-à-dire l'ensemble des fractions irréductibles dont le dénominateur est d. L'union des Δ_d sera donc disjointe.

Toujours à l'application $x \mapsto e^{2i\pi x}$ près, le groupe U_n est donné par

$$\mathrm{U}_n = \{\frac{k}{n} \text{ tel que } k = 0, \dots, n-1\}. \tag{18.15}$$

L'égalité (18.12) revient maintenant à dire que toute fraction de la forme $\frac{k}{n}$ s'écrit de façon irréductible avec un dénominateur qui divise n.

La relation (18.13) consiste à prendre le cardinal des deux côtés de (18.12). Nous avons $\mathrm{Card}(\mathrm{U}_n) = n$ et l'union étant disjointe, à droite nous avons la somme des cardinaux.

Pour chaque diviseur d de n nous considérons l'ensemble

$$\Phi_n(d) = \{m \in \mathbb{N} \text{ tel que } \mathrm{pgcd}(m,n) = \frac{n}{d}\}. \tag{18.16}$$

Étant donné que tous les entiers entre 1 et n ont un pgcd avec n qui est automatiquement un quotient de n nous avons

$$\{1, \dots, n\} = \bigcup_{d\,|\,n} \Phi_n(d) \tag{18.17}$$

où l'union est disjointe. Par ailleurs nous savons que si $\mathrm{pgcd}(a,b) = 1$, alors $\mathrm{pgcd}(ka, kb) = k$. Donc si $m \in \Delta_d$, alors $m \cdot \frac{n}{d}$ appartient à $\Phi_n(d)$. En d'autres termes, $a \mapsto \frac{n}{d}a$ est une bijection entre Δ_d et $\Phi_n(d)$.

Nous avons donc $\mathrm{Card}(\Phi_n(d)) = \mathrm{Card}(\Delta_d) = \varphi(d)$ et finalement

$$\mathrm{Card}\{1, \dots, n\} = \sum_{d\,|\,n} \mathrm{Card}(\Phi_n(d)) = \sum_{d\,|\,n} \varphi(d). \tag{18.18}$$

□

Lemme 18.11.
Si p est un nombre premier, alors $\varphi(p^n) = p^n - p^{n-1}$.

Démonstration. Les éléments de $\{1, \dots, p^n\}$ qui ont un pgcd différent de 1 avec p^n sont des nombres qui s'écrivent sous la forme qp avec $q \leqslant p^{n-1}$ [5]. Il y a évidemment p^{n-1} tels nombres.

Par conséquent le cardinal de P_{p^n} est $\varphi(p^n) = p^n - p^{n-1}$. □

Problèmes et choses à faire

P_n n'a pas été défini.

Définition proposée (et vue par après) : $P_n = \{m \in \mathbb{N} \text{ tel que } \mathrm{pgcd}(m,n) = 1\}$. À mettre donc en lien avec Δ_d.

5. Corolaire 3.32.

18.1.4 Générateurs

Proposition 18.12.
Soit $n \in \mathbb{N}^*$ et le groupe (additif) $\mathbb{Z}/n\mathbb{Z}$. L'élément $[x]_n$ est un générateur de $\mathbb{Z}/n\mathbb{Z}$ si et seulement si $x \in P_n$. En particulier $\mathbb{Z}/n\mathbb{Z}$ est un groupe contenant $\varphi(n)$ générateurs.

Démonstration. Nous avons $\operatorname{gr}\big([1]_n\big) = \mathbb{Z}/n\mathbb{Z}$. L'élément $[x]_n$ sera générateur si et seulement s'il génère $[1]_n$, c'est-à-dire s'il existe u tel que $u[x]_n = [1]_n$. Cette dernière égalité étant une égalité de classes dans $\mathbb{Z}/n\mathbb{Z}$, elle sera vraie si et seulement s'il existe v tel que

$$ux + vn = 1. \tag{18.19}$$

Cela signifie entre autres que[6] $x\mathbb{Z} + n\mathbb{Z} = \mathbb{Z}$, et aussi que $\operatorname{pgcd}(x, n) = 1$ par le théorème de Bézout 3.13, et donc que $x \in P_n$. □

Corolaire 18.13.
Un groupe monogène d'ordre n possède $\varphi(n)$ générateur où φ est la fonction indicatrice d'Euler définie en 18.9.

Démonstration. Le théorème 5.16 nous dit qu'un groupe monogène d'ordre n est isomorphe à $\mathbb{Z}/n\mathbb{Z}$. La proposition 18.12 nous indique que $\mathbb{Z}/n\mathbb{Z}$ possède $\varphi(n)$ générateurs. □

18.1.5 Fonction indicatrice d'Euler (propriétés)

Corolaire 18.14.
L'indicatrice d'Euler est multiplicative : si p est premier avec q, alors $\varphi(pq) = \varphi(p)\varphi(q)$. De plus si p et q sont premiers entre eux,

$$\varphi(pq) = (p-1)(q-1). \tag{18.20}$$

Démonstration. Nous savons que si p et q sont premiers entre eux, alors le théorème 5.25 nous donne l'isomorphisme de groupe

$$(\mathbb{Z}/pq\mathbb{Z}, +) \simeq (\mathbb{Z}/p\mathbb{Z}, +) \times (\mathbb{Z}/q\mathbb{Z}, +). \tag{18.21}$$

Un élément (x, y) est générateur du produit si et seulement si x est générateur de $\mathbb{Z}/p\mathbb{Z}$ et y est générateur de $\mathbb{Z}/q\mathbb{Z}$. Par la proposition 18.12, il y a $\varphi(p)\varphi(q)$ tels éléments. Par ailleurs le nombre de générateurs de $\mathbb{Z}/pq\mathbb{Z}$ est $\varphi(pq)$, d'où l'égalité.

Si p est premier, nous avons $\varphi(p) = p - 1$ parce que tous les entiers de $\{1, \dots, p-1\}$ sont premiers avec p. □

18.1.6 Décomposition en fractions simples

Il est possible de décomposer une fraction rationnelle en fractions dites « simples ». Si $|z| < 1$ nous avons par exemple la décomposition

$$\frac{1}{1 - z^r} = \sum_{\omega \in U_r} \frac{A_\omega}{\omega - z} \tag{18.22}$$

où U_r est le groupe des racines r^{e} de l'unité définit en (18.1). Les nombres A_ω peuvent alors être déterminés en effectuant la somme. Le dénominateur commun sera $1 - z^r$ tandis que les A_ω sont déterminés en égalant le numérateur à 1.

Exemple 18.15.
Pour décomposer la fraction $\frac{1}{1-x^2}$ nous savons que les racines sont ± 1. Donc nous écrivons

$$\frac{1}{1 - x^2} = \frac{A}{1 - x} + \frac{B}{1 + x}. \tag{18.23}$$

6. Corolaire 3.14

Nous trouvons les valeurs de A et B en effectuant la somme :

$$\frac{A(1+x)+B(1-x)}{1-x^2}=\frac{A+B+(A-B)x}{1-x^2}. \tag{18.24}$$

Les coefficients A et B doivent donc vérifier $A+B=1$ et $A-B=0$. Au final,

$$\frac{1}{1-x^2}=\frac{1}{2(1-x)}+\frac{1}{2(1+x)}. \tag{18.25}$$

△

18.2 Chiffrement RSA

Ce passage sur RSA provient en bonne partie de la page de Wikipédia[369].

Alice veut envoyer un message à Bob. L'idée est que Bob va donner à Alice une clef publique qui va permettre de chiffrer le message tandis que Bob va garder pour lui une clef privée qui permet de déchiffrer.

18.2.1 Mise en place par Bob

Bob se crée une paire de clef publique, clef privée de la façon suivante.

(1) Bob choisit deux nombres premiers distincts p, q.

(2) Il calcule $n = pq$.

(3) Par le corolaire 18.14, l'indicatrice d'Euler $\varphi(n) = (p-1)(q-1)$ est facile à calculer pour Bob.

(4) Bob choisit $e \in \mathbb{N}$ premier avec $\varphi(n)$, puis d tel que $ed \in [1]_{\varphi(n)}$.

Maintenant la paire est : clef publique (n, e) et clef privée (n, d) [7].

Bob envoie la paire (n, e) à Alice.

Remarque 18.16.
Ici nous ne supposons pas que la communication soit sure. Une tierce personne peut intercepter le message. D'ailleurs en principe les gens publient leurs clef publique sur leurs sites, voire sur des sites dédiés. Le problème de l'identification reste à résoudre à l'ancienne.

18.2.2 Chiffrement

Nous chiffrons en utilisant la clef publique (n, e). D'abord Alice se débrouille pour transformer son message en un nombre plus petit que n. Soit M ce message. Alice code M en

$$C = M^e \mod n. \tag{18.26}$$

Tout le truc est que nous allons voir que l'application $x \mapsto x^e$ est une bijection de $\mathbb{F}_n$ et que l'inverse est facile à calculer par Bob et difficile pour les autres. Alice envoie C à Bob. Encore une fois, nous ne supposons pas que cette communication soit privée. Le nombre C peut être intercepté.

18.2.3 Déchiffrement

Nous allons montrer que $M = C^d \mod n$, et donc que Bob, connaissant (n, d), peut déchiffrer. D'abord

$$C^d = (M^e)^d = M^{ed}, \tag{18.27}$$

mais nous savons qu'il existe k tel que

$$ed = 1 + k\varphi(n) = 1 + k(p-1)(q-1). \tag{18.28}$$

7. Le fait que e soit public et d soit privé est une convention. e comme *encryption* et d comme *decryption*.

L'étape astucieuse est de remarquer que

$$M^{1+k(p-1)(q-1)} \in [M]_p \cap [M]_q. \tag{18.29}$$

Pour montrer cela nous utilisons le petit théorème de Fermat 6.13 et la remarque 6.14.

— Si M est premier avec p, alors $M^{p-1} \in [1]_p$.

— Si M n'est pas premier avec p, alors M est multiple de p et on sait que $M^{p-1} \in [0]_p = [M]_p$.

Dans les deux cas nous avons (18.29). Le nombre $M^{1+k\varphi(n)} - M$ est donc à la fois multiple de p et de q.

Le lemme chinois 6.24 nous dit immédiatement [8] qu'alors

$$M^{1+k\varphi(n)} - M \tag{18.30}$$

est un multiple de $pq = n$, c'est-à-dire que

$$C^d = M^{ed} \in [M]_n. \tag{18.31}$$

Si on ne croit pas au lemme chinois, on peut utiliser le lemme de Gauss. Posons

$$M^{1+k\varphi(n)} - M = ap = bq. \tag{18.32}$$

Dans ce cas p divise bq, mais q est premier avec p, donc le lemme de Gauss 3.134 nous enseigne [9] que p divise b.

18.2.4 Une imprudence à ne pas commettre

Nous avons pris deux cas selon que M soit ou non premier avec p. Une question qui se pose est la suivante : est-ce que c'est une bonne idée d'envoyer un message qui ne soit pas premier avec p ?

Si nous savons que M n'est pas premier avec p, alors nous avons $M^e = l^e p^e$ et $n = pq$ qui sont publics. Donc un calcul de PGCD permettrait de trouver p.

Il faut cependant savoir que

— La probabilité que ça arrive est infime : vu que M est entre 0 et $n = pq$, les multiples de p possibles sont $p, 2p, \cdots pq$. Il y a dont une chance sur p que cela arrive. Typiquement avec des p de l'ordre de 10^{120}, on peut utiliser RSA chaque milliseconde sur chaque atome de l'univers depuis le début des temps que ça ne se serait presque certainement pas encore produit.

— De toutes façons Alice ne sait pas vérifier si son message est premier avec p parce qu'elle ne connaît pas p.

— En conclusion la partie de la preuve qui montre que $M^{1+\varphi(n)} \in [M]_p \cap [M]_q$ dans le cas M non premier avec p est, à toutes fins pratiques, inutile parce que ce cas de figure ne se présentera jamais dans toutes l'histoire de l'univers, même pas avec une civilisation intelligente autour de chaque étoile.

Problèmes et choses à faire

Est-ce que ces trois points sont corrects ?

18.2.5 Problèmes calculatoires

Pour implémenter RSA, il faut pouvoir faire (au moins) trois choses :

(1) Trouver de grands nombres premiers.

(2) Trouver des couples de Bézout.

(3) Calculer M^e lorsque e est très grand.

8. C'est ici qu'il est important que p ne soit pas égal à q. Si $p = q$, alors le lemme chinois ne fonctionne pas.
9. Ici aussi, si $p = q$, ça ne marche pas.

En ce qui concerne le problème de trouver des nombres premiers, c'est compliqué, mais il faut savoir qu'il y en a plein. À 120 chiffres, il y a environ autant de nombres premiers que d'atomes dans 10^{20} fois l'univers connu. Cela rend impossible toute tentative de factoriser un grand nombre en essayant toutes les possibilités. Même pas en science-fiction [10].

Trouver des nombres u et v tels que $Au + Bv = \text{pgcd}(A, B)$ est un problème expliqué en 3.3.4.

En ce qui concerne le calcul de M^e lorsque e est grand, il n'est évidemment pas pensable de faire $M \cdot M \cdot \ldots M$ avec e facteurs. Un truc pour calculer en moins d'étapes est l'**exponentiation rapide**. Si $e = 2k$ est pair, nous calculons

$$M^e = (M^k)^2; \tag{18.33}$$

si $e = 2k + 1$ alors nous calculons

$$M^e = M(M^k)^2. \tag{18.34}$$

Le calcul prend alors seulement environ $\log_2(e)$ étapes. Pour donner une idée,

$$\log_2(10^{120}) \simeq 400. \tag{18.35}$$

Très raisonnable, mais un ordinateur reste indispensable.

18.2.6 La solidité de RSA

La solidité de la méthode repose sur deux conjectures (non démontrées ! !) :

— Pour déchiffrer il *faut* connaitre p et q.

— La difficulté de trouver p et q en partant de $n = pq$ est exponentielle en n.

Dans la méthode de déchiffrage proposée ici, p et q sont utilisés pour calculer d qui est solution de $ed = [1]_{\varphi(n)}$. La seule formule connue pour calculer $\varphi(n)$ est $\varphi(n) = (p - 1)(q - 1)$. Si on trouve plus simple, alors RSA peut être craqué.

18.2.7 Note non mathématique pour doucher l'enthousiasme

Il est souvent dit[370] que différents systèmes de chiffrement peuvent aider à avoir des discussions « discrètes » dans les régimes totalitaires. La technologie au service de la démocratie, voilà qui enthousiasme la jeunesse [11]. La réalité est qu'il est souvent possible de craquer un système de chiffrement arbitrairement complexe, même sans connaitre le petit théorème de Fermat ...
... tout dépends du contexte.

18.3 Polynômes cyclotomiques

18.3.1 Définitions et propriétés

Définition 18.17.
*Le **polynôme cyclotomique** d'indice n est le polynôme*

$$\phi_n(X) = \prod_{z \in \Delta_n} (X - z) \tag{18.36}$$

où Δ_n est l'ensemble des racines primitives de l'unité de la définition 18.8 :

$$\Delta_n = \{e^{2ik\pi/n} \text{ tel que } 0 \leqslant k \leqslant n - 1 \text{ tel que } \text{pgcd}(k, n) = 1\}, \tag{18.37}$$

10. Cela donne une idée des connaissances en math des klingons, dont le docteur Spock parvient à craquer le code mentalement en deux heures.

11. Cela dit, le navigateur Tor[371], qui est un pur produit de RSA, permet effectivement d'accéder en France aux sites bloqués pour apologie du terrorisme (mars 2015).

Le polynôme ϕ_n est un polynôme unitaire de degré $\varphi(n)$ où φ est l'indicatrice d'Euler[12] $\varphi(n)$. Nous avons par exemple

$$\Delta_1 = \{1\} \tag{18.38a}$$
$$\Delta_2 = \{-1\} \tag{18.38b}$$
$$\Delta_3 = \{e^{2\pi i/3}, e^{4\pi i/3}\} \tag{18.38c}$$

et les premiers polynômes cyclotomiques sont donnés par

$$\phi_1(X) = X - 1 \tag{18.39a}$$
$$\phi_2(X) = X + 1 \tag{18.39b}$$
$$\phi_3(X) = X^2 + X + 1. \tag{18.39c}$$

Pour le dernier nous avons utilisé le fait que $e^{6\pi i/3} = 1$ et $e^{4\pi i/3+e^{2\pi i/3}} = -1$.

Lemme 18.18.
Le polynôme $X^n - 1$ se factorise des diverses manières suivantes :

$$X^n - 1 = \prod_{z\in \mathrm{U}_n}(X - z) = \prod_{d\,|\,n}\prod_{z\in\Delta_d}(X - z) = \prod_{d\,|\,n}\phi_d(X) \tag{18.40}$$

Démonstration. En ce qui concerne la première égalité, tous les éléments de U_n sont des racines simples de $X^n - 1$. Donc le théorème 3.194 dit qu'il existe un nombre k (polynôme de degré zéro) tel que $X^n - 1 = k\prod_{z\in\mathrm{U}_n}(X - z)$. Vu le coefficient du terme de plus haut degré, ce k ne peut être que 1.

Pour la suite nous utilisons l'union disjointe $\mathrm{U}_n = \bigcup_{d\,|\,n}\Delta_d$ du lemme 18.10 et la définition (18.36) des polynômes cyclotomiques. □

Remarque 18.19.
Notons juste pour le plaisir que dans le produit $\prod_{d\,|\,n}\prod_{z\in\Delta_d}$, il y a bien n termes parce que $\mathrm{Card}(\Delta_d) = \varphi(d)$ et $\sum_{d\,|\,n}\varphi(d) = n$ (définition 18.8 et lemme 18.10).

Proposition 18.20.
Les polynômes cyclotomiques sont à coefficients entiers : $\phi_n \in \mathbb{Z}[X]$.

Démonstration. Nous devons démontrer que les coefficients de ϕ_n sont dans $\mathbb{Z}$ alors qu'ils sont à priori dans $\mathbb{C}$. Nous démontrons cela par récurrence. D'abord $\phi_1(X) = X - 1$, d'accord. Ensuite

$$X^{n+1} - 1 = \prod_{d\,|\,n+1}\phi_d(X) = \phi_{n+1}(X)\cdot\underbrace{\prod_{\substack{d\,|\,n+1\\ d\leqslant n}}\phi_d(X)}_{\in\mathbb{Z}[X]\text{ par récurrence}} \tag{18.41}$$

Le lemme 6.43 conclut que $\phi_{n+1} \in \mathbb{Z}[X]$. Nous avons vu $\mathbb{Z}$ comme sous anneau du corps $\mathbb{C}$. □

Proposition 18.21.
Soient $1 \leqslant m \leqslant n$ deux entiers et

$$T(X) = \frac{X^n - 1}{X^m - 1} \in \mathbb{Z}(X). \tag{18.42}$$

Alors :

(1) si $m \mid n$ alors $T \in \mathbb{Z}[X]$,

(2) si $m \mid n$ et si $m < n$ alors ϕ_n divise T dans $\mathbb{Z}[X]$.

12. Définie par l'équation 18.11.

Démonstration. Nous prouvons point par point.

(1) Si m divise n alors les diviseurs de n sont l'union des diviseurs de m et des diviseurs de n qui ne divisent pas m. Soit

$$Q = \{\text{diviseurs de } n \text{ ne divisant pas } m\}. \tag{18.43}$$

Nous avons alors

$$X^n - 1 = \prod_{d|n} \phi_d(X) = \prod_{d|m} \phi_d(X) \cdot \prod_{q\in Q} \phi_q(X) = (X^m - 1) \cdot \prod_{q\in Q} \phi_q(X). \tag{18.44}$$

Nous avons donc

$$T(X) = \frac{X^n - 1}{X^m - 1} = \prod_{q\in Q} \phi_q(X) \in \mathbb{Z}[X]. \tag{18.45}$$

(2) Nous venons de montrer que

$$T = \prod_{q\in Q} \phi_q \in \mathbb{Z}[X]. \tag{18.46}$$

Étant donné que $m < n$ nous avons $n \in Q$ et donc

$$T = \phi_n \cdot \prod_{q\in Q\setminus\{n\}} \phi_q. \tag{18.47}$$

Par conséquent ϕ_n divise T dans $\mathbb{Z}[X]$.

□

Corolaire 18.22.
Si p est premier alors le polynôme cyclotomique ϕ_p a une bonne tête :

$$\phi_p(X) = 1 + X + \cdots + X^{p-1}. \tag{18.48}$$

Démonstration. Nous utilisons la formule du lemme 18.18 en remarquant que seuls p et 1 divisent p :

$$X^p - 1 = \prod_{d|p} \phi_d(X) = \phi_1(X)\phi_p(X) = (X - 1)\phi_p(X). \tag{18.49}$$

Nous pouvons simplifier par $X - 1$ en utilisant la formule du lemme 3.198(2) :

$$1 + X + \cdots + X^{p-1} = \phi_p(X) \tag{18.50}$$

□

Proposition 18.23 (Irréductibilité des polynômes cyclotomiques[372])**.**
Les polynômes cyclotomiques sont irréductibles sur $\mathbb{Q}$.

Démonstration. Pour rappel, nous savons déjà que pour tout $n \in \mathbb{N}$, $\phi_n \in \mathbb{Z}[X]$. Vu que les racines de ϕ_n sont les racines primitives de l'unité, nous devons montrer que toutes les racines primitives de l'unité ont même polynôme minimal (qui sera alors ϕ_n) ; en effet vu que ces polynômes divisent ϕ_n, s'ils sont distincts, la proposition 6.98 s'applique et le produit des polynômes minimaux diviserait ϕ_n. Dans le cas inverse, ϕ_n est polynôme minimal des racines primitives de l'unité et est donc irréductible. Soit donc ξ, une telle racine primitive. Une autre racine primitive est de la forme ξ^l où l est un nombre premier tel que $\text{pgcd}(l, n) = 1$.

Soient f et g, les polynômes minimaux dans $\mathbb{Z}[X]$ de ξ et ξ^l. Nous allons montrer que $f = g$ et donc que $f = g = \phi_n$. Supposons par l'absurde que $f \neq g$. Dans ce cas ils seraient des facteurs irréductibles distincts de ϕ_n et il existerait un polynôme h tel que $\phi_n = fgh$. À priori, $h \in \mathbb{Q}[X]$ parce que nous sommes justement en train de prouver que ϕ_n est irréductible dans $\mathbb{Q}[X]$. Quoi

qu'il en soit, le lemme de Gauss 6.45 nous montre que $h \in \mathbb{Z}[X]$ parce que ϕ_n, f et g ont des coefficients entiers. Nous avons

$$f(\xi) = g(\xi^l) = 0. \tag{18.51}$$

Considérons le polynôme $\psi(X) = g(X^l)$. Ce polynôme ψ est dans $\mathbb{Z}[X]$ et ψ est annulateur de ξ, donc f divise ψ en tant que polynôme minimal de ξ. Il y a un polynôme unitaire à coefficients entiers (lemme de Gauss forever) k tel que

$$\psi = fk \tag{18.52}$$

Nous considérons maintenant les projections sur $\mathbb{F}_l[X]$: étant donné que $\phi_n = fgh$, nous savons que $\bar{f}\bar{g}$ divise $\bar{\phi}_n$. En même temps, $\bar{f}$ divise $\bar{\psi}$. En utilisant le morphisme de Frobenius (c'est ici que la projection sur $\mathbb{F}_l$ joue), nous avons aussi

$$\bar{\psi}(X) = \bar{g}(X^l) = \bar{g}(X)^l. \tag{18.53}$$

Par conséquent dire que $\bar{f}$ divise $\bar{\psi}$ revient à dire que $\bar{f}(X)$ divise $\bar{g}(X)^l$. En particulier tous facteur irréductible de $\bar{f}$ divise $\bar{g}$. Un facteur irréductible de $\bar{f}$ serait donc à la fois dans $\bar{f}$ et dans $\bar{g}$ et donc deux fois (au moins) dans $\bar{\phi}_n$ parce que $\bar{f}\bar{g}$ divise ϕ_n. Dans un corps de décomposition de ce facteur, ϕ_n aurait une racine double, alors que ce n'est pas le cas. Contradiction. Nous concluons que $f = g$. □

Le corolaire suivant va être utilisé pour déterminer les polygones constructibles à la règle et au compas, théorème de Gauss-Wantzel 18.87.

Corolaire 18.24.
Soit p un nombre premier et α un entier non nul. Nous posons $q = p^\alpha$. Alors le polynôme minimal de $e^{2i\pi/q}$ sur $\mathbb{Q}$ est le polynôme cyclotomique ϕ_q.

Démonstration. Le polynôme ϕ_q est irréductible par la proposition 18.23, il est unitaire par définition et contient le monôme $X - e^{2i\pi/q}$, donc il est annulateur. Annulateur, irréductible et unitaire, la proposition 6.72(2) en fait le polynôme minimal de ω. □

Théorème 18.25.
Soit $P \in \mathbb{Z}[X]$ un polynôme unitaire irréductible non constant tel que toutes les racines dans $\mathbb{C}$ soient de module $\leqslant 1$. Alors $P = X$ ou P est un polynôme cyclotomique.

Démonstration. Nous supposons que $X \neq 0$, et nous notons $P = \sum_i a_i X^i$. Étant donné que P est irréductible et différent de X, nous avons $a_0 \neq 0$ (sinon $x = 0$ serait une racine). Nous allons montrer que les racines de P sont toutes des racines N-ièmes de l'unité (avec le même N pour toutes).

Soient $\{\xi_i\}_{i=1,\dots,d}$ les racines de P ; on a

$$P = \prod_{i=1}^{d}(X - \xi_i) \tag{18.54}$$

avec $\prod_{i=1}^{d} \xi_i = a_0$. Par hypothèse, $|\xi_i| \leqslant 1$ et donc $0 < |a_0| \leqslant 1$. Vu que $P \in \mathbb{Z}[X]$ nous avons donc $a_0 = 1$ et donc $|\xi_i| = 1$ pour tout i.

Nous introduisons les polynômes

$$g_q(X) = \prod_{i=1}^{d} \big(X - (\xi_i)^q\big), \tag{18.55}$$

et en particulier $g_1 = P$, et nous développons

$$g_q(X) = X^n + C_{1,q}X^{n-1} + \cdots + C_{n,q} \tag{18.56}$$

où

$$C_{k,q} = (-1)^k \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant d} (\xi_{i_1} \ldots \xi_{i_k})^q. \tag{18.57}$$

Nous introduisons aussi les polynômes

$$F_{k,q}(X_1, \ldots, X_n) = (-1)^k \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant d} (X_{i_1} \ldots X_{i_k})^q \tag{18.58}$$

qui sont des polynômes symétriques. Ils vérifient deux propriétés. La première est que

$$C_{r,q} = F_{r,q}(\xi_1, \ldots, \xi_n), \tag{18.59}$$

et la seconde est que les polynômes $F_{r,1}$ sont les polynômes symétriques élémentaires à un coefficient près. Le théorème 6.165 nous donne alors des polynômes $G_{k,q} \in \mathbb{Z}[X_1, \ldots, X_n]$ tels que

$$F_{k,q}(X_1, \ldots, X_n) = G_{k,q}\big(F_{1,1}(X_1, \ldots, X_n), \ldots, F_{k,1}(X_1, \ldots, X_n)\big). \tag{18.60}$$

Nous savons que

$$|C_{k,q}| \leqslant \sum_{1 \leqslant i_1 < \ldots < i_k < d} 1 = \binom{d}{k}. \tag{18.61}$$

Donc g_q fait partie de l'ensemble fini des polynômes dans $\mathbb{Z}[q]$ dont tous les coefficients sont bornés en valeur absolue par

$$\max_{k=1,\ldots,d} \binom{d}{k}. \tag{18.62}$$

Il existe un certain nombre d'ensembles $\{\xi_i\}$ qui sont racines de polynômes vérifiant les conditions du théorème. À chacun de ces ensembles est associé une suite de polynômes g_q et donc des coefficients $C_{k,q}$. Ce que nous avons vu est que l'ensemble de tous les coefficients $C_{k,q}$ possibles (pour un choix donné des $\{\xi_i\}$) est fini, en particulier, vu que $C_{1,q} = \sum_i \xi_i^q$, pour chaque k, l'ensemble

$$\{\xi_k^q \text{ tel que } q \in \mathbb{N}\}. \tag{18.63}$$

Par le principe des tiroirs, il existe q_1 et q_2 tels que $\xi_k^{q_1} = \xi_k^{q_2}$. Ici, q_1 et q_2 dépendent de k et nous notons $N_k = q_1 - q_2$; nous avons donc $\xi_k^{N_k} = 1$.

En posant $N = \operatorname{ppcm}(N_1, \ldots, N_d)$, nous avons

$$\xi_k^N = 1 \tag{18.64}$$

pour tout k.

Mais P est irréductible dans $\mathbb{Z}[X]$; s'il a ± 1 comme racines, alors c'est que $P = X + 1$ ou $P = X - 1$ et ce sont des polynômes cyclotomiques. Si P n'a pas ± 1 parmi ses racines, alors P n'a pas de racines dans $\mathbb{Q}$ parce que ± 1 sont les seules racines de $X^N - 1$ dans $\mathbb{Q}$.

Par conséquent P est un facteur irréductible de $X^N - 1$ dans $\mathbb{Q}[X]$. Mais étant donné que

$$X^N - 1 = \prod_{d \mid N} \phi_d(X), \tag{18.65}$$

les polynômes cyclotomiques sont les seuls facteurs irréductibles de $X^N - 1$. Donc P est un polynôme cyclotomique. □

18.3.2 Nombres premiers

Lemme 18.26 ([373]).
Soit $n \geqslant 1$. Il existe un nombre premier p et un entier a tels que

(1) p divise $\phi_n(a)$,

(2) p ne divise aucun de $\phi_d(a)$ avec $d \mid n$ et $d \neq n$.

De tels p et a vérifient automatiquement

(1) p divise $a^n - 1$,

(2) p ne divise aucun des $a^d - 1$ pour $d \mid n$, $d \neq n$.

Démonstration. Nous posons

$$B(X) = \prod_{\substack{d \mid n \\ d \neq n}} \phi_d(X), \tag{18.66}$$

et nous commençons par montrer que ϕ_n est premier avec B. Nous avons $X^n - 1 = B\phi_n$, donc B et ϕ_n n'ont pas de racines communes (même pas dans $\mathbb{C}$) parce que ce serait une racine double de $X^n - 1$. Notons que par définition 18.36, les polynômes cyclotomiques sont scindés (dans $\mathbb{C}$), donc en particulier les polynômes ϕ_n et B sont scindés et donc premiers entre eux, dans $\mathbb{C}$ et a fortiori dans $\mathbb{Q}$. Par Bézout (corolaire 3.133), il existe $U, V \in \mathbb{Q}[X]$ tels que

$$U\phi_n + VB = 1. \tag{18.67}$$

Si nous prenons $a \in \mathbb{Z}$ tel que $U' = aU$ et $V' = aV$ soient tous deux dans $\mathbb{Z}[X]$, alors nous avons

$$U'\phi_n + V'B = a, \tag{18.68}$$

égalité dans $\mathbb{Z}[X]$. Quitte à prendre un multiple assez grand de a, nous pouvons choisir a de telle sorte que $|\phi_n(a)| \geqslant 2$. Nous prenons alors un nombre premier p divisant $\phi_n(a)$.

Montrons que le a et le p ainsi construits satisfont aux exigences.

Vu que $X^n - 1 = B\phi_n$, si p divise $\phi_n(a)$, il divise automatiquement $a^n - 1$ et donc $[a^n]_p = 1$, ce qui signifie entre autres que a et p sont premiers entre eux. Évaluons l'équation (18.68) en a :

$$U'(a)\phi_n(a) + V'(a)B(a) = a. \tag{18.69}$$

Le nombre p ne divisant pas a, mais divisant $\phi_n(a)$, il ne peux pas diviser $B(a)$ [13]. Étant donné que p ne divise pas $B(a)$, il ne divise aucun des $\phi_d(a)$ avec $d \mid n$ et $d \neq n$.

Nous passons maintenant à la seconde partie de la preuve. Nous supposons avoir a et p tels que p soit un nombre premier divisant $\phi_n(a)$ et tels que p ne divise aucun des $\phi_d(a)$ avec $d \mid n$, $d \neq n$. Le fait de diviser $\phi_n(a)$ entraine le fait de diviser $a^n - 1$ parce que ϕ_n est un des facteurs de $X^n - 1$. Soit maintenant $d \neq n$ divisant n ; nous avons

$$X^d - 1 = \prod_{d' \mid d} \phi_{d'}, \tag{18.70}$$

et cela est une partie du produit

$$\prod_{\substack{d \mid n \\ d \neq n}} \phi_d. \tag{18.71}$$

Vu que p ne divise aucun des $\phi_d(a)$ de ce dernier produit, a fortiori, il ne divise pas le produit 18.70, et donc pas $a^d - 1$. □

Lemme 18.27.
Si $n \geqslant 1$, alors il existe un nombre premier dans $[1]_n$, c'est-à-dire un nombre premier de la forme $1 + kn$ avec $k \in \mathbb{N}^*$.

Démonstration. Soient $n \geqslant 1$ et p, a les nombres donnés par le lemme 18.26. Vu que p divise $\phi_n(a)$, p divise $a^n - 1$ et donc $[a]_p$ a un ordre qui divise n dans $(\mathbb{Z}/p\mathbb{Z})^*$ parce que $[a]_p^n = [1]_p$.

Prenons $d \neq n$ divisant n. Nous savons que

$$a^d - 1 = \prod_{d' \mid d} \phi_{d'}(a). \tag{18.72}$$

13. C'est pour pouvoir dire ça que l'on a choisi $V' \in \mathbb{Z}[X]$ de telle sorte que $V'(a)$ soit dans $\mathbb{Z}$

Par construction de a et p, nous avons

$$[\phi_{d'}(a)]_p \neq 0 \tag{18.73}$$

Vu que $\mathbb{Z}/p\mathbb{Z}$ est intègre, le produit est également non nul, c'est-à-dire

$$\Big[\prod_{d'\mid d}\phi_{d'}(a)\Big]_p \neq 0, \tag{18.74}$$

et donc $[a]_p^a \neq 1$. Nous avons donc montré que si $d \neq n$ divise n, alors nous avons en même temps

$$[a]_p^n = 1 \tag{18.75}$$

et

$$[a]_p^d \neq 1. \tag{18.76}$$

Cela prouve que $[a]_p$ est d'ordre exactement n. Oui, mais l'ordre de $[a]_p$ doit diviser l'ordre du groupe $\mathbb{Z}/p\mathbb{Z}$ qui est $p-1$, donc n divise $p-1$ et nous écrivons $p = kn+1$ avec k entier. □

Théorème 18.28 (Forme faible du théorème de Dirichlet [71]).
Pour tout $n \geqslant 1$, il existe une infinité de nombres premiers dans $[1]_n$.

Démonstration. Le lemme 18.27 nous donne déjà l'existence de nombres premiers dans $[1]_n$. Il faut maintenant voir qu'il y en a une infinité. Nous supposons qu'il y en ait seulement un nombre fini : $p_1, \dots, p_r$, et nous notons

$$N = np_1 \dots p_r. \tag{18.77}$$

Nous utilisons maintenant le lemme 18.27 avec ce N, c'est-à-dire qu'on a un nombre premier de la forme

$$p = 1 + kN = 1 + knp_1 \dots p_r. \tag{18.78}$$

Cela est un nombre premier plus grand que tous les p_i et de la forme $1 + \lambda n$. Cela contredit l'exhaustivité de la liste $p_1, \dots, p_r$. □

18.4 Dénombrement des solutions d'une équation diophantienne

Théorème 18.29 ([71, 268]).
Soient des entiers naturels premiers dans leur ensemble[14] *$\alpha_1, \dots, \alpha_p$ et l'équation*

$$\alpha_1 n_1 + \dots + \alpha_p n_p = n \tag{18.79}$$

pour les naturels n_i où n est un naturel donné. Nous notons S_n le nombre de solutions de cette équation. Alors :

(1) Il existe un algorithme (en temps fini) pour calculer S_n en fonction des α_i et de n.

(2) Nous avons le comportement asymptotique

$$S_n \sim \frac{1}{\alpha_1 \dots \alpha_p}\frac{n^{p-1}}{(p-1)!}. \tag{18.80}$$

Démonstration. Pour $|z| < 1$ dans $\mathbb{C}$, utilisant le lemme 14.28, nous écrivons le développement

$$F(z) = \prod_{i=1}^{p}\frac{1}{1-z^{\alpha_i}} = \prod_{i=1}^{p}\sum_{n\geqslant 0} z^{n\alpha_i}. \tag{18.81}$$

14. Définition 3.11.

Nous allons maintenant à la pêche au terme de degré k dans ce produit de sommes en utilisant p fois le produit de Cauchy de la formule (14.38). Nous avons

$$F(z) = \sum_{k\geqslant 0}\left(\sum_{n_1\alpha_1+\cdots+n_p\alpha_p=n} 1\right) z^k = \sum_{k\geqslant 0} S_k z^k. \tag{18.82}$$

La technique pour déterminer la valeur de S_n est alors de développer $F(z)$ en série de façon un peu explicite et d'identifier le coefficient de z^n parce que nous venons de voir que ce coefficient est S_n. Nous commençons par une décomposition en éléments simples, expliquée autour de l'équation (18.22) :

$$\frac{1}{1-z^{\alpha_i}} = \sum_{\alpha\in U_{\alpha_i}} \frac{A_{\omega,i}}{\omega - z}. \tag{18.83}$$

où U_{α_i} est le groupe des racines α_i^e^de l'unité décrit en 18.1. La raison de ce développement est que, comme mentionné dans le lemme 18.18, $\prod_{\omega\in \mathrm{U}_{\alpha_i}}(z-\omega) = z^{\alpha_1}-1$. Lorsque nous effectuons la somme, le dénominateur commun est donc bien [15] $1-z^{\alpha_i}$. En récrivant le produit :

$$F(z) = \prod_{i=1}^{p} \frac{1}{1-z^{\alpha_i}} = \prod_{i=1}^{p} \sum_{\omega\in U_{\alpha_i}} \frac{A_{\omega,i}}{\omega - z} \tag{18.84}$$

Les coefficients $A_{\omega,i}$ sont calculables explicitement, en temps fini.

Vu que 1 est dans tous les U_{α_i}, le produit fait intervenir au dénominateur des puissances de $(1-z)$ jusqu'à la puissance p. Les autres racines de l'unité appartiennent au maximum à $p-1$ des groupes U_{α_i} parce que les nombres α_i sont premiers dans leur ensemble, voir la proposition 18.7.

La fonction F peut alors s'écrire sous la forme

$$F(z) = \frac{A}{(1-z)^p} + G(z) \tag{18.85}$$

où $G(z)$ est une somme de termes de la forme

$$\frac{a_{i,1}}{1-\omega_i} + \cdots + \frac{a_{i,p}}{(1-\omega_i)^{p-1}} \tag{18.86}$$

où les ω_i sont les racines α_i^e^de l'unité et $a_{k,r}$ sont des nombres complexes. Trouvons A. D'abord grâce au lemme 3.198(2) nous avons

$$F(z)(1-z)^p = \prod_{l=1}^{p} \frac{1-z}{1-z^{\alpha_i}} = \prod_{i=1}^{p} \frac{1}{1+z+\cdots+z^{\alpha_i-1}}, \tag{18.87}$$

et donc

$$\lim_{z\to 1} F(z)(1-z)^p = \prod_{i=1}^{p} \frac{1}{\alpha_i}. \tag{18.88}$$

Mais vu ce que contient $G(z)$, nous avons aussi $\lim_{z\to 1} F(z) = A$. Nous avons donc déjà déterminé $A = \frac{1}{\alpha_1\ldots\alpha_p}$.

Pour la suite nous avons besoin des développements du lemme 14.28. Nous utiliserons en particulier celle-ci :

$$\frac{1}{(\omega-z)^k} = \frac{1}{(k-1)!}\sum_{s=0}^{\infty} \omega^{-s-1-k}\frac{(s+k-1)!}{s!} z^s. \tag{18.89}$$

En particulier le module du coefficient de z^n là dedans est : $\frac{(n+k-1)!}{n!(k-1)!}$. Dans la partie G de la décomposition (18.85), k est majoré par $p-2$ et la dépendance en n est donc au maximum du type

$$\frac{(n+p-2)!}{n!(p-2)!} \sim \frac{n^{n+p-2}}{n^n(p-2)!} = \frac{n^{p-2}}{(p-2)!}. \tag{18.90}$$

15. Pour le signe, c'est ajustable avec le signe de $A_{\omega,i}$.

Dans le premier terme par contre, il y a des termes jusqu'à $k = p$. Le terme dominant est alors en $\frac{n^{p-1}}{(p-1)!}$ et son coefficient est A qui est déjà calculé. Au final le terme dominant du coefficient de z^n dans $F(z)$ est

$$S_n \sim \frac{A}{(p-1)!} n^{p-1} = \frac{1}{\alpha_1 \dots \alpha_p} \frac{n^{p-1}}{(p-1)!}. \tag{18.91}$$

□

Exemple 18.30.
Pour $p = 1$, l'équation est $\alpha x = n$, qui possède au maximum une solution, quel que soit n. Et de plus pour avoir une solution il faut et suffit que α divise n, c'est-à-dire que n soit un multiple de α. Il n'y a que un nombre sur α à être multiple de α. D'où le comportement en $\frac{1}{\alpha}$.

Pour $p = 2$, c'est l'équation (3.62) déjà étudiée. Il y a une famille à un paramètre de solutions dont seulement un certain nombre sont positives. À priori, le nombre de solutions positives croît linéairement en n. △

18.5 Corps finis

Si vous cherchez des choses à propos de RSA, c'est à la section 18.2.

18.5.1 Théorème de Wedderburn

Théorème 18.31 (Théorème de Wedderburn[100])**.**
Tout corps fini est commutatif.

Démonstration. Soit $\mathbb{K}$ un corps fini et Z, le centre de $\mathbb{K}$. Ce dernier est un corps fini et un sous corps de $\mathbb{K}$. Si $q = \text{Card}(Z)$ alors par le lemme 6.51 nous avons

$$\text{Card}(\mathbb{K}) = q^n \tag{18.92}$$

pour un certain n.

Nous supposons maintenant que $\mathbb{K}$ est non commutatif. Dans ce cas $Z \neq \mathbb{K}$ et nous avons $n \geqslant 2$. Nous considérons aussi

$$Z_x = \{a \in \mathbb{K} \text{ tel que } ax = xa\}. \tag{18.93}$$

Le centre Z est un sous corps de Z_x, donc il existe $d(x)$ tel que

$$\text{Card}(Z_x) = q^{d(x)}. \tag{18.94}$$

De la même manière, Z_x est un sous corps de $\mathbb{K}$, donc il existe $m(x)$ tel que

$$\text{Card}(\mathbb{K}) = \text{Card}(Z_x)^{m(x)}. \tag{18.95}$$

En mettant bout à bout nous avons

$$q^n = \text{Card}(Z_x)^{m(x)} = q^{d(x)m(x)}, \tag{18.96}$$

et par conséquent $n = d(x)m(x)$. Le point important à retenir est que $d(x)$ divise n pour tout $x \in \mathbb{K}$.

Nous considérons maintenant l'action adjointe du groupe $\mathbb{K}^*$ sur lui-même :

$$\varphi(k)x = kxk^{-1}. \tag{18.97}$$

Nous notons $\mathcal{O}_x$ l'orbite de $x \in \mathbb{K}^*$ pour cette action, et $\text{Fix}(x)$ son stabilisateur. Nous avons

$$Z_y = \text{Fix}(y) \cup \{0\} \tag{18.98}$$

parce que Z_y et $\mathrm{Fix}(y)$ ont les mêmes définitions, sauf que $\mathrm{Fix}(y)$ est dans $\mathbb{K}^*$ alors que Z_y est dans $\mathbb{K}$. Nous avons donc

$$\mathrm{Card}\big(\mathrm{Fix}(y)\big) = \mathrm{Card}(Z_y) - 1 = q^{d(y)} - 1. \tag{18.99}$$

Nous avons $\mathrm{Card}(\mathcal{O}_x) = 1$ si et seulement si $\mathcal{O}_x = \{x\}$ si et seulement si $\mathrm{Fix}(x) = \mathbb{K}^*$ si et seulement si $z \in Z^*$. Soient $z_0, \dots, z_{q-1}$ les éléments de Z avec $z_0 = 0$. Ce sont les éléments qui auront une orbite réduite à un point. Les orbites qui coupent Z^* sont

$$\{z_1\}, \dots, \{z_{q-1}\} \tag{18.100}$$

et il y en a $q-1$. Soient $\mathcal{O}_{y_1}, \dots, \mathcal{O}_{y_r}$, les autres orbites. Nous utilisons l'équation des classes (2.88) :

$$\mathrm{Card}(\mathbb{K}^*) = \mathrm{Card}(Z^*) + \sum_{i=1}^{r} \frac{\mathrm{Card}(\mathbb{K}^*)}{\mathrm{Card}(\mathrm{Fix}(y_i))}, \tag{18.101}$$

mais $\mathrm{Card}(Z^*) = q - 1$, $\mathrm{Card}(\mathbb{K}^*) = q^n - 1$ et $\mathrm{Card}\big(\mathrm{Fix}(y_i)\big) = q^{d(y_i)} - 1$, donc

$$q^n - 1 = (q-1) + \sum_{i=1}^{r} \frac{q^n - 1}{q^{d(y_i)} - 1}. \tag{18.102}$$

Nous considérons la fraction rationnelle

$$F(X) = (X^n - 1) - \sum_{i=1}^{r} \frac{X^n - 1}{X^{d(y_i)} - 1}. \tag{18.103}$$

Étant donné que $d(y_i)$ divise n, nous avons, contrairement aux apparences, que $F \in \mathbb{Z}[X]$ par la proposition 18.21(1).

Nous pouvons exploiter un peu mieux la proposition 18.21 en remarquant que $d(y_i) < n$ parce que sinon $\mathrm{Card}(Z_{y_i}) = \mathrm{Card}(\mathbb{K})$, ce qui signifierait que $y_i \in Z$, ce qui nous avions exclu. Par conséquent le polynôme cyclotomique ϕ_n divise

$$\frac{X^n - 1}{X^{d(y_i)} - 1} \tag{18.104}$$

dans $\mathbb{Z}[X]$. Le polynôme cyclotomique ϕ_n divise également $X^n - 1$ et par conséquent ϕ_n divise F. Il existe donc $Q \in \mathbb{Z}[X]$ tel que $F = Q\phi_n$. En particulier en évaluant en q :

$$F(q) = Q(q)\phi_n(q) = q - 1. \tag{18.105}$$

En effet nous avons $F(q) = q - 1$ par construction : comparer (18.102) avec (18.103). Évidemment $q \neq 1$ parce que si $q = 1$ alors $\mathrm{Card}(\mathbb{K}) = 1$ et le théorème est trivial. Par ailleurs $Q(q)$ est un entier (parce que $Q \in \mathbb{Z}[X]$ et $q \in \mathbb{N}$) et $Q(q) \neq 0$, parce qu'à droite de (18.105) nous avons $q - 1 \neq 0$. Nous avons donc $|Q(q)| \geqslant 1$ et donc

$$|\phi_n(q)| \leqslant q - 1. \tag{18.106}$$

Par définition du polynôme cyclotomique nous avons

$$|\phi_n(q)| = \prod_{z \in \Delta_n} |q - z|. \tag{18.107}$$

Étant donné que ce produit doit être inférieur à $q-1$, au moins un des termes doit l'être : il existe $z_0 \in \Delta_n$ tel que $|z_0 - q| \leqslant q - 1$. Étant donné que $n \geqslant 2$ nous avons $z_0 \neq 1$.

Mais d'autre part, comme indiqué sur la figure 18.1, la distance entre z_0 et q doit être strictement plus grande que $q-1$ parce que $q-1$ est le minimum de la distance entre le cercle trigonométrique et q, et n'est atteint qu'en $z = 1$.

Nous avons ainsi obtenu une contradiction, et nous concluons que le corps $\mathbb{K}$ est commutatif. □

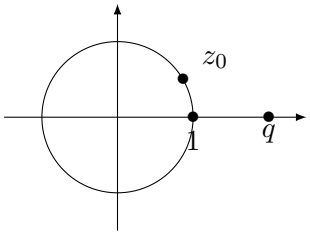


FIGURE 18.1 – Nous devons avoir $|z_0 - q| > q - 1$.

18.5.2 Existence, unicité

Nous avons déjà défini le corps fini $\mathbb{F}_p$ lorsque p est un nombre premier dans la section 6.1.3. Le théorème suivant sert à définir $\mathbb{F}_{p^n}$ lorsque p est premier.

Théorème 18.32.
Soit p un nombre premier, soit $n \in \mathbb{N}^*$ et $q = p^n$. Alors il existe un unique corps $\mathbb{K}$ de cardinal q. Ce corps est le corps de décomposition du polynôme $X^q - X$ sur $\mathbb{F}_p$.

Démonstration. Montrons l'unicité. Soit $\mathbb{K}$ un corps fini de cardinal $q = p^n$. Le groupe multiplicatif $\mathbb{K}^*$ est de cardinal $q - 1$, et par le corolaire 2.35 tous les éléments de $\mathbb{K}^*$ vérifient $g^{q-1} = e$, c'est-à-dire que dans $\mathbb{K}[X]$, les éléments de $\mathbb{K}^*$ sont des racines du polynôme

$$X^{q-1} - 1 \tag{18.108}$$

Par conséquent $\mathbb{K}$ est un corps de décomposition pour le polynôme $Q(X) = X^q - X = X(X^{q-1} - 1)$ parce que $Q(X) = 0$ dans $\mathbb{K}$. Il est unique par la proposition 6.135.

Montrons maintenant que le corps de décomposition de $P = X^q - X$ sur $\mathbb{F}_p$ est un corps de cardinal q. Pour ce faire nous considérons $\mathbb{K}$ ce corps de décomposition et $\mathbb{E}$, l'ensemble des racines de P dans $\mathbb{K}$. Nous allons montrer que $\mathbb{E} = \mathbb{K}$ et que $\mathbb{E}$ est un corps contenant q éléments.

Montrons que $\mathbb{E}$ est un corps. Pour $\alpha, \beta \in \mathbb{E}$ nous avons

$$(\alpha\beta)^q = \alpha^q \beta^q = \alpha\beta \tag{18.109}$$

parce que $\alpha^q = \alpha$. Le produit $\alpha\beta$ est donc encore dans $\mathbb{E}$. Pour la somme,

$$(\alpha + \beta)^q = (\alpha + \beta)^{p^n} = \Big((\alpha + \beta)^p\Big)^{p^{n-1}} = (\alpha^p + \beta^p)^{p^{n-1}} = \ldots = \alpha^{p^n} + \beta^{p^n} = \alpha + \beta. \tag{18.110}$$

En ce qui concerne l'inverse,

$$(\alpha^{-1})^q = (\alpha^q)^{-1} = \alpha^{-1}. \tag{18.111}$$

Donc $\mathbb{E}$ est un corps. Évidemment $\mathbb{E}$ est un corps de décomposition de P au sens où $\mathbb{E}$ est une extension de $\mathbb{F}_p$ sur lequel P est scindé (parce qu'il est scindé sur $\mathbb{K}$ et $\mathbb{E}$ est le sous corps de $\mathbb{K}$ contenant les racines de P) et tel que $\mathbb{E} = \mathbb{F}_p(\{\alpha_i\})$ où les α_i sont les racines de P. Notons que $\mathbb{F}_p \subset \mathbb{E}$ parce que dans $\mathbb{F}_p$ on a $x^q = x$.

Par unicité, nous avons $\mathbb{K} = \mathbb{E}$. Nous devons montrer que P possède exactement q racines distinctes, afin d'avoir $\mathrm{Card}(\mathbb{E}) = q$. Pour cela remarquons que

$$P'(X) = qX^{q-1} - 1 = -1 \tag{18.112}$$

dans $\mathbb{F}_p$. En effet $P \in \mathbb{F}_p$ et $q = 0$ dans $\mathbb{F}_p$. Par conséquent P' ne s'annule pas et P n'a pas de racines doubles. Toutes les racines étant simples, il y en a exactement q.

□

Le théorème 18.32 ne permet pas de *construire* le corps à $q = p^n$ éléments. Nous allons maintenant voir un certain nombre de résultats donnant des façons de construire. Ces résultats proviennent de [374, 375, 376] et de wikipedia

Proposition 18.33 ([376]).
Soit $\mathbb{K}$ un corps fini. Alors le groupe multiplicatif $\mathbb{K}^*$ est cyclique.

Démonstration. Soit $\mathbb{K}$ un corps ayant q éléments. Le groupe $\mathbb{K}^*$ en a $q-1$; ergo l'ordre des éléments de $\mathbb{K}^*$ sont des diviseurs de $q-1$; c'est le corolaire 2.35. Soit d un diviseur de $q-1$ et

$$H_d^* = \{x \text{ d'ordre } d \text{ dans } \mathbb{K}^*\} \tag{18.113a}$$
$$H_d = \{\text{racines de } X^d - 1 \text{ dans } \mathbb{K}\}. \tag{18.113b}$$

Ici le polynôme $X^d - 1$ est vu dans $\mathbb{K}[X]$. Notons que nous avons automatiquement $H_d^* \subset H_d$, mais l'inclusion inverse n'est pas assurée parce que les éléments d'ordre $d/2$ par exemple sont aussi dans H_d. Supposons $H_d^* \neq \emptyset$ et considérons $a \in H_d^*$. Alors l'application

$$\begin{aligned} \phi\colon \mathbb{Z}/d\mathbb{Z} &\to H_d \\ n &\mapsto a^n \end{aligned} \tag{18.114}$$

est un isomorphisme d'anneaux. En effet étant donné que $a \in H_d^* \subset H_d$, l'ensemble H_d contient le groupe cyclique engendré par a. Ce dernier contient, par construction, d éléments. Mais $\mathrm{Card}(H_d) \leqslant d$ parce que H_d est l'ensemble des racines d'un polynôme de degré d. Par conséquent $\mathrm{Card}(H_d) = d$ et l'ensemble H_d est bien engendré par a et ϕ est bien un isomorphisme. Par conséquent tous les éléments de H_d^* sont des générateurs de H_d.

Inversement soit x un générateur de H_d. L'ordre de H_d étant d, l'ordre de x doit être un diviseur de d. Supposons donc que x soit d'ordre d/k. Dans ce cas nous devrions avoir $\mathrm{Card}(H_d) = d/k$, ce qui contredit l'isomorphisme ϕ.

En conclusion, H_d^* est l'ensemble des générateurs du groupe H_d. Le nombre de générateurs de $\mathbb{Z}/d\mathbb{Z}$ étant $\varphi(d)$ par la proposition 18.12, et H_d étant isomorphe à $\mathbb{Z}/d\mathbb{Z}$ nous avons

$$\mathrm{Card}(H_d^*) = \varphi(d). \tag{18.115}$$

Par conséquent si H_d^* n'est pas vide, son cardinal est $\varphi(d)$. Nous avons

$$\begin{aligned} q-1 &= \mathrm{Card}(\mathbb{K}^*) && (18.116a)\\ &= \mathrm{Card}\Big(\bigcup_{d\,|\,q-1} H_d^*\Big) && (18.116b)\\ &= \sum_{d\,|\,q-1} \mathrm{Card}(H_d^*) && (18.116c)\\ &\leqslant \sum_{d\,|\,q-1} \varphi(d) && (18.116d)\\ &= q-1 && (18.116e) \end{aligned}$$

où nous avons utilisé le lemme 18.10. Par conséquent pour tout d divisant $q-1$ nous avons $\mathrm{Card}(H_d^*) = \varphi(d)$ et il y a au moins un élément d'ordre $q-1$ dans $\mathbb{K}$. Cet élément engendre $\mathbb{K}^*$ parce que $\mathbb{K}^*$ contient exactement $q-1$ éléments. Par conséquent $\mathbb{K}$ est cyclique. □

Corolaire 18.34.
Si p est un nombre premier, alors

$$(\mathbb{Z}/p\mathbb{Z})^* \simeq \mathbb{Z}/(p-1)\mathbb{Z}. \tag{18.117}$$

L'isomorphisme est un isomorphisme de groupes (abéliens). À gauche multiplicatif et à droite additif.

Démonstration. La proposition 18.33 nous enseigne que le le groupe multiplicatif d'un corps fini est cyclique et donc isomorphe à un certain $\mathbb{Z}/n\mathbb{Z}$. Donc $(\mathbb{Z}/p\mathbb{Z})^*$ est un groupe cyclique d'ordre p, et donc isomorphe à $\mathbb{Z}/n\mathbb{Z}$ avec $n = p-1$. □

Lorsque $\mathbb{K}$ est un corps les éléments du groupe $\mathbb{K}^*$ sont les **éléments primitifs** de $\mathbb{K}$.

Proposition 18.35.
Soit $\mathbb{K}$ un corps contenant q éléments. Alors

(1) $x^q = x$ pour tout $x \in \mathbb{K}$,

(2) $X^q - X = \prod_{a\in\mathbb{K}}(X - a)$.

Démonstration. Le groupe $\mathbb{K}^*$ ayant $q-1$ éléments, ses éléments vérifient $a^{q-1} = 1$ par le corolaire 2.35 et par conséquent $a^q = aa^{q-1} = a$.

Soit $a \in \mathbb{K}$. Étant donné que $a^q - a = 0$, le polynôme $(X - a)$ divise $X^q - X$ dans $\mathbb{K}[X]$. Par conséquent

$$\prod_{a\in\mathbb{K}}(X - a) \tag{18.118}$$

divise également $X^q - X$. Les polynômes $X^q - X$ et $\prod_{a\in\mathbb{K}}(X - a)$ étant deux polynômes unitaires de même degré, le fait que l'un divise l'autre montre qu'ils sont égaux. □

Exemple 18.36.
Soit $\mathbb{K} = \mathbb{Q}$ et $\mathbb{L} = \mathbb{Q}(\sqrt{2}, \sqrt{3})$. Afin de montrer que $\mathbb{L} = \mathbb{Q}(\alpha)$ avec $\alpha = \sqrt{2} + \sqrt{3}$ nous devons montrer que $\sqrt{2}$ et $\sqrt{3}$ sont des polynômes en α. △

Une conséquence du fait que $x^q = x$ est qu'il ne faut pas regarder le théorème 6.101 trop rapidement en disant « s'il s'annule partout, alors c'est le polynôme nul ». En effet dans un corps fini, « partout » n'est pas forcément très grand.

Exemple 18.37.
Si $\mathbb{F}_3 = \mathbb{Z}/3\mathbb{Z}$ est le[16] corps à 3 éléments, alors le polynôme $P(X) = X^3 - X$ s'évalue à zéro pour tout $x \in \mathbb{F}_3$ (proposition 18.35.) mais il n'est pas le polynôme nul. △

18.5.3 Symboles de Legendre et carrés

Source : [377].

Nous disons que $a \in \mathbb{F}_p$ est un **carré** s'il existe $b \in \mathbb{F}_p$ tel que $a = b^2$.

Définition 18.38.
*Soit $n \in \mathbb{N}$ et $p > 2$ un nombre premier. Le **symbole de Legendre** par*

$$\left(\frac{n}{p}\right) = \begin{cases} 0 & \text{si } p \text{ divise } n \\ 1 & \text{si } n \text{ est un carré dans } \mathbb{F}_p \\ -1 & \text{sinon.} \end{cases} \tag{18.119}$$

Note que -1 peut être un carré, et pas que dans $\mathbb{C}$. Par exemple dans $\mathbb{F}_5$ nous avons $4 = -1$ et donc -1 est un carré.

Proposition 18.39.
Soit un nombre premier $p > 2$. Le corps $\mathbb{F}_p^*$ contient autant de carrés que de non carrés. De plus pour tout $n \in \mathbb{N}$ nous avons

$$\left(\frac{n}{p}\right) = n^{(p-1)/2} \mod p. \tag{18.120}$$

Démonstration. Nous considérons l'application

$$\begin{aligned} \psi\colon \mathbb{F}_p^* &\to \mathbb{F}_p^* \\ x &\mapsto x^2. \end{aligned} \tag{18.121}$$

16. Le singulier est justifié par le théorème 18.32, mais ça n'a pas d'importance ici.

C'est un morphisme de groupes multiplicatifs et $\ker\psi = \{-1, 1\}$. Étant donné que $p > 2$, nous avons alors

$$\mathrm{Card}(\ker\psi) = 2 \tag{18.122}$$

parce que $1 \neq -1$. Évidemment l'ensemble des carrés dans $\mathbb{F}_p^*$ est l'image de ψ. Le premier théorème d'isomorphisme 2.28(3) nous permet alors de conclure que

$$\mathrm{Card}(\mathrm{Image}(\psi)) = \frac{\mathrm{Card}(\mathbb{F}_p^*)}{2}. \tag{18.123}$$

Ceci prouve la première assertion.

Par le petit théorème de Fermat (théorème 6.13), nous avons $x^{p-1} = 1$ pour tout $x \in \mathbb{F}_p^*$. Les $(p-1)$ éléments de $\mathbb{F}_p^*$ sont donc tous racines d'un des deux polynômes

$$X^{(p-1)/2} = \pm 1. \tag{18.124}$$

Mais chacun des deux ne peut avoir, au maximum, que $(p-1)/2$ solutions. Ils ont donc chacun exactement $(p-1)/2$ racines.

Nous pouvons maintenant prouver la formule (18.120). D'abord si $n = 0$, elle est évidente. Si n est un carré dans $\mathbb{F}_p$, nous posons $n = x^2$ et nous avons

$$n^{(p-1)/2} = n^{p-1} = 1 = \left(\frac{n}{p}\right). \tag{18.125}$$

Si n n'est pas un carré, c'est que n n'est pas une racine de $X^{(p-1)/2} = 1$. Le nombre n est alors une racine de $X^{(p-1)/2} = -1$. Nous avons alors

$$n^{(p-1)/2} = -1 = \left(\frac{n}{p}\right). \tag{18.126}$$

□

Corolaire 18.40.
Si $a, b \in \mathbb{N}$ et si $p > 2$ est un nombre premier, alors

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right). \tag{18.127}$$

Démonstration. Par la formule (18.120),

$$\left(\frac{ab}{p}\right) = (ab)^{(p-1)/2} = a^{(p-1)/2}b^{(p-1)/2} = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right). \tag{18.128}$$

□

Soit un nombre premier $q > 2$ et $\mathbb{A}$, un anneau de caractéristique p. Si $\alpha \in \mathbb{A}$ vérifie

$$1 + \alpha + \cdots + \alpha^{q-1} = 0, \tag{18.129}$$

nous définissons la **somme de Gauss** par

$$\tau = \sum_{x \in \mathbb{F}_q} \left(\frac{i}{q}\right) \alpha^i = \sum_{x=1}^{q-1} \left(\frac{x}{q}\right) \alpha^i. \tag{18.130}$$

Notons que la somme de Gauss dépend de q et du α choisis.

Proposition 18.41.
Les sommes de gauss vérifient les propriétés suivantes.

(1) $\tau^2 = \left(\frac{-1}{q}\right) q$. *Nous allons noter* $\epsilon(q) = \left(\frac{-1}{q}\right)$.

(2) Si $\mathbb{A}$ est de caractéristique $p \geqslant 3$ et si $p \neq q$ alors

$$\tau^p = \left(\frac{p}{q}\right)\tau. \tag{18.131}$$

(3) Si $\mathbb{A}$ est de caractéristique p et si q est premier avec p, alors τ est inversible dans $\mathbb{A}$.

Démonstration. D'abord nous notons que

$$\alpha^q - 1 = (\alpha - 1)(1 + \alpha + \cdots + \alpha^{q-1}) = 0 \tag{18.132}$$

par définition de α. Nous calculons

$$\begin{aligned}
\epsilon(q)\tau^2 &= \epsilon(q) \sum_{x,y\in\mathbb{F}_q} \left(\frac{x}{q}\right)\left(\frac{y}{q}\right)\alpha^{x+y} && \text{(18.133a)}\\
&= \sum_{x,y\in\mathbb{F}_q} \left(\frac{-xy}{q}\right)\alpha^{x+y}. && \text{(18.133b)}\\
&= \sum_{z\in\mathbb{F}_q}\sum_{y\in\mathbb{F}_q} \left(\frac{-(z-y)y}{q}\right)\alpha^{z} && \text{(18.133c)}\\
&= \sum_{z\in\mathbb{F}_q} s_z\alpha^z && \text{(18.133d)}
\end{aligned}$$

Justifications :

— Pour obtenir (18.133b) nous avons utilisé le corolaire 18.40.

— (18.133c) est un changement de variable $z = x + y$ dans la somme sur x.

— Pour (18.133d) nous avons posé

$$s_z = \sum_{y\in\mathbb{F}_q} \left(\frac{-(z-y)y}{q}\right). \tag{18.134}$$

Nous avons

$$s_0 = \sum_{y\in\mathbb{F}_q} \left(\frac{y^2}{q}\right). \tag{18.135}$$

Dans cette somme, tous les termes sont 1 sauf celui avec $y = 0$ qui vaut zéro. Nous avons donc $s_0 = q - 1$. Voyons maintenant s_y avec $y \neq 0$. L'application

$$\begin{aligned}
\mathbb{F}_q^* &\to \mathbb{F}_q\backslash\{1\}\\
k &\mapsto 1 - zy^{-1}
\end{aligned} \tag{18.136}$$

étant une bijection nous pouvons effectuer le changement de variables $t = y^{-1}z - 1$ pour la somme sur y en notant y^{-1} l'inverse de y dans $\mathbb{F}_q^*$, nous trouvons alors

$$\begin{aligned}
\sum_{y\in\mathbb{F}_q}\left(\frac{y(z-y)}{q}\right) &= \sum_{y\in\mathbb{F}_q}\left(\frac{y^2(y^{-1}z-1)}{q}\right) && \text{(18.137a)}\\
&= \sum_{y\in\mathbb{F}_q}\left(\frac{y^{-1}z-1}{q}\right) && \text{(18.137b)}\\
&= \sum_{t\in\mathbb{F}_q\backslash\{1\}}\left(\frac{t}{q}\right) && \text{(18.137c)}\\
&= \underbrace{\sum_{t\in\mathbb{F}_q}\left(\frac{y}{q}\right)}_{=0} - \left(\frac{1}{1}\right) && \text{(18.137d)}\\
&= -1 && \text{(18.137e)}
\end{aligned}$$

parce qu'il y a autant de carrés que de non carrés dans $\mathbb{F}_q^*$ (proposition 18.39). En résumé nous avons

$$\epsilon(q)\tau^2 = \sum_{z\in\mathbb{F}_q} s_z\alpha^z \tag{18.138}$$

où

$$s_z = \begin{cases} q-1 & \text{si } z=0 \\ -1 & \text{sinon.} \end{cases} \tag{18.139}$$

Cela donne

$$\epsilon(q)\tau^2 = (q-1) - \underbrace{(\alpha+\cdots+\alpha^{q-1})}_{=-1} = q \tag{18.140}$$

où nous avons utilisé l'hypothèse sur α. Donc $\epsilon(q)\tau^2 = q$, et étant donné que $\epsilon(q) = \pm 1$ nous concluons

$$\tau^2 = \epsilon(q)q. \tag{18.141}$$

Nous prouvons maintenant la seconde partie. Vu que $\mathbb{A}$ est de caractéristique p en utilisant le fait que le morphisme de Frobenius est un morphisme,

$$\tau^p = \left(\sum_{x\in\mathbb{F}_q}\left(\frac{x}{q}\right)\alpha^x\right)^p = \sum_{x\in\mathbb{F}_q}\left(\frac{x}{q}\right)^p\alpha^{px}. \tag{18.142}$$

Étant donné que $\left(\frac{x}{q}\right) = \pm 1$ et que p est impair, nous avons

$$\left(\frac{x}{q}\right)^p = \left(\frac{x}{q}\right). \tag{18.143}$$

Du coup nous avons

$$\left(\frac{p}{q}\right)\tau^p = \sum_{x\in\mathbb{F}_p}\left(\frac{xp}{q}\right)\alpha^{px}. \tag{18.144}$$

Mais p étant inversible dans $\mathbb{F}_q$, l'application $x\mapsto px$ est une bijection et nous pouvons sommer sur px au lieu de x :

$$\left(\frac{p}{q}\right)\tau^p = \sum_{x\in\mathbb{F}_p}\left(\frac{x}{q}\right)\alpha^x = \tau. \tag{18.145}$$

Nous trouvons alors que

$$\tau^p = \left(\frac{p}{q}\right)\tau. \tag{18.146}$$

Étant donné la formule du τ^2 que nous venons de démontrer, nous avons $\tau^2 = \pm q$. Les nombres p et q étant premiers entre eux, la relation de Bézout (théorème 3.13) nous donne a et b tels que

$$ap + ba = 1. \tag{18.147}$$

Cela montre que b est un inverse de q modulo p. Donc τ^2 est inversible, et il en découle que τ lui-même est inversible. □

Théorème 18.42 (Loi de réciprocité quadratique)**.**
Soient deux nombres premiers distincts $p, q \geqslant 3$. Alors

$$\left(\frac{p}{q}\right) = (-1)^{\frac{(p-1)(q-1)}{4}}\left(\frac{q}{p}\right). \tag{18.148}$$

Démonstration. Soit ϕ_q le polynôme $1 + X + \cdots + X^{q-1}$ et l'anneau

$$\mathbb{A} = \mathbb{F}_p[X]/(\phi_q). \tag{18.149}$$

Cela est un anneau de caractéristique p parce que son unité est le polynôme constant 1. Nous nommons $\alpha = X/(\phi_q)$, c'est-à-dire que $\phi_q(\alpha) = 0$ dans $\mathbb{A}$, et nous pouvons considérer la somme de Gauss

$$\tau = \sum_{i\in\mathbb{F}_q} \left(\frac{i}{q}\right) \alpha^i. \tag{18.150}$$

Notons que cela est un élément de $\mathbb{A}$ et plus précisément un polynôme de degré zéro dans $\mathbb{A}$, et encore plus précisément, une classe d'un tel polyôme. Donc les coefficients de α doivent être compris comme des éléments de $\mathbb{F}_p$. Nous savons (proposition 18.41) que

$$\tau^2 = \left(\frac{-1}{q}\right), \tag{18.151}$$

et en utilisant la formule (18.120) nous trouvons

$$\left(\frac{\tau^2}{p}\right) = (\tau^2)^{(p-1)/2} \mod p = \tau^{p-1} \mod p \tag{18.152}$$

En réalité sur cette dernière ligne, nous ne devrions pas préciser le « $\mod p$ » parce que, comme mentionné plus haut, ce sont des éléments de $\mathbb{F}_p$. En utilisant cela ainsi que (18.131) nous avons

$$\underbrace{\left(\frac{\tau^2}{p}\right)}_{\tau^{p-1}} \tau = \tau^p = \left(\frac{p}{q}\right) \tau \tag{18.153}$$

Vu que τ est inversible, nous écrivons

$$\left(\frac{\tau^2}{p}\right) = \left(\frac{p}{q}\right) \tag{18.154}$$

Nous utilisons maintenant la formule (18.120) sur le membre de gauche avec $n = \tau^2 = \left(\frac{-1}{q}\right)$:

$$\left(\frac{p}{q}\right) = \left(\frac{-1}{q}\right)^{\frac{q-1}{2}} \left(\frac{q}{p}\right). \tag{18.155}$$

Toujours avec la même formule nous pouvons substituer $\left(\frac{-1}{q}\right)$ par $(-1)^{(q-1)/2}$ et obtenir

$$\left(\frac{p}{q}\right) = (-1)^{\frac{(q-1)}{2}\frac{(p-1)}{2}}. \tag{18.156}$$

□

Lemme 18.43.
Si p est un nombre premier $p \geqslant 3$, alors le symbole de Legendre $x \mapsto \left(\frac{x}{p}\right)$ est l'unique morphisme non trivial de $\mathbb{F}_p^*$ dans $\{-1, 1\}$.

Démonstration. Le fait que le symbole de Legendre soit non trivial est simplement le fait qu'il y ait des carrés et des non carrés dans $\mathbb{F}_p^*$; voir la proposition 18.39. Pour l'unicité, soit $\alpha\colon \mathbb{F}_p^* \to \{-1, 1\}$ un morphisme surjectif (c'est-à-dire non trivial). Étant donné que

$$\mathbb{F}_p^* = \ker(\alpha) \cup -\ker(\alpha), \tag{18.157}$$

le groupe $\mathbb{F}_p^*/\ker(\alpha)$ ne contient que deux éléments : $[1]$ et $[-1]$. Autrement dit, $\ker(\alpha)$ est d'indice 2 dans $\mathbb{F}_p^*$.

Or $\mathbb{F}_p^*$ ne possède qu'un seul sous-groupe d'indice 2. En effet soit S un tel sous-groupe et a, un générateur de $\mathbb{F}_p^*$ (qui est cyclique par la proposition 18.33), alors $a^2 \in S$ par le lemme 3.37. Par conséquent S contient le groupe des puissances paires de a. Le groupe S ne peut rien contenir de plus parce qu'il est d'indice 2 et que l'ordre de $\mathbb{F}_p^*$ est pair.

Bref, le sous-groupe $\ker(\alpha)$ est l'unique sous-groupe d'indice 2 dans $\mathbb{F}_p^*$. Mais la proposition 18.39 nous indique que $|(\mathbb{F}_p^*)^2| = \frac{p-1}{2}$, c'est-à-dire que le groupe des carrés est d'indice 2. Nous avons donc, par l'unicité,

$$\ker(\alpha) = (\mathbb{F}_p^*)^2. \tag{18.158}$$

Au final, pour $y \in \mathbb{F}_p^*$,

$$\alpha(y) = \begin{cases} 1 & \text{si } y \text{ est un carré} \\ -1 & \text{sinon.} \end{cases} \tag{18.159}$$

Cela est bien la définition des symboles de Legendre. □

Proposition 18.44.
Pour p premier nous avons

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & \textit{si } p \in [1]_8 \textit{ ou } p \in [7]_8 \\ -1 & \textit{sinon.} \end{cases} \tag{18.160}$$

Démonstration. Soit le polynôme

$$X^4 + 1 \in \mathbb{F}_p[X] \tag{18.161}$$

et α, une racine dans une extension [17] de $\mathbb{F}_p$ [18]. Nous posons $\theta = \alpha + \alpha^{-1}$ et nous calculons

$$\theta^2 = (\alpha + \alpha^{-1})(\alpha + \alpha^{-1}) = \alpha^2 + 2 + (\alpha^2)^{-1} = \alpha^2 + 2 - \alpha^2 = 2 \tag{18.162}$$

parce que α^2 étant -1, nous avons $(\alpha^2)^{-1} = -\alpha^2$. Bref, $\theta^2 = 2$.

Dire que 2 est un carré modulo p revient à dire que θ est dans $\mathbb{F}_p$. C'est-à-dire que pour calculer le symbole de Legendre $\left(\frac{2}{p}\right)$, nous étudions pour quels p, l'élément θ est vraiment dans $\mathbb{F}_p$ et non seulement dans l'extension $\mathbb{F}_p(\alpha)$. En tenant compte de l'exemple 6.96, il faut distinguer deux cas : $\alpha^p = \alpha$ et $\alpha^p \neq \alpha$. Autrement dit, si $\alpha^k = \alpha$ pour un certain nombre premier k, alors le cas $p = k$ est à traiter à part. La liste des puissances de α est :

$$1, \alpha, \alpha^2, \alpha^3, -1, -\alpha, -\alpha^2, -\alpha^3, 1, \alpha, \ldots \tag{18.163}$$

Nous avons donc automatiquement $\alpha^{9k} = \alpha$, mais $p = 9k$ est exclu parce que p est premier. Nous devons donc vérifier si une des possibilités

$$\alpha^2 = \alpha \tag{18.164a}$$
$$\alpha^3 = \alpha \tag{18.164b}$$
$$-\alpha = \alpha \tag{18.164c}$$
$$-\alpha^3 = \alpha \tag{18.164d}$$

est possible. Il est aisément vérifiable, au cas par cas, que ces possibilités sont toutes incompatibles avec $\alpha^4 = -1$. Nous avons donc certainement $\alpha^p \neq \alpha$ et compte tenu de l'exemple 6.96, l'équation $x^p = x$ caractérise les éléments de $\mathbb{F}_p$ dans $\mathbb{F}_p(\alpha)$.

L'équation $X^2 = 2$ a exactement deux solutions qui sont $\pm\theta$. Nous avons donc $2 \in \mathbb{F}_p^2$ si et seulement si $\theta \in \mathbb{F}_p$ si et seulement si $\theta^p = \theta$. Nous avons réduit notre problème à déterminer pour quels p nous avons $\theta^p = \theta$. D'abord nous avons, par le morphisme de Frobenius,

$$\theta^p = (\alpha + \alpha^{-1})^p = \alpha^p + \alpha^{-p}. \tag{18.165}$$

17. Dans la source que je suivais (je ne sais plus où), on parlait ici de « fermeture » de $\mathbb{F}_p j$ et non d'exetension. Il me semble que parler simplement d'extension suffit. Vous confirmez ?

18. Voir par exemple la proposition 6.132 pour l'existence d'une extension comme il faut.

Nous pouvons maintenant conclure facilement. Un nombre premier étant impair (sauf $p = 2$ qui peut être traité à part), p est automatiquement dans un des ensembles $[1]_8$, $[3]_8$, $[5]_8$ ou $[7]_8$. Nous avons quatre petites vérifications à faire. Dans tous les cas $\alpha^{8k} = 1$. Si $p = 1 + 8k$, alors

$$\theta^p = \alpha^{1+8k} + (\alpha^{-1})^{1+8k} = \alpha + \alpha^{-1} = \theta, \tag{18.166}$$

donc 2 est un carré dans $\mathbb{F}_p$. Si $p \in [3]_8$, alors $\theta^p = \alpha^3 + \alpha^{-3}$. Si cela était égal à $\alpha + \alpha^{-1}$, alors nous aurions

$$\alpha^6 + 1 = \alpha^4 + \alpha^2, \tag{18.167}$$

et donc $\alpha^2 = 1$, ce qui est impossible. Les vérifications pour $p \in [5]_8$ et $p \in [7]_8$ sont du même style. □

18.5.4 Théorème de Chevalley-Warning

Lemme 18.45.
Soit $\mathbb{K}$ un corps de caractéristique p et de cardinal q. Pour $m \in \mathbb{N}$ nous définissons

$$S_m = \sum_{x \in \mathbb{K}} x^m. \tag{18.168}$$

Alors nous avons

$$S_m \mod p = \begin{cases} -1 & \text{si } m \geqslant 1 \text{ et } m \text{ divisible par } q-1 \\ 0 & \text{sinon.} \end{cases} \tag{18.169}$$

Démonstration. Si $m = 0$, alors $x^0 = 1$ et $S_m = q$. Par conséquent $S_m \mod p = 0$ parce que la caractéristique d'un corps divise son ordre (proposition 3.62).

Nous prenons maintenant $m \geqslant 1$ et nous voyons séparément les cas où $q - 1$ divise m ou non. Si $q - 1$ divise m, alors pour tout $x \neq 0$ nous avons

$$x^m = x^{k(q-1)} = 1 \tag{18.170}$$

parce que $\mathbb{K}^*$ est cyclique et $x^{q-1} = 1$ par le petit théorème de Fermat (théorème 6.13). Par conséquent nous avons

$$\sum_{x \in \mathbb{K}} x^m = \sum_{x \in \mathbb{K}^*} 1 = q - 1. \tag{18.171}$$

Si le nombre $m \geqslant 1$ n'est pas divisible par $q - 1$ alors nous prenons un générateur y du groupe $\mathbb{K}^*$. Un tel élément vérifie $y^m \neq 1$. En effet, si y vérifiait $y^m = 1$ alors cela signifierait que l'ordre de $\mathbb{K}^*$ est un diviseur de m, ce qui n'est pas le cas ici parce que l'ordre de $\mathbb{K}^*$ est $q - 1$. Pour un tel y, l'application

$$\begin{aligned} \varphi \colon \mathbb{K}^* &\to \mathbb{K}^* \\ x &\mapsto yx \end{aligned} \tag{18.172}$$

est une bijection [19]. En ce qui concerne l'injectivité, $ya = yb$ implique $a = b$. En ce qui concerne la surjectivité, si a est un générateur, si $z = a^l$ et si $y = a^k$, alors

$$z = \varphi(a^{l-k}). \tag{18.173}$$

Nous pouvons maintenant faire le calcul.

$$S_m = \sum_{x \in \mathbb{K}^*} x^n = \sum_{x \in \mathbb{K}^*} (yx)^m = y^m \sum_{x \in \mathbb{K}^*} x^m = y^m S_m. \tag{18.174}$$

Étant donné que $y^m \neq 1$, la seule solution est $S_m = 0$. □

19. Notons que nous n'avons pas réellement besoin que y soit un générateur. Nous n'utilisons seulement le fait que $y^m \neq 1$ et $y \neq 0$.

Théorème 18.46 (Chevalley-Warninig[378]).
Soit $\mathbb{K}$ un corps fini de cardinal q et de caractéristique p. Soient $P_1, \dots, P_r$ des éléments de $\mathbb{K}[X_1, \dots, X_n]$ tels que $\sum_{i=1}^r \deg(P_i) < n$. Nous considérons l'ensemble des zéros communs à tous les polynômes :

$$V = \{x \in \mathbb{K}^n \text{ tel que } P_1(x) = \dots = P_r(x) = 0\}. \tag{18.175}$$

Alors $\mathrm{Card}(V) = 0 \mod p$.

Démonstration. Nous considérons le polynôme

$$P = \prod_{i=1}^r (1 - P_i^{q-1}). \tag{18.176}$$

Montrons que

$$P(x) = \begin{cases} 1 & \text{si } x \in V \\ 0 & \text{sinon.} \end{cases} \tag{18.177}$$

La première ligne est facile : étant donné que tous les $P_i(x)$ sont nuls pour $x \in V$, nous avons $P(x) = 1$. Si x n'est pas dans V, alors nous avons un i tel que $P_i(x) \in \mathbb{K}^*$. Mais dans ce cas (toujours la cyclicité de $\mathbb{K}^*$) nous avons $P_i(x)^{q-1} = 1$ et donc le produit est nul.

En utilisant l'hypothèse sur le degré des P_i, nous trouvons que

$$\deg(P) = \sum_{i=1}^r (q-1)\deg(P_i) < n(q-1). \tag{18.178}$$

Pour un polynôme $Q \in \mathbb{K}[X_1, \dots, X_n]$, nous définissons

$$\int Q = \sum_{x \in \mathbb{K}^n} Q(x). \tag{18.179}$$

Nous avons immédiatement

$$\int P = \sum_{x \in \mathbb{K}^n} P(x) = \sum_{x \in V} 1 = \mathrm{Card}(V) \mod p. \tag{18.180}$$

Nous insistons sur le « modulo p » parce que dans la formule $P(x) = 1$, le membre de droite est le 1 de $\mathbb{K}$; il est donc automatiquement modulo la caractéristique de $\mathbb{K}$.

Il nous reste à prouver que $\int P = 0$. Pour cela nous décomposons

$$P = \sum_m c_m X_1^{m_1} \dots X_n^{m_n} \tag{18.181}$$

où la somme s'étend sur les $m \in \mathbb{N}^n$ tels que $c_m \neq 0$. Nous avons

$$\int P = \sum_{\in \mathbb{K}^n} \sum_m c_m x_1^{m_1} \dots x_n^{m_n} \tag{18.182a}$$

$$= \sum_m c_m \left(\sum_{x \in \mathbb{K}^n} x_1^{m_1} \dots x_n^{m_n} \right) \tag{18.182b}$$

$$= \sum_m c_m S_{m_1} \dots S_{m_n}. \tag{18.182c}$$

Le terme de plus haut degré dans la décomposition (18.181) est celui du m tel que $\sum_i m_i$ est le plus grand, mais étant donné que nous savons que ce degré est plus petit que $n(q-1)$, nous avons pour tous les m entrant dans la somme que

$$\sum_{i=1}^n m_i < n(q-1). \tag{18.183}$$

En particulier pour tout $m \in \mathbb{N}^n$, il existe i tel que $m_i < q-1$, et dans ce cas $S_{m_i} = 0$. Donc tous les termes de la somme

$$\sum_{m\in\mathbb{N}^n} c_m S_{m_1} \dots S_{m_n} \tag{18.184}$$

ont un facteur nul. □

Corolaire 18.47.
Soit P_i des polynômes à n variables avec $\sum_{i=1}^r \deg(P_i) < n$. Si les P_i n'ont pas de termes constants, alors ils ont un zéro commun non trivial.

Démonstration. Nous reprenons les notations du théorème 18.46. Étant donné que les P_i n'ont pas de termes constants, $0 \in V$, mais $\mathrm{Card}(V) = 0 \mod p$. Par conséquent nous devons avoir $\mathrm{Card}(V) > p$. □

Exemple 18.48.
Nous considérons les polynômes

$$P_1(x,y,t,u) = xy + x + ux \tag{18.185a}$$
$$P_2(x,y,t,u) = x + y - 3t. \tag{18.185b}$$

La somme de leurs degrés est 3 et ce sont des polynômes à 4 variables. Nous devons donc avoir, en vertu du corolaire 18.47, des autres racines que la racine triviale $(x,y,t,u) = (0,0,0,0)$.

Le corolaire nous donne aussi une borne inférieure nombre de racines à chercher : plus que la caractéristique du corps sur lequel nous travaillons. Nous pouvons dire cela sans avoir la moindre idée de la façon dont on pourrait résoudre le système $P_1 = P_2 = 0$. △

18.5.5 Contenu d'un polynôme

Lemme 18.49 (de Gauss[54, 379]).
Soient $P, Q \in \mathbb{Z}[X]$. Alors

$$c(PQ) = c(P)c(Q) \tag{18.186}$$

où c est le contenu, définition 3.181.

Démonstration. Afin de fixer les notations, nous posons $P = \sum_i a_i X^i$ et $Q = \sum_j b_j Y^j$.

Pour les polynômes primitifs Nous commençons par supposer que $c(P) = c(Q) = 1$. Dans ce cas si $c(PQ) \neq 1$, nous considérons un nombre premier p divisant $c(PQ)$. Vu que le contenu de P et de Q sont 1, le nombre p ne peut pas diviser tous leurs coefficients. Nous définissons i_0 de façon que a_{i_0} soit le premier à ne pas être divisible par p et j_0 de telle façon que b_{j_0} soit le premier à ne pas être divisible par p. Autrement dit :

$$p \mid a_0, p \mid a_1, \dots, p \mid a_{i_0-1}, p \nmid a_{i_0} \tag{18.187}$$

et similaire pour j_0. Donc p ne divise ni a_{i_0} ni b_{j_0}. Nous nous demandons alors avec malice quel est le coefficient de $X^{i_0+j_0}$ dans PQ. La réponse est :

$$a_{i_0}b_{j_0} + \sum_{\substack{i+j=i_0+j_0\\ i<i_0 \text{ ou } j<j_0}} a_i b_j. \tag{18.188}$$

Par définition p divise soit a_i soit b_j pour chacun des termes de la grande somme. Vu que p ne divise pas $a_{i_0}b_{j_0}$, il ne divise pas le coefficient de $X^{i_0+j_0}$ dans PQ, alors que nous étions partis en disant que p divisait tous les coefficients de PQ.

Nous concluons donc que $c(PQ) = 1$.

Cas général Si P et Q sont maintenant des polynômes sans conditions particulières dans $\mathbb{Z}[X]$, nous considérons $P_1 = \frac{P}{c(P)}$ et $Q_1 = \frac{Q}{c(Q)}$; ces deux polynômes sont primitifs et nous avons alors, en utilisant la première partie :

$$c(P_1Q_1) = 1. \tag{18.189}$$

Étant donné que

$$P_1Q_1 = \frac{1}{c(P)c(Q)}PQ, \tag{18.190}$$

nous avons

$$c(PQ) = c(P)c(Q)c(P_1Q_1) = c(P)c(Q). \tag{18.191}$$

□

18.5.6 Théorème de l'élément primitif

Définition 18.50.
*Soit $\mathbb{K}$ un corps. Une extension $\mathbb{L}$ de $\mathbb{K}$ est dite **finie** si $\mathbb{L}$ est un espace vectoriel de dimension finie sur $\mathbb{K}$.*

Notez que la définition d'extension finie ne suppose ni que $\mathbb{K}$ ni que $\mathbb{L}$ soient finis en tant qu'ensembles.

Théorème 18.51 (de l'élément primitif)**.**
Si $\mathbb{K}$ est un corps fini, toute extension finie de $\mathbb{K}$ est simple[20].

Si $\mathbb{K}$ est un corps quelconque alors toute extension séparable finie est simple.

Démonstration. Nous ne donnons la preuve que dans le cas où $\mathbb{K}$ est fini. Dans ce cas nous savons par la proposition 18.33 que le groupe $\mathbb{K}^*$ est cyclique. Si de plus $\mathbb{L}$ est une extension finie alors $\mathbb{L}$ est fini en tant qu'ensemble. Par conséquent $\mathbb{L}^*$ est un groupe cyclique. Si α est un générateur de $\mathbb{L}$ alors $\mathbb{L} = \mathbb{K}(\alpha)$ et l'extension est donc simple.

Une preuve de l'assertion dans le cas où $\mathbb{K}$ est infini peut être trouvée sur wikipédia. □

Proposition 18.52.
L'ordre d'un polynôme P vérifie les propriétés suivantes :

(1) L'ordre de P est l'ordre multiplicatif de ses racines

(2) L'ordre de P divise $p^n - 1$.

Lemme 18.53.
Soit p un nombre premier et P un polynôme irréductible unitaire de degré n. Si $\alpha, \beta \in \mathbb{F}_p[X]/P$, alors $(\alpha + \beta)^p = \alpha^p + \beta^p$.

Démonstration. La preuve est exactement la preuve classique :

$$(\alpha + \beta)^p = \sum_k \binom{k}{p} \alpha^k \beta^{p-k} \tag{18.192}$$

où les coefficients binomiaux sont dans $\mathbb{F}_p$ et donc nuls pour les k différents de p et de 0. □

Cette proposition est encore vraie avec $\alpha, \beta \in \mathbb{F}_{p^n}$ et $(\alpha + \beta)^{p^n}$.

Lemme 18.54.
Si $\alpha \in \mathbb{F}_q$ est une racine d'ordre k de P (de degré n) alors les racines de $X^k - 1$ sont $\{\alpha^i$ tel que $i = 0, \dots, k-1\}$.

20. Définition 6.92.

Nous serions donc intéressé à construire $\mathbb{F}_q$ comme quotient de $\mathbb{F}_p[X]$ par un polynôme primitif. Le théorème suivant donne une description abstraite de $\mathbb{F}_q$ qui va nous servir de point de départ pour la construction.

Théorème 18.55 (Théorème de l'élément primitif)**.**
Soit p un nombre premier, $n \in \mathbb{N}$ et $q = p^n$. Soit $\mathbb{K}$ un corps à q éléments. Alors

(1) Il existe $\alpha \in \mathbb{K}$ tel que $\mathbb{K} = \mathbb{F}_p[\alpha]$.

(2) Il existe une polynôme irréductible $P \in \mathbb{F}_p[X]$ de degré n tel que

$$\begin{aligned} \phi\colon \mathbb{F}_p[X]/(P) &\to \mathbb{K} \\ \bar{X} &\mapsto \alpha \end{aligned} \tag{18.193}$$

soit un isomorphisme de corps.

Soit α et P choisis pour avoir les propriétés citées plus haut. Alors nous avons les propriétés suivantes.

(1) P est primitif[21]*.*

(2) P est scindé dans $\mathbb{K}$.

(3) L'ensemble des racines de P est $\{\alpha, \alpha^p, \dots, \alpha^{p^{n-1}}\}$.

(4) Le polynôme P divise $X^q - X$ dans $\mathbb{F}_p[X]$.

Démonstration. Le corps $\mathbb{K}$ étant fini, il est cyclique par la proposition 18.33. Si α un générateur de $\mathbb{K}^*$ alors

$$\mathbb{K} = \mathbb{F}_p[\alpha]. \tag{18.194}$$

Soit ℓ le plus grand entier tel que l'ensemble

$$\{1, \alpha, \cdots, \alpha^{\ell-1}\} \subset \mathbb{K} \tag{18.195}$$

soit libre. Pour rappel $\mathbb{K}$ est une espace vectoriel sur $\mathbb{F}_p$. Il existe des $a_i \in \mathbb{F}_p$ tels que

$$\alpha^\ell + a_{\ell-1}\alpha^{\ell-1} + \cdots + a_0 = 0. \tag{18.196}$$

De façon équivalente, il existe un polynôme unitaire $P \in \mathbb{F}_p[X]$ de degré ℓ tel que $P(\alpha) = 0$. Étant donné que α est générateur de $\mathbb{K}$,

$$\mathbb{K} = \operatorname{Span}\{1, \alpha, \dots, \alpha^{\ell-1}\} \tag{18.197}$$

parce que $\mathbb{K}$ est généré par les puissances de α alors que les puissances de α plus hautes que $\ell - 1$ peuvent être générées par $1, \alpha, \dots, \alpha^{\ell-1}$. L'espace $\mathbb{K}$ est donc un $\mathbb{F}_p$-espace vectoriel de dimension ℓ ; par conséquent

$$\operatorname{Card}(\mathbb{K}) = p^n = q \tag{18.198}$$

et $\ell = n$.

Montrons que P est irréductible dans $\mathbb{F}_p$. Si P était réductible dans $\mathbb{F}_p$, l'élément $\alpha \in \mathbb{K}$ serait une racine d'un des facteurs, c'est-à-dire qu'il serait racine d'un polynôme de degré inférieur à n, ce qui contredirait le fait que

$$\{\alpha^{\ell-1}, \dots, 1\} \tag{18.199}$$

est libre.

Montrons que l'application

$$\begin{aligned} \phi\colon \mathbb{F}_p[X]/(P) &\to \mathbb{K} \\ \bar{X} &\mapsto \alpha \end{aligned} \tag{18.200}$$

21. Définition 3.183.

est un isomorphisme. Pour l'injectivité, deux éléments $Q_1, Q_2 \in \mathbb{F}_p[X]/(P)$ s'écrivent

$$Q_1 = \sum_{k=0}^{n-1} a_k \bar{X}^k \tag{18.201a}$$

$$Q_2 = \sum_{k=0}^{n-1} b_k \bar{X}^k. \tag{18.201b}$$

Dans ce cas si $\phi(Q_1) = \phi(Q_2)$ alors

$$\phi(Q_1) = \sum_{k=0}^{n-1} a_k \alpha^k = \phi(Q_2) = \sum_{k=0}^{n-1} b_k \alpha^k. \tag{18.202}$$

Mais l'ensemble $\{1, \alpha, \dots, \alpha^{n-1}\}$ étant libre sur $\mathbb{F}_p$, cela implique $a_k = b_k$. La surjectivité de ϕ provient du fait que α génère $\mathbb{K}$.

Nous passons maintenant à la seconde partie de la démonstration. Soient $\alpha \in \mathbb{K}$ tel que $\mathbb{K} = \mathbb{F}_p[\alpha]$ et $P \in \mathbb{F}_p[X]$ un polynôme irréductible de degré n tel que $\alpha \mapsto \bar{X}$ soit un isomorphisme entre $\mathbb{K}$ et $\mathbb{F}_p[X]/(P)$.

Le polynôme P est primitif parce que α est d'ordre p^n dans $\mathbb{K}$ alors que $\bar{X} \mapsto \alpha$ est un isomorphisme. Par conséquent $\bar{X}$ est d'ordre p^n dans $\mathbb{F}_p[X]/P$.

Nous commençons par prouver que l'ensemble

$$\{\alpha, \alpha^p, \alpha^{p^2}, \dots, \alpha^{p^{n-1}}\} \tag{18.203}$$

est l'ensemble des racines distinctes de P. Pour cela nous posons

$$P(X) = \sum_{k=0}^{n} a_k X^k \tag{18.204}$$

avec $a_k \in \mathbb{F}_p$. D'abord α est une racine de P. En effet

$$P(\bar{X}) = \sum_k a_k \bar{X}^k = 0 \tag{18.205}$$

parce que cette somme est calculée dans $\mathbb{F}_p[X]/(P)$. En appliquant l'isomorphisme ϕ à l'égalité (18.205) nous trouvons

$$0 = \phi\big(P(\bar{X})\big) = \sum_k a_k \phi(\bar{X}^k) = \sum_k a_k \alpha^k. \tag{18.206}$$

Donc α est bien une racine de P dans $\mathbb{F}_p[X]$. Nous devons montrer qu'il en est de même pour les autres puissances dans l'ensemble (18.203). Étant donné que pour tout x dans $\mathbb{F}_p$ nous avons $x^p = x$, nous avons aussi

$$P(X^p) = \sum_k a_k (X^p)^k = \sum_k a_k^p (X^p)^k = \sum_k (a_k X^k)^p \tag{18.207}$$

alors que nous savons que $x \mapsto x^p$ est un automorphisme de $\mathbb{F}_p$ par la proposition 3.65. Par conséquent

$$P(X^p) = \sum_k (a_k X^k)^p = \left(\sum_k a_k X^k\right)^p = P(X)^p. \tag{18.208}$$

Nous avons montré que si β est une racine de P, alors β^p est également une racine de P. Nous savons déjà que α est une racine de P, et que α est également générateur de $\mathbb{K}$, c'est-à-dire que α est d'ordre $q-1$. Les puissances

$$\alpha, \alpha^p, \alpha^{p^2}, \dots, \alpha^{p^{n-1}} \tag{18.209}$$

sont donc distinctes ($\alpha^{p^n} = \alpha^q = 1$) et sont toutes des racines de P. Étant donné que P est de degré n il ne peut pas y avoir d'autres racines. Nous concluons que l'ensemble

$$\{\alpha, \alpha^p, \alpha^{p^2}, \dots, \alpha^{p^{n-1}}\} \tag{18.210}$$

est l'ensemble des racines distinctes de P dans $\mathbb{K}$. Le polynôme P est alors scindé dans $\mathbb{K}[X]$.

Le dernier point du théorème est de montrer que P divise $X^q - X$. Pour cela nous allons montrer que toutes les racines de P sont des racines de $X^q - X$. Soit β une racine de P ; il s'écrit $\beta = \alpha^k$ pour un certain k. Étant donné que $\alpha^{q-1} = e = \alpha^{p^n-1}$,

$$\begin{aligned}
\beta^q &= (\alpha^{p^n})^k && (18.211a)\\
&= \big(\alpha^{p^n-1}\alpha\big)^k && (18.211b)\\
&= \big(\alpha^{q-1}\alpha\big)^k && (18.211c)\\
&= \alpha^k && (18.211d)\\
&= \beta. && (18.211e)
\end{aligned}$$

Cela signifie que $\beta^q = \beta$ et donc que β est racine de $X^q - X$. □

Corolaire 18.56.
Le corps fini à $q = p^n$ éléments est de caractéristique p.

Démonstration. Nous considérons le corps fini $\mathbb{K}$ à q éléments sous la forme $\mathbb{K} = \mathbb{F}_p[X]/P$ comme indiqué par l'équation (18.193). Soit 1_q la classe du polynôme 1 modulo P, nous considérons le morphisme

$$\begin{aligned}
\mu\colon \mathbb{Z} &\to \mathbb{F}_q\\
n &\mapsto n1_q.
\end{aligned} \tag{18.212}$$

Le noyau de cette application est $\ker\mu = \mathbb{Z}_p$ parce que $p1_q = 0$, les coefficients étant à comprendre dans $\mathbb{F}_p$. □

Définition 18.57.
Soit P, un polynôme de degré n et p, un nombre premier. Un élément $\alpha \in \mathbb{F}_p[X]/(P)$ est une ***racine primitive*** *si les puissances de α parcourent tout le groupe multiplicatif $(\mathbb{F}_p[X]/P)^*$.*

Lemme 18.58.
Soit p un nombre premier et P, un polynôme de degré n. Si $\alpha \in \mathbb{F}_p[X]/P$ est une racine primitive de P alors les autres racines de P sont également primitives.

Démonstration. Soit $\alpha \in \mathbb{F}_p[X]/P$ une racine primitive de P. L'élément α^p est également une racine parce que si $P = \sum_k a_k X^k$,

$$P(\alpha^p) = \sum_k (a_k\alpha^k)^p = \big(\sum_k a_k\alpha^k\big)^p = 0 \tag{18.213}$$

où nous avons utilisé le fait que $a_k^p = a_k$ étant donné que $a_k \in \mathbb{F}_p$. Par hypothèse α est une racine primitive ; cela implique que les éléments $\alpha, \alpha^p, \alpha^{p^2}, \dots, \alpha^{p^n-1}$ sont distincts dans $\mathbb{F}_p[X]/P$. Ces éléments constituent donc *toutes* les racines de P.

Soit $\beta = \alpha^{p^k}$ une racine de P. Montrons que α est une puissance de β. Étant donné que $(\mathbb{F}_p[X]/P)^*$ est un groupe à $p^n - 1$ éléments, le corolaire 2.35 indique que $\alpha^{p^n} = \alpha$. En particulier avec $r = p^{n-k}$ nous avons

$$\beta^r = \alpha^{rp^k} = \alpha^{p^n} = \alpha. \tag{18.214}$$

Par suite toutes les puissances de α sont des puissances de β, ce qui implique que β est générateur du groupe cyclique $(\mathbb{F}_p[X]/P)^*$. □

Lemme 18.59.
Soit p un nombre premier et n, un entier. Un polynôme de degré d irréductible dans $\mathbb{F}_p[X]$ divise $X^{p^n} - X$ si et seulement si d divise n.

Théorème 18.60.
Soient P et Q deux polynômes irréductibles de degré n dans $\mathbb{F}_p[X]$. Alors les quotients $\mathbb{F}_p[X]/P$ et $\mathbb{F}_p[X]/Q$ sont isomorphes en tant que corps.

En guise de démonstration de ce théorème, nous allons démontrer la proposition suivante.

Proposition 18.61.
Si $\mathbb{K}$ et $\mathbb{L}$ sont deux corps à $q = p^n$ éléments, alors ils sont isomorphes.

Démonstration. Soit a un élément primitif de $\mathbb{K}$ et P son polynôme minimal. Nous savons que $\mathbb{K} \simeq \mathbb{F}_p[X]/P$ par le théorème de l'élément primitif 18.55. L'élément a est en particulier une racine de $X^q - X$. Par ailleurs P divise $X^q - X$ par le lemme 18.59.

Nous avons aussi

$$X^q - X = \prod_{b \in \mathbb{L}} (X - b) \tag{18.215}$$

par la proposition 18.35. Étant donné que P divise $X^q - X$, un des éléments de $\mathbb{L}$ annule P. Soit $b \in \mathbb{L}$ tel que $P(b) = 0$. Soit Q le polynôme minimal de b. Par définition nous avons que Q divise P, mais P étant irréductible et unitaire nous avons immédiatement $P = Q$. En particulier nous avons

$$\mathbb{F}_p[X]/P \simeq \mathbb{F}_p[X]/Q \simeq \mathbb{K}. \tag{18.216}$$

Nous montrons maintenant que $\mathbb{F}_p[X]/Q \simeq \mathbb{L}$ par l'application

$$\begin{aligned} \phi \colon \mathbb{F}_p[X]/Q &\to \mathbb{L} \\ \bar{X} &\mapsto b \end{aligned} \tag{18.217}$$

qui se prolonge en $R(\bar{X}) \mapsto R(b)$ pour tout $R \in \mathbb{F}_p[X]$. Cette application est bien définie parce que $Q(b) = 0$. Elle est injective parce que $R(b) = 0$ ne peut pas avoir lieu avec $R \in \mathbb{F}_p[X]/Q$ parce que Q est le polynôme minimal de b. La surjectivité vient alors du fait que les deux corps ont le même nombre d'éléments. □

18.5.7 Construction de $\mathbb{F}_q$

Soit p un nombre premier et $n \in \mathbb{N}$. Nous souhaitons construire un corps à $q = p^n$ éléments. Nous savons déjà que ce corps est unique (théorème 18.32) et que nous le notons $\mathbb{F}_q$. Le théorème 18.55 nous incite à l'écrire sous la forme

$$\mathbb{F}_q = \mathbb{F}_p[X]/(P) \tag{18.218}$$

pour un certain polynôme irréductible $P \in \mathbb{F}_p[X]$.

18.5.7.1 La version du faignant

Nous pouvons construire le corps à p^n éléments en prenant le quotient de $\mathbb{F}_p[X]$ par n'importe quel polynôme irréductible de degré n. Le résultat est le suivant.

Proposition 18.62.
Soit P une polynôme unitaire irréductible dans $\mathbb{F}_p[X]$. Nous posons $\mathbb{K} = \mathbb{F}_p[X]/(P)$. Alors

(1) $\mathbb{K}$ est un corps à q éléments.

(2) $\alpha = \bar{X}$ est une racine de P dans $\mathbb{K}$.

(3) $\mathbb{K} = \mathbb{F}_p[\alpha]$.

Démonstration. (1) En vertu du corolaire 6.38, $\mathbb{K}$ est un corps. Il est aussi un espace vectoriel de dimension n sur $\mathbb{F}_p$, et contient donc $p^n = q$ éléments.

(2) Nous avons $P(\bar{X}) = 0$ par construction de $\mathbb{K} = \mathbb{F}_p[X]/(P)$.

(3) En tant que quotient de $\mathbb{F}_p[X]$, les éléments de $\mathbb{K}$ sont des polynômes en $\bar{X}$.

□

18.5.7.2 La version plus élaborée

Construire $\mathbb{F}_q$ comme quotient de $\mathbb{F}_p[X]$ par un polynôme irréductible quelconque ne donne pas d'informations sur les générateurs de $\mathbb{F}_q^*$, et en particulier il n'est pas toujours vrai que $\bar{X}$ est générateur.

Exemple 18.63.
Construisons $\mathbb{F}_4$. Le polynôme $X^2 + X + 1$ est irréductible dans $\mathbb{F}_2$ parce qu'il n'a pas de racines (c'est vite vu : dans $\mathbb{F}_2$ il n'y a que deux candidats). Donc $\mathbb{F}_4 = \mathbb{F}_2[X]/(X^2 + X + 1)$. △

Remarque 18.64.
Le corps $\mathbb{F}_2$ n'est pas un sous-corps de $\mathbb{C}$ parce que leurs caractéristiques ne sont pas les mêmes. Une conséquence est que les racines de polynômes peuvent être très différentes. Par exemple le polynôme $X^2 + 1$ accepte $x = 1$ comme racine dans $\mathbb{F}_2$ tandis qu'il a pour racines $\pm i$ dans $\mathbb{C}$.

En changeant de corps, les racines peuvent donc complètement changer. Ce n'est pas juste qu'il y a des racines dans l'un et pas dans l'autre.

Exemple 18.65.
Cherchons à construire $\mathbb{F}_{16}$ comme quotient de $\mathbb{F}_2$ par un polynôme de degré 4.

```
----------------------------------------------------------------------
| Sage Version 4.7.1, Release Date: 2011-08-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: x=polygen(GF(2))
sage: -x-1
x + 1
sage: Q=x**15-1
sage: Q.factor()
(x + 1) * (x^2 + x + 1) * (x^4 + x + 1) * (x^4 + x^3 + 1)
            * (x^4 + x^3 + x^2 + x + 1)
```

Les polynômes candidats a avoir des racines génératrices sont donc au nombre de 3 :

$$P_1 = X^4 + X + 1 \tag{18.219a}$$
$$P_2 = X^4 + X^3 + 1 \tag{18.219b}$$
$$P_3 = X^4 + X^3 + X^2 + X + 1. \tag{18.219c}$$

Dans le quotient $\mathbb{F}_2[X]/P_3$, l'élément $\bar{X}$ n'est pas générateur. En effet nous avons $X^4 = X^3 + X^2 + X + 1$ et par conséquent les puissances successives de X sont

$$X \tag{18.220a}$$
$$X^2 \tag{18.220b}$$
$$X^3 \tag{18.220c}$$
$$X^4 = X^3 + X^2 + X + 1 \tag{18.220d}$$
$$1. \tag{18.220e}$$

La classe de X dans $\mathbb{F}_2[X]/P_3$ n'est donc pas génératrice du groupe $(\mathbb{F}_2[X]/P_3)^*$.

Le polynôme $P_1 = X^4 + X + 1$ par contre est primitif parce que les puissances de X dans

$\mathbb{F}_2[X]/P_1$ sont

$$X \tag{18.221a}$$
$$X^2 \tag{18.221b}$$
$$X^3 \tag{18.221c}$$
$$X+1 \tag{18.221d}$$
$$X^2+X \tag{18.221e}$$
$$X^3+X^2 \tag{18.221f}$$
$$X+1+X^3 \tag{18.221g}$$
$$X^2+1 \tag{18.221h}$$
$$X^3+X \tag{18.221i}$$
$$X+1+X^2 \tag{18.221j}$$
$$X^2+X+X^3 \tag{18.221k}$$
$$X^3+X^2+X+1 \tag{18.221l}$$
$$1+X^2+X^2 \tag{18.221m}$$
$$1+X^3 \tag{18.221n}$$
$$1 \tag{18.221o}$$

Cela fait 15 puissances distinctes, ce qui prouve que P_1 est primitif. Nous verrons plus loin comment alléger un peu la vérification de la primitivité de P_1. △

Proposition 18.66 ([1]).
Soit P un polynôme irréductible unitaire primitif dans $\mathbb{F}_p[X]$. Nous considérons $\mathbb{K}=\mathbb{F}_p[X]/P$ et $\alpha=\bar{X}\in\mathbb{K}$. Alors

(1) Les racines de P sont $\{\alpha,\alpha^p,\dots,\alpha^{p^{n-1}}\}$ et $\alpha^q=\alpha$.

(2) P est le polynôme minimal de α.

(3) P est scindé dans $\mathbb{K}$.

(4) P divise X^q-X dans $\mathbb{K}$.

(5) La famille $\{1,\alpha,\alpha^2,\dots,\alpha^{n-1}\}$ est une base de $\mathbb{K}$ en tant qu'espace vectoriel sur $\mathbb{F}_p$.

(6) En tant qu'ensemble,
$$\mathbb{F}_q=\{0,\alpha,\alpha^2,\alpha^3,\dots,\alpha^{q-1}\}, \tag{18.222}$$
et les α^k sont distincts pour $k=1,\dots,q-1$.

Démonstration. La plupart des assertions sont des corolaires ou des paraphrases de résultats contenus dans les propositions précédentes.

(1) L'assertion à propos des racines de P est contenue dans le lemme 18.58. D'autre part le groupe $(\mathbb{F}_p[X]/P)^*$ est cyclique d'ordre $q-1$. Par conséquent le corolaire 2.35 indique que $\alpha^{q-1}=1$ et donc $\alpha^q=\alpha$.

(2) Soit $\tilde{P}$ un polynôme annulateur de α. Nous voyons que si β est racine de $\tilde{P}$ alors β^p est également racine de $\tilde{P}$ en utilisant les techniques habituelles. Par conséquent toutes les racines de P sont racines de $\tilde{P}$, ce qui implique que $\tilde{P}$ est de degré au moins égal à celui de P.

(3) Possédant n racines distinctes dans $\mathbb{K}$, le polynôme P est scindé.

(4) D'après le lemme 18.35 un polynôme irréductible de degré n divise le polynôme $X^{p^n}-X$. Une autre façon de montrer ce point est de remarquer que le polynôme P est scindé et que toutes ses racines sont également racines de X^q-X.

(5) Une combinaison linéaire nulle entre les éléments de $\{1,\alpha,\alpha^2,\dots,\alpha^{n-1}\}$ serait un polynôme annulateur de degré $n-1$ de α. Cet ensemble est donc libre. Par ailleurs un ensemble libre de n éléments dans un espace vectoriel de dimension n est générateur.

(6) Si $\alpha^l = \alpha^k$ avec $k < l$ et $k, l \leqslant q$ alors nous avons $\alpha^r = 1$ avec $r = l - k < q$, ce qui contredirait la primitivité de P. Les éléments $0, \alpha, \ldots, \alpha^{q-1}$ étant distincts et au nombre de q, ils forment tout l'ensemble $\mathbb{F}_q$.

□

18.5.8 Exemple : étude de $\mathbb{F}_{16}$

Dans cette sous section nous voulons construire $\mathbb{F}_{16}$. Nous considérons donc $p = 2$ et $n = 4$. Des polynôme irréductibles de degré 4 dans $\mathbb{F}_2[X]$ ne sont pas très difficiles à trouver. Par exemple $X^4 + X^3 + X^2 + X + 1$, plus généralement un polynôme contenant un nombre impair de termes non nuls dont le terme indépendant.

Les polynômes primitifs par contre doivent être trouvés parmi les diviseurs irréductibles de $X^{15} - 1$. Montrons que

$$P = X^4 + X^3 + 1 \tag{18.223}$$

est primitif. Nous posons $\omega = \bar{X} \in \mathbb{F}_2[X]/P$. L'ordre de ω dans le groupe $(\mathbb{F}_2[X]/P)^*$ doit être un diviseur de 15 et donc seulement 1, 3, 5 ou 15. Le fait que l'ordre ne soit ni 1 ni 3 est trivial parce que le degré de P est 4. Montrons que l'ordre de ω n'est pas 5 non plus :

$$\omega^5 = \omega^4\omega = (\omega^3 + 1)\omega = \omega^4 + \omega = \omega^3 + \omega + 1 \neq 1. \tag{18.224}$$

Dans ce calcul nous avons abondamment utilisé le fait que $-1 = 1$.

À partir de maintenant nous posons $\mathbb{K} = \mathbb{F}_2[X]/P$. Les racines de P sont $\omega, \omega^2, \omega^4$ et ω^8. En effet si β est une racine de P, alors β^2 est une racine en vertu de

$$P(\beta^2) = (\beta^2)^4 + (\beta^2)^3 + 1 = (\beta^4)^2 + (\beta^3)^2 + 1^2 = (\beta^4 + \beta^3 + 1)^2 = 0. \tag{18.225}$$

Ici nous avons implicitement utilisé le lemme 18.53. D'autre part P ne peut pas avoir plus de 4 racines.

Proposition 18.67.
L'ensemble $\{\omega, \omega^2, \omega^4, \omega^8\}$ est une base de $\mathbb{F}_{16}$ sur $\mathbb{F}_2$.

Démonstration. Nous savons que $\{1, \omega, \omega^2, \omega^3\}$ est une base. En effet cet ensemble est libre (sinon ω aurait un polynôme annulateur de degré 3) et générateur parce que l'espace engendré par 4 vecteurs indépendants sur $\mathbb{F}_2$ contient $2^4 = 16$ éléments.

Nous posons $e_0 = 1$, $e_1 = \omega$, $e_2 = \omega^2$, $e_3 = \omega^3$ et $f_1 = \omega$, $f_2 = \omega^2$, $f_3 = \omega^4$, $f_4 = \omega^8$. En utilisant le calcul modulo $\omega^4 + \omega^3 + 1 = 0$ et $2 = 0$ nous trouvons

$$f_1 = \omega \tag{18.226a}$$
$$f_2 = \omega^2 \tag{18.226b}$$
$$f_3 = \omega^3 + 1 \tag{18.226c}$$
$$f_4 = \omega^3 + \omega^2 + \omega. \tag{18.226d}$$

Ensuite nous montrons que les vecteurs e_i peuvent être construits comme combinaisons linéaires des vecteurs f_j :

$$f_1 + f_2 + f_3 + f_4 = e_0 \tag{18.227a}$$
$$f_1 = e_1 \tag{18.227b}$$
$$f_2 = e_2 \tag{18.227c}$$
$$f_1 + f_2 + f_4 = e_3. \tag{18.227d}$$

Les quatre vecteurs f_j forment donc bien un base parce qu'ils sont générateurs d'un espace de dimension 4. □

Exemple 18.68. (1) Résoudre dans $\mathbb{F}_{16}$ l'équation $x^5 = a$ en discutant éventuellement en fonction de la valeur de a.

(2) Montrer qu'il existe quatre éléments $\gamma \in \mathbb{F}_{16}$ tels que pour chacun d'eux l'ensemble $B_\gamma = \{\gamma, \gamma^2, \gamma^4, \gamma^8\}$ est une base de $\mathbb{F}_{16}$ sur $\mathbb{F}_2$ telle que le produit de deux éléments de B_γ est soit un élement de B_γ soit 1.

C'est parti !

(1) Si $a = 0$, alors $x = 0$ est la seule solution. Si $a \neq 0$ alors a est une puissance de ω ; nous posons $a = \omega^l$. Nous cherchons x sous la forme $x = \omega^k$. L'équation à résoudre pour k est

$$\omega^{5k} = \omega^l \tag{18.228}$$

où l est donné. Cette équation revient à

$$5k = l \mod 15. \tag{18.229}$$

Si l n'est pas un multiple de 5, alors il n'y a pas de solutions. Il n'y a des solutions uniquement pour $l = 0, 5, 10$ et elles sont :

$$k = \begin{cases} 3, 6, 9, 12 & \text{si } l{=}0 \\ 1 & \text{si } l{=}5 \\ 2 & \text{si } l{=}10 \end{cases} \tag{18.230}$$

(2) Nous cherchons γ sous la forme $\gamma = \omega^k$. Parmi les nombreuses contraintes liées à l'énoncé nous devons avoir

$$\gamma^5 = 1, \gamma, \gamma^2, \gamma^4, \gamma^8. \tag{18.231}$$

Les possibilités $\gamma^5 = \gamma, \gamma^2, \gamma^4, \gamma^5$ ne sont pas bonnes parce qu'elles impliqueraient que B_γ n'est pas une base. Reste à explorer $\gamma^5 = 1$.

Étant donné le premier point nous restons avec les possibilités

$$\gamma = 1, \omega^3, \omega^6, \omega^9, \omega^{12}. \tag{18.232}$$

Évidemment $\gamma = 1$ ne produit pas une base. Avec $\gamma = \omega^3$ nous trouvons

$$B_\gamma = \{\omega^3, \omega^6, \omega^{12}, \omega^{24}\} = \{\omega^3, \omega^6, \omega^{12}, \omega^9\} \tag{18.233}$$

où nous avons utilisé le fait que $\omega^k = \omega^{k \mod 15}$. En utilisant le fait que $\omega^4 = \omega^3 + 1$ nous trouvons

$$\omega^5 = \omega^3 + \omega + 1 \tag{18.234a}$$
$$\omega^6 = \omega^3 + \omega^2 + \omega + 1 \tag{18.234b}$$
$$\omega^9 = \omega^2 + 1 \tag{18.234c}$$
$$\omega^{12} = \omega + 1. \tag{18.234d}$$

L'ensemble B_γ est alors formé des éléments

$$f_1 = \omega^3 \tag{18.235a}$$
$$f_2 = \omega^3 + \omega^2 + \omega + 1 \tag{18.235b}$$
$$f_3 = \omega + 1 \tag{18.235c}$$
$$f_4 = \omega^2 + 1. \tag{18.235d}$$

Il est assez simple de vérifier que cela est une base en remarquant que $f_1 + f_2 + f_2 + f_4 = 1$. Les possibilités $\gamma = \omega^6, \omega^9, \omega^{12}$ produisent les mêmes ensemble B_γ.

△

18.5.9 Polynômes irréductibles sur $\mathbb{F}_q$

Définition 18.69.
*La **fonction de Möbius** est la fonction $\mu\colon \mathbb{N}^* \to \{-1,0,1\}$ définie par*

$$\mu(n) = \begin{cases} 0 & \text{si } n \text{ est divisible par un carré différent de } 1, \\ 1 & \text{si } n \text{ est le produit d'un nombre pair de nombres premiers distincts,} \\ -1 & \text{si } n \text{ est le produit d'un nombre impair de nombres premiers distincts,} \end{cases} \tag{18.236}$$

Proposition 18.70 ([380]).
Si m et n sont strictement positifs et premiers entre eux, alors

$$\mu(mn) = \mu(m)\mu(n). \tag{18.237}$$

De plus nous avons

$$\sum_{d\,|\,n} \mu(d) = \begin{cases} 1 & \text{si } n = 1 \\ 0 & \text{si } n > 1 \end{cases} \tag{18.238}$$

Proposition 18.71 (Formule d'inversion de Möbius[380]).
Soient $f, g\colon \mathbb{N} \to \mathbb{C}$ telles que pour tout $n \geqslant 1$,

$$g(n) = \sum_{d\,|\,n} f(d). \tag{18.239}$$

Alors

$$f(n) = \sum_{d\,|\,n} \mu\left(\frac{n}{d}\right) g(d) \tag{18.240}$$

où μ est la fonction de Möbius pour tout $n \geqslant 1$.

Lemme 18.72 ([381]).
Soient $P, Q \in \mathbb{K}[X]$ ayant une racine commune dans une extension $\mathbb{L}$ de $\mathbb{K}$. Si P est irréductible, alors $P \mid Q$.

Démonstration. Si P ne divise pas Q, alors P et Q sont premiers entre eux parce que dans la décomposition en irréductibles de Q, il n'y a pas de P tandis que dans celle de P, il n'y a que P. Par conséquent, il existe $a, b \in \mathbb{K} \subset \mathbb{L}$ tels que[22] $aP + bQ = 1$. Cette dernière égalité est encore valable dans $\mathbb{L}$ et donc rend impossible l'existence d'une racine commune. □

Proposition 18.73 ([381, 54]).
Soit p un nombre premier, $n \geqslant 1$ et $r \in \mathbb{N}^*$. Nous notons $q = p^r$, $A(n,q)$, l'ensemble des polynômes unitaires irréductibles de degré n sur $\mathbb{F}_q$. Nous notons aussi $I(n,q) = \operatorname{Card}\big(A(n,q)\big)$. Alors :

(1) Le polynôme $X^{q^n} - X$ se décompose en irréductibles de la façon suivante :

$$X^{q^n} - X = \prod_{d\,|\,n} \prod_{p \in A(d,q)} P. \tag{18.241}$$

(2) Le nombre d'irréductibles est donné par

$$I(n,q) = \frac{1}{n} \sum_{d\,|\,n} \mu\left(\frac{n}{d}\right) q^d \tag{18.242}$$

où μ est la fonction de Möbius (définition 18.69).

(3) Nous avons l'équivalence de suite

$$I(n,q) \sim_{n\to\infty} \frac{q^n}{n}. \tag{18.243}$$

22. Théorème de Bézout, 6.41.

Démonstration. (1) Soit un diviseur d de n et $P \in A(d,q)$. Montrons que P divise $X^{q^n} - X$. Nous considérons la corps $\mathbb{K} = \mathbb{F}_q[X]/(P)$, qui est une extension de degré $\deg(P)$ de $\mathbb{F}_q$ parce qu'il s'agit des polynômes de degré au maximum $\deg(P)$ a coefficients dans $\mathbb{F}_q$. Ce corps possède donc q^d éléments et est isomorphe à $\mathbb{F}_{q^d}$ par la proposition 18.61. Par construction dans $\mathbb{K}$, l'élément $\alpha = [X]$ (la classe de X dans le quotient par P) est une racine de P. Cet élément est également une racine de $X^{q^d} - X$ parce que tout élément de $\mathbb{F}_{q^d}$ est une racine de ce polynôme. Ce dernier point est la proposition 18.35.

Nous sommes donc dans la situation où P et $X^{q^d} - X$ ont une racine commune dans l'extension $\mathbb{F}_q[X]/(P)$. Nous en déduisons que α est aussi une racine de $X^{q^n} - X$. En effet en utilisant le fait que $\alpha^{q^d} = \alpha$, nous avons

$$\alpha^{q^n} = \alpha^{q^{kd}} = \alpha^{q^d q^{(k-1)d}} = \left(\alpha^{q^d}\right)^{q^{(k-1)d}} = \alpha^{q^{(k-1)d}}, \tag{18.244}$$

donc par récurrence, on a encore $\alpha^{q^n} = \alpha$, et α est racine de $X^{q^n} - X$. Vu que P est irréductible, le lemme 18.72 nous indique que P divise $X^{q^n} - X$. Nous en déduisons que P divise $X^{q^n} - X$.

Étant donné que tous les éléments de $A(d,q)$ divisent $X^{q^n} - X$ et sont irréductibles, leur produit divise encore $X^{q^n} - X$:

$$\prod_{d\,|\,n} \prod_{P \in A(d,q)} P \mid X^{q^n} - X. \tag{18.245}$$

Nous devons à présent montrer que tous les facteurs irréductibles de $X^{q^n} - X$ sont dans un $A(d,q)$ avec $d \mid n$. Soit donc P un facteur irréductible de $X^{q^n} - X$ de degré $d \geqslant 1$. Nous posons encore $\mathbb{K} = \mathbb{F}_q[X]/(P)$ et nous utilisons la propriété de multiplication sur les degrés (proposition 6.56) :

$$[\mathbb{F}_{q^n} : \mathbb{K}][\mathbb{K} : \mathbb{F}_q] = [\mathbb{F}_{q^n} : \mathbb{F}_q] = n, \tag{18.246}$$

donc $[\mathbb{K} : \mathbb{F}_q]$, qui vaut $\deg(P)$ est un diviseur de n.

Étant donné que $X^{q^n} - X$ n'a que des racines simples sur $\mathbb{F}_{q^n}$ (à nouveau la proposition 18.35), dans sa décomposition en irréductibles sur $\mathbb{F}_q$, il n'a pas de facteurs carrés ; il n'a donc qu'une fois chacun des $P \in A(d,q)$ avec $d \mid n$. Autrement dit, tous les facteurs irréductibles de $X^{q^n} - X$ sont dans le produit $\prod_{d\,|\,n} \prod_{P \in A(d,q)} P$ et donc $X^{q^n} - X$ divise ce gros produit :

$$X^{q^n} - X \mid \prod_{d\,|\,n} \prod_{P \in A(d,q)} P. \tag{18.247}$$

Ayant déjà obtenu la divisibilité inverse et les polynômes étant unitaires, nous avons égalité.

(2) Nous passons au degré dans l'expression que nous venons de démontrer :

$$q^n = \sum_{d\,|\,n} d \operatorname{Card}\big(A(d,q)\big) = \sum_{d\,|\,n} dI(d,q). \tag{18.248}$$

Nous pouvons utiliser la formule d'inversion de Möbius (proposition 18.71) pour les fonctions $g(n) = q^n$ et $f(n) = dI(n,q)$. Nous écrivons alors

$$f(n) = \sum_{d\,|\,n} \mu\left(\frac{n}{d}\right) q^d, \tag{18.249}$$

ou encore

$$I(n,q) = \frac{1}{n} \sum_{d\,|\,n} \mu\left(\frac{n}{d}\right) q^d, \tag{18.250}$$

ce qu'il fallait.

(3) Nous posons

$$r_n = \sum_{\substack{d|n \\ d<n}} \mu\left(\frac{n}{d}\right) q^d, \tag{18.251}$$

mais sachant que les diviseurs de n, outre n lui-même, sont tous plus petit ou égal à $n/2$ et qu'en valeur absolue, la fonction de Möbius est toujours plus petite ou égale à [23] 1,

$$|r_n| \leqslant \sum_{d=1}^{\lfloor n/2 \rfloor} q^d = \frac{q - q^{\lfloor n/2 \rfloor}}{1-q} = q\frac{q^{\lfloor n/2 \rfloor} - 1}{q-1} \leqslant \frac{q^{\lfloor n/2 \rfloor + 1}}{q-1}. \tag{18.252}$$

D'autre part en reprenant la formule déjà prouvée,

$$I(n,q) = \frac{1}{n}\sum_{d|n} \mu\left(\frac{n}{d}\right) q^d = \frac{1}{n}\left(r_n + \mu\left(\frac{n}{n}\right) q^n\right) = \frac{r_n + q^n}{n}. \tag{18.253}$$

Au numérateur, le plus haut degré en n est q^n parce que r_n est en $q^{\lfloor n/2 \rfloor}$. Donc nous avons bien l'équivalence de suite pour $n \to \infty$:

$$\frac{q^n + r_n}{n} \sim_{n\to\infty} \frac{q^n}{n}. \tag{18.254}$$

□

18.5.10 Matrices

Proposition 18.74.
Nous avons

$$|\mathrm{GL}(n, \mathbb{F}_p)| = (p^n - 1)(p^n - p)\dots(p^n - p^{n-1}). \tag{18.255}$$

Démonstration. Par construction il existe une bijection entre $\mathrm{GL}(n, \mathbb{F}_p)$ et l'ensemble des bases de $\mathbb{F}_p^n$. Nous devons donc seulement compter le nombre de bases. Pour le premier vecteur de base nous avons le choix entre les $p^n - 1$ éléments non nuls de $\mathbb{F}_p^n$. Pour le second nous avons le choix entre $p^n - p$ éléments, et ainsi de suite. □

Lemme 18.75.
Soit $\mathbb{K}$ un corps fini autre que $\mathbb{F}_2$ [24], soit un groupe abélien M et un morphisme $\varphi\colon \mathrm{GL}(n, \mathbb{K}) \to M$. Alors il existe un unique morphisme $\delta\colon \mathbb{K}^* \to M$ tel que $\varphi = \delta \circ \det$.

Démonstration. D'abord le groupe dérivé de $\mathrm{GL}(n, \mathbb{K})$ est $\mathrm{SL}(n, \mathbb{K})$ parce que les éléments de $D\big(\mathrm{GL}(n, \mathbb{K})\big)$ sont de la forme $ghg^{-1}h^{-1}$ dont le déterminant est 1.

De plus le groupe $\mathrm{SL}(n, \mathbb{K})$ est normal dans $\mathrm{GL}(n, \mathbb{K})$. Par conséquent $\mathrm{GL}(n, \mathbb{K})/\mathrm{SL}(n, \mathbb{K})$ est un groupe et nous pouvons définir l'application relevée

$$\tilde{\varphi}\colon \frac{\mathrm{GL}(n, \mathbb{K})}{\mathrm{SL}(n, \mathbb{K})} \to M \tag{18.256}$$

vérifiant $\varphi = \tilde{\varphi} \circ \pi$ où π est la projection.

Nous pouvons faire la même chose avec l'application

$$\det\colon \mathrm{GL}(n, \mathbb{K}) \to \mathbb{K}^* \tag{18.257}$$

qui est un morphisme de groupes dont le noyau est $\mathrm{SL}(n, \mathbb{K})$. Cela nous donne une application

$$\tilde{\det}\colon \frac{\mathrm{GL}(n, \mathbb{K})}{\mathrm{SL}(n, \mathbb{K})} \to \mathbb{K}^* \tag{18.258}$$

23. Dans [54], ma dernière inégalité arrive comme une égalité.
24. Je ne comprends pas très bien à quel moment joue cette hypothèse.

telle que $\det = \tilde{\det} \circ \pi$. Cette application $\tilde{\det}$ est un isomorphisme. En effet elle est surjective parce que le déterminant l'est et elle est injective parce que son noyau est précisément ce par quoi on prend le quotient. Par conséquent $\tilde{\det}$ possède un inverse et nous pouvons écrire

$$\varphi = \tilde{\varphi} \circ \tilde{\det}^{-1} \circ \tilde{\det} \circ \pi. \tag{18.259}$$

État donné que $\tilde{\det} \circ \pi = \det$, nous avons alors $\varphi = \delta \circ \det$ avec $\delta = \tilde{\varphi} \circ \tilde{\det}^{-1}$. Ceci conclut la partie existence de la preuve.

En ce qui concerne l'unicité, nous considérons $\delta' \colon \mathbb{K}^* \to M$ telle que $\varphi = \delta' \circ \det$. Pour tout $u \in \mathrm{GL}(n, \mathbb{K})$ nous avons $\delta'(\det(u)) = \varphi(u) = \delta(\det(u))$. L'application det étant surjective depuis $\mathrm{GL}(n, \mathbb{K})$ vers $\mathbb{K}^*$, nous avons $\delta' = \delta$. □

Théorème 18.76.
Soit $p \geqslant 3$ un nombre premier et V, un $\mathbb{F}_p$-espace vectoriel de dimension finie n. Pour tout $u \in \mathrm{GL}(V)$ nous avons

$$\epsilon(u) = \left(\frac{\det(u)}{p}\right). \tag{18.260}$$

Ici ϵ est la signature de u vue comme une permutation des éléments de $\mathbb{F}_p$.

Démonstration. Commençons par prouver que

$$\epsilon \colon \mathrm{GL}(V) \to \{-1, 1\}. \tag{18.261}$$

est un morphisme. Si nous notons $\bar{u} \in S(V)$ l'élément du groupe symétrique correspondant à la matrice $u \in \mathrm{GL}(V)$, alors nous avons $\overline{uv} = \bar{u} \circ \bar{v}$, et la signature étant un homomorphisme (proposition 2.76),

$$\epsilon(uv) = \epsilon(\bar{u} \circ \bar{v}) = \epsilon(\bar{u})\epsilon(\bar{v}). \tag{18.262}$$

Par ailleurs $\{-1, 1\}$ est abélien, donc le lemme 18.75 s'applique et nous pouvons considérer un morphisme $\delta \colon \mathbb{F}_p^* \to \{-1, 1\}$ tel que $\epsilon = \delta \circ \det$.

Nous allons utiliser le lemme 18.43 pour montrer que δ est le symbole de Legendre. Pour cela il nous faudrait trouver un $x \in \mathbb{F}_p^*$ tel que $\delta(x) = -1$. Étant donné que det est surjective, nous cherchons ce x sous la forme $x = \det(u)$. Par conséquent nous aurions

$$\delta(x) = (\delta \circ \det)(u) = \epsilon(u), \tag{18.263}$$

et notre problème revient à trouver une matrice $u \in \mathrm{GL}(V)$ dont la permutation associée soit de signature -1.

Soit $n = \dim V$; en conséquence de la proposition 18.62(3), l'espace $\mathbb{E}_q = \mathbb{F}_{p^n}$ est un $\mathbb{F}_p$-espace vectoriel de dimension n et est donc isomorphe en tant qu'espace vectoriel à V. Étant donné que $\mathbb{F}_q$ est un corps fini, nous savons que $\mathbb{F}_q^*$ est un groupe cyclique à $q-1$ éléments. Soit y, un générateur de $\mathbb{F}_q^*$ et l'application

$$\begin{aligned} \beta \colon \mathbb{F}_q &\to \mathbb{F}_q \\ x &\mapsto yx. \end{aligned} \tag{18.264}$$

Cela est manifestement $\mathbb{F}_p$-linéaire (ici y et x sont des classes de polynômes et $\mathbb{F}_p$ est le corps des coefficients). L'application β fixe zéro et à part zéro, agit comme le cycle

$$(1, y, y^2, \ldots, y^{q-2}). \tag{18.265}$$

Nous savons qu'un cycle de longueur n est de signature $(-1)^{n+1}$. Ici le cycle est de longueur $q-1$ qui est pair (parce que $p \geqslant 3$) et par conséquent, l'application β est de signature -1. □

18.6 Constructions à la règle et au compas

Définition 18.77 ([382]).
Soit E une partie de $\mathbb{R}^2$. Un point de $\mathbb{R}^2$ est constructible en une étape à partir de E s'il est une point de E ou une intersection de deux objets parmi

— *les droites passant par deux points distincts de E ;*

— *les cercles centrés en un point de E et dont le rayon est la distance entre deux points de E.*

Nous notons $C_1(E)$ l'ensemble des points constructibles en une étape à partir de E.

*Les points constructibles en n étapes à partir de E sont définis par récurrence : $C_{n+1}(E) = C_1\big(C_n(E)\big)$. Enfin un point de $\mathbb{R}^2$ est **constructible** à partir de E s'il appartient à*

$$C(E) = \bigcup_{n=1}^{\infty} C_n(E). \tag{18.266}$$

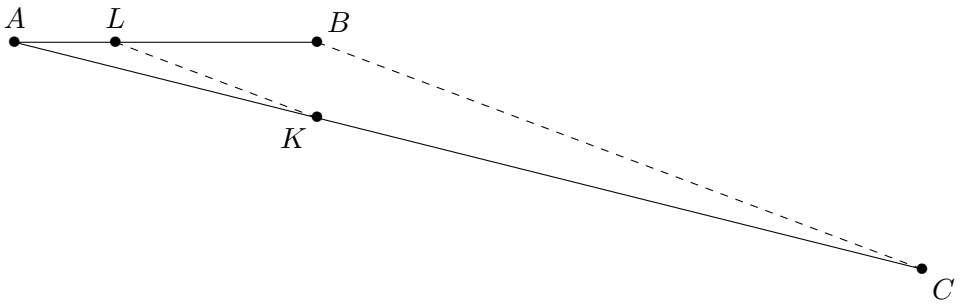
*Un réel est **constructible** s'il est l'abscisse d'un point constructible.*

Pour toute la suite nous allons considérer les points et réels constructibles à partir de l'ensemble $E = \{(0,0),(0,1)\}$.

18.6.1 Quelques constructions

Proposition 18.78 ([383]).
Les nombres rationnels sont tous constructibles.

Démonstration. Si le réel r est constructible, alors kr est également constructible pour tout $k \in \mathbb{Z}$. Nous devons donc seulement pouvoir construire le nombre $1/n$ pour tout $n \in \mathbb{N}^*$.

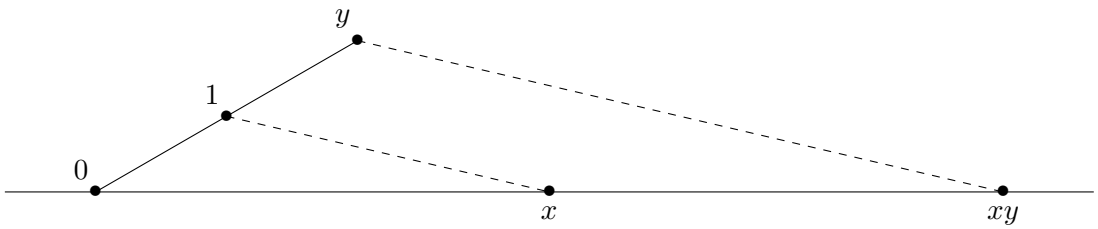


La méthode pour construire le nombre $1/n$ est la suivante. Soit $[AB]$ un segment de longueur 1 (par exemple $A = (0,0)$ et $B = (1,0)$) et un point C tel que $[AC]$ ait une longueur n. Nous plaçons sur $[AC]$ le point K situé à une distance 1 de A en pointant le compas en A et en traçant le cercle de rayon $[AB]$.

La droite passant par K et parallèle à (BC) coupe $[AB]$ en un point L. Maintenant $[AL]$ a une longueur $1/n$ par le théorème de Thalès.

□

Exemple 18.79 (Multiplication à la règle et au compas[382]).
Soient x et y deux nombres constructibles. Montrons qu'il est possible de construire le nombre xy. La construction est la suivante :



— On trace deux droites sécantes en A.
— Sur la première nous plaçons le point Y à distance y de A et le point P à distance 1 de A.
— Sur la seconde on place le point X à distance x de A.
— On trace la droite (PX)
— Puis la parallèle à (PX) passant par Y.
— Le point d'intersection entre cette dernière droite et (AX) est le point B.

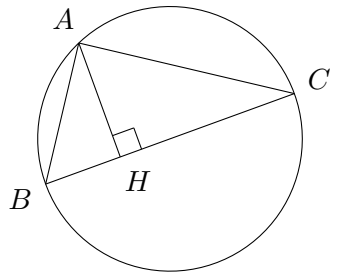
La longueur AB est égale à xy. △

Exemple 18.80 (Racine carrée à la règle et au compas[382])**.**
Nous supposons que le nombre x est constructible, et nous voulons une construction qui donne un segment de longueur $\sqrt{x}$. Nous traçons un segment $[BC]$ dont la longueur correspond à la plus grande des valeurs entre x et 1, puis le cercle de diamètre BC, ensuite le point H sur $[BC]$ tel que BH corresponde à la plus petite des valeurs entre x et 1, enfin la perpendiculaire à (BC) menée par H, qui rencontre le cercle en un point A. D'après le théorème de Thalès sur le cercle, le triangle ABC est rectangle en A.

Les triangles ABC et ABH sont donc semblables parce qu'ils sont rectangles avec un angle (autre que l'angle droit) égal. Nous avons donc proportionnalité des longueurs des côtés :

$$\frac{AB}{BH} = \frac{BC}{AB}, \tag{18.267}$$

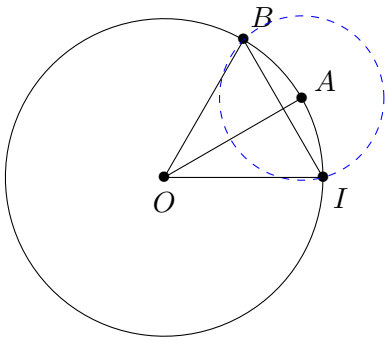
ce qui donne $AB^2 = BC \times BH = x$ (BC et BH valent respectivement 1 et x ou le contraire).



△

Exemple 18.81 (Duplication d'un angle)**.**
Si un angle α est constructible, nous allons construire les angles 2α, 3α, etc. Pour cela nous considérons un cercle de centre O et les points A et I sur le cercle tels que $\widehat{AOI} = \alpha$. Le cercle de centre A et de rayon AI intersecte le cercle de départ en les points I et B.

Le point A est à égale distance de B et I ; le point O également. Donc la droite (OA) est médiatrice du segment $[BI]$. Par conséquent elle est la hauteur du triangle isocèle OBI. L'angle $\widehat{BOA}$ est alors le même que $\widehat{AOI}$; par conséquent $\widehat{BOI} = 2\alpha$.



En traçant le cercle de centre B et de rayon BA, nous continuons et nous construisons 3α.

△

L'exemple suivant qui permet d'additionner des angles repose sur le fait que deux cordes de mêmes longueurs sous-tendent des angles égaux, et est une adaptation simple de la duplication d'angle.

Exemple 18.82 (Addition d'angles)**.**
Quitte à soustraire ou additionner un certain nombre de fois 90°, nous supposons que les deux angles donnés sont entre 0° et 90°.

Soient A, B, I sur un cercle de centre O, et nous notons $\alpha = \widehat{OAI}$, $\beta = \widehat{BOI}$. Nous traçons le cercle de centre B et de rayon AI ; il intersecte le cercle en des points K_1 et K_2. Les angles $\widehat{K_1OB}$ et $\widehat{K_2OB}$ sont tout deux égaux à α.

Les angles $\widehat{K_1OA}$ et $\widehat{K_2OA}$ sont égaux à $\beta - \alpha$ et $\beta + \alpha$. △

18.6.2 Nombres constructibles

Théorème 18.83 (Wantzel[384])**.**
Le réel a est constructible si et seulement s'il existe une suite finie de corps $\mathbb{L}_i$ tels que

(1) $\mathbb{L}_0 = \mathbb{Q}$,
(2) $\mathbb{L}_{i+1}$ *est un extension quadratique*[25] *de* $\mathbb{L}_i$
(3) $a \in \mathbb{L}_n$.

Démonstration. Soit $\mathbb{K}$ un corps de nombres constructibles (par exemple $\mathbb{Q}$) ; nous notons $E_{\mathbb{K}}$ l'ensemble des points de $\mathbb{R}^2$ dont les coordonnées sont dans $\mathbb{K}$. Ce sont des points forcément constructibles.

Intersection de droites Si $A, B \in E_{\mathbb{K}}$ alors la droite (AB) a pour équation $ax + by + c = 0$ avec $a, b, c \in \mathbb{K}$, et le point d'intersection entre deux droites est donné par la solution du système

$$\begin{cases} ax + by + c = 0 & (18.268a) \\ a'x + b'y + c' = 0, & (18.268b) \end{cases}$$

dont les solutions sont encore dans $\mathbb{K}$.

Intersection droite-cerle L'équation d'un cercle est de la forme

$$(x - u)^2 + (y - v)^2 = r^2 \tag{18.269}$$

où $(u, v) \in E_{\mathbb{K}}$ et r est la distance entre deux points de $E_{\mathbb{K}}$; donc $r^2 \in \mathbb{K}$. En développant et en redéfinissant u, v nous voyons que tous les cercles à considérer ont une équation de la forme

$$x^2 + y^2 + ux + vy + t = 0 \tag{18.270}$$

avec $u, v, t \in \mathbb{K}$. Il s'agit de voir où sont les solutions du système

$$\begin{cases} ax + by + c = 0 & (18.271a) \\ x^2 + y^2 + ux + vy + t = 0. & (18.271b) \end{cases}$$

Si $a \neq 0$ alors nous pouvons faire la substitution $x = -(c + by)/a$ et obtenir l'équation suivante pour y :

$$\left(\frac{-c - by}{a}\right)^2 + y^2 + u\left(\frac{-c - by}{a}\right) + vy + t = 0. \tag{18.272}$$

Cela est de la forme $P(y) = 0$ où P est un polynôme du second degré à coefficients dans $\mathbb{K}$. Si $a = 0$ nous substituons y au lieu de x et le résultat est le même.

C'est le moment de relire la proposition 6.87 qui nous assure que si le réel α est une solution de $P(y) = 0$ hors de $\mathbb{K}$, alors l'extension $\mathbb{K}(\alpha)$ est de degré 2 parce que le polynôme minimal de α est de degré 2 :

$$\big[\mathbb{K}[\alpha] : \mathbb{K}\big] = 2. \tag{18.273}$$

De plus $\mathbb{K}[\alpha] = \mathbb{K}(\alpha)$.

25. C'est-à-dire une extension finie de degré 2.

Intersection cercle-cercle Le système

$$\begin{cases} x^2 + y^2 + ux + vy + y = 0 & (18.274a) \\ x^2 + y^2 + ax + by + c = 0 & (18.274b) \end{cases}$$

est équivalent au système (en substituant la seconde équation par la différence entre les deux)

$$\begin{cases} x^2 + y^2 + ux + vy + y = 0 & (18.275a) \\ (u-a)x + (v-b)y + t - c = 0 & (18.275b) \end{cases}$$

qui est à nouveau une intersection entre un cercle et une droite.

Passons à la conclusion. Si α est un nombre constructible, alors il apparaît dans les coordonnées d'un point de $C_m(E_{\mathbb{Q}})$ pour un certain m. Nous supposons (pour la récurrence) que pour chaque réel constructible en $m-1$ étapes possède sa pile d'extension quadratiques en partant de $\mathbb{Q}$. Le nombre α vérifie donc

$$a\alpha^2 + b\alpha + c = 0 \tag{18.276}$$

avec a, b, c trois réels constructibles en $m-1$ étapes. L'hypothèse de récurrence donne donc des piles d'extensions

$$\begin{array}{ccc} \mathbb{L}_0 = \mathbb{Q} & \mathbb{L}'_0 = \mathbb{Q} & \mathbb{L}''_0 = \mathbb{Q} \\ L_{i+1} = \mathbb{L}_i(a_i) & \mathbb{L}'_{i+1} = \mathbb{L}'_i(b_i) & \mathbb{L}''_{i+1} = \mathbb{L}''_i(c_i) \\ a \in \mathbb{L}_n & b \in \mathbb{L}'_{n'} & c \in \mathbb{L}''_{n''}. \end{array} \tag{18.277}$$

Nous considérons donc la suite d'extensions de $\mathbb{Q}$ qui consiste à étendre successivement par les nombres $a_1, \dots, a_n, b_1, \dots, b_{n'}, c_1, \dots, c_{n''}$ tout en excluant les doublons : il est possible que par exemple b_3 soit déjà dans $\mathbb{Q}(a_1, \dots, a_n)$. Cela nous fournit une suite d'extensions

$$\begin{array}{c} \mathbb{M}_0 = \mathbb{Q} \\ \mathbb{M}_{i+1} = \mathbb{M}_i(\alpha_i) \\ a, b, c \in \mathbb{M}_n. \end{array} \tag{18.278}$$

Ici le n n'est pas spécialement le même que celui plus haut. Maintenant, α est dans un extension quadratique de $\mathbb{M}_n$.

Pour la réciproque nous supposons avoir une tour d'extensions quadratiques $\mathbb{L}_0 = \mathbb{Q}$, $\mathbb{L}_i$ et $\alpha \in \mathbb{L}_n$ et nous voulons prouver que α est constructible. Nous y allons par récurrence : si $n = 0$ alors α est rationnel et il est constructible par la proposition 18.78.

Supposons que tous les points à coordonnées dans $\mathbb{L}_i$ sont constructibles. Alors nous allons prouver que les éléments de $\mathbb{L}_{i+1}$ le sont également. Vu que α est solution d'une équation de degré 2 à coefficients dans $\mathbb{L}_i$, nous avons

$$a\alpha^2 + b\alpha + c = 0 \tag{18.279}$$

pour certains $a, b, c \in \mathbb{L}_i$ et donc

$$\alpha = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{18.280}$$

Les exemples 18.79 et 18.80 montrent que les produits et les racines carrées de nombres constructibles sont constructibles. Donc α est constructible.

□

18.6.3 Polygones constructibles

Définition 18.84.
*Un angle α est **constructible** si le nombre $\cos(\alpha)$ est constructible.*

La raison est que si qu'il suffit de prendre la perpendiculaire à l'axe horizontal

Lemme 18.85 ([54]).
Soient m et n, deux nombres premiers entre eux. L'angle $\frac{2\pi}{mn}$ est constructible si et seulement si les angles $\frac{2\pi}{m}$ et $\frac{2\pi}{n}$ sont constructibles.

Démonstration. **Sens direct** Il suffit de pouvoir multiplier un angle par un entier, ce qui est fait dans l'exemple 18.81.

Sens réciproque Le théorème de Bézout 3.13 nous donne $a, b \in \mathbb{Z}$ tels que $an + bm = 1$. Cela donne immédiatement

$$\frac{1}{mn} = \frac{a}{m} + \frac{b}{n}, \tag{18.281}$$

et donc

$$\frac{2\pi}{mn} = a\frac{2\pi}{m} + b\frac{2\pi}{n}, \tag{18.282}$$

ce qui fait que l'angle $2\pi/mn$ est une combinaison entière d'angles constructibles. Il est donc constructible par l'exemple 18.82.

□

Lemme 18.86.
Soit un entier $n \geqslant 3$ ayant

$$n = \prod_{i=1}^{k} p_i^{\alpha_i} \tag{18.283}$$

comme décomposition en facteurs premiers. Les polynôme régulier à n côtés est constructible si et seulement si les angles $\frac{2\pi}{p_i^{\alpha_i}}$ sont constructibles.

Démonstration. Le polygone régulier à n côtés est constructible si et seulement si l'angle $\frac{2\pi}{n}$ est constructible, c'est-à-dire si l'angle

$$\frac{2\pi}{\prod_{i=1}^{k} p_i^{\alpha_i}} \tag{18.284}$$

est constructible. Par récurrence sur le lemme 18.85, cet angle est constructible si et seulement si les angles $\frac{2\pi}{p_i^{\alpha_i}}$ sont constructibles. □

Théorème 18.87 (Gauss-Wantzel[54]).
Soit $\alpha \in \mathbb{N}^*$.

(1) L'angle $\frac{2\pi}{2^\alpha}$ est constructible.

*(2) Si p est premier $p \neq 2$ alors l'angle $\frac{2\pi}{p^\alpha}$ est constructible si et seulement si $\alpha = 1$ et p est un **nombre de Fermat**, c'est-à-dire de la forme $1 + 2^{(2^\beta)}$ pour $\beta \in \mathbb{N}$.*

(3) Le polygone régulier à n côtés est constructible si et seulement si n est le produit d'une puissance de 2 et d'un nombre fini de nombres de Fermat premiers distincts.

Démonstration. Le point (1) est une construction de bissectrice, et le point (3) consistera à remettre en place différents morceaux. Le gros de la preuve est donc consacré à (2).

Sens direct Nous supposons que l'angle $\frac{2\pi}{p^\alpha}$ est constructible ; alors le nombre $\cos\big(2\pi/p^\alpha\big)$ l'est également et le théorème de Wantzel 18.83 nous indique que

$$\Big[\mathbb{Q}\Big(\cos(2\pi/p^\alpha)\Big) : \mathbb{Q}\Big] = 2^m \tag{18.285}$$

pour un certain m. Posons $q = p^\alpha$ et $\omega = e^{2i\pi/q}$. Grâce au corolaire 18.24, nous savons que le polynôme minimal de ω est le polynôme cyclotomique ϕ_q dont le degré est [26]

$$\varphi(q) = \varphi(p^\alpha) = p^{\alpha-1}(p-1). \tag{18.286}$$

26. Voir juste en dessous de la définition 18.17.

Par conséquent

$$\big[\mathbb{Q}(\omega):\mathbb{Q}\big]=p^{\alpha-1}(p-1). \tag{18.287}$$

Mais par ailleurs

$$\omega+\omega^{-1}=2\cos(2\pi/q), \tag{18.288}$$

donc $\cos(2\pi/q)\in\mathbb{Q}(\omega)$. Et en multipliant (18.288) par ω nous trouvons le polynôme annulateur suivant pour ω :

$$\omega^2-2\cos\left(\frac{2\pi}{q}\right)\omega+1=0. \tag{18.289}$$

Cela signifie que ω est de degré 2 dans $\mathbb{Q}\big(\cos(2\pi/q)\big)$, c'est-à-dire

$$\big[\mathbb{Q}(\omega):\mathbb{Q}\big(\cos(2\pi/q)\big)\big]=2. \tag{18.290}$$

En remettant bout à bout et en utilisant la propriété multiplicative des degrés des extensions [27],

$$\underbrace{\big[\mathbb{Q}(\omega):\mathbb{Q}\big]}_{p^{\alpha-1}(p-1)}=\underbrace{\Big[\mathbb{Q}(\omega):\mathbb{Q}\big(\cos(2\pi/p^\alpha)\big)\Big]}_{2}\underbrace{\Big[\mathbb{Q}\big(\cos(2\pi/p^\alpha)\big):\mathbb{Q}\Big]}_{2^m}, \tag{18.291}$$

donc $p^{\alpha-1}(p-1)=2^{m+1}$, mais comme p est premier et impair, $\alpha=1$ et $p=2^{m+1}+1$. Par ailleurs $m+1$ est un entier et nous nous proposons de mettre en facteur la puissance de 2 dans son développement en facteurs premiers : $m+1=\lambda 2^\beta$ avec $\beta\in\mathbb{N}$ et un certain nombre impair $\lambda\in\mathbb{N}^*$. Nous avons :

$$p=2^{\lambda 2^\beta}+1=\left(2^{2^\beta}\right)^\lambda+1, \tag{18.292}$$

mais le lemme 3.198(1) nous indique alors que $1+2^{2^\beta}$ divise $1+\left(2^{2^\beta}\right)^\lambda=p$. Mais vu que p est premier, il ne peut être divisé que par p lui-même et donc λ soit être égal à 1 et nous avons

$$p=1+2^{2^\beta}, \tag{18.293}$$

ce qui signifie que p est un nombre de Fermat premier.

Sens réciproque Nous supposons que p est un nombre premier de la forme $p=1+2^{2^\beta}$ et nous devons prouver que l'angle $\frac{2\pi}{p}$ est constructible. Pour cela nous posons tout de suite $n=2^\beta$ et $\omega=e^{2i\pi/p}$. Comme dans la première partie nous nous souvenons que le polynôme minimal de ω est le polynôme cyclotomique ϕ_p de degré $p-1$; donc

$$[\mathbb{Q}(\omega):\mathbb{Q}]=p-1. \tag{18.294}$$

Un groupe d'automorphismes Nous considérons le groupe $G=\mathrm{Aut}_{\mathbb{Q}}\big(\mathbb{Q}(\omega)\big)$ des automorphismes du corps $\mathbb{Q}(\omega)$ agissant sur $\mathbb{Q}$ comme l'identité. Tous les éléments de $\mathbb{Q}(\omega)$ étant des polynômes en ω (proposition 6.87), un élément $g\in G$ est uniquement déterminé par $g(\omega)$, et de plus $g(0)=0$ ainsi que $\phi_p(\omega)=0$ et que ϕ_p commute avec g, donc

$$\phi_p\big(g(\omega)\big)=g\big(\phi_p(\omega)\big)=g(0)=0, \tag{18.295}$$

ce qui signifie que $g(\omega)$ est une racine du polynôme cyclotomique ϕ_p. Par définition 18.17, les racines sont $\{\omega,\omega^2,\dots,\omega^{p-1}\}$, ce qui signifie que l'action de g consiste à élever ω à une certaine puissance entre 1 et $p-1$. Le groupe G est donc d'ordre $|G|=p-1$ et a pour éléments

$$g_k:\omega\mapsto\phi^k \tag{18.296}$$

27. Proposition 6.56.

avec $k = 1, \dots, p-1$. L'application[28]

$$\begin{aligned} \psi\colon G &\to (\mathbb{Z}/p\mathbb{Z})^* \\ g_k &\mapsto [k]_p. \end{aligned} \tag{18.297}$$

est un morphisme surjectif entre deux groupes finis de même cardinal, donc c'est un isomorphisme. Le corolaire 18.34 nous donne de plus l'isomorphisme $(\mathbb{Z}/p\mathbb{Z})^* \simeq \mathbb{Z}/(p-1)\mathbb{Z}$, ce qui fait que G a un élément d'ordre $p-1$ (parce que $\mathbb{Z}/(p-1)\mathbb{Z}$ en a un). Nous notons g_0 cet élément.

La tour d'extensions À partir de cet élément $g_0 \in G$ nous définissons avec $0 \leqslant i \leqslant n$:

$$\mathbb{K}_i = \{z \in \mathbb{Q}(\omega) \text{ tel que } g_0^{2^i}(z) = z\}. \tag{18.298}$$

Ces ensembles sont bien des corps parce que g_0 est un morphisme : $g_0(zz') = g_0(z)g_0(z')$; on en déduit immédiatement que $g_0^2(zz') = g_0^2(z)g_0^2(z')$.

$\mathbb{K}_n = \mathbb{Q}(\omega)$ Cela est dû au fait que g_0 est par définition d'ordre 2^n, ce qui signifie que $g_0^{2^n}(z) = z$ pour tout $z \in \mathbb{Q}(\omega)$.

$\mathbb{K}_0 = \mathbb{Q}$ Vu que les éléments de G laissent $\mathbb{Q}$ invariant nous avons forcément $\mathbb{Q} \subset \mathbb{K}_0$. Nous nous attelons maintenant à prouver l'inclusion inverse. Nous savons par la proposition 6.87(2) que $\{\omega^i\}_{i=0,\dots,p-2}$ est une base[29] de $\mathbb{Q}(\omega)$. Vu que $g_0\colon \mathbb{Q}(\omega) \to \mathbb{Q}(\omega)$ est un isomorphisme, l'image d'une base est une base :

$$\{g_0(\omega)^i\}_{i=0,\dots,p-2} \tag{18.299}$$

est également une base de $\mathbb{Q}(\omega)$. Soit $z \in \mathbb{K}_0$ et décomposons-le dans cette base[30] :

$$z = \sum_{i=0}^{p-2} \lambda_i g_0(\omega)^i; \tag{18.300}$$

nous appliquons g_0 à cette égalité en tenant compte du fait que $g_0(z) = z$:

$$z = \sum_{i=0}^{p-2} \lambda_i g_0(\omega)^{i+1}. \tag{18.301}$$

En identifiant les coefficients (et en remarquant que le dernier terme dans (18.300) est le premier dans (18.301)) nous voyons que tous les coefficients sont égaux :

$$z = \lambda_0 g_0(\omega + \dots + \omega^{p-1}) = \lambda_0 g_0\big(\phi_p(\omega) - 1\big) = -\lambda_0 \in \mathbb{Q}. \tag{18.302}$$

Dans cette chaine d'égalités nous avons utilisé le fait que $\phi_p(X) = 1 + X + \dots + X^{p-1}$ (corolaire 18.22) et que $\phi_p(\omega) = 0$ par définition.

$\mathbb{K}_i \subset \mathbb{K}_{i+1}$ **strictement** Nous avons une inclusion pour la simple raison que si $z \in \mathbb{K}_i$, alors $g_0^{2^i}(z) = z$. Par conséquent :

$$g_0^{2^{i+1}}(z) = g_0^{2\times 2^i}(z) = (g_0^{2^i})^2(z) = z. \tag{18.303}$$

Afin de voir que l'inclusion est stricte nous montrons que l'élément

$$x = \sum_{k=0}^{2^{n-i}-1} g_0^{k2^{i+1}}(\omega) \tag{18.304}$$

28. Pour rappel, la notation $[k]_p$ est la classe de l'entier k modulo p, qui est un élément de $\mathbb{Z}/p\mathbb{Z}$.
29. Ici [54] parle de $\mathbb{Q}(\omega)/\mathbb{Q}$. Dans ce cas il me semble qu'il faille faire partir les valeurs de i de 0 et non de 1.
30. Il s'agit d'une base en tant qu'espace vectoriel sur $\mathbb{Q}$, donc les λß sont dans $\mathbb{Q}$.

est dans $\mathbb{K}_{i+1}\backslash\mathbb{K}_i$. Dans la base $\{1,\omega,\dots,\omega^{p-1}\}$ l'élément x a une composante ω avec le terme $k=0$, mais si on lui applique $g_0^{2^i}$ nous obtenons

$$g_0^{2^i}(x)=\sum_{k=0}^{2^{n-i-1}-1}g_0^{2^i(1+2k)}(\omega). \tag{18.305}$$

Une composante ω pour cela demanderait d'avoir $2^i(1+2k)=\lambda 2^n$ avec $\lambda\in\mathbb{N}$; cela demanderait

$$k=\lambda 2^{n-i-1}-\frac{1}{2} \tag{18.306}$$

ce qui est impossible. Donc $x\notin\mathbb{K}_i$.

Nous montrons que $x\in\mathbb{K}_{i+1}$ de la même manière :

$$g_0^{2^{i+1}}(x)=\sum_{k=0}^{2^{n-i-1}-1}g^{2^{i+1}(k+1)}(\omega). \tag{18.307}$$

Soit k_0 entre 0 et $2^{n-i-1}-1$; nous voulons trouver un k entre 0 et $2^{n-i-1}-1$ tel que $k_0 2^{i+1}=2^{i+1}(k+1)$; cela est très simple : il suffit de prendre $k=k_0-1$ tant que $k_0\neq 0$. Si $k_0=0$ alors cela correspond au terme ω dans (18.304), et il se trouve dans (18.307) avec $k=2^{n-i-1}-1$. Donc $g_0^{2^{i+1}}(x)=x$ et $x\in\mathbb{K}_{i+1}$.

Les degrés dans la tour Par définition de n et de ω, et par la propriété multiplicative des degrés[31] nous avons :

$$2^n=p-1=\big[\mathbb{Q}(\omega):\mathbb{Q}\big]=\underbrace{\big[\mathbb{Q}(\omega):\mathbb{K}_{n-1}\big]\dots\big[\mathbb{K}_1:\mathbb{Q}\big]}_{n\text{ facteurs}}. \tag{18.308}$$

Un produit de n facteurs entiers tous strictement plus grand que zéro doit valoir 2^n. Ils doivent donc tous valoir 2 et nous avons en particulier

$$\big[\mathbb{K}_{n-1}:\mathbb{Q}\big]=2^{n-1}. \tag{18.309}$$

$\cos(2\pi/p)\in\mathbb{K}_{n-1}$ Soit $f=g_0^{2^n-1}$; vu que l'action de g_0 (comme de tous les éléments de G) est de décaler les puissances de ω, il existe $\lambda\in\mathbb{N}$ tel que $f(\omega)=\omega^\lambda$. Mais $f^2=\mathrm{Id}$ donc

$$\omega=f^2(\omega)=\omega^{2\lambda}, \tag{18.310}$$

ce qui donne $1=\omega^{\lambda^2-1}$ ou encore que p divise λ^2-1 parce que ω est une racine primitive p^{e}de l'unité. Par conséquent $[\lambda^2-1]_p=0$ ce qui donne $[\lambda]_p=\pm[1]_p$, mais comme $f\neq\mathrm{Id}$ nous avons

$$[\lambda]_p=-[1]_p. \tag{18.311}$$

Cela fait $f(\omega)=\omega^{-1}$.

Par ailleurs $\mathbb{Q}(\omega)$ est un corps, donc il contient

$$\cos\left(\frac{2\pi}{p}\right)=\frac{1}{2}(\omega+\omega^{-1}). \tag{18.312}$$

Nous en déduisons que $\cos(2\pi/p)$ est un point fixe de $f=g_0^{2^{-1}}$:

$$f\big(\cos(2\pi/p)\big)=\frac{1}{2}\big(f(\omega)+f(\omega^{-1})\big)=\cos(2\pi/p). \tag{18.313}$$

Être un point fixe de $g_0^{2^{n-1}}$ signifie être un élément de $\mathbb{K}_{n-1}$:

$$\cos\left(\frac{2\pi}{p}\right)\in\mathbb{K}_{n-1}. \tag{18.314}$$

31. Proposition 6.56

Questions de degrés Nous avons alors les (non)inclusions suivantes :

$$\mathbb{Q}\left(\cos(2\pi/p)\right) \subset \mathbb{K}_{n-1} \subsetneq \mathbb{K}_n = \mathbb{Q}(\omega), \tag{18.315}$$

ce qui fait que

$$1 < \big[\mathbb{Q}(\omega) : \mathbb{K}_{n-1}\big] \leqslant \big[\mathbb{Q}(\omega) : \mathbb{Q}\big(\cos(2\pi/p)\big)\big] = 2 \tag{18.316}$$

La première inégalité est le fait que $\mathbb{Q}(\omega)$ n'est pas égal à $\mathbb{K}_{n-1}$. La dernière égalité se démontre de la même manière que (18.290) (ici $\alpha = 1$). Les inégalités de (18.316) sont donc en réalité des égalités :

$$\big[\mathbb{Q}(\omega) : \mathbb{K}_{n-1}\big] = \big[\mathbb{Q}(\omega) : \mathbb{Q}\big(\cos(2\pi/p)\big)\big] = 2. \tag{18.317}$$

Cela, combiné au fait que $\mathbb{Q}\big(\cos(2\pi/p)\big) \subset \mathbb{K}_{n-1}$ donne

$$\mathbb{K}_{n-1} = \mathbb{Q}\big(\cos(2\pi/p)\big). \tag{18.318}$$

Mais nous savons déjà que $\big[\mathbb{K}_{n-1} : \mathbb{Q}\big] = 2^{n-1}$, donc

$$\big[\mathbb{Q}\big(\cos(2\pi/p)\big) : \mathbb{Q}\big] = 2^{n-1} \tag{18.319}$$

et le nombre $\cos(2\pi/p)$ est bien sûr le sommet d'une tour d'extensions quadratiques partant de $\mathbb{Q}$. Il est donc constructible par le théorème de Wantzel 18.83.

Il est maintenant l'heure de conclure en prouvant le point (3). D'abord si n est un produit de nombres premiers de Fermat distincts, alors $n = p_1 \dots p_m$ et l'angle $\frac{2\pi}{n}$ est constructible si et seulement si chacun des angles $\frac{2\pi}{p_k}$ est constructible (lemme 18.86), ce qui est le cas d'après la partie (2). Si n est le produit d'une puissance de 2 avec un produit de nombres premiers de Fermat, le polygone à n côtés est tout autant constructible : il suffit de bissecter les angles.

À l'inverse si le polynôme à n avec n impair côtés est constructible alors $n = \prod_i p_i^{\alpha_i}$, dont l'angle est constructible si et seulement si $\alpha = 1$ et p_i est un nombre premier de Fermat. Et si le polygone à n côtés (n pair) est constructible, alors $n = 2^\lambda \prod_i p_i^{\alpha_i}$. Ce polygone est constructible si et seulement si le polygone à $\prod_i p_i^{\alpha_i}$ est constructible : il suffit de regrouper les côtés par deux.

□

Remarque 18.88.
D'après Wikipédia[385], les seuls nombres de Fermat connus pour être premiers sont

$$F_0 = 3 \tag{18.320a}$$
$$F_1 = 5 \tag{18.320b}$$
$$F_2 = 17 \tag{18.320c}$$
$$F_3 = 257 \tag{18.320d}$$
$$F_4 = 65537. \tag{18.320e}$$

Pour les autres, essentiellement on ne sait pas. Il n'est même pas sûr qu'il y en ait d'autres. Le problème des polygones constructibles n'est donc pas encore tout à fait terminé.

Remarque 18.89.
Les angles $\frac{2\pi}{p^\alpha}$ avec p premier et $\alpha > 1$ ne sont pas constructibles. Il n'est donc pas possible de trisecter l'angle $\frac{2\pi}{3}$. Voilà qui règle un des vieux problèmes de l'antiquité.

Chapitre 19

Intégration sur des variétés

19.1 Variétés

19.1.1 Introduction

Soit $f : S^2 \to \mathbb{R}$ une fonction définie sur la sphère usuelle $S^2 \subset \mathbb{R}^3$. Une question naturelle est d'estimer la régularité de f ; est-elle continue, dérivable, différentiable ? Il n'existe pas de dérivée directionnelle étant donné que le quotient différentiel

$$\frac{f(x+\epsilon u_1, y+\epsilon u_2) - f(x,y)}{\epsilon}$$

n'a pas de sens pour un point $(x+\epsilon u_1, y+\epsilon u_2)$ qui n'est pas –sauf valeurs particulières– dans la surface. Pour la même raison il n'est pas possible de parler de différentiabilité de cette manière. Comment faire, sans devoir étendre le domaine de définition de f à un voisinage de la sphère ? Une solution possible est de parler de la notion de variété.

Une variété est un objet qui ressemble, vu de près, à $\mathbb{R}^m$ pour un certain m. En d'autres termes, on imagine une variété comme un recollement de morceaux de $\mathbb{R}^m$ vivant dans un espace plus grand $\mathbb{R}^n$. Ces morceaux sont appelés des ouverts de carte, et l'application qui exprime la ressemblance à $\mathbb{R}^m$ est l'application de carte.

19.1.2 Définition, carte

Définition 19.1.
*Une **variété** de dimension m de classe C^1 est une partie M de $\mathbb{R}^n$ ($n \geqslant m$) munie d'un ensemble de paires $\{U_\alpha, \varphi_\alpha\}$ où U_α est un ouvert de $\mathbb{R}^n$ et $\varphi_\alpha \colon U_\alpha \to \mathbb{R}^n$ est une application vérifiant*

(1) φ_α est une bijection entre U_α et $\varphi_\alpha(U_\alpha)$;

(2) pour tout $a \in M$, il existe un α tel que $a \in \varphi_\alpha(U_\alpha)$;

(3) pour tout α, β, la partie $\varphi_\alpha^{-1}\big(\varphi_\alpha(U_\alpha) \cap \varphi_\beta(U_\beta)\big)$ est un ouvert de U_α ;

(4) L'application

$$\varphi_\beta^{-1} \circ \varphi_\alpha \colon \varphi_\alpha^{-1}\big(\varphi_\alpha(U_\alpha) \cap \varphi_\beta(U_\beta)\big) \to U_\beta \tag{19.1}$$

est un C^1-difféomorphisme vers son image.

Notez que cette définition n'utilise pas du tout la structure de $\mathbb{R}^n$ dans lequel se trouve M. Seulement la structure différentielle du $\mathbb{R}^m$ depuis lequel les cartes partent. En particulier, nous n'avons pas dit si $\varphi_\alpha \colon U_\alpha \to M$ était de classe C^1 pour une topologie induite de $\mathbb{R}^n$ vers M. Nous pouvons en réalité définir plus généralement une variété en remplaçant « une partie de $\mathbb{R}^n$ » par « un ensemble » sans rien y changer.

La difficulté qui apparaît lorsque nous voulons dire que M est un ensemble quelconque au lieu d'une partie de $\mathbb{R}^n$ est pour la mesure.

Nous verrons plus loin comment les cartes φ_α peuvent transporter la mesure de U_α (celle de Lebesgue dans $\mathbb{R}^m$) vers une partie de $\mathbb{R}^n$ en utilisant le déterminant de $d\varphi_\alpha$. Cela n'est possible

que parce que $\mathbb{R}^n$ a une structure différentielle qui permet de définir $d\varphi_\alpha$. Si M est un ensemble quelconque, cette façon de faire n'est plus possible. Dans ce cas, il faut définir des fibrés tangents et cotangents puis en extraire une « forme volume ». Le choix de la forme volume n'étant pas canonique, il n'y a pas d'intégrale canonique sur une variété générale comme il y en a une sur une variété « dans $\mathbb{R}^n$ ».

Définition 19.2.
Soit une variété M de dimension m. Une ***carte*** *pour M est un couple (U, φ) où U est un ouvert de $\mathbb{R}^m$ et $\varphi\colon U \to M$ est une application*

(1) qui est une bijection entre U et son image,

(2) pour tout α,

$$\varphi_\alpha^{-1} \circ \varphi\colon \varphi^{-1}\big(\varphi(U) \cap \varphi_\alpha(U_\alpha)\big) \to \varphi_\alpha^{-1}\big(\varphi(U) \cap \varphi_\alpha(U_\alpha)\big) \tag{19.2}$$

est un C^1-difféomorphisme.

Proposition 19.3 ([1])**.**
Si (U_1, φ_1) et (U_2, φ_2) sont des cartes pour la variété M, alors en posant $S = \varphi_1(U_1) \cap \varphi_2(U_2)$, l'application

$$\varphi_2^{-1} \circ \varphi_1\colon \varphi_1^{-1}(S) \to \varphi_2^{-1}(S) \tag{19.3}$$

est un C^1-difféomorphisme.

Démonstration. Le fait que $\varphi_2^{-1} \circ \varphi_1$ soit une bijection est dû au fait que nous ayons choisit les espaces de départ et d'arrivée pour que ce soit une bijection.

Soit $a \in \varphi_1(U_1) \cap \varphi_2(U_2)$. Soient α tel que $a \in \varphi_\alpha(U_\alpha)$ et un ouvert A autour de $\varphi_\alpha^{-1}(a)$ tel que $\varphi_\alpha(A) \subset \varphi_1(U_1) \cap \varphi_2(U_2)$. Nous allons montrer que

$$\varphi_2^{-1} \circ \varphi_1\colon \varphi_1\big(\varphi_\alpha(A)\big) \to \varphi_2^{-1}\big(\varphi_\alpha(A)\big) \tag{19.4}$$

est un C^1-difféomorphisme. Nous avons

$$\varphi_2 \circ \varphi_1 = \varphi_2^{-1} \circ \varphi_\alpha \circ \varphi_\alpha^{-1} \circ \varphi_1, \tag{19.5}$$

mais vu que φ_1 et φ_2 sont des cartes, les applications $\varphi_2^{-1} \circ \varphi_\alpha$ et $\varphi_\alpha \circ \varphi_1$ sont de classe C^1 et d'inverses C^1. La composition l'est encore. □

Nous avons demandé qu'une variété n'admette que des cartes partant d'ouverts de $\mathbb{R}^m$. Aurions-nous pu admettre, dans la définition 19.2 que la carte parte d'un ouvert de $\mathbb{R}^p$ avec $p \neq m$? Non. Voici une résultat qui dit que si une carte part de $\mathbb{R}^m$, alors toutes les cartes doivent partir de $\mathbb{R}^m$.

Proposition 19.4.
Soit une variété M de dimension m et une carte $\varphi_\alpha\colon U_\alpha \to M$. Si U est un ouvert de $\mathbb{R}^p$ et si $\varphi\colon U \to M$ est telle que $\varphi_\alpha^{-1} \circ \varphi$ soit un difféomorphisme, alors $p = m$.

Démonstration. Soit $A = \varphi_\alpha(U_\alpha) \cap \varphi(U)$. Les parties $\varphi_\alpha^{-1}(A)$ et $\varphi^{-1}(U)$ sont des ouverts de $\mathbb{R}^m$ et de $\mathbb{R}^p$ respectivement. Par hypothèse, l'application

$$\varphi_\alpha^{-1} \circ \varphi\colon \varphi^{-1}(A) \to \varphi_\alpha^{-1}(A) \tag{19.6}$$

est un difféomorphisme entre l'ouvert $\varphi^{-1}(A) \subset \mathbb{R}^p$ et l'ouvert $\varphi_\alpha^{-1}(A) \subset \mathbb{R}^m$. La proposition 10.138 implique que $m = p$. □

19.1.3 Ancienne définition

Cette section est à recycler.

Définition 19.5.
Soit $\emptyset \neq M \subset \mathbb{R}^n$, $1 \leqslant m < n$ et $k \geqslant 1$. M est une variété de classe C^k de dimension m *si pour tout $a \in M$, il existe un voisinage ouvert U de a dans $\mathbb{R}^n$, et un ouvert V de $\mathbb{R}^m$ tel que $U \cap M$ soit le graphe d'une fonction $f : V \subset \mathbb{R}^m \to \mathbb{R}^{n-m}$ de classe C^1, c'est-à-dire qu'il existe un réagencement des coordonnées $(x_{i_1}, \dots, x_{i_m}, x_{i_{m+1}}, \dots, x_{i_n})$ avec*

$$M \cap U = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n \text{ tel que } (x_{i_1}, \dots, x_{i_m}) \in V \quad \left\{ \begin{array}{rcl} x_{i_{m+1}} & = & f_1(x_{i_1}, \dots, x_{i_m}) \\ \vdots & = & \vdots \\ x_{i_n} & = & f_{n-m}(x_{i_1}, \dots, x_{i_m}) \end{array} \right. \right\}$$

où V est un voisinage ouvert de $(a_{i_1}, \dots, a_{i_m}) \in \mathbb{R}^m$.

La littérature regorge de théorèmes qui proposent des conditions équivalentes à la définition d'une variété. Celle que nous allons le plus utiliser est la suivante

Proposition 19.6.
Soit $M \subset \mathbb{R}^n$ et $1 \leqslant m \leqslant n-1$. L'ensemble M est une variété si et seulement si $\forall a \in M$, il existe un voisinage ouvert $\mathcal{U}$ de a dans $\mathbb{R}^n$ et une application $F \colon W \subset \mathbb{R}^m \to \mathbb{R}^n$ où W est un ouvert tels que

(1) F est un homéomorphisme de W vers $M \cap \mathcal{U}$,

(2) $F \in C^1(W, \mathbb{R}^n)$,

(3) Le rang de $dF(w) \in L(\mathbb{R}^m, \mathbb{R}^n)$ est de rang maximum (c'est-à-dire m) en tout point $w \in W$.

Pour rappel, si $T \colon \mathbb{R}^m \to \mathbb{R}^n$ est une application linéaire, son rang[1] est la dimension de son image. Si A est la matrice d'une application linéaire, alors le rang de cette application linéaire est égal à la taille de la plus grande matrice carrée de déterminant non nul contenue dans A[2].

La condition de rang maximum sert à éviter le genre de cas de la figure 19.1 qui représente l'image de l'ouvert $]-1, 1[$ par l'application $F(t) = (t^2, t^3)$.

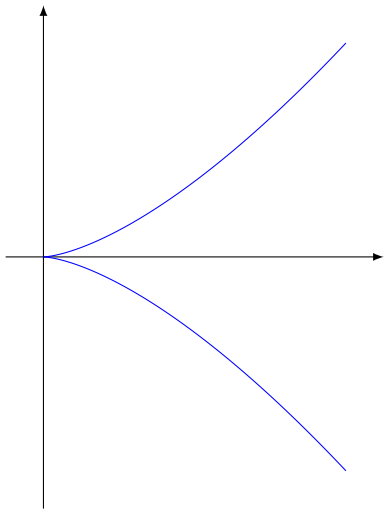


FIGURE 19.1 – Quelque chose qui n'est pas de rang maximum et qui n'est pas une variété.

La différentielle a pour matrice

$$dF(t) = (2t, 3t^2). \tag{19.7}$$

Le rang maximum est 1, mais en $t = 0$, la matrice vaut $(0, 0)$ et son rang est zéro. Pour toute autre valeur de t, c'est bon.

Une autre caractérisation des variétés est donnée par la proposition suivante

1. Définition 4.38.
2. Proposition 4.128

Proposition 19.7.
Soit $M \in \mathbb{R}^n$ et $1 \leqslant m \leqslant n-1$. L'ensemble M est une variété si et seulement si $\forall a \in M$, il existe un voisinage ouvert $\mathcal{U}$ de a dans $\mathbb{R}^n$ tel et une application $G \in C^1(\mathcal{U}, \mathbb{R}^{n-m})$ tel que

(1) le rang de $dG(a) \in L(\mathbb{R}^n, \mathbb{R}^{n-m})$ soit maximum (c'est-à-dire $n-m$) en tout $a \in M$,

(2) $M \cap \mathcal{U} = \{x \in \mathcal{U}$ tel que $G(x) = 0\}$.

19.1.4 Espace tangent

Soit M, une variété dans $\mathbb{R}^n$, et considérons un chemin $\gamma\colon I \to \mathbb{R}^n$ tel que $\gamma(t) \in M$ pour tout $t \in I$ et tel que $\gamma(0) = a$ et que γ est dérivable en 0. La **tangente** au chemin γ au point $a \in M$ est la droite

$$s \mapsto a + s\gamma'(0). \tag{19.8}$$

L'**espace tangent** de M au point a est l'ensemble décrit par toutes les tangentes en a pour tous les chemins γ possibles.

Proposition 19.8.
Une variété de dimension m dans $\mathbb{R}^n$ a un espace tangent de dimension m en chacun de ses points.

19.2 Intégration

19.2.1 Résultats préliminaires

Soient une variété M et une partie A de M. Si nous avons une fonction $f\colon A \to \mathbb{R}$, nous voudrions définir une intégrale de f sur A en utilisant les cartes.

Il y a une petite complication. Supposons que A soit dans $\varphi_\alpha(U_\alpha) \cap \varphi_\beta(U_\beta)$. Pour la carte φ_α nous aurions tendance à vouloir écrire

$$\int_A f = \int_{\varphi_\alpha^{-1}(A)} f \circ \varphi_\alpha \tag{19.9}$$

pour la mesure de Lebesgue induite de $\mathbb{R}^m$ vers son ouvert U_α. Pas de problème à ce que (19.9) soit bien définie. Le problème est que α n'est pas à priori spécial par rapport à β et qu'il faudrait également que

$$\int_A f = \int_{\varphi_\beta^{-1}(A)} f \circ \varphi_\beta \tag{19.10}$$

Mais en utilisant le changement de variable (théorème 13.255(3)) pour le C^1-difféomorphisme $\phi = \varphi_\alpha^{-1} \circ \varphi_\beta$, nous avons

$$\int_{\varphi_\alpha^{-1}(A)} f \circ \varphi_\alpha = \int_{\varphi_\beta(A)} (f \circ \varphi_\alpha \circ \varphi_\alpha^{-1} \circ \varphi_\beta)|J_\phi| = \int_{\varphi_\beta(A)} (f \circ \varphi_\beta)|J_\phi|. \tag{19.11}$$

À moins d'une coïncidence extraordinaire sur les valeurs du jacobien, il n'y aura pas égalité entre (19.9) et (19.10).

Pour faire mieux, il faudra ajouter quelque chose qui compense l'arrivée du jacobien. C'est ce que nous allons faire maintenant.

Proposition-définition 19.9 ([1]).
Soient des ouverts U_1 et U_2 de $\mathbb{R}^m$ ainsi que $A \subset \mathbb{R}^n$. Nous supposons que les applications $\varphi_1\colon U_1 \to A$ et $\varphi_2\colon U_2 \to A$ sont des bijections telles que $\varphi_1^{-1} \circ \varphi_2\colon U_2 \to U_1$ soit un C^1-difféomorphisme.

Soit une fonction mesurable $f\colon A \to [0, +\infty]$ [3].

3. Si f est seulement mesurable, allez voir les hypothèses du théorème 13.255(3).

Alors[4]

$$\int_{U_1}(f\circ\varphi_1)(x)\sqrt{\det\big((d\varphi_1)_x^*(d\varphi_1)_x\big)}dx=\int_{U_2}(f\circ\varphi_2)(y)\sqrt{\det\big((d\varphi_2)_y^*(d\varphi_2)_y\big)}dy \tag{19.12}$$

où les deux intégrales sont au sens de la mesure de Lebesgue sur des ouverts de $\mathbb{R}^m$.

Ce nombre est ***l'intégrale*** *de* f *sur* A*, et est noté*

$$\int_A f. \tag{19.13}$$

Démonstration. Nous considérons le C^1-difféomorphisme $\phi=\varphi_1^{-1}\circ\varphi_2$. Notre but est d'utiliser le théorème de changement de variables 13.255(2).

Vu que φ_1 est une bijection, $(\varphi_1\circ\varphi_1^{-1})(A)=A$ et nous avons

$$\phi^{-1}\big(\varphi_1^{-1}(A)\big)=\varphi_2^{-1}(A). \tag{19.14}$$

Ensuite, $f\circ\varphi_1\circ\phi=f\circ\varphi_2$.

Jusqu'ici, rien de drôle me diriez-vous. Ok. Alors voyons un peu comment se passe le jacobien et ce qui se trouve sous la racine carré.

L'application $x\mapsto\det\big((d\varphi_1)_x^*\circ(d\varphi_1)_x\big)$ doit être composée avec ϕ pour donner

$$y\mapsto\det\big((d\varphi_1)_{\phi(y)}^*\circ(d\varphi_1)_{\phi(y)}\big). \tag{19.15}$$

Attardons-nous un peu sur $(d\varphi_1)_{\phi(y)}$. Nous savons par la règle de différentiation en chaine 10.135 que

$$d(\varphi_1\circ\phi)_y=(d\varphi_1)_{\phi(y)}\circ(d\phi)_y. \tag{19.16}$$

Le lemme 10.137 nous dit que l'application linéaire $d\phi_y\colon\mathbb{R}^m\to\mathbb{R}^m$ est inversible et que $d\phi_y^{-1}=(d\phi^{-1})_{\phi(y)}$. En composant les deux côtés de (19.16) par cela nous trouvons

$$(d\varphi_1)_{\phi(y)}=d(\varphi_1\circ\phi)_y\circ(d\phi)_y^{-1}=(d\varphi_2)_y\circ(d\phi^{-1})_{\phi(y)}. \tag{19.17}$$

C'est me moment d'utiliser la proposition 9.33 à propos de la composition des applications adjointes. Ce qui est dans le déterminant est :

$$\big((d\varphi_2)_y\circ(d\phi^{-1})_{\phi(y)}\big)^*\circ\big((d\varphi_2)_y\circ(d\phi^{-1})_{\phi(y)}\big)=\underbrace{(d\phi^{-1})_{\phi(y)}^*}_{\mathbb{R}^m\to\mathbb{R}^m}\circ\underbrace{(d\varphi_2)_y^*\circ(d\varphi_2)_y}_{\mathbb{R}^m\to\mathbb{R}^m}\circ\underbrace{(d\phi^{-1})_{\phi(y)}}_{\mathbb{R}^m\to\mathbb{R}^m}. \tag{19.18}$$

Nous utilisons à présent un combo entre les propriétés des déterminants et de l'adjoint : les propositions 9.10(2) et 9.32 font sortir $|\det\big((d\phi^{-1})_{\phi(y)}\big)|$ de la racine carré. Ne pas oublier la valeur absolue parce que $\sqrt{x^2}=|x|$ et non $\sqrt{x^2}=x$. Cela pour dire que

$$\sqrt{\det\big((d\varphi_1)_{\phi(y)}^*\circ(d\varphi_1)_{\phi(y)}\big)}=|\det\big((d\phi^{-1})_{\phi(y)}\big)|\sqrt{\det\big((d\varphi_2)_y^*\circ(d\varphi_2)_y\big)}. \tag{19.19}$$

En ce qui concerne l'intégrale que nous voulons calculer, en mettant les bouts ensemble,

$$\begin{aligned}\int_{\varphi_1^{-1}(A)}&(f\circ\varphi_1)(x)\sqrt{\det(\cdots)}dx\\&=\int_{\varphi_2^{-1}(A)}(f\circ\varphi_2)(y)|\det\big((d\phi^{-1})_{\phi(y)}\big)|\sqrt{\det\big((d\varphi_2)_y^*\circ(d\varphi_2)_y\big)}|\det(d\phi_y)|dy.\end{aligned} \tag{19.20}$$

Enfin, les déterminants se suppriment : dans la valeur absolue nous avons :

$$\det\big((d\phi^{-1})_{\phi(y)}\big)\det(d\phi_y)=\det\big((d\phi^{-1})_{\phi(y)}\circ d\phi_y\big)=\det(\mathrm{Id})=1. \tag{19.21}$$

□

4. Voir la définition de l'adjoint 9.30.

Si nous voulons que la formule de la proposition 19.9 fournisse une définition raisonnable de $\int_A f$, il faut que si (U, φ) est une carte,

$$\det\left(d\varphi_x^* \circ d\varphi_x\right) \geqslant 0 \tag{19.22}$$

pour tout x, et ne soit nul sur aucun ouvert de U.

Lemme 19.10 ([1]).
Soit une carte (U, φ) d'une variété M de dimension $\mathbb{R}^m$. Soit $x \in U$. Alors

(1) L'opérateur $d\varphi_x^* \circ d\varphi_x$ est autoadjoint semi-défini positif.

(2) Il existe une base de $\mathbb{R}^m$ formée de vecteurs propres de $d\varphi_x^* \circ d\varphi_x$. Les valeurs propres sont réelles et positives.

(3) Pour tout $x \in U$ nous avons

$$\det(d\varphi_x^* \circ d\varphi_x) > 0. \tag{19.23}$$

Démonstration. Point par point.

(1) Soit $x \in U$. Nous notons $A = d\varphi_x$. Par la proposition 9.33 nous avons $(AA^*)^* = AA^*$ du fait que $(A^*)^* = A$. Donc A^*A est autoadjoint. Sa matrice est donc symétrique par la proposition 9.30(2).

De plus, pour tout $u \in \mathbb{R}^m$ nous avons

$$\langle A^*Au, u\rangle = \langle Au, Au\rangle \geqslant 0. \tag{19.24}$$

Cela implique que la matrice de A^*A est semi-définie positive par le lemme 9.157(2). L'opérateur A^*A est également semi-défini positif par la proposition 9.156.

(2) L'existence de la base de vecteurs propres est le théorème 9.150. Ce théorème dit de plus que les valeurs propres sont réelles. Le fait qu'elles soient positives est le fait déjà prouvé que A^*A est semi-défini positif.

(3) Dans la base de vecteurs propres, la matrice est diagonale avec des nombres positifs sur la diagonale. Le déterminant est alors le produit des valeurs propres. Voila pourquoi le déterminant est positif.

Il nous reste à montrer que ce déterminant ne peut pas être nul. Si $\det(d\varphi_a^* \circ d\varphi_a) = 0$, alors l'application $d\varphi_a^* \circ d\varphi$ a un noyau (proposition 9.10(2)), c'est-à-dire un vecteur u tel que $d\varphi_a^* d\varphi_a(u) = 0$ en particulier,

$$0 = \langle d\varphi_a^* d\varphi_a u, u\rangle = \langle d\varphi_a u, d\varphi_a u, \rangle = \|d\varphi_a u\|^2. \tag{19.25}$$

Donc $d\varphi_a(u) = 0$.

Mais $\varphi\colon U \to \varphi(U)$ est bijective, donc l'application $\varphi^{-1}\colon \varphi(U) \to U$ l'est aussi et vu que φ est une carte, l'application $\varphi^{-1} \circ \varphi\colon U \to U$ est égale à l'identité. L'identité est une application linéaire dont la différentielle est elle-même, c'est-à-dire que

$$d(\varphi^{-1} \circ \varphi)_a = (d\varphi^{-1})_{\varphi(a)} \circ d\varphi_a = \mathrm{Id}\,. \tag{19.26}$$

Si donc $u \neq 0$, nous devons avoir $d\varphi_a(u) \neq 0$.

□

19.2.2 Quelques expressions pour l'élément de volume

Ce que nous appelons « élément de volume » (pour des raisons qui apparaîtront plus tard) est le coefficient $\sqrt{\det(d\varphi_x^* \circ d\varphi_x)}$, introduit pour le besoin de compenser l'arrivée du jacobien en cas de changement de variables. Mais à quoi ressemble de coefficient pour les variétés de petite dimension ?

C'est ce que nous allons voir maintenant.

19.2.2.1 En dimension un

Nous supposons une variété de dimension un dans $\mathbb{R}^n$. Une carte est donc une application $\varphi\colon \mathbb{R} \to \mathbb{R}^n$ dont nous notons φ_i les composantes. Nous avons, pour $u \in \mathbb{R}$,

$$d\varphi_a(u) = \frac{d}{dt}\Big[\varphi(a+tu)\Big]_{t=0} = \frac{d}{dt}\begin{pmatrix}\varphi_1(a+tu)\\ \vdots \\ \varphi_n(a+tu)\end{pmatrix} = u\begin{pmatrix}\varphi_1'(a)\\ \vdots \\ \varphi_n'(a)\end{pmatrix} = u\varphi'(a). \tag{19.27}$$

Là dedans, vous vous souviendrez que $a \in \mathbb{R}$ et que $\varphi'(a) \in \mathbb{R}^n$.

Nous devons savoir ce que vaut $d\varphi_a^* x$ lorsque $a \in \mathbb{R}$ et $x \in \mathbb{R}^m$. Pour tout $u \in \mathbb{R}$ nous avons

$$\langle d\varphi_a x, u\rangle_{\mathbb{R}} = \langle x, d\varphi_a(u)\rangle_{\mathbb{R}^n} = u\langle x, \varphi'(a)\rangle_{\mathbb{R}^n} = u\varphi'(a) \cdot x = \langle \varphi'(a) \cdot x, u\rangle_{\mathbb{R}}. \tag{19.28}$$

Donc $d\varphi_a^*(x) = \varphi'(a) \cdot x$.

Question notation, nous avons noté $\langle a, b\rangle_{\mathbb{R}} = ab$ et indifféremment $\langle x, y\rangle_{\mathbb{R}^n} = x \cdot y$.

L'application $d\varphi_a^* \circ d\varphi_a$ est une application linéaire bijective entre $\mathbb{R}$ et $\mathbb{R}$. Sa matrice est donc 1×1 et elle se calcule en appliquant l'application au vecteur de base 1 de $\mathbb{R}$:

$$(d\varphi_a^* \circ d\varphi_a)(1) = d\varphi_a^*\big(\varphi'(a)\big) = \varphi'(a) \cdot \varphi'(a) = \|\varphi'(a)\|^2. \tag{19.29}$$

Nous avons donc

$$\det\big(d\varphi_a^* \circ d\varphi_a\big) = \|\varphi'(a)\|^2. \tag{19.30}$$

19.2.2.2 En dimension quelconque

Soit une carte $\varphi\colon \mathbb{R}^m \to \mathbb{R}^n$ pour une variété de dimension m. Les éléments de matrice de la différentielle d'une application sont donnés par la proposition 11.188 :

$$(d\varphi_a)_{ij} = \frac{\partial \varphi_i}{\partial x_j}(a). \tag{19.31}$$

Les éléments de la matrice de $d\varphi_a^*$ sont ceux de la matrice transposée. En ce qui concerna la composition, c'est le produit des matrices :

$$(d\varphi_a^* \circ d\varphi_a)_{ij} = \sum_k (d\varphi_a^*)_{ik}(d\varphi_a)_{kj} = \sum_k (d\varphi_a)_{ki}(d\varphi_a)_{kj} = \sum_k \frac{\partial \varphi_k}{\partial x_i}(a)\frac{\partial \varphi_k}{\partial x_j}(a). \tag{19.32}$$

Cela pour écrire cette bonne formule :

$$(d\varphi_a^* \circ d\varphi_a)_{ij} = \frac{\partial \varphi}{\partial x_i}(a) \cdot \frac{\partial \varphi}{\partial x_j}(a). \tag{19.33}$$

19.2.2.3 En dimension deux

Nous considérons maintenant une carte $\varphi\colon U \to \mathbb{R}^3$ avec $U \subset \mathbb{R}^2$. En notant $v_i = \partial_i\varphi(a)$ nous notons la formule (19.33) sous la forme

$$\det\big(d\varphi_a^* \circ d\varphi_a\big) = \det\begin{pmatrix} v_1 \cdot v_1 & v_1 \cdot v_2 \\ v_2 \cdot v_1 & v_2 \cdot v_2 \end{pmatrix} = \|v_1\|^2\|v_2\|^2 - (v_1 \cdot v_2)^2. \tag{19.34}$$

L'identité de Lagrange de la proposition 7.212 nous donne alors

$$\sqrt{\det(d\varphi_a^* \circ d\varphi_a)} = \|\frac{\partial \varphi}{\partial x} \times \frac{\partial \varphi}{\partial y}\|. \tag{19.35}$$

19.2.2.4 En dimension trois

Nous considérons maintenant une carte $\varphi\colon U \to \mathbb{R}^3$ avec $U \subset \mathbb{R}^3$. Nous notons $u = \frac{\partial\varphi}{\partial\partial x}(a)$, $v = \frac{\partial\varphi}{\partial y}(a)$ et $w = \frac{\partial\varphi}{\partial z}(a)$.

En utilisant l'expression du lemme 7.210 à propos du produit mixte, la formule 19.33 donne

$$\det(d\varphi_a^* \circ d\varphi_a) = \det \begin{pmatrix} \|u\|^2 & u \cdot v & u \cdot w \\ v \cdot u & \|v\|^2 & v \cdot w \\ w \cdot u & w \cdot v & \|w\|^2 \end{pmatrix} = \big|(u \times v) \cdot w\big|. \tag{19.36}$$

Donc

$$\det(d\varphi_a^* \circ d\varphi_a) = \big|(\frac{\partial\varphi}{\partial x}(a) \times \frac{\partial\varphi}{\partial y}(a)) \cdot \frac{\partial\varphi}{\partial z}(a)\big|. \tag{19.37}$$

19.3 Intégrale sur une variété

19.3.1 Mesure sur une carte

Nous considérons dans cette section uniquement des variétés M de dimension 2 dans $\mathbb{R}^3$. Une particularité de $\mathbb{R}^3$ (par rapport aux autres $\mathbb{R}^n$) est qu'il existe le produit vectoriel.

Si v, $w \in \mathbb{R}^3$, alors le vecteur $v \times w$ est une vecteur normal au plan décrit par v et w qui jouit de l'importante propriété suivante :

$$\text{aire du parallélogramme} = \|v \times w\|. \tag{19.38}$$

L'aire du parallélogramme construit sur v et w est donnée par la norme du produit vectoriel. Afin de donner une mesure infinitésimale en un point $p \in M$, nous voudrions prendre deux vecteurs tangents à M en p, et puis considérer la norme de leur produit vectoriel. Cette idée se heurte à la question du choix des vecteurs tangents à considérer.

Dans $\mathbb{R}^2$, le choix est évident : nous choisissons e_x et e_y, et nous avons $\|e_x \times e_y\| = 1$. L'idée est donc de choisir une carte $F\colon W \to F(w)$ autour du point $p = F(w)$, et de choisir les vecteurs tangents qui correspondent à e_x et e_y via la carte, c'est-à-dire les vecteurs

$$\frac{\partial F}{\partial x}(w), \quad \text{et} \quad \frac{\partial F}{\partial y}(w). \tag{19.39}$$

L'**élément infinitésimal de surface** sur M au point $p = F(w)$ est alors défini par

$$d\sigma_F = \|\frac{\partial F}{\partial x}(w) \times \frac{\partial F}{\partial y}(w)\| dw, \tag{19.40}$$

et si la partie $A \subset M$ est entièrement contenue dans $F(W)$, nous définissons la **mesure** de A par

$$\mu_2(A) = \int_{F^{-1}(A)} d\sigma_F = \int_{F^{-1}(A)} \|\frac{\partial F}{\partial x}(w) \times \frac{\partial F}{\partial y}(w)\| dw. \tag{19.41}$$

Remarque 19.11.
Afin que cette définition ait un sens, nous devons prouver qu'elle ne dépend pas du choix de la carte F. En effet, les vecteurs $\partial_x F$ et $\partial_y F$ dépendent de la carte F, donc leur produit vectoriel (et sa norme) dépendent également de la carte F choisie. Il faudrait donc un petit miracle pour que le nombre $\mu_2(A)$ donné par (19.41) soit indépendant du choix de F. Nous allons bientôt voir comme cas particulier du théorème 19.14 que c'est en fait le cas. C'est-à-dire que si F et $\tilde{F}$ sont deux cartes qui contiennent A, alors

$$\int_{F^{-1}(A)} d\sigma_F = \int_{\tilde{F}^{-1}(A)} d\sigma_{\tilde{F}}. \tag{19.42}$$

19.3.1.1 Exemple : la mesure de la sphère

Nous nous proposons maintenant de calculer la surface de la sphère $S^2 = x^2 + y^2 + z^2 = R^2$. L'application $F\colon B((0,0),R) \to R^3$ donnée par

$$F(x,y) = \begin{pmatrix} x \\ y \\ \sqrt{R^2 - x^2 - y^2} \end{pmatrix} \tag{19.43}$$

est une carte pour une demi-sphère. Ses dérivées partielles sont

$$\frac{\partial F}{\partial x} = \begin{pmatrix} 1 \\ 0 \\ -\frac{x}{\sqrt{R^2-x^2-y^2}} \end{pmatrix}, \quad \frac{\partial F}{\partial y} = \begin{pmatrix} 0 \\ 1 \\ -\frac{y}{\sqrt{R^2-x^2-y^2}} \end{pmatrix}. \tag{19.44}$$

Le produit vectoriel de ces deux vecteurs tangents donne

$$\frac{\partial F}{\partial x}(x,y) \times \frac{\partial F}{\partial y}(x,y) = \frac{x}{\alpha}e_1 + \frac{y}{\alpha}e_2 + e_3 \tag{19.45}$$

où $\alpha = \sqrt{R^2 - x^2 - y^2}$. En calculant la norme, nous trouvons

$$\|\frac{\partial F}{\partial x}(x,y) \times \frac{\partial F}{\partial y}(x,y)\| = \sqrt{\frac{R^2}{R^2 - x^2 - y^2}}, \tag{19.46}$$

et en passant aux coordonnées polaires, nous écrivons l'intégrale (19.41) sous la forme

$$\int_B \|\partial_x F \times \partial_y F\| = \int_0^{2\pi} d\theta \int_0^R r\sqrt{\frac{R^2}{R^2 - x^2 - y^2}}dr = 2\pi R^2, \tag{19.47}$$

qui est bien la mesure de la demi-sphère.

19.3.2 Intégrale sur une carte

Nous pouvons maintenant définir l'intégrale d'une fonction sur une carte de la variété M.

Définition 19.12.
*Soit $F\colon W \subset \mathbb{R}^2 \to \mathbb{R}^3$, une carte pour une variété M. Soit A, une partie de $F(W)$ telle que $A = F(B)$ où $B \subset W$ est mesurable. Soit encore $f\colon A \to \mathbb{R}$, une fonction continue. L'**intégrale** de f sur A est le nombre*

$$\int_A f = \int_A f d\sigma_F = \int_{F^{-1}(A)} (f \circ F)(w)\|\frac{\partial F}{\partial x}(w) \times \frac{\partial F}{\partial y}(w)\|dw \tag{19.48}$$

Remarque 19.13.
L'intégrale (19.48) n'est pas toujours bien définie. Étant donné que F est C^1 et que f est continue, l'intégrante est continue. L'intégrale sera donc bien définie par exemple lorsque B est borné et si la fermeture $\bar{A}$ est un compact contenu dans $F(w)$.

Le théorème suivant montre que le travail que nous avons fait jusqu'à présent ne dépend en fait pas du choix de carte F effectué.

Théorème 19.14.
Soient $F\colon W \to F(w)$ et $\tilde{F}\colon \tilde{W} \to \tilde{F}(\tilde{W})$, deux cartes de la variété M. Soit une partie $A \subset F(W) \cap \tilde{F}(\tilde{W})$ telle que $A = F(B)$ avec $B \subset W$ mesurable. Alors $A = \tilde{F}(\tilde{B})$ avec $\tilde{B} \subset \tilde{W}$ mesurable.

Si f est une fonction continue, et si $\int_A f d\sigma_F$ existe, alors $\int_A f d\sigma_{\tilde{F}}$ existe et

$$\int_A f d\sigma_F = \int_A f d\sigma_{\tilde{F}}. \tag{19.49}$$

19.3.3 Exemples

Intégrons la fonction $f(x,y,z)$ sur le carré $K =]0,1[\times]0,2[\times\{1\}$. La première carte que nous pouvons utiliser est

$$\begin{aligned} F\colon]0,1[\times]0,2[&\to K \\ (x,y) &\mapsto (x,y,1). \end{aligned} \tag{19.50}$$

Nous trouvons aisément les vecteurs tangents qui forment l'élément de surface :

$$\frac{\partial F}{\partial x} = \begin{pmatrix}1\\0\\0\end{pmatrix}, \quad \frac{\partial F}{\partial y} = \begin{pmatrix}0\\1\\0\end{pmatrix}, \tag{19.51}$$

donc $d\sigma_F = 1\cdot dxdy$, et

$$\int_K f d\sigma_F = \int_{]0,1[\times]0,2[} f(x,y,1)\cdot 1\cdot dxdy. \tag{19.52}$$

Nous pouvons également utiliser la carte

$$\begin{aligned} \tilde{F}\colon]0,\frac{1}{2}[\times]0,6[&\to K \\ (\tilde{x},\tilde{y}) &\mapsto (2\tilde{x},\frac{\tilde{y}}{3},1). \end{aligned} \tag{19.53}$$

Les vecteurs tangents sont maintenant

$$\frac{\partial \tilde{F}}{\partial \tilde{x}} = \begin{pmatrix}2\\0\\0\end{pmatrix}, \quad \frac{\partial \tilde{F}}{\partial \tilde{y}} = \begin{pmatrix}0\\1/3\\0\end{pmatrix}, \tag{19.54}$$

et nous avons donc $d\sigma_{\tilde{F}} = \|\frac{2}{3}e_3\| = \frac{2}{3}$. Cette fois, l'intégrale de f sur K s'écrit

$$\int_K f d\sigma_{\tilde{F}} = \int_{]0,\frac{1}{2}[\times]0,6[} f\big(2\tilde{x},\frac{\tilde{y}}{3},1\big)\cdot\frac{2}{3}\cdot d\tilde{x}s\tilde{y}. \tag{19.55}$$

Conformément au théorème 19.14, cette dernière intégrale est égale à l'intégrale (19.52) parce qu'il s'agit juste d'un changement de variable.

19.3.4 Orientation

Soient $F\colon W\to F(w)$ et $\tilde{F}\colon \tilde{W}\to\tilde{F}(\tilde{W})$, deux cartes de la variété M. Nous pouvons considérer la fonction $h = \tilde{F}^{-1}\circ F$, définie uniquement sur l'intersection des cartes :

$$h\colon F^{-1}\big(F(W)\cap\tilde{F}(\tilde{W})\big)\to\tilde{F}^{-1}\big(F(W)\cap\tilde{F}(\tilde{W})\big). \tag{19.56}$$

Nous disons que F et $\tilde{F}$ ont même **orientation** si

$$J_h(w) > 0 \tag{19.57}$$

pour tout $w\in F^{-1}\big(F(W)\cap\tilde{F}(\tilde{W})\big)$.

Considérons les deux cartes suivantes pour le même carré :

$$\begin{aligned} F\colon]0,1[\times]0,1[&\to \mathbb{R}^3 \\ (x,y) &\mapsto (x,y,0) \end{aligned} \tag{19.58}$$

et

$$\begin{aligned} \tilde{F}\colon]0,\frac{1}{2}[\times]0,\frac{1}{3}[&\to \mathbb{R}^3 \\ (x,y) &\mapsto (2x,3y,0) \end{aligned} \tag{19.59}$$

Ici, $h(x,y) = \left(\frac{x}{2}, \frac{y}{3}\right)$ et nous avons $J_h = \frac{1}{6} > 0$. Ces deux cartes ont même orientation. Notez que

$$\frac{\partial F}{\partial x} \times \frac{\partial F}{\partial y} = e_3, \tag{19.60}$$

tandis que

$$\frac{\partial \tilde{F}}{\partial x} \times \frac{\partial \tilde{F}}{\partial y} = 6e_3. \tag{19.61}$$

Les vecteurs normaux à le paramétrage pointent dans le même sens.

Si par contre nous prenons le paramétrage

$$\begin{aligned} G\colon]0,1[\times]0,1,[&\to \mathbb{R}^2 \\ (x,y) &\mapsto (x,(1-y),0), \end{aligned} \tag{19.62}$$

nous avons

$$\frac{\partial G}{\partial x} \times \frac{\partial G}{\partial y} = -e_3, \tag{19.63}$$

et si $g = G^{-1} \circ F$, alors $J_g = -1$.

L'orientation d'une carte montre donc si le vecteur normal à la surface pointe d'un côté ou de l'autre de la surface.

Définition 19.15.
*Une variété M est **orientable** s'il existe un atlas de M tel que deux cartes quelconques ont toujours même orientation. Une variété est **orientée** lorsque qu'un tel choix d'atlas est fait.*

Proposition 19.16.
Soit M, une variété orientable et un atlas orienté $\{F_i\colon W_i \to \mathbb{R}^3\}$. Alors le vecteur unitaire

$$\frac{\frac{\partial F}{\partial x}(x,y) \times \frac{\partial F}{\partial y}(x,y)}{\|\frac{\partial F}{\partial x}(x,y) \times \frac{\partial F}{\partial y}(x,y)\|} \tag{19.64}$$

ne dépend pas du choix de F parmi les F_i.

Démonstration. Considérons deux cartes F_1 et F_2, ainsi que l'application $h = F_2^{-1} \circ F_1$. Écrivons le vecteur $\partial_x F_1 \times \partial_y F_1$ en utilisant $F_1 = F_2 \circ h$. D'abord, par la règle de dérivation de fonctions composées,

$$\frac{\partial(F_2 \circ h)}{\partial x} = \frac{\partial F_2}{\partial x}\frac{\partial h_1}{\partial x} + \frac{\partial F_2}{\partial y}\frac{\partial h_2}{\partial x}. \tag{19.65}$$

Après avoir fait le même calcul pour $\frac{\partial(F_2 \circ h)}{\partial y}$, nous pouvons écrire

$$\partial_x(F_2 \circ h) \times \partial_y(F_2 \circ h) = (\partial_x h_1 \partial_x F_2 + \partial_x h_2 \partial_y F_2) \times (\partial_y h_1 \partial_x F_2 + \partial_y h_2 \partial_y F_2). \tag{19.66}$$

Dans cette expression, les facteurs $\partial_i h_j$ sont des nombres, donc ils se factorisent dans les produits vectoriels. En tenant compte du fait que $\partial_x F_2 \times \partial_x F_2 = 0$ et $\partial_y F_2 \times \partial_y F_2 = 0$, ainsi que de l'antisymétrie du produit vectoriel, l'expression se réduit à

$$\left(\frac{\partial F_2}{\partial x} \times \frac{\partial F_2}{\partial y}\right)(\partial_x h_1 \partial_y h_2 - \partial_x h_2 \partial_y h_2). \tag{19.67}$$

Par conséquent,

$$\frac{\partial F_1}{\partial x} \times \frac{\partial F_1}{\partial y} = \frac{\partial(F_2 \circ h)}{\partial x} \times \frac{\partial(F_2 \circ h)}{\partial y} = \left(\frac{\partial F_2}{\partial x} \times \frac{\partial F_2}{\partial y}\right) \det J_h. \tag{19.68}$$

Donc, tant que J_h est positif, les vecteurs unitaires correspondants au membre de gauche et de droite sont égaux. □

Corolaire 19.17.
Si nous avons choisi un atlas orienté pour la variété M, nous avons une fonction continue $G\colon M \to \mathbb{R}^3$ telle que $\|G(p)\| = 1$ pour tout $p \in M$. Cette fonction est donnée par

$$G(F(x,y)) = \frac{\frac{\partial F}{\partial x}(x,y) \times \frac{\partial F}{\partial y}(x,y)}{\|\frac{\partial F}{\partial x}(x,y) \times \frac{\partial F}{\partial y}(x,y)\|} \tag{19.69}$$

sur l'image de la carte F.

Démonstration. La fonction G est construite indépendamment sur chaque carte $F(W)$ en utilisant la formule (19.69). Cette fonction est une fonction bien définie sur tout M parce que nous venons de démontrer que sur $F_1(W_1) \cap F_2(W_2)$, les fonctions construites à partir de F_1 et à partir de F_2 sont égales. □

Il est possible que prouver, bien que cela soit plus compliqué, que la réciproque est également vraie.

Proposition 19.18.
Une variété M de dimension 2 dans $\mathbb{R}^3$ est orientable si et seulement s'il existe une fonction continue $G\colon M \to \mathbb{R}^3$ telle que pour tout $p \in M$, le vecteur $G(p)$ soit de norme 1 et normal à M au point p.

19.3.5 Formes différentielles

Nous allons donner une toute petite introduction aux formes différentielles sur des variétés compactes.

Lemme 19.19 ([386])**.**
Soit ω une k-forme sur $\mathbb{R}^n$ et f, une fonction C^∞ sur $\mathbb{R}^n$. Alors $d(f^*\omega) = f^*d\omega$.

Démonstration. Nous effectuons la preuve par récurrence sur le degré de la forme. Soit d'abord une 0-forme, c'est-à-dire une fonction $g\colon \mathbb{R}^n \to \mathbb{R}$. Nous avons

$$d(d^*g)X = d(g \circ f)X = (dg \circ df)X = dg\big(dfX\big) = (f^*dg)(X). \tag{19.70}$$

Supposons maintenant que le résultat soit exact pour toutes les $p-1$-formes et montrons qu'il reste valable pour les p-formes. Par linéarité de la différentielle nous pouvons nous contenter de considérer la forme différentielle

$$\omega = g\,dx^1 \wedge \dots dx^p \tag{19.71}$$

où g est une fonction C^∞. Pour soulager les notations nous allons noter $dx^I = dx^1 \wedge \dots dx^{p-1}$. Nous avons

$$\begin{aligned}
d(f^*\omega) &= d\big(f^*(gdx^I \wedge dx^p)\big) && (19.72a)\\
&= d\big(f^*(gdx^I) \wedge f^*dx^p\big) && (19.72b)\\
&= d\big(f^*(gdx^I)\big) \wedge f^*dx^p + (-1)^{p-1} f^*(gdx^I) \wedge (f^*dx^p) && (19.72c)\\
&= f^*\big(d(gdx^I)\big) \wedge f^*dx^p && (19.72d)\\
&= f^*\big(d(gdx^I) \wedge dx^p\big) && (19.72e)\\
&= f^*d\omega && (19.72f)
\end{aligned}$$

Justifications : (19.72c) est la formule de Leibnitz. (19.72d) est parce que le second terme est nul : $d(f^*dx^p) = f^*(d^2x^p) = 0$. Nous avons utilisé l'hypothèse de récurrence et le fait que $d^2 = 0$. L'étape (19.72f) est une utilisation à l'envers de la règle de Leibnitz en tenant compte que $d^2x^p = 0$. □

Lemme-définition 19.20.
Soient une variété M munie d'un recouvrement par un nombre fini d'ouverts[5] *$\{U_i\}$. Alors il existe une famille de fonctions $f_i \in C^\infty(M)$ telle que*

5. Si M n'est pas compacte, alors il faut utiliser une version un peu plus élaborée du lemme[386].

(1) $\operatorname{supp} f_i \subset U_i$,

(2) pour tout i, nous avons $f_i \geqslant 0$,

(3) pour tout $p \in M$ nous avons $\sum_i f_i(p) = 1$.

La famille (f_i) est une **partition de l'unité** subordonnée au recouvrement $\{U_i\}$.

Soit M une variété de dimension n et ω une n-forme différentielle

$$\omega_p = f(p)dx_1 \wedge \ldots \wedge dx_n. \tag{19.73}$$

Si (U, φ) est une carte ($U \subset \mathbb{R}^n$ et $\varphi\colon U \to M$) alors nous définissons

$$\int_{\varphi(U)} \omega = \int_U f\big(\varphi(x)\big)dx_1 \ldots dx_n. \tag{19.74}$$

Lorsque nous voulons intégrer sur une partie plus grande qu'une carte nous utilisons une partition de l'unité du lemme 19.20.

Définition 19.21 (Intégrale d'une forme sur une variété)**.**
Si $\{f_i\}$ est une partition de l'unité subordonnée à un atlas de M nous définissons

$$\int_M \omega = \sum_i \int_{U_i} f\omega. \tag{19.75}$$

Il est possible de montrer que cette définition ne dépend pas du choix de la partition de l'unité.

Remarque 19.22.
Nous ne définissons pas d'intégrale de k-forme différentielle sur une variété de dimension $n \neq k$. Le seul cas où cela se fait est le cas de 0-formes (les fonctions), mais cela n'est pas vraiment un cas particulier vu que les 0-formes sont associées aux n-formes de façon évidente.

Remarque 19.23.
La définition 19.21 permet d'intégrer des formes sur des variétés, pas sur des sous-variétés. Autrement dit, ce n'est pas cette définition là qu'il faut utiliser pour comprendre des objets comme

$$\int_\gamma f \tag{19.76}$$

où γ est un chemin dans $\mathbb{R}^n$. Nous définirons ce genre de choses plus bas.

19.3.6 Intégrale d'une fonction sur une sous-variété

Nous allons nous restreindre au cas d'une sous-variété de $\mathbb{R}^n$. C'est-à-dire que nous considérons l'espace $\mathbb{R}^n$ comme espace ambiant et nous allons intégrer sur des parties de $\mathbb{R}^n$. Ces parties peuvent être de dimension plus basses que n, et c'est justement ça qui fait la différence entre ce que nous faisons maintenant et la définition 19.21.

L'exemple typique est l'intégrale sur une surface dans $\mathbb{R}^3$, ou de volumes.

Nous supposons à présent que M est une variété compacte de dimension 2 dans $\mathbb{R}^3$. La compacité fait que M possède un atlas contenant un nombre fini de cartes $F_i\colon W_i \to F_i(W_i)$.

Si $A \subset M$ est tel que pour chaque i, $A \cap F_i(W_i) = F_i(V_i)$ pour une ensemble V_i mesurable dans $\mathbb{R}^2$, alors nous considérons

$$A_1 = A \cap F_1(W_2) = F_1(V_1). \tag{19.77}$$

Ensuite, nous construisons A_2 en considérant $F_A(W_2)$ et en lui retranchant A_1 :

$$A_2 = \big(A \cap F_2(W_2)\big) \cap F_1(V_1). \tag{19.78}$$

En continuant de la sorte, nous construisons la décomposition

$$A = A_1 \cup \ldots \cup A_p \tag{19.79}$$

de A en ouverts disjoints, chacun de ouverts A_p étant compris dans une carte.

Il est possible de prouver que dans ce cas, la définition suivante a un sens et ne dépend pas du choix de l'atlas effectué.

Définition 19.24.
*Si $f\colon A \to \mathbb{R}$ est une fonction continue, alors l'**intégrale** est le nombre*

$$\int_A f = \sum_{i=1}^{p} \int_{A_i} f d\sigma_{F_i}. \tag{19.80}$$

19.4 Longueur, aire, volumes etc.

Grâce à la mesure de Lebesgue (définition 13.213), nous avons la définition d'aires dans $\mathbb{R}^2$ et de volumes dans $\mathbb{R}^3$. Dans tout ce qui suit, nous considérons toujours la tribu de Lebesgue, la mesure de Lebesgue, et l'intégrale de Lebesgue correspondante.

Définition 19.25.
*L'**aire** de la partie mesurable S de $\mathbb{R}^2$ est le nombre*

$$\int_{\mathbb{R}^2} \mathbb{1}_S \tag{19.81}$$

au sens de l'intégrale pour la mesure de Lebesgue de la fonction caractéristique de la partie S.

Définition 19.26.
*Le **volume** de la partie mesurable V de $\mathbb{R}^3$ est le nombre*

$$\int_{\mathbb{R}^R} \mathbb{1}_V \tag{19.82}$$

au sens de l'intégrale pour la mesure de Lebesgue de la fonction caractéristique de la partie V.

Ceci est bien beau, mais ne permet pas de définir l'aire d'une surface dans $\mathbb{R}^3$, ni une longueur dans $\mathbb{R}^2$. Nous n'avons pas encore défini ce que nous appelons une surface dans $\mathbb{R}^3$, mais selon toute définition raisonnable, si S en est une, elle sera négligeable au sens de la mesure de Lebesgue dans $\mathbb{R}^3$ et nous aurons toujours

$$\int_{\mathbb{R}^3} \mathbb{1}_S = 0. \tag{19.83}$$

L'objet de ce chapitre sera de donner un sens aux notions de longueurs dans $\mathbb{R}^2$, de surfaces dans $\mathbb{R}^3$ et d'y définir des intégrales permettant de définir longueurs et aires.

19.4.1 Quelques aires faciles

Nous nous souvenons de la proposition 7.207 qui donnait une bonne propriété du produit vectoriel dans $\mathbb{R}^3$. Nous en donnons une autre.

Proposition 19.27.
Si u et v sont des vecteurs de $\mathbb{R}^2$ alors le vecteur $u \times v$ (où nous avons prolongé u et v par zéro dans $\mathbb{R}^3$) est de norme égale à l'aire[6] du parallélogramme construit sur u et v dans $\mathbb{R}^2$.

La subtilité quant au prolongement de u et v par zéro est due au fait que nous n'avons encore défini d'aires que celles de parties de $\mathbb{R}^2$. Officiellement, nous ne savons pas encore ce que serait l'aire du parallélogramme construit sur deux vecteurs de $\mathbb{R}^3$. Officieusement, nous nous attendons à ce que ce soit égal à la norme du produit vectoriel.

6. Définition 19.25.

19.5 Autres théorèmes de points fixes

En termes de théorème de points fixes nous avons déjà vu le théorème de Picard 16.34. Voir aussi le thème 19.

19.5.1 Brouwer

Proposition 19.28.
Soit $f\colon [a,b] \to [a,b]$ une fonction continue. Alors f accepte un point fixe.

Démonstration. En effet si nous considérons $g(x) = f(x)-x$ alors nous avons $g(a) = f(a)-a \geqslant 0$ et $g(b) = f(b)-b \leqslant 0$. Si $g(a)$ ou $g(b)$ est nul, la proposition est démontrée ; nous supposons donc que $g(a) > 0$ et $g(b) < 0$. La proposition découle à présent du théorème des valeurs intermédiaires 11.50. □

Exemple 19.29.
La fonction $x \mapsto \cos(x)$ est continue entre $[-1,1]$ et $[-1,1]$. Elle admet donc un point fixe. Par conséquent il existe (au moins) une solution à l'équation $\cos(x) = x$. △

Proposition 19.30 (Brouwer dans $\mathbb{R}^n$ version C^∞ via Stokes)**.**
Soit B la boule fermée de centre 0 et de rayon 1 de $\mathbb{R}^n$ et $f\colon B \to B$ une fonction C^∞. Alors f admet un point fixe.

Démonstration. Supposons que f ne possède pas de points fixes. Alors pour tout $x \in B$ nous considérons la demi-droite issue de $f(x)$ et passant par x (cette demi-droite existe parce que x et $f(x)$ sont supposés distincts). Cette demi-droite intersecte ∂B en un point que nous appelons $g(x)$. Prouvons que cette fonction est C^k dès que f est C^k (y compris avec $k = \infty$).

Le point $g(x)$ est la solution du système

$$\begin{cases} g(x) - f(x) = \lambda\big(x - f(x)\big) & \text{(19.84a)} \\ \|g(x)\|^2 = 1 & \text{(19.84b)} \\ \lambda \geqslant 0. & \text{(19.84c)} \end{cases}$$

En substituant nous obtenons l'équation

$$P_x(\lambda) = \|\lambda\big(x - f(x)\big) + f(x)\|^2 - 1 = 0, \tag{19.85}$$

ou encore

$$\lambda^2\|x - f(x)\|^2 + 2\lambda\big(x - f(x)\big) \cdot f(x) + \|f(x)\|^2 - 1 = 0. \tag{19.86}$$

En tenant compte du fait que $\|f(x)\| < 1$ (parce que les images de f sont dans $\mathcal{B}$), nous trouvons que $P_x(0) \leqslant 0$ et $P_x(1) \leqslant 0$. De même $\lim_{\lambda\to\infty} P_x(\lambda) = +\infty$. Par conséquent le polynôme de second degré P_x a exactement deux racines distinctes $\lambda_1 \leqslant 0$ et $\lambda_2 \geqslant 1$. La racine que nous cherchons est la seconde. Le discriminant est strictement positif, donc pas besoin d'avoir peur de la racine dans

$$\lambda(x) = \frac{-\big(x - f(x)\big) \cdot f(x) + \sqrt{\Delta_x}}{\|x - f(x)\|^2} \tag{19.87}$$

où

$$\Delta_x = 4\Big(\big(x - f(x)\big) \cdot f(x)\Big)^2 - 4\|x - f(x)\|^2\big(\|f(x)\|^2 - 1\big). \tag{19.88}$$

Notons que la fonction $\lambda(x)$ est C^k dès que f est C^k ; et en particulier elle est C^∞ si f l'est.

En résumé la fonction g ainsi définie vérifie deux propriétés :

(1) elle est C^∞ ;

(2) elle est l'identité sur ∂B.

La suite de la preuve consiste à montrer qu'une telle rétraction sur B ne peut pas exister[7].

Nous considérons une forme de volume ω sur ∂B : l'intégrale de ω sur ∂B est l'aire de ∂B qui est non nulle. Nous avons alors la contradiction suivante :

$$0 < \int_{\partial B} \omega = \int_{\partial B} g^*\omega = \int_B d(g^*\omega) = \int_B g^*(d\omega) = 0 \tag{19.89}$$

Justifications :

— L'intégrale $\int_{\partial B} \omega$ est l'aire de ∂B et est donc non nulle.

— La fonction g est l'identité sur ∂B. Nous avons donc $\omega = g^*\omega$.

— Le lemme 19.19.

— La forme ω est de volume, par conséquent de degré maximum et $d\omega = 0$.

□

Un des points délicats est de se ramener au cas de fonctions C^∞. Pour la régularisation par convolution, voir [387] ; pour celle utilisant le théorème de Weierstrass, voir [388].

Théorème 19.31 (Brouwer dans $\mathbb{R}^n$ version continue)**.**
Soit B la boule fermée de centre 0 et de rayon 1 de $\mathbb{R}^n$ et $f\colon B \to B$ une fonction continue[8]*. Alors f admet un point fixe.*

Démonstration. Nous commençons par définir une suite de fonctions

$$f_k(x) = \frac{f(x)}{1 + \frac{1}{k}}. \tag{19.90}$$

Nous avons $\|f_k - f\|_\infty \leqslant \frac{1}{1+k}$ où la norme est la norme uniforme sur B. Par le théorème de Weierstrass 11.318 il existe une suite de fonctions $C^\infty(B, \mathbb{R})$ que nous nommons g_k telles que

$$\|g_k - f_k\|_\infty \leqslant \frac{1}{1+k}. \tag{19.91}$$

Vérifions que cette fonction g_k soit bien une fonction qui prend ses valeurs dans B :

$$\begin{aligned}
\|g_k(x)\| &\leqslant \|g_k(x) - f_k(x)\| + \|f_k(x)\| && (19.92a)\\
&\leqslant \frac{1}{1+k} + \frac{\|f(x)\|}{1+\frac{1}{k}} && (19.92b)\\
&\leqslant \frac{1}{1+k} + \frac{1}{1+\frac{1}{k}} && (19.92c)\\
&= 1. && (19.92d)
\end{aligned}$$

Par la version C^∞ du théorème (proposition 19.30), g_k admet un point fixe que l'on nomme x_k.

Étant donné que x_k est dans le compact B, quitte à prendre une sous-suite nous supposons que la suite (x_k) converge vers un élément $x \in B$. Nous montrons maintenant que x est un point fixe de f :

$$\begin{aligned}
\|f(x) - x\| &= \|f(x) - g_k(x) + g_k(x) - x_k + x_k - x\| && (19.93a)\\
&\leqslant \|f(x) - g_k(x)\| + \underbrace{\|g_k(x) - x_k\|}_{=0} + \|x_k - x\| && (19.93b)\\
&\leqslant \frac{1}{1+k} + \|x_k - x\|. && (19.93c)
\end{aligned}$$

En prenant le limite $k \to \infty$ le membre de droite tend vers zéro et nous obtenons $f(x) = x$. □

7. Notons qu'il n'existe pas non plus de rétractions continues sur B, mais pour le montrer il faut utiliser d'autres méthodes que Stokes, ou alors présenter les choses dans un autre ordre.

8. Une fonction continue sur un fermé de $\mathbb{R}^n$ est à comprendre pour la topologie induite.

19.5.2 Théorème de Schauder

Une conséquence du théorème de Brouwer est le théorème de Schauder qui est valide en dimension infinie.

Théorème 19.32 (Théorème de Schauder[389])**.**
Soit E, un espace vectoriel normé, K un convexe compact de E et $f\colon K \to K$ une fonction continue. Alors f admet un point fixe.

Démonstration. Étant donné que $f\colon K \to K$ est continue, elle y est uniformément continue. Si nous choisissons ϵ alors il existe $\delta > 0$ tel que

$$\|f(x) - f(y)\| \leqslant \epsilon \tag{19.94}$$

dès que $\|x - y\| \leqslant \delta$. La compacité de K permet de choisir un recouvrement fini par des ouverts de la forme

$$K \subset \bigcup_{1 \leqslant i \leqslant p} B(x_j, \delta) \tag{19.95}$$

où $\{x_1, \dots, x_p\} \subset K$. Nous considérons maintenant $L = \text{Span}\{f(x_j) \text{ tel que } 1 \leqslant j \leqslant p\}$ et

$$K^* = K \cap L. \tag{19.96}$$

Le fait que K et L soient convexes implique que K^* est convexe. L'ensemble K^* est également compact parce qu'il s'agit d'une partie fermée de K qui est compact (lemme 7.64). Notons en particulier que K^* est contenu dans un espace vectoriel de dimension finie, ce qui n'est pas le cas de K.

Nous allons à présent construire une sorte de partition de l'unité subordonnée au recouvrement (19.95) sur K (voir le lemme 19.20). Nous commençons par définir

$$\psi_j(x) = \begin{cases} 0 & \text{si } \|x - x_j\| \geqslant \delta \\ 1 - \frac{\|x - x_j\|}{\delta} & \text{sinon.} \end{cases} \tag{19.97}$$

pour chaque $1 \leqslant j \leqslant p$. Notons que ψ_j est une fonction positive, nulle en-dehors de $B(x_j, \delta)$. En particulier la fonction suivante est bien définie :

$$\varphi_j(x) = \frac{\psi_j(x)}{\sum_{k=1}^p \psi_k(x)} \tag{19.98}$$

et nous avons $\sum_{j=1}^p \varphi_j(x) = 1$. Les fonctions φ_j sont continues sur K et nous définissons finalement

$$g(x) = \sum_{j=1}^p \varphi_j(x) f(x_j). \tag{19.99}$$

Pour chaque $x \in K$, l'élément $g(x)$ est une combinaison des éléments $f(x_j) \in K^*$. Étant donné que K^* est convexe et que la somme des coefficients $\varphi_j(x)$ vaut un, nous avons que g prend ses valeurs dans K^* par la proposition 8.27.

Nous considérons seulement la restriction $g\colon K^* \to K^*$ qui est continue sur un compact contenu dans un espace vectoriel de dimension finie. Le théorème de Brouwer nous enseigne alors que g a un point fixe (proposition 19.31). Nous nommons y ce point fixe. Notons que y est fonction du ϵ choisi au début de la construction, via le δ qui avait conditionné la partition de l'unité.

Nous avons

$$f(y) - y = f(y) - g(y) \tag{19.100a}$$

$$= \sum_{j=1}^p \varphi_j(y) f(y) - \sum_{j=1}^p \varphi_j(y) f(x_j) \tag{19.100b}$$

$$= \sum_{j=1}^p \varphi(j)(y)\big(f(y) - f(x_j)\big). \tag{19.100c}$$

Par construction, $\varphi_j(y) \neq 0$ seulement si $\|y - x_j\| \leqslant \delta$ et par conséquent seulement si $\|f(y) - f(x_j)\| \leqslant \epsilon$. D'autre par nous avons $\varphi_j(y) \geqslant 0$; en prenant la norme de (19.100) nous trouvons

$$\|f(y) - y\| \leqslant \sum_{j=1}^{p} \|\varphi_j(y)\big(f(y) - f(x_j)\big)\| \leqslant \sum_{j=1}^{p} \varphi_j(y)\epsilon = \epsilon. \tag{19.101}$$

Nous nous souvenons maintenant que y était fonction de ϵ. Soit y_m le y qui correspond à $\epsilon = 2^{-m}$. Nous avons alors

$$\|f(y_m) - y_m\| \leqslant 2^{-m}. \tag{19.102}$$

L'élément y_m est dans K^* qui est compact, donc quitte à choisir une sous-suite nous pouvons supposer que y_m est une suite qui converge vers $y^* \in K$ [9]. Nous avons les majorations

$$\|f(y^*) - y^*\| \leqslant \|f(y^*) - f(y_m)\| + \|f(y_m) - y_m\| + \|y_m - y^*\|. \tag{19.103}$$

Si m est assez grand, les trois termes du membre de droite peuvent être rendus arbitrairement petits, d'où nous concluons que

$$f(y^*) = y^* \tag{19.104}$$

et donc que f possède un point fixe. □

19.5.3 Théorème de Cauchy-Arzella

Théorème 19.33 (Cauchy-Arzela[318]).
Nous considérons le système d'équation différentielles

$$\begin{cases} y' = f(t, y) & (19.105a) \\ y(t_0) = y_0. & (19.105b) \end{cases}$$

avec $f\colon U \to \mathbb{R}^n$, continue où U est ouvert dans $\mathbb{R} \times \mathbb{R}^n$. Alors il existe un voisinage fermé V de t_0 sur lequel une solution C^1 du problème (19.105) existe.

Idée de la démonstration. Nous considérons $M = \|f\|_\infty$ et K, l'ensemble des fonctions M-Lipschitz sur U. Nous prouvons que $(K, \|.\|_\infty)$ est compact. Ensuite nous considérons l'application

$$\begin{aligned} \Phi\colon K &\to K \\ \Phi(f)(t) &= x_0 + \int_{t_0}^{t} f\big(u, f(u)\big)du. \end{aligned} \tag{19.106}$$

Après avoir prouvé que Φ était continue, nous concluons qu'elle a un point fixe par le théorème de Schauder 19.32. □

Remarque 19.34.
Quelques remarques.

(1) Les théorème de Cauchy-Lipschitz et Cauchy-Arzella donnent des existences pour des équations différentielles du type $y' = f(t, y)$. Et si nous avons une équation du second ordre ? Alors il y a la méthode de la réduction de l'ordre qui permet de transformer une équation différentielle d'ordre élevé en un système d'ordre 1.

(2) Ces théorèmes posent des *conditions initiales* : la valeur de y est donnée en un point, et la méthode de la réduction de l'ordre permet de donner l'existence de solutions d'un problème d'ordre k en donnant les valeurs de $y(0)$, $y'(0)$, $\dots y^{(k-1)}(0)$. C'est-à-dire de la fonction et de ses dérivées en un point. Rien n'est dit sur l'existence de *conditions aux bords*.

Ces deux points sont illustrés dans les exemples 31.15 et 31.16.

9. Notons que même dans la sous-suite nous avons $\|f(y_m) - y_m\| \leqslant 2^{-m}$, avec le même « m » des deux côtés de l'inégalité.

19.5.4 Théorème de Markov-Kakutani et mesure de Haar

Définition 19.35.
Soit G un groupe topologique. Une ***mesure de Haar*** *sur G est une mesure positive*[10] *μ telle que*

(1) $\mu(gA) = \mu(A)$ pour tout mesurable A et tout $g \in G$,

(2) $\mu(K) < \infty$ pour tout compact $K \subset G$.

Si de plus le groupe G lui-même est compact nous demandons que la mesure soit normalisée : $\mu(G) = 1$.

Le théorème suivant nous donne l'existence d'une mesure de Haar sur un groupe compact.

Théorème 19.36 (Markov-Katutani[390])**.**
Soit E un espace vectoriel normé et L, une partie non vide, convexe, fermée et bornée de E'. Soit $T\colon L \to L$ une application continue. Alors T a un point fixe.

Démonstration. Nous considérons un point $x_0 \in L$ et la suite

$$x_n = \frac{1}{n+1}\sum_{i=0}^{n} T^i x_0. \tag{19.107}$$

La somme des coefficients devant les $T^i(x_0)$ étant 1, la convexité de L montre que $x_n \in L$. Nous considérons l'ensemble

$$C = \bigcap_{n\in\mathbb{N}} \overline{\{x_m \text{ tel que } m \geqslant n\}}. \tag{19.108}$$

Le lemme 7.306 indique que C n'est pas vide, et de plus il existe une sous-suite de (x_n) qui converge vers un élément $x \in C$. Nous avons

$$\lim_{n\to\infty} x_{\sigma(n)}(v) = x(v) \tag{19.109}$$

pour tout $v \in E$. Montrons que x est un point fixe de T. Nous avons

$$\|(Tx_{\sigma(k)} - x_{\sigma(k)})v\| = \Big\|T\frac{1}{1+\sigma(k)}\sum_{i=0}^{\sigma(k)} T^i x_0(v) - \frac{1}{1+\sigma(k)}\sum_{i=0}^{\sigma(k)} T^i x_0(v)\Big\| \tag{19.110a}$$

$$= \Big\|\frac{1}{1+\sigma(k)}\sum_{i=0}^{\sigma(k)} T^{i+1} x_0(v) - T^i x_0(v)\Big\| \tag{19.110b}$$

$$= \frac{1}{1+\sigma(k)}\big\|T^{\sigma(k)+1}x_0(v) - x_0(v)\big\| \tag{19.110c}$$

$$\leqslant \frac{2M}{\sigma(k)+1} \tag{19.110d}$$

où $M = \sum_{y\in L}\|y(v)\| < \infty$ parce que L est borné. En prenant $k \to \infty$ nous trouvons

$$\lim_{k\to\infty}\big(Tx_{\sigma(k)} - x_{\sigma(k)}\big)v = 0, \tag{19.111}$$

ce qui signifie que $Tx = x$ parce que T est continue. □

Le théorème suivant est une conséquence du théorème de Markov-Katutani.

Théorème 19.37.
Si G est un groupe topologique compact possédant une base dénombrable de topologie alors G accepte une unique mesure de Haar normalisée. De plus elle est unimodulaire :

$$\mu(Ag) = \mu(gA) = \mu(A) \tag{19.112}$$

pour tout mesurables $A \subset G$ et tout élément $g \in G$.

10. Définition 13.20.

19.6 Intégrales curvilignes

19.6.1 Chemins de classe C^1

Définition 19.38.
*Soit $p, q \in \mathbb{R}^n$. Un **chemin** C^1 par morceaux joignant p à q est une application continue*

$$\gamma : [a, b] \to \mathbb{R}^n \qquad \gamma(a) = p, \gamma(b) = q \tag{19.113}$$

pour laquelle il existe une subdivision $a = t_0 < t_1 < \ldots < t_{r-1} < t_r = b$ telle que :

(1) la restriction de γ sur chaque ouvert $]t_i, t_{i+1}[$ est de classe C^1 ;

(2) pour tout $0 \leqslant i \leqslant r$, γ' possède une limite à gauche (sauf pour $i = 0$) et une limite à droite (sauf pour $i = r$) en t_i.

*Le **chemin** γ **est (globalement) de classe** C^1 si la subdivision peut être choisie de « longueur » $r = 1$.*

Remarque 19.39.
Si a et b sont des points de $\mathbb{R}^n$, on peut créer le chemin particulier

$$\gamma : [0, 1] \to \mathbb{R}^n : t \mapsto (1 - t)a + tb \tag{19.114}$$

qui relie ces points par un segment de droite.

19.6.2 Intégrer une fonction

Définition 19.40.
*Soit $f : D \subset \mathbb{R}^n \to \mathbb{R}$ une fonction continue, et $\gamma : [a, b] \to D$ un chemin C^1. On définit **l'intégrale de f sur γ** par*

$$\int_\gamma f ds = \int_\gamma f = \int_a^b f(\gamma(t)) \left\|\gamma'(t)\right\| dt. \tag{19.115}$$

Exemple 19.41.
Soit l'hélice

$$\begin{aligned} \sigma \colon [0, 2\pi] &\to \mathbb{R}^3 \\ t &\mapsto \begin{pmatrix} \cos(t) \\ \sin(t) \\ t \end{pmatrix}, \end{aligned} \tag{19.116}$$

et la fonction $f(x, y, z) = x^2 + y^2 + z^2$. L'intégrale de f sur σ est

$$\begin{aligned} \int_\sigma f &= \int_0^{2\pi} (\cos^2 t + \sin^2 t + t^2) \|\sigma'(t)\| dt \\ &= \int_0^{2\pi} (1 + t^2)\sqrt{2} dt \\ &= \sqrt{2} \left[t + \frac{t^3}{3} \right]_0^{2\pi} \\ &= \sqrt{2} \left(2\pi + \frac{8\pi^3}{8} \right). \end{aligned} \tag{19.117}$$

△

Remarque 19.42.
Si $f = 1$, alors nous tombons sur

$$\int_\gamma ds = \int_a^b \|\gamma'(t)\| dt, \tag{19.118}$$

Nous verrons par le théorème 20.10 que cette dernière intégrale est la longueur de la courbe. Il est un fait général que l'intégrale de la fonction 1 sur un ensemble en donne la « mesure ». Cela est à mettre en rapport avec le lemme 13.157 en gardant en tête que $\int_\gamma 1$ n'est pas la mesure de l'image de γ dans $\mathbb{R}^2$.

Proposition 19.43 (Indépendence en le paramétrage)**.**
La valeur de l'intégrale de f sur γ ne dépend pas du paramétrage (équivalent ou pas) choisi.

Démonstration. Soit donc un chemin $\gamma\colon [c,d] \to \mathbb{R}^3$ ainsi que $\varphi\colon [c,d] \to [a,b]$, un reparamétrage de classe C^1, strictement monotone et le chemin σ définit par $\gamma(s) = \sigma\big(\varphi(s)\big)$ avec $s \in [c,d]$. En supposant que $\varphi'(s) \geqslant 0$, nous avons

$$\begin{aligned} I = \int_\gamma f &= \int_c^d f\big(\gamma(s)\big)\|\gamma'(s)\|ds \\ &= \int_c^d f\Big(\sigma\big(\varphi(s)\big)\Big)\|\sigma'\big(\varphi(s)\big)\||\varphi'(s)|ds. \end{aligned} \tag{19.119}$$

Pour cette intégrale, nous posons $t = \varphi(s)$, et par conséquent $dt = \varphi'(s)ds$. Étant donné que $\varphi'(s) \geqslant 0$, nous pouvons supprimer les valeurs absolues, et obtenir

$$\begin{aligned} I &= \int_{\varphi(c)}^{\varphi(d)} f\big(\sigma(t)\big)\|\sigma'(t)\|dt \\ &= \int_a^b f\big(\sigma(t)\big)\|\sigma'(t)\|dt \\ &= \int_\sigma f. \end{aligned} \tag{19.120}$$

Essayez de faire le cas $\varphi'(s) \leqslant 0$. □

Remarque 19.44.
Attention : les intégrales sur des chemins dans $\mathbb{C}$ ne sont la même chose. En effet $\mathbb{C}$ doit être souvent plutôt traité comme $\mathbb{R}$ que comme $\mathbb{R}^2$. Si γ est un chemin dans $\mathbb{C}$, l'intégrale

$$\int_\gamma f \tag{19.121}$$

doit être comprise comme une généralisation de $\int_a^b f(x)dx$ et non comme l'intégrale sur un chemin. La différence est qu'en retournant les bornes d'une intégrale usuelle sur $\mathbb{R}$ on change le signe, alors qu'en retournant un chemin dans $\mathbb{R}^2$, on ne change pas. Bref, la définition est que si $\gamma\colon [a,b] \to \mathbb{C}$ est un chemin, alors

$$\int_\gamma f = \int_\gamma f(z)dz = \int_a^b f\big(\gamma(t)\big)\gamma'(t)dt, \tag{19.122}$$

sans valeur absolue autour de $\gamma'(t)$.

19.6.3 Intégrer un champ de vecteurs

Définition 19.45.
*Un **champ de vecteur** est une application $G : \mathbb{R}^n \to \mathbb{R}^n$. On définit l'intégrale de G sur un chemin $\gamma : [a,b] \to \mathbb{R}^n$ par*

$$\int_\gamma G \stackrel{def}{=} \int_a^b \big\langle G(\gamma(t)), \gamma'(t)\big\rangle\, dt.$$

Remarque 19.46.
Cette définition ne dépend pas de le paramétrage choisie, mais le signe change selon le sens du chemin.

Si $\sigma'(t) \neq 0$, nous pouvons considérer le vecteur unitaire tangent à la courbe :

$$T(t) = \frac{\sigma'(t)}{\|\sigma'(t)\|}. \tag{19.123}$$

Si F est un champ de vecteurs sur $\mathbb{R}^3$, la circulation de F le long de σ sera donnée par

$$\int_\sigma F \cdot ds = \int_a^b F\big(\sigma(t)\big) \cdot \sigma'(t)dt = \int_a^b F\big(\sigma(t)\big) \cdot \frac{\sigma'(t)}{\|\sigma'(t)\|}dt = \int_\sigma F \cdot T ds \tag{19.124}$$

où dans la dernière expression, $F \cdot T$ est vu comme fonction $(x, y, z) \mapsto F(x, y, z) \cdot T(x, y, z)$. L'intégrale d'un champ de vecteurs sur une courbe n'est donc rien d'autre que l'intégrale de la composante tangentielle du champ de vecteurs.

19.6.4 Intégrer une forme différentielle sur un chemin

La formule d'intégration d'un champ de vecteur [11],

$$\int_\gamma G = \int_{[a,b]} \langle G(\gamma(t)), \gamma'(t)\rangle dt, \tag{19.125}$$

contient quelque chose d'intéressant : la combinaison $\langle G(\gamma(t)), \gamma'(t)\rangle$. Cette combinaison sert à transformer le vecteur tangent $\gamma'(t)$ en un nombre en utilisant le produit scalaire avec le vecteur $G(\gamma(t))$.

Si G est un champ de vecteur sur $\mathbb{R}^n$, et si $x \in \mathbb{R}^n$, nous pouvons utiliser l'isomorphisme musical (définition 11.433)

$$\begin{aligned} G_x^\flat \colon \mathbb{R}^n &\to \mathbb{R} \\ v &\mapsto \langle G(x), v\rangle \end{aligned} \tag{19.126}$$

pour écrire de façon plus compacte :

$$\int_\gamma G = \int_{[a,b]} G^\flat_{\gamma(t)}\big(\gamma'(t)\big)dt. \tag{19.127}$$

Définition 19.47.
*Soient une forme différentielle ω sur $\mathbb{R}^n$ et un chemin de classe C^1 $\gamma\colon [a, b] \to \mathbb{R}^n$. L'**intégrale** de ω sur γ est définie par*

$$\int_\gamma \omega = \int_a^b \omega_{\gamma(t)}\gamma'(t)dt \tag{19.128}$$

Remarque 19.48.
Cette définition ne dépend pas de le paramétrage choisie, mais le signe change selon le sens du chemin.

19.6.5 Intégration d'une forme différentielle sur un chemin

Les formes intégrales que nous avons déjà vues sont celles de fonctions et de champs de vecteur sur des chemins. Si $\gamma\colon [a, b] \to \mathbb{R}^n$ est le chemin, les formules sont

$$\begin{aligned} \int_\gamma f &= \int_{[a,b]} f\big(\gamma(t)\big)\|\gamma'(t)\|dt \\ \int_\gamma G &= \int_{[a,b]} \langle G\big(\gamma(t)\big), \gamma'(t)\rangle dt. \end{aligned} \tag{19.129}$$

Dans les deux cas, le principe est que nous disposons de quelque chose (la fonction f ou le vecteur G), et du vecteur tangent $\gamma'(t)$, et nous essayons d'en tirer un nombre que nous intégrons. Lorsque nous avons une 1-forme, la façon de l'utiliser pour produire un nombre avec le vecteur tangent est évidemment d'appliquer la 1-forme au vecteur tangent. La définition suivante est donc naturelle.

11. Définition 19.45.

Définition 19.49.
*Soit $\gamma\colon [a,b] \to \mathbb{R}^n$, un chemin de classe C^1 tel que son image est contenue dans le domaine D. Si ω es une 1-forme différentielle sur D, nous définissons l'**intégrale de ω le long de γ** le nombre*

$$\begin{aligned}\int_\gamma \omega &= \int_a^b \omega_{\gamma(t)}\big(\gamma'(t)\big)dt \\ &= \int_a^b \Big[a_1\big(\gamma(t)\big)\gamma_1'(t) + \cdots + a_n\big(\gamma(t)\big)\gamma_n'(t)\Big]dt.\end{aligned} \tag{19.130}$$

Cette définition est une bonne définition parce que si on change le paramétrage du chemin, on ne change pas la valeur de l'intégrale, c'est la proposition suivante.

Proposition 19.50.
Si γ et β sont des chemins équivalents, alors

$$\int_\gamma \omega = \int_\beta \omega, \tag{19.131}$$

c'est-à-dire que l'intégrale est invariante sous les reparamétrages du chemin.

Démonstration. Deux chemins sont équivalents quand il existe un difféomorphisme C^1 $h\colon [a,b] \to [c,d]$ tel que $\gamma(t) = (\beta \circ h)(t)$. En remplaçant γ par $(\beta \circ h)$ dans la définition de $\int_\gamma \omega$, nous trouvons

$$\int_a^b \omega_{\gamma(t)}\big(\gamma'(t)\big)dt = \int_a^b \omega_{(\beta\circ h)(t)}\big((\beta \circ h)'(t)\big)dt. \tag{19.132}$$

Un changement de variable $u = h(t)$ transforme cette dernière intégrale en $\int_\beta \omega$, ce qui prouve la proposition. □

Remarque 19.51.
Si γ est une somme de chemins, $\gamma = \gamma^{(1)} + \cdots + \gamma^{(n)}$, où chacun des $\gamma^{(i)}$ est un chemin, alors

$$\int_\gamma \omega = \sum_{i=1}^n \int_{\gamma_i} \omega \tag{19.133}$$

parce que ω est linéaire.

Remarque 19.52.
Si $-\gamma$ est le chemin

$$\begin{aligned}-\gamma\colon [a,b] &\to \mathbb{R}^n \\ t &\mapsto \gamma\big(b - (t-a)\big),\end{aligned} \tag{19.134}$$

alors

$$\int_{-\gamma} \omega = -\int_\gamma \omega, \tag{19.135}$$

c'est-à-dire que si l'on parcours le chemin en sens inverse, alors on change le signe de l'intégrale.

L'intégrale d'une forme différentielle sur un chemin est compatible avec l'intégrale déjà connue d'un champ de vecteur sur le chemin parce que si G est un champ de vecteurs,

$$\int_\gamma G^\flat = \int_\gamma G. \tag{19.136}$$

En effet,

$$\begin{aligned}\int_{\gamma G^\flat} &= \int_a^b G^\flat_{\gamma(t)}(\gamma'(t))\\ &= \int_a^b \big[G_1(\gamma(t))dx_1 + \dots G_n(\gamma(t))dx_n\big]\big(\gamma_1'(t),\dots,\gamma_n'(t)\big)\\ &= \int_a^b \langle G(\gamma(t)),\gamma'(t)\rangle\\ &= \int_\gamma G.\end{aligned} \tag{19.137}$$

Proposition 19.53.
Soit $\omega = df$, une 1-forme exacte et continue sur le domaine D. Alors la valeur de $\int_\gamma df$ ne dépend que des valeurs de f aux extrémités de γ.

Démonstration. Nous avons

$$\begin{aligned}\int_\gamma \omega = \int_\gamma df &= \int_a^b \sum_{i=1} n\frac{\partial f}{\partial x_i}(\gamma(t))\gamma_i'(t)dt\\ &= \int_a^b \frac{d}{dt}\Big((f\circ\gamma)(t)\Big)dt\\ &= (f\circ\gamma)(b) - (f\circ\gamma(a)).\end{aligned} \tag{19.138}$$

□

19.6.6 Interprétation physique : travail

Définition 19.54.
Une force $F\colon D\subset\mathbb{R}^n\to\mathbb{R}^n$ est ***conservative*** *si elle dérive d'un potentiel, c'est-à-dire s'il existe une fonction $V\in C^1(D,\mathbb{R})$ telle que*

$$F(x) = (\nabla V)(x). \tag{19.139}$$

Étant donné que F est un champ de vecteurs, nous avons une forme différentielle associée $F^\flat$,

$$F^\flat_x\colon x\mapsto \langle F(x),v\rangle. \tag{19.140}$$

Lemme 19.55.
Le champ F est conservatif si et seulement si la 1-forme différentielle $F^\flat$ est exacte.

Démonstration. Supposons que la force F soit conservative, c'est-à-dire qu'il existe une fonction V telle que $F = \nabla V$. Dans ce cas, il est facile de prouver que $F^\flat$ est exacte et est donnée par $F^\flat_x = dV(x)$. En effet,

$$\begin{aligned}F^\flat_x(v) &= \langle F(x),v\rangle\\ &= F_1(x)v_1 + \dots + F_n(x)v_n\\ &= \frac{\partial V}{\partial x_1}(x)v_1 + \dots \frac{\partial V}{\partial x_n}(x)v_n\\ &= dV(x)v.\end{aligned} \tag{19.141}$$

Pour le sens inverse, supposons que $F^\flat$ soit exacte. Dans ce cas, nous avons une fonction V telle que $F^\flat = dV$. Il est facile de prouver qu'alors, $F = \nabla V$. □

En résumé, nous avons deux façons équivalentes d'exprimer que la force F dérive du potentiel V : soit nous disons $F = \nabla V$, soit nous disons $F^\flat = dV$.

Proposition 19.56.
Si F est une force conservative, alors le travail de F lors d'un déplacement ne dépend pas du chemin suivit.

Démonstration. Le travail d'une force le long d'un chemin n'est autre que l'intégrale de la force le long du chemin, et le calcul est facile :

$$W_\gamma(F) = \int_\gamma F = \int_\gamma dV = V\big(\gamma(b)\big) - V\big(\gamma(a)\big). \tag{19.142}$$

Donc si β est un autre chemin tel que $\beta(a) = \gamma(a)$ et $\beta(b) = \gamma(b)$, nous avons $W_\beta(F) = W_\gamma(F)$. □

19.6.7 Intégrer un champ de vecteurs sur un bord en $2D$

Si $D \subset \mathbb{R}^2$ est tel que ∂D est une variété de dimension 1 et tel que D accepte un champ de vecteur normal extérieur unitaire ν. Si nous voulons définir

$$\int_{\partial D} G, \tag{19.143}$$

le mieux est de prendre un paramétrage $\gamma\colon [0,1] \to \mathbb{R}^2$ et de calculer

$$\int_0^1 \langle G_{\gamma(t)}, \frac{\dot\gamma(t)}{\|\dot\gamma(t)\|}\rangle dt. \tag{19.144}$$

Hélas, cette définition ne fonctionne pas parce que son signe dépend du sens de le paramétrage γ. Si le paramétrage tourne dans l'autre sens, il y a un signe de différence.

Nous allons définir

$$\int_{\partial D} G = \int_0^1 \langle G_{\gamma(t)}, T(t)\rangle dt \tag{19.145}$$

où $T(t) = \dot\gamma(t)/\|\dot\gamma(t)\|$ et où γ est choisi de telle façon que la rotation d'angle $\frac{\pi}{2}$ amène ν sur T. Cela fixe le choix de sens.

Ce choix de sens aura des répercussions dans l'application de la formule de Green et du théorème de Stokes.

19.6.8 Intégrer une forme différentielle sur un bord en $2D$

Nous n'allons pas chercher très loin :

$$\int_{\partial D} \omega = \int_{\partial D} \omega^\sharp, \tag{19.146}$$

c'est-à-dire que l'intégrale de la forme différentielle est celle du champ de vecteur associé. Le membre de droite est définit par (19.145), avec le choix d'orientation qui va avec.

19.6.9 Intégrer une forme différentielle sur un bord en $3D$

Nous allons maintenant intégrer une forme différentielle sur certains chemins fermés dans $\mathbb{R}^3$. Soit $F(D) \subset \mathbb{R}^3$, une variété de dimension 2 dans $\mathbb{R}^3$ où $F\colon D \subset \mathbb{R}^2 \to \mathbb{R}^3$ est la carte. Nous supposons que D vérifie les hypothèses de la formule de Green. Alors nous définissons

$$\int_{F(\partial D)} \omega = \int_{\partial D} F^*\omega \tag{19.147}$$

où $F^*\omega$ est la forme différentielle définie sur ∂D par $(F^*\omega)(v) = \omega\big(dF(v)\big)$.

Cette définition est très abstraite, mais nous n'allons, en pratique, jamais l'utiliser, grâce au théorème de Stokes.

19.6.10 Intégrer un champ de vecteurs sur un bord en $3D$

Encore une fois, nous n'allons pas chercher bien loin :

$$\int_{F(\partial D)} G = \int_{F(\partial D)} G^\flat \tag{19.148}$$

où $G^\flat$ est la forme différentielle associée au champ de vecteur. Le membre de droite est définit par l'équation (19.147).

19.6.11 Dérivées croisées et forme différentielle exacte

Nous considérons le problème suivant : trouver une fonction $f\colon \mathbb{R}^2 \to \mathbb{R}$ telle que

$$\begin{cases} \dfrac{\partial f}{\partial x} = a(x,y) & (19.149a) \\ \dfrac{\partial f}{\partial y} = b(x,y) & (19.149b) \end{cases}$$

où a et b sont des fonctions supposées suffisamment régulières. Nous savons que ce problème n'a pas de solutions lorsque

$$\frac{\partial a}{\partial y} \neq \frac{\partial b}{\partial x} \tag{19.150}$$

parce que cela impliquerait $\partial^2_{xy} f \neq \partial^2_{yx} f$. Nous sommes en droit de nous demander si la condition

$$\frac{\partial a}{\partial y} = \frac{\partial b}{\partial x} \tag{19.151}$$

impliquerait qu'il existe une solution au problème (19.149). La réponse est oui, et nous allons brièvement la justifier. Pour plus de détails nous vous demandons de chercher un peu sur internet les mots-clefs *forme différentielles exacte*. Vous consulterez également avec profit [391].

Proposition 19.57.
Si a et b sont des fonctions qui satisfont à la condition

$$\frac{\partial a}{\partial y} = \frac{\partial b}{\partial x}, \tag{19.152}$$

alors la fonction

$$f(x,y) = \int_0^x a(t,0)dt + \int_0^y b(x,t)dt \tag{19.153}$$

répond au problème

$$\begin{cases} \dfrac{\partial f}{\partial x} = a(x,y) & (19.154a) \\ \dfrac{\partial f}{\partial y} = b(x,y) & (19.154b) \end{cases}$$

La preuve qui suit n'en est pas complètement une parce qu'il manque des justifications, notamment au moment de permuter la dérivée et l'intégrale.

Démonstration. La clef de la preuve est le théorème fondamental de l'analyse :

$$\int_0^x \frac{\partial f}{\partial x}(t,y)dt = f(x,y) \tag{19.155}$$

et son pendant par rapport à y :

$$\int_0^y \frac{\partial f}{\partial y}(x,t)dt = f(x,y). \tag{19.156}$$

En appliquant ces version du théorème fondamental, nous obtenons immédiatement.

$$\frac{\partial f}{\partial y} = b(x,y). \tag{19.157}$$

En ce qui concerne la dérivée par rapport à y,

$$\begin{aligned} \frac{\partial f}{\partial x} &= a(x,0) + \int_0^y \frac{\partial b}{\partial x}(x,t)dt & (19.158a) \\ &= a(x,0) + \int_0^y \frac{\partial a}{\partial y}(x,t)dt & (19.158b) \\ &= a(x,0) + [a(x,t)]_{t=0}^{t=y} & (19.158c) \\ &= a(x,y). & (19.158d) \end{aligned}$$

□

En ce qui concerne l'unicité, supposons que f et g soient deux solutions au problème. L'équation

$$\frac{\partial f}{\partial x} = a(x,y) = \frac{\partial g}{\partial x} \tag{19.159}$$

implique que

$$f(x,y) = g(x,y) + C(y) \tag{19.160}$$

où C est une fonction seulement de y. L'autre équation implique

$$f(x,y) = g(x,y) + D(x) \tag{19.161}$$

où D est seulement une fonction de x. En égalisant nous voyons que les fonctions C et D doivent être des constantes.

Par conséquent la fonction f est donnée à une constante près et en réalité la fonction (19.153) est suffisante pour répondre au problème de trouver toutes les fonctions dont les dérivées partielles sont données par les fonctions a et b.

La fonction f ainsi créée est un **potentiel** pour le champ de force

$$F(x,y) = \begin{pmatrix} a(x,y) \\ b(x,y) \end{pmatrix}. \tag{19.162}$$

Notez que ce champ de vecteurs est le gradient de f. La question initiale aurait donc pu être posée en les termes suivants : trouver une fonction f dont le gradient est donné par

$$\nabla f = \begin{pmatrix} a(x,y) \\ b(x,y) \end{pmatrix}. \tag{19.163}$$

19.7 Surfaces paramétrées

De la même façon qu'un chemin dans $\mathbb{R}^3$ est décrit comme une application $\sigma\colon \mathbb{R} \to \mathbb{R}^3$, une surface dans $\mathbb{R}^3$ sera vue comme une application $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^3$. Une **surface paramétrée** dans $\mathbb{R}^3$ est une application

$$\begin{aligned} \varphi\colon D \subset \mathbb{R}^2 &\to \mathbb{R}^3 \\ (u,v) &\mapsto \varphi(u,v) = \begin{pmatrix} x(u,v) \\ y(u,v) \\ z(u,z) \end{pmatrix}. \end{aligned} \tag{19.164}$$

Nous allons parler de la « surface φ » pour désigner l'image de φ dans $\mathbb{R}^3$.

Si on fixe le paramètre $u - u_0$, alors l'application

$$v \mapsto \varphi(u_0, v) \tag{19.165}$$

est un chemin dans la surface. Un vecteur tangent à ce chemin sera tangent à la courbe :

$$\frac{\partial \varphi}{\partial v}(u_0, v_0) = \begin{pmatrix} \frac{\partial x}{\partial v}(u_0, v_O) \\ \frac{\partial y}{\partial v}(u_0, v_O) \\ \frac{\partial z}{\partial v}(u_0, v_O) \end{pmatrix}. \tag{19.166}$$

De même, en fixant v_0, on considère le chemin

$$u \mapsto \varphi(u, v_0). \tag{19.167}$$

Le vecteur tangent à ce chemin est égalent tangent à la surface :

$$\frac{\partial \varphi}{\partial u} = \begin{pmatrix} \frac{\partial x}{\partial u}(u_0, v_O) \\ \frac{\partial y}{\partial u}(u_0, v_O) \\ \frac{\partial z}{\partial u}(u_0, v_O) \end{pmatrix} \tag{19.168}$$

Définition 19.58.
Nous disons que la surface est ***régulière*** *si les vecteurs $\partial_u\varphi(u_0, v_0)$ et $\partial_v\varphi(u_0, v_0)$ sont non nuls et non colinéaires.*

Si la surface est régulière, les vecteurs tangents à le paramétrage forment le plan tangent à la surface au point $\varphi(u_0, v_0)$.

Un vecteur orthogonal à la surface (et donc au plan tangent) est donc donné par le produit vectoriel :

$$n(u_0, v_0) = \frac{\partial\varphi}{\partial u}(u_0, v_0) \times \frac{\partial\varphi}{\partial v}(u_0, v_0). \tag{19.169}$$

L'équation du plan tangent est alors obtenue par

$$\begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix} \cdot n(u_0, v_0) = 0 \tag{19.170}$$

où $x_0 = x(u_0, v_0)$, $y_0 = y(u_0, v_0)$, $z_0 = z(u_0, v_0)$.

19.7.1 Graphe d'une fonction

Soit la fonction $f\colon D \subset \mathbb{R}^2 \to \mathbb{R}$. Le graphe de f est l'ensemble des points de la forme

$$\big(x, y, f(x, y)\big) \tag{19.171}$$

tels que $(x, y) \in D$. Cela est une surface paramétrée par

$$\begin{aligned} \varphi\colon D &\to \mathbb{R}^3 \\ (x, y) &\mapsto \begin{pmatrix} x \\ y \\ f(x, y) \end{pmatrix}. \end{aligned} \tag{19.172}$$

Les vecteurs tangents sont

$$\frac{\partial\varphi}{\partial x} = \begin{pmatrix} 1 \\ 0 \\ \frac{\partial f}{\partial x} \end{pmatrix}, \quad \frac{\partial\varphi}{\partial y} = \begin{pmatrix} 0 \\ 1 \\ \frac{\partial\varphi}{\partial y} \end{pmatrix}. \tag{19.173}$$

La surface est donc partout régulière parce que ces deux vecteurs ne sont jamais nuls ou colinéaires. Un vecteur normal à cette surface au point $(x_0, y_0, f(x_0, y_0))$ est donné par le produit vectoriel

$$n = \begin{vmatrix} e_x & e_y & e_z \\ 1 & 0 & \partial_x f(x_0, y_0) \\ 0 & 1 & \partial_y f(x_0, y_0) \end{vmatrix} = -\frac{\partial f}{\partial x}(x_0, y_0)e_x - \frac{\partial f}{\partial y}(x_0, y_0)e_y + e_z. \tag{19.174}$$

En suivant l'équation (19.170), nous avons l'équation suivante pour le plan :

$$\begin{pmatrix} x - x_0 \\ y - y_0 \\ z - z_0 \end{pmatrix} \cdot \begin{pmatrix} -\frac{\partial f}{\partial x}(x_0, y_0) \\ -\frac{\partial f}{\partial y}(x_0, y_0) \\ 1 \end{pmatrix} = 0, \tag{19.175}$$

c'est-à-dire

$$-(x - x_0)\frac{\partial f}{\partial x}(x_0, y_0) - (y - y_0)\frac{\partial f}{\partial y}(x_0, y_0) + z - f(x_0, y_0) = 0, \tag{19.176}$$

ce qui revient à

$$z - f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0). \tag{19.177}$$

Nous retrouvons donc l'équation du plan tangent à un graphe.

Exemple 19.59.
La sphère de rayon R peut être paramétrée par les angles sphériques :

$$\phi(\theta,\varphi) = \begin{pmatrix} R\sin\theta\cos\varphi \\ R\sin\theta\sin\varphi \\ R\cos\theta \end{pmatrix} \tag{19.178}$$

avec $(\theta,\varphi) \in [0,\pi] \times [0,2\pi]$.

Tentons d'en trouver le plan tangent au point $(x,y,z) = (R,0,0)$. Un petit dessin nous montre que c'est un plan vertical d'équation $x = R$. Montrons cela en utilisant la théorie que nous venons de découvrir. D'abord le point $(R,0,0)$ correspond à $\theta_0 = \frac{\pi}{2}$ et $\varphi = 0$. Les vecteurs tangents sont

$$T_\theta = \frac{\partial\phi}{\partial\theta}(R,\frac{\pi}{2},0) = \begin{pmatrix} R\cos\theta\cos\varphi \\ R\cos\theta\sin\varphi \\ -R\sin\theta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -R \end{pmatrix}, \tag{19.179}$$

et

$$T_\varphi = \frac{\partial\phi}{\partial\varphi}(R,\frac{\pi}{2},0) = \begin{pmatrix} -R\sin\theta\sin\varphi \\ R\sin\theta\cos\varphi \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ R \\ 0 \end{pmatrix}. \tag{19.180}$$

Cela sont de toute évidence bien les deux vecteurs tangents à la sphère au point $(x,y,z) = (R,0,0)$. Le vecteur normal est

$$\begin{vmatrix} e_x & e_y & e_z \\ 0 & 0 & -R \\ 0 & R & 0 \end{vmatrix} = R^2 e_x. \tag{19.181}$$

Ici encore, nous avons le vecteur que nous attendions sur un dessin. L'équation du plan tangent est maintenant

$$\begin{pmatrix} x-R \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} R^2 \\ 0 \\ 0 \end{pmatrix} = 0, \tag{19.182}$$

c'est-à-dire $R^2(x-R) = 0$ et donc $x = R$. △

19.7.2 Intégrale sur une partie de $\mathbb{R}^m$

Soit M une variété de dimension n dans $\mathbb{R}^m$. Soit $F : W \subset \mathbb{R}^n \to M$ un paramétrage d'un ouvert relatif de M.

Si f est une fonction définie sur un sous-ensemble $A \subset F(W)$ tel que $F^{-1}(A)$ est mesurable, l'*intégrale de f sur A* est définie par

$$\int_A f = \int_{F^{-1}(A)} f(F(w))\sqrt{\det(J_F(w)^t J_F(w))}dw$$

où l'intégrale est l'intégration usuelle (de Lebesgue) sur $F^{-1}(A) \subset \mathbb{R}^n$. On écrit parfois cette intégrale $\int_{F^{-1}(A)} f(F(w))d\sigma$ où

$$d\sigma = \sqrt{\det(J_F(w)^t J_F(w))}dw \tag{19.183}$$

est l'*élément infinitésimal de volume* de la variété.

Si $m = 3$ et $n = 2$, l'élément infinitésimal de volume vaut

$$d\sigma = \left\| \frac{\partial F}{\partial w_1} \times \frac{\partial F}{\partial w_2} \right\| dw$$

où $\times$ représente le produit vectoriel dans $\mathbb{R}^3$, et (w_1,w_2) sont les coordonnées sur $W \subset \mathbb{R}^2$. Dans la suite, nous ne regarderons plus que ce cas.

19.8 Intégrales de surface

19.8.1 Intégrale d'un champ de vecteurs

Dans l'intégration curviligne, on a noté que si l'intégrale d'une fonction ne dépendait pas de l'orientation du chemin, l'intégrale d'un champ de vecteurs ou d'une forme différentielle en dépendait. Ce problème d'orientation apparait également dans l'intégration sur des surfaces de l'espace.

Définition 19.60.
Une ***orientation*** *sur une surface $S \subset \mathbb{R}^3$ est le choix d'un champ de vecteurs continu $\nu : S \to \mathbb{R}^3$ dont la norme en tout point de S vaut 1.*

On remarque qu'ayant fait un tel choix d'orientation $\nu(x)$ en un point x, le seul autre choix possible en x est $-\nu(x)$. Si S est le bord d'un ouvert $D \subset \mathbb{R}^3$, l'*orientation induite par D sur S* est, si elle existe, l'orientation qui pointe hors de D en tout point de S. Plus précisément, il faut que pour tout $x \in D$ il existe $\epsilon > 0$ vérifiant, pour tout $0 < t < \epsilon$, la relation $t\nu(x) \notin D$. Dans ce cas, le champ de vecteurs ν est appelé le *vecteur normal unitaire extérieur* à D et il est forcément unique.

Soit G un champ de vecteurs défini sur une surface orientée par un champ ν. L'intégrale de G sur S, aussi appelée le *flux de G à travers S*, est

$$\iint_S G \cdot dS \stackrel{\text{def}}{=} \iint_S \langle G, \nu \rangle \, d\sigma. \tag{19.184}$$

Si on suppose que la surface est paramétrée par une application

$$F : W \subset \mathbb{R}^2 \to \mathbb{R}^3 : (u, v) \mapsto (F_1(u, v), F_2(u, v), F_3(u, v))$$

alors un vecteur unitaire ν peut s'écrire sous la forme

$$\nu = \frac{\frac{\partial F}{\partial u} \times \frac{\partial F}{\partial v}}{\left\| \frac{\partial F}{\partial u} \times \frac{\partial F}{\partial v} \right\|}$$

et grâce à ce paramétrage l'intégrale (19.184) devient

$$\iint_S G \cdot dS = \iint_W \left\langle G(F(u, v)), \frac{\partial F}{\partial u} \times \frac{\partial F}{\partial v} \right\rangle du dv.$$

où on utilise l'expression de $d\sigma$ obtenue précédemment dans le cas qui nous intéresse (surface dans l'espace).

19.9 Intégrales de surface

19.9.1 Aire d'une surface paramétrée

Lorsque nous avions vu la longueur d'une courbe paramétrée, nous avions pris comme « élément de longueur » la norme du vecteur tangent. Il est donc naturel de prendre comme « élément de surface » une petite surface que l'on peut construire à partir des deux vecteurs tangents à la surface.

Au point $\varphi(u_0, v_0)$, nous avons les deux vecteurs tangents

$$T_u = \frac{\partial \varphi}{\partial u}(u_0, v_0) \quad T_v = \frac{\partial \varphi}{\partial v}(u_0, v_0). \tag{19.185}$$

L'élément de surface que nous pouvons construire à partir de ces deux vecteurs est la surface du parallélogramme, donnée par la norme du produit vectoriel :

$$dS = \|T_u \times T_v\|. \tag{19.186}$$

L'aire de la surface donné par $\varphi\colon D \subset \mathbb{R}^2 \to \mathbb{R}^3$ sera donc donnée par

$$Aire\big(\varphi(D)\big) = \iint_D \|T_u \times T_v\| du\, dv. \tag{19.187}$$

Exemple 19.61.
Calculons l'aire de la sphère. Les vecteurs tangents ont déjà été calculés aux équations (19.179) et (19.180) :

$$T_\theta = \begin{pmatrix} R\cos\theta\cos\varphi \\ R\cos\theta\sin\varphi \\ -R\sin\theta \end{pmatrix}, \quad T_\varphi = \begin{pmatrix} -R\sin\theta\sin\varphi \\ R\sin\theta\cos\varphi \\ 0 \end{pmatrix}. \tag{19.188}$$

Le produit vectoriel vaut

$$\begin{aligned} T_\theta \times T_\varphi &= \begin{vmatrix} e_x & e_y & e_z \\ R\cos\theta\cos\varphi & R\cos\theta\sin\varphi & -R\sin\theta \\ -R\sin\theta\sin\varphi & R\sin\theta\cos\varphi & 0 \end{vmatrix} \\ &= (R^2\sin^2\theta\cos\varphi)e_x + (R^2\sin^2\theta\sin\varphi)e_y \\ &\quad + (R^2\cos\theta\sin\theta\cos^2\varphi + R^2\sin\theta\cos\theta\sin^2\varphi)e_z. \end{aligned} \tag{19.189}$$

La norme demande quelques calculs et mises en évidences. Le résultat est :

$$\|T_\theta \times T_\varphi\| = R^2\sin\theta. \tag{19.190}$$

L'aire de la sphère est donc donnée par

$$Aire = \int_0^{2\pi} d\varphi \int_0^\pi R^2\sin\theta d\theta = 2\pi R^2[-\cos\theta]_0^\pi = 4\pi R^2. \tag{19.191}$$

Il est bon de se souvenir que, en coordonnées sphériques,

$$\|T_\theta \times T_\varphi\| = R^2\sin\theta. \tag{19.192}$$

Or nous savons que ce vecteur est dirigé dans le sens de e_r parce que ce dernier est le vecteur qui est constamment dirigé radialement. En coordonnées sphériques nous avons donc

$$T_\theta \times T_\varphi = R^2\sin(\theta)e_r. \tag{19.193}$$

△

Remarque 19.62.
L'équation (19.190) donne l'élément de surface pour la sphère. Notez que cela est justement l'expression du jacobien des coordonnées sphériques. Cela n'est évidemment pas une coïncidence.

Exemple 19.63.
Nous pouvons donner l'aire du graphe d'une fonction quelconque. La surface est paramétrée par

$$\varphi(x,y) = \begin{pmatrix} x \\ y \\ f(x,y) \end{pmatrix}. \tag{19.194}$$

Les vecteurs tangents sont

$$T_x = \begin{pmatrix} 1 \\ 0 \\ \partial_x f \end{pmatrix}, \quad T_y = \begin{pmatrix} 0 \\ 1 \\ \partial_y f \end{pmatrix}. \tag{19.195}$$

Le produit vectoriel est donné par

$$T_x \times T_y = \begin{vmatrix} e_x & e_y & e_z \\ 1 & 0 & \partial_x f \\ 0 & 1 & \partial_y f \end{vmatrix} = (-\partial_x f)e_x - (\partial_y f)e_y + e_z. \tag{19.196}$$

L'élément de surface est par conséquent

$$dS = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + 1}, \tag{19.197}$$

et la surface du graphe sera

$$Aire = \iint_D \sqrt{\left(\frac{\partial f}{\partial x}(x,y)\right)^2 + \left(\frac{\partial f}{\partial y}(x,y)\right)^2 + 1}\, dx\, dy \tag{19.198}$$

△

19.9.2 Intégrale d'une fonction sur une surface

Si S est une surface dans $\mathbb{R}^3$ paramétrée par

$$\begin{aligned} \varphi\colon D &\to \mathbb{R}^3 \\ (u,v) &\mapsto \varphi(u,v) \in S, \end{aligned} \tag{19.199}$$

et si f est une fonction $f\colon \mathbb{R}^3 \to \mathbb{R}$ définie au moins sur S, l'intégrale de f sur S est logiquement définie par

$$\int_S f\, dS = \iint_D f\big(\varphi(u,v)\big)\|T_u(u,v) \times T_v(u,v)\| du dv \tag{19.200}$$

où $T_u = \frac{\partial \varphi}{\partial u}$ et $Y_v = \frac{\partial \varphi}{\partial v}$. La quantité

$$\|T_u(u,v) \times T_v(u,v)\| du dv \tag{19.201}$$

est appelé **élément de surface**.

Encore une fois, si on prend $f = 1$, alors on retrouve la surface de S :

$$\int_S dS = Aire(S). \tag{19.202}$$

Remarque 19.64.
Le nombre $\int_S f dS$ ne dépend pas de le paramétrage choisie pour S.

19.9.3 Intégrale d'une 2-forme

Nous considérons ω, une 2-forme différentielle sur $\mathbb{R}^2$.

Définition 19.65.
Si $\omega_{(x,y)} = u(x,y)dx \wedge dy$ et si D est un ouvert de $\mathbb{R}^2$ alors nous définissons

$$\int_D \omega = \int_D u(x,y) dx\, dy. \tag{19.203}$$

Nous voulons maintenant intégrer une 2-forme sur une surface dans $\mathbb{R}^3$. Soit $S \subset \mathbb{R}^3$, une surface orientée (c'est-à-dire que nous avons un choix continu d'un vecteur normal unitaire n). Nous supposons de plus avoir un paramétrage $\phi\colon D \to S$ de S avec D ouvert dans $\mathbb{R}^2$ compatible avec l'orientation, c'est-à-dire que pour tout $(t,s) \in D$,

$$n\big(\phi(t,s)\big) = \frac{\partial \phi}{\partial t}(t,s) \times \frac{\partial \phi}{\partial s}(t,s). \tag{19.204}$$

Définition 19.66.
Pour intégrer ω sur S nous faisons

$$\int_S \omega = \int_D \phi^* \omega \tag{19.205}$$

où $\phi^*\omega$ est de la forme $F(t,s)dt \wedge ds$.

Montrons ce que cela fait. Soient u, v des vecteurs de D et calculons

$$(\phi^*\omega)(u,v) = \omega\big(d\phi(u), d\phi(v)\big) \tag{19.206a}$$

$$= \omega\left(u_1\frac{\partial\phi}{\partial x_1} + u_2\frac{\partial\infty}{\partial x_2}, v_1\frac{\partial\phi}{\partial x_1} + v_2\frac{\partial\infty}{\partial x_2}\right). \tag{19.206b}$$

Les termes en u_1v_1 et u_2v_2 sont nuls ; par exemple :

$$\omega\left(u_1\frac{\partial\phi}{\partial x_1}, v_1\frac{\partial\phi}{\partial x_1}\right) = u_1v_1\omega\left(\frac{\partial\infty}{\partial x_1}, \frac{\partial\phi}{\partial x_1}\right) = 0 \tag{19.207}$$

parce que ω est antisymétrique. Il nous reste donc

$$(\phi^*\omega)(u,v) = (u_1v_2 - u_2v_1)\omega\left(\frac{\partial\phi}{\partial x_1}, \frac{\partial\phi}{\partial x_2}\right) \tag{19.208a}$$

$$= \omega\left(\frac{\partial\phi}{\partial x_1}, \frac{\partial\phi}{\partial x_2}\right)(dt\wedge ds)(u,v). \tag{19.208b}$$

Cette dernière ligne est bien de la forme $\phi^*\omega = F(t,s)dt\wedge ds$.

19.10 Flux d'un champ de vecteurs à travers une surface

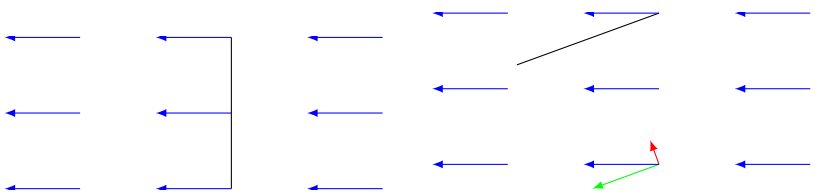
Nous voulons construire un moulin à eau. Comment placer les pales pour maximiser le travail de la pression de l'eau ? On n'a pas attendu l'invention du calcul intégral pour répondre à cette question. Trois paramètres rentrent en ligne de compte :

(1) plus il y a d'eau, plus ça pousse ;

(2) plus la surface de la palle est grande, plus on va utiliser d'eau ;

(3) plus la palle est perpendiculaire au courant, plus on va en profiter.

Nous voyons sur la figure 19.2 que lorsque la palle du moulin est inclinée, non seulement elle prend moins d'eau sur elle, mais qu'en plus elle la prend avec un moins bon angle : une partie de la force ne sert pas à la faire tourner.



(a) L'eau pousse bien perpendiculairement à la palle du moulin.

(b) Lorsque l'eau ne pousse pas perpendiculairement, une partie de la force est perdue. Ici la partie normale (en rouge) est très petite.

FIGURE 19.2 – La partie rouge de la force est perdue si l'eau ne pousse pas perpendiculairement. De plus lorsque la palle est inclinée, elle prend moins d'eau sur elle.

L'idée du flux d'un champ de vecteurs à travers une surface est de savoir quelle est la quantité « utile » de vecteurs qui traverse la surface. Ce sera simplement l'intégrale sur la surface de la composante du champ de vecteurs normale à la surface. Il reste deux problèmes à régler : le premier est de savoir quel est le vecteur normal à la surface, et le second est de savoir comment « sélectionner » la composante normale d'un champ de vecteurs F.

Le problème de trouver un vecteur normal est résolu par le produit vectoriel des vecteurs tangents. Si la surface est donnée par $\varphi\colon D\subset\mathbb{R}^2\to\mathbb{R}^3$, les vecteurs tangents sont $T_u = \partial_u\varphi(u,v)$ et $T_v = \partial_v\varphi(u,v)$. Le normal de norme 1 est donné par :

$$n(u,v) = \frac{T_u\times T_v}{\|T_u\times T_v\|}. \tag{19.209}$$

Si p est un point de la surface $\varphi(D)$, la composante de $F(p)$ qui est normale à la surface au point p est donnée par le produit scalaire

$$F(p)_\perp = F(p) \cdot n(p). \tag{19.210}$$

C'est ce nombre là que nous intégrons sur la surface.

Définition 19.67.
*Le **flux du champ de vecteurs** à travers la surface $S = \varphi(D)$ est*

$$\int F \cdot dS = \int F \cdot n \, dudv. \tag{19.211}$$

Une petite simplification se produit lorsqu'on veut calculer effectivement cette intégrale. En effet $F \cdot n$ est, en soi, une fonction sur S. Pour l'intégrer, il faut donc la multiplier par $\|T_u \times T_v\|$ (c'est la définition de l'intégrale d'une fonction sur une surface). Donc, étant donné que $n = (T_u \times T_v)/\|T_u \times T_v\|$, nous avons

$$\int F \cdot dS = \iint_D F\big(\varphi(u,v)\big) \cdot (T_u \times T_v) \, dudv \tag{19.212}$$

où $T_u = \frac{\partial \phi}{\partial u}$ et $T_v = \frac{\partial \varphi}{\partial v}$.

Exemple 19.68.
Soit le champ de vecteurs

$$F = \begin{pmatrix} 2x \\ 2y \\ 2z \end{pmatrix}. \tag{19.213}$$

Calculons son flux au travers de la sphère de rayon R.

Nous choisissons de paramétrer la sphère en coordonnées sphériques avec $\phi(\theta, \varphi)$. Nous pouvons reprendre le résultat (19.193) :

$$T_\theta \times T_\varphi = R^2 \sin(\theta). \tag{19.214}$$

Nous savons aussi que

$$F\big(\phi(\theta,\varphi)\big) = 2e_r. \tag{19.215}$$

L'intégrale à calculer est donc

$$I = \int_0^\pi d\theta \int_0^{2\pi} d\varphi \, 2e_r \cdot \big(R^2 \sin(\theta) e_r\big). \tag{19.216}$$

Vu que le produit scalaire $e_r \cdot e_r$ vaut 1, nous calculons

$$I = 4\pi R^2 \int_0^\pi \sin(\theta) d\theta = 8\pi R^2. \tag{19.217}$$

△

Exemple 19.69.
Calculons le flux du champ de force de gravitation d'une masse au travers de la sphère de centre R centrée autour la masse. À un coefficient constant près, le champ vaut

$$G(r,\theta,\varphi) = \frac{1}{r^2} e_r. \tag{19.218}$$

Sur la sphère de rayon R, nous avons

$$G\big(\phi(\theta,\varphi)\big) = \frac{1}{R^2} e_r. \tag{19.219}$$

L'intégrale est donc

$$\int_0^\pi d\theta \int_0^{2\pi} \frac{1}{R^2} e_r \cdot \big(R^2 \sin(\theta) e_r\big) d\varphi = 8\pi. \tag{19.220}$$

Ce flux ne dépend pas de R. △

Exemple 19.70.
Soit S le disque de rayon 5 placé horizontalement à la hauteur 12. Calculer le flux du champ de vecteurs

$$F(x,y,z) = xe_x + ye_y + ze_z. \tag{19.221}$$

Les équations de la surface sont $z = 12$, $x^2+y^2 \leqslant 25$. Nous prenons le paramétrage en coordonnées cylindriques :

$$\varphi(r,\theta) = \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \\ 12 \end{pmatrix}. \tag{19.222}$$

Les vecteurs tangents sont

$$T_r = \frac{\partial\varphi}{\partial r} = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix} \quad T_\theta = \frac{\partial\varphi}{\partial\theta} = \begin{pmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{pmatrix}. \tag{19.223}$$

Le vecteur normal est alors

$$T_r \times T_\theta = re_z. \tag{19.224}$$

Sur la surface, le champ de vecteurs s'écrit

$$F\big(\varphi(r,\theta)\big) = r\cos(\theta)e_x + r\sin(\theta)e_y + 12e_z. \tag{19.225}$$

Par conséquent

$$F \cdot (T_r \times T_\theta) = 12r. \tag{19.226}$$

L'intégrale à calculer est

$$\begin{aligned} \int_0^5 dr \int_0^{2\pi} 12r\,d\theta &= 12 \cdot 2\pi \int_0^5 r\,dr \\ &= \frac{25}{2}24\pi \\ &= 300\pi. \end{aligned} \tag{19.227}$$

△

19.11 Divergence, Green, Stokes

Le théorème de Stokes (et ses variations) peut se voir comme une généralisation du théorème fondamental du calcul différentiel et intégral qui stipule que

$$\int_a^b f'(x)dx = f(b) - f(a)$$

c'est-à-dire qui relie l'intégrale de f' sur $I = [a,b]$ aux valeurs de f sur le bord $\partial I = \{a,b\}$. Le signe $-$ qui apparait vient de l'orientation ; celle-ci requiert de la prudence dans l'utilisation des théorèmes.

Voici, pour votre culture générale, un énoncé général :

Théorème 19.71.
Si M est une variété orientable de dimension n avec un bord noté ∂M, alors pour toute forme différentielle ω de degré $n-1$ on a

$$\int_M d\omega = \int_{\partial M} \omega.$$

où $d\omega$ désigne la différentielle extérieure de ω.

Nous allons maintenant voir quelques cas particuliers.

Une des nombreuses formes du théorème de Stokes (théorème 19.71) est que si la forme différentielle ω est exacte alors son intégrale est facile.

Théorème 19.72 ([1]).
Si γ est une chemin de classe C^1 dans un ouvert Ω et si ω est la forme différentielle exacte $\omega = df$, alors

$$\int_\gamma df = f\big(\gamma(1)\big) - f\big(\gamma(0)\big). \tag{19.228}$$

Démonstration. C'est une application du lemme 11.223 et du théorème fondamental du calcul intégral 13.236. □

19.11.1 Théorème de la divergence

Si nous considérons une surface dans $\mathbb{R}^n$ et un champ de vecteurs, il est bon de se demander quelle « quantité de vecteurs » traverse la surface. Soit D, un ouvert borné de $\mathbb{R}^n$ telle que ∂D soit une variété de dimension $n-1$, et G, un champ de vecteurs défini sur $\bar{D}$. Afin de compter combien de G traverse ∂D, il faudra faire en sorte de ne considérer que la composante de G normale à ∂D : pas question d'intégrer par exemple la norme de G sur ∂D.

Comme nous le savons, la composante du vecteur v dans la direction w est le produit scalaire $v \cdot 1_w$ où 1_w est le vecteur de norme 1 dans la direction w. Nous allons donc introduire le concept de vecteur normal extérieur.

Définition 19.73.
Soit $x \in \partial D$ et $\nu \in \mathbb{R}^n$, nous disons que ν est un ***vecteur normal extérieur*** *de ∂D si*

(1) $\langle \nu, v\rangle = 0$ pour tout vecteur tangent v à ∂D au point x. Pour rappel, ∂D étant une variété de dimension $n-1$, il y a $n-1$ tels vecteurs v linéairement indépendants.

(2) Il existe un $\delta > 0$ tel que $\forall t \in \,]0, \delta[$, nous avons $c + t\nu \notin \bar{D}$ et $x - t\nu \in D$.

Nous pouvons maintenant définir le concept de flux. Soit $D \subset \mathbb{R}^n$ tel que ∂D soit une variété de dimension $n-1$ qui admette un vecteur normal extérieur $\nu(x)$ en chaque point. Soit aussi $G\colon \bar{D} \to \mathbb{R}^n$, un champ de vecteur de classe C^1. Le **flux** de G au travers de ∂D est le nombre

$$\int_{\partial D} \langle G(x), \nu(x)\rangle d\sigma(x). \tag{19.229}$$

Cette intégrale est en général très compliquée à calculer parce qu'il faut trouver le champ de vecteur normal, puis un paramétrage de la surface ∂D et ensuite appliquer la méthode décrite au point 19.8.

Heureusement, il y a un théorème qui nous permet de calculer plus facilement : sans devoir trouver de vecteurs normaux.

Il n'est pas plus contraignant d'énoncer ce théorème dans le cadre d'une hypersurface de $\mathbb{R}^n$, ce que nous faisons donc :

Théorème 19.74 (Formule de la divergence).
Soit D un ouvert borné de $\mathbb{R}^n$ dont le bord est « assez régulier par morceaux », c'est-à-dire :

$$\partial D = A_1 \cup \ldots A_p \cup N \tag{19.230}$$

où

(1) $A_1, \ldots, A_p, N$ sont deux à deux disjoints,

(2) pour tout $i \leqslant p$, A_i est un ouvert relatif d'une certaine variété M_i de dimension $(n-1)$

(3) $\bar{A}_i \subset M_i$

(4) N est un compact contenu dans une réunion finie de variétés de dimensions $(n-2)$.

Supposons également qu'en chaque point de $A_1 \cup \ldots \cup A_p$ il existe un vecteur normal extérieur ν.

Si G est un champ de vecteurs de classe C^1 sur $\bar{D}$ alors

$$\int_D \nabla \cdot G = \sum_{i=1}^{p} \int_{A_i} \langle G, \nu\rangle. \tag{19.231}$$

L'intégrale du membre de gauche est l'intégrale sur un ouvert d'une simple fonction.

19.11.2 Formule de Green

Pour rappel, une chemin $\gamma\colon [0,1] \to \mathbb{R}^n$ est **régulier** s'il est C^1 et si $\gamma(t) \neq 0$ pour tout r. Le chemin est de **Jordan** si $\gamma(1) = \gamma(0)$ et si $\gamma\colon [a,b[\to \mathbb{R}^n$ est injective.

La formule de Green est un cas particulier du théorème de la divergence dans le cas $n = 2$, légèrement reformulé :

Théorème 19.75.
Soit $D \subset \mathbb{R}^2$ ouvert borné tel que son bord est est la réunion finie d'un certain nombre de chemins de classe C^1 de Jordan réguliers. Supposons qu'en chaque point de son bord, D possède un vecteur normal unitaire extérieur ν. Soient P et Q deux fonctions réelles de classe C^1 sur $\bar{D}$. Alors

$$\iint_D (\partial_x Q - \partial_y P) dx\, dy = \oint_{\partial D} P dx + Q dy \tag{19.232}$$

où chaque chemin γ formant le bord de D est orienté de sorte que $T\nu = \frac{\dot{\gamma}}{\|\dot{\gamma}\|}$ où T représente la rotation d'angle $+\frac{\pi}{2}$.

Justifions le fait que cela soit un cas particulier de la formule de Stokes du théorème 19.71. Nous considérons la forme différentielle

$$\omega = P dx + Q dy, \tag{19.233}$$

et sa différentielle

$$d\omega = \sum_i d\omega_i \wedge dx_i \tag{19.234a}$$

$$= \left(\frac{\partial P}{\partial x} dx + \frac{\partial P}{\partial y} dy\right) \wedge dx + \left(\frac{\partial Q}{\partial x} dx + \frac{\partial Q}{\partial y} dy\right) \wedge dy \tag{19.234b}$$

$$= \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy. \tag{19.234c}$$

Intégrons cette forme $d\omega$ sur le domaine ouvert D que nous paramétrons de façon triviale par

$$\begin{aligned} \varphi\colon D &\to \mathbb{R}^2 \\ (u,v) &\mapsto (u,v). \end{aligned} \tag{19.235}$$

Ce que nous avons est

$$\iint_D d\omega = \iint_D d\omega_{(u,v)} \left(\frac{\partial \varphi}{\partial u}, \frac{\partial \varphi}{\partial v}\right) du dv \tag{19.236}$$

Nous avons aussi $T_u = \frac{\partial \varphi}{\partial u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ et $T_v = \frac{\partial \varphi}{\partial v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ et donc

$$(dx \wedge dy)(T_u, T_v) = dx(T_u) dy(T_v) - dx(T_v) dy(T_u) = 1 - 0 = 1. \tag{19.237}$$

L'intégrale (19.236) se développe donc en

$$\iint_D d\omega = \iint_D \left(\frac{\partial Q}{\partial x}(u,v) - \frac{\partial P}{\partial y}(u,v)\right) (dx \wedge dy)(T_u, T_v) du dv = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) du dv. \tag{19.238}$$

Par conséquent la formule de Stokes nous donne la formule (19.232).

La formule de Green nous permet de calculer l'aire de la surface délimitée par une courbe fermée en termes de l'intégrale d'une forme bien choisie le long du contour. Pour cela nous prenons la forme

$$\omega = -\frac{y}{2} dx + \frac{x}{2} dy, \tag{19.239}$$

de telle sorte que $\partial_x Q - \partial_y Q = 1$ et que

$$\iint_D d\omega = \iint_D ddudv = S, \tag{19.240}$$

et au final l'aire est donnée par

$$S = \int_{\partial D} \Big(-\frac{y}{2} dx + \frac{x}{2} dy \Big). \tag{19.241}$$

Lorsque le bord de D est paramétré par

$$\begin{aligned} \gamma \colon [a,b] &\to \mathbb{R}^2 \\ u &\mapsto \begin{pmatrix} x(u) \\ y(u) \end{pmatrix}, \end{aligned} \tag{19.242}$$

nous avons

$$(Pdx + Qdy)\gamma'(u) = Px' + Qy', \tag{19.243}$$

et alors

$$\int_{\partial D} Pdx + Qdy = \int_a^b P\big(x(u), y(u)\big)x'(u) + Q\big(x(u), y(u)\big)y'(u)du. \tag{19.244}$$

En ce qui concerne l'aire de la surface, nous prenons les P et Q de la forme 19.239 :

$$S = \frac{1}{2} \int_a^b \Big(-y(u)x'(u) + x(u)y'(u) \Big) du. \tag{19.245}$$

19.11.3 Formule de Stokes

La formule de Stokes est la version classique, qui permet d'exprimer la circulation d'un champ de vecteur le long d'une courbe de $\mathbb{R}^3$ comme le flux de son rotationnel à travers n'importe quel surface dont le bord est la courbe. La version présentée ici suppose que la surface peut se paramétrer en un seul morceau :

Théorème 19.76 (Formule de Stokes)**.**
Soit $F : W \subset \mathbb{R}^2 \to \mathbb{R}^3$ un paramétrage (carte) d'une surface dans $\mathbb{R}^3$, supposée de classe C^2. Soit D un ouvert de $\mathbb{R}^2$ vérifiant les hypothèses de la formule de Green, et tel que $\bar{D} \subset W$. Soit G un champ de vecteurs de classe C^1 défini sur $F(\bar{D})$, et soit N le champ normal unitaire donné par le paramétrage

$$N = \frac{\frac{\partial F}{\partial u} \wedge \frac{\partial F}{\partial v}}{\Big\| \frac{\partial F}{\partial u} \wedge \frac{\partial F}{\partial v} \Big\|} \tag{19.246}$$

alors

$$\iint_{F(D)} \langle \nabla \times G, N \rangle \, d\sigma_F = \int_{F(\partial D)} G \tag{19.247}$$

où les chemins formant le bord ∂D sont orientés comme dans le théorème de Green.

Notons, juste pour avoir une bonne nouvelle de temps en temps, que

$$d\sigma_F = \left\| \frac{\partial F}{\partial u} \times \frac{\partial F}{\partial v} \right\| dudv, \tag{19.248}$$

mais cette norme apparaît exactement au dénominateur de N. Il ne faut donc pas la calculer parce qu'elle se simplifie.

Sous forme un peu plus physicienne [12], la formule (19.247) s'écrit

$$\int_{F(D)} \langle \nabla \times G, N(x) \rangle \, d\sigma_F(x) = \int_{F(\gamma)} \langle G, T \rangle \, ds \tag{19.249}$$

où T est le vecteur unitaire tangent à $F(\gamma)$.

12. et surtout plus explicite.

19.11.3.1 Quelle est la bonne orientation ?

Le signe du vecteur normal N dépend du choix de l'ordre des coordonnées dans la carte. Supposons que je veuille paramétrer la surface $x^2 + y^2 = 1$, $z = 1$. Nous prenons naturellement comme carte le cercle C de rayon 1 dans $\mathbb{R}^2$ et la carte

$$F(r, \theta) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \\ 1 \end{pmatrix}. \tag{19.250}$$

Mais nous aurions aussi pu mettre les coordonnées r et θ dans l'autre ordre :

$$\tilde{F}(\theta, r) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \\ 1 \end{pmatrix}. \tag{19.251}$$

Les vecteurs normaux ne sont pas les même : la carte F donnera $\partial_r F \times \partial_\theta F$, tandis que l'autre donnera $\partial_\theta \tilde{F} \times \partial_r \tilde{F}$. Le signe change !

Il faut savoir laquelle choisir. Le cercle $C \subset \mathbb{R}^2$ a une orientation donnée par le théorème de Green. Nous choisissons l'ordre des coordonnées pour que 1_θ et 1_r soient dans la même orientation que les vecteurs ν et T tels que donnés par le théorème de Green, et tels que dessinés sur la figure 19.3.

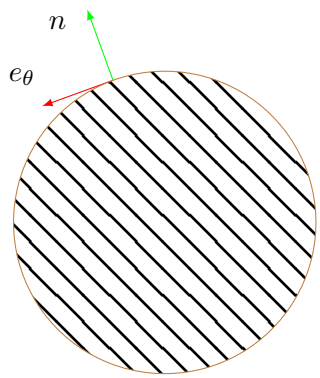


FIGURE 19.3 – L'orientation sur le cercle. Si nous les prenons dans l'ordre, les vecteurs $(1_r, 1_\theta)$ ont la même orientation que celle donnée par les vecteurs (ν, T) donnés par la convention de Green.

Plus généralement, nous choisissons l'ordre des coordonnées u et v pour que la base $(1_u, 1_v)$ ait la même orientation que (ν, T) où T a le sens convenu dans le théorème de Green.

19.12 Résumé des intégrales vues

Nous sommes maintenant capables de revoir tous les types d'intégrales vues jusqu'ici de façon très cohérentes. Nous commencerons par les intégrales de fonctions et nous ferons ensuite les intégrales de champs de vecteurs.

19.12.1 L'intégrale d'une fonction sur les réels

Si $f\colon [a, b] \subset \mathbb{R} \to \mathbb{R}$ est une fonction usuelle, sont intégrale est

$$\int_a^b f(x)dx = F(b) - F(a) \tag{19.252}$$

où F est une primitive de f.

19.12.2 Intégrale d'une fonction sur un chemin

Si f est une fonction sur $\mathbb{R}^3$ et si $\sigma\colon [a,b] \to \mathbb{R}^3$ est un chemin dans $\mathbb{R}^3$, l'intégrale de f sur σ est, par définition,

$$\int f\,d\sigma = \int_a^b f\big(\sigma(t)\big)\|\sigma'(t)\|dt. \tag{19.253}$$

19.12.3 Intégrale d'une fonction sur une surface

Nous devons paramétrer la surface S par une application $\varphi\colon D \subset \mathbb{R}^2 \to \mathbb{R}^3$. À partir d'un tel paramétrage, nous construisons un élément de surface en prenant le produit vectoriel des deux vecteurs tangents :

$$dS = \frac{\partial\varphi}{\partial u} \times \frac{\partial\varphi}{\partial v} dudv. \tag{19.254}$$

L'intégrale est

$$\int f\,dS = \iint_D f\big(\varphi(u,v)\big)\left\|\frac{\partial\varphi}{\partial u} \times \frac{\partial\varphi}{\partial v}\right\| dudv. \tag{19.255}$$

Il ne faut pas rajouter de jacobien : la norme du produit vectoriel *est* le jacobien.

Remarque 19.77.
La formule (19.255) est autant valable pour des surfaces dans $\mathbb{R}^2$ que dans $\mathbb{R}^3$. Si nous considérons une surface dans $\mathbb{R}^2$, nous la voyons dans $\mathbb{R}^3$ en ajoutant un zéro comme troisième composante.

Exemple 19.78.
Les coordonnées polaires sont données par

$$\varphi(r,\theta) = \begin{pmatrix} r\cos\theta \\ r\sin\theta \\ 0 \end{pmatrix}. \tag{19.256}$$

Les vecteurs tangents à ce paramétrage sont

$$T_r = \frac{\partial\varphi}{\partial r} = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}, \quad T_\theta = \frac{\partial\varphi}{\partial\theta} = \begin{pmatrix} -r\sin\theta \\ r\cos\theta \\ 0 \end{pmatrix}. \tag{19.257}$$

Le vecteur normal est

$$\frac{\partial\varphi}{\partial r} \times \frac{\partial\varphi}{\partial\theta} = \begin{vmatrix} e_x & e_y & e_z \\ \cos\theta & \sin\theta & 0 \\ -r\sin\theta & r\cos\theta & 0 \end{vmatrix} = re_z. \tag{19.258}$$

Nous trouvons donc que l'élément de surface est la norme de re_z, c'est-à-dire r, le jacobien connu. △

19.12.4 Intégrale d'une fonction sur un volume

Si V est un volume dans $\mathbb{R}^3$, nous effectuons la même procédure : nous trouvons un paramétrage, et nous formons un élément de volume avec les vecteurs tangents de le paramétrage. Nous avons donc un volume déterminé par l'application

$$\varphi\colon D \subset \mathbb{R}^3 \to \mathbb{R}^3, \tag{19.259}$$

et ses trois vecteurs tangents

$$\begin{aligned} T_u &= \frac{\partial\varphi}{\partial u} \\ T_v &= \frac{\partial\varphi}{\partial v} \\ T_w &= \frac{\partial\varphi}{\partial w}. \end{aligned} \tag{19.260}$$

Comment former un volume avec trois vecteurs ? Réponse : le produit mixte. L'intégrale de f sur V sera

$$\int f\,dV = \iiint\limits_D f\big(\varphi(u,v,w)\big)\left\|\frac{\partial\varphi}{\partial u}\cdot\left(\frac{\partial\varphi}{\partial v}\times\frac{\partial\varphi}{\partial w}\right)\right\|dudv. \tag{19.261}$$

Encore une fois, le produit mixte *est* le jacobien. Prenons les coordonnées sphériques :

$$\begin{aligned} x(r,\theta,\varphi) &= r\sin(\theta)\cos(\varphi)\\ y(r,\theta,\varphi) &= r\sin(\theta)\sin(\varphi)\\ z(r,\theta,\varphi) &= r\cos(\theta) \end{aligned} \tag{19.262}$$

Les trois vecteurs tangents seront

$$T_r = \begin{pmatrix}\frac{\partial x}{\partial r}\\ \frac{\partial y}{\partial r}\\ \frac{\partial z}{\partial r}\end{pmatrix} = \begin{pmatrix}\sin(\theta)\cos(\varphi)\\ \sin(\theta)\sin(\varphi)\\ \cos(\theta)\end{pmatrix} \tag{19.263a}$$

$$T_\theta = \begin{pmatrix}\frac{\partial x}{\partial \theta}\\ \frac{\partial y}{\partial \theta}\\ \frac{\partial z}{\partial \theta}\end{pmatrix} = \begin{pmatrix}r\cos(\theta)\cos(\varphi)\\ r\cos(\theta)\sin(\varphi)\\ -r\sin(\theta)\end{pmatrix} \tag{19.263b}$$

$$T_\varphi = \begin{pmatrix}\frac{\partial x}{\partial \varphi}\\ \frac{\partial y}{\partial \varphi}\\ \frac{\partial z}{\partial \varphi}\end{pmatrix} = \begin{pmatrix}-r\sin(\theta)\sin(\varphi)\\ r\sin(\theta)\cos(\varphi)\\ 0\end{pmatrix} \tag{19.263c}$$

Nous avons vu que le produit mixte revient à mettre toutes les composantes dans une matrice. Ici nous avons donc

$$\frac{\partial\phi}{\partial r}\cdot\left(\frac{\partial\phi}{\partial\theta}\times\frac{\partial\phi}{\partial\varphi}\right) = \begin{vmatrix}\frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} & \frac{\partial z}{\partial r}\\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} & \frac{\partial z}{\partial \theta}\\ \frac{\partial x}{\partial \varphi} & \frac{\partial y}{\partial \varphi} & \frac{\partial z}{\partial \varphi}\end{vmatrix} \tag{19.264}$$

Cela est précisément le jacobien dont nous parlions plus haut.

19.12.5 Conclusion pour les fonctions

Lorsque nous intégrons une fonction sur un chemin, une surface ou un volume, la technique est toujours la même :

(1) Trouver un paramétrage à une, deux ou trois variables.

(2) Dériver le paramétrage par rapport à ses variables.

(3) Construire un élément de longueur, surface ou volume à partir des vecteurs que l'on a. Cela se fait en prenant la norme, le produit vectoriel ou le produit mixte.

19.12.6 Circulation d'un champ de vecteurs

Pour les champs de vecteurs, nous faisons la même chose, mais au lieu de *multiplier* par l'élément de longueur ou de surface, nous prenons le produit scalaire. Si nous considérons la courbe paramétrée $\sigma\colon [a,b]\to\mathbb{R}^3$ et le champ de vecteurs F, nous avons donc

$$\int_\sigma F = \int F\cdot d\sigma = \int_a^b F\big(\sigma(t)\big)\cdot\sigma'(t)dt. \tag{19.265}$$

19.12.7 Flux d'un champ de vecteurs

Si la surface $S\subset\mathbb{R}^3$ est paramétrée par

$$\begin{aligned}\phi\colon D\subset\mathbb{R}^2&\to\mathbb{R}^3\\ (u,v)&\mapsto\phi(u,v),\end{aligned} \tag{19.266}$$

et si F est un champ de vecteurs, alors on a

$$\int_S F = \int_S F \cdot dS = \iint_D F\big(\phi(u,v)\big) \cdot \left(\frac{\partial \phi}{\partial u} \times \frac{\partial \phi}{\partial v}\right) dudv. \tag{19.267}$$

19.12.8 Conclusion pour les champs de vecteurs

La circulation et le flux ne représentent pas tout à fait la même chose. En effet pour la circulation, nous sélectionnons la composante *tangente* à la courbe, c'est-à-dire la partie du vecteurs qui « circule » le long de la courbe. Une force perpendiculaire au mouvement ne travaille pas.

La situation est exactement le contraire pour le flux. Étant donné que le vecteur

$$\frac{\partial \phi}{\partial u} \times \frac{\partial \phi}{\partial v} \tag{19.268}$$

est normal à la surface, le fait de prendre le produit scalaire du champ de vecteurs avec lui sélectionne la composante *normale* à la surface, c'est-à-dire la partie du vecteur qui traverse la surface.

19.12.9 Attention pour les surfaces fermées !

Si nous considérons une surface fermée, il faut faire attention à choisir une *orientation*. Les vecteurs normaux doivent soit tous pointer vers l'intérieur soit tous vers l'extérieur. En effet, en tant que vecteur normal, nous avons choisi de prendre

$$T_u \times T_v. \tag{19.269}$$

Mais le vecteur $T_v \times T_u$ est tout aussi normal ! Il n'y a pas à priori de façon standard pour choisir l'un ou l'autre. Il faut juste être cohérent : il faut que si on divise la surface en plusieurs morceaux, tous les vecteurs pointent dans le même sens.

Notez que si vous faites un choix et que votre voisin fait le choix inverse, vous obtiendrez des réponses qui diffèrent d'un signe. Sans plus de précisions [13], les deux réponses sont correctes.

Un exemple de ce problème est donné dans l'exemple 19.79.

Exemple 19.79.
Calculer le flux du champ de vecteurs

$$F(x,y,z) = e_x \tag{19.270}$$

au travers du cylindre de rayon R et de hauteur h autour de l'axe z.

Même question si le cylindre est autour de l'axe x.

Remarque : ces cylindres sont considérés *avec* leur « couvercles ».

Un paramétrage du cylindre autour de l'axe z est

$$\phi(\theta, z) = \begin{pmatrix} R\cos\theta \\ R\sin\theta \\ z \end{pmatrix}. \tag{19.271}$$

Les vecteurs tangents sont

$$T_\theta = \begin{pmatrix} -R\sin\theta \\ R\cos\theta \\ 0 \end{pmatrix}, T_z = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{19.272}$$

Le vecteur normal est donc

$$T_\theta \times T_z = R\cos(\theta)e_x + R\sin(\theta)e_y. \tag{19.273}$$

13. Il faudrait définir ce qu'est une surface *orientable* et choisir une orientation.

C'est un vecteur dirigé vers l'extérieur.

Le champ de vecteurs considéré est constant : $F(\theta, z) = e_x$. Nous avons donc

$$F(\theta, z) \cdot (T_\theta \times T_z) = R\cos(\theta) \tag{19.274}$$

et le flux vaut

$$\Phi = \int_0^{2\pi} d\theta \int_0^h R\cos(\theta)dz = 0. \tag{19.275}$$

En ce qui concerne les couvercles haut au bas, ils sont paramétrés par

$$\phi_1(r, \theta) = \begin{pmatrix} R\cos(\theta) \\ R\sin(\theta) \\ h \end{pmatrix}, \phi_2(r, \theta) \quad = \begin{pmatrix} R\cos(\theta) \\ R\sin(\theta) \\ 0 \end{pmatrix}. \tag{19.276}$$

Les vecteurs normaux correspondants sont dans la direction de e_z, de façon que le produit scalaire avec $F(r, \theta)$ soit nul. Le flux total est donc nul.

Regardons maintenant le cylindre le long de l'axe x. Un paramétrage est

$$\phi(\theta, x) = \begin{pmatrix} x \\ R\cos(\theta) \\ R\sin(\theta) \end{pmatrix}, \tag{19.277}$$

et le vecteurs tangents sont

$$T_\theta = \begin{pmatrix} 0 \\ -R\sin\theta \\ R\cos\theta \end{pmatrix}, T_x \quad = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \tag{19.278}$$

Le vecteur normal est alors donné par

$$T_\theta \times T_x = R\cos(\theta)e_y + R\sin(\theta)e_z. \tag{19.279}$$

Nous avons par conséquent $F(\theta, x) \cdot (T_\theta \times T_x) = 0$. Pas de flux par le côté du cylindre.

Regardons les « couvercles ». Le premier est donné par le paramétrage

$$\phi_1(r, \theta) = \begin{pmatrix} 0 \\ r\cos(\theta) \\ r\sin(\theta) \end{pmatrix}. \tag{19.280}$$

Le vecteur normal serait $T_r \times T_\theta = re_x$, et le flux

$$\Phi = \int_0^{2\pi} d\theta \int_0^R r\,dr = \pi R^2. \tag{19.281}$$

Le second couvercle est donné par

$$\phi_2(r, \theta) = \begin{pmatrix} h \\ r\cos(\theta) \\ r\sin(\theta) \end{pmatrix}. \tag{19.282}$$

Le vecteur normal est encore re_x, et le flux est à nouveau πR^2.

Le flux total serait donc $2\pi R^2$.

Cela n'est pas possible parce que tous les vecteurs qui « rentrent » d'un côté doivent « sortir » de l'autre côté. L'erreur est le le premier vecteur normal est un vecteur qui pointe vers l'intérieur du cylindre, tandis que le second pointe vers l'extérieur. Si nous choisissons, par convention, de prendre uniquement les vecteurs extérieurs, il faut changer le vecteur normal du premier couvercle en $-re_x$. Le premier flux vaudra donc

$$-\pi R^2, \tag{19.283}$$

de telle sorte que le flux total sera nul. △

19.13 Formes différentielles exactes et fermées

Nous avons déjà parlé de formes différentielles et de leurs intégrales sur un chemin dans la section 11.19.

Définition 19.80.
*La forme différentielle ω est **exacte** s'il existe une fonction f telle que $\omega = df$; elle est dite **fermée** si $d\omega = 0$.*

Dire que la forme différentielle $\omega = f dx + g dy$ est fermée, c'est dire que

$$\frac{\partial g}{\partial x} = \frac{\partial f}{\partial y}. \tag{19.284}$$

Il est naturel de se demander si toutes les formes différentielles sont des différentielles de fonctions. Une réponse complète est délicate à établir, mais a d'innombrables conséquences en physique, notamment en ce qui concerne l'existence d'un potentiel vecteur pour le champ magnétique dans les équations de Maxwell.

Le fait qu'une forme exacte soit fermée est relativement facile à établir ; c'est la proposition suivante. La question plus délicate est la réciproque : sous quelles conditions une forme fermée est-elle exacte ?

Proposition 19.81.
Si ω est une 1-forme exacte de classe C^1, alors ω est fermée.

Démonstration. Le fait que ω soit exacte implique l'existence d'une fonction f telle que $\omega = df$, c'est-à-dire

$$\omega_x = \sum_i a_i(x) dx_i = \sum_i \frac{\partial f}{\partial x_i}(x) dx_i, \tag{19.285}$$

c'est-à-dire que $a_i(x) = \frac{\partial f}{\partial x_i}(x)$. L'hypothèse que ω est C^1 implique que f est C^2, et donc que nous pouvons inverser l'ordre de dérivation pour les dérivées secondes $\partial^2_{ij} f = \partial^2_{ji} f$. Nous pouvons donc faire le calcul suivant :

$$\frac{\partial a_i}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j} = \frac{\partial a_j}{\partial x_i}, \tag{19.286}$$

ce qu'il fallait démontrer. □

La réciproque est vraie sur un ouvert simplement connexe.

Théorème 19.82.
Supposons que $D \subset \mathbb{R}^n$ soit un ouvert simplement connexe. Alors toute forme différentielle de degré 1 et de classe C^1 sur D qui est fermée est exacte.

Nous allons prouver ce théorème dans un cas un peu moins général : celui d'un domaine étoilé de $\mathbb{R}^2$ plutôt que simplement connexe de $\mathbb{R}^n$.

Théorème 19.83.
Soit $D \subset \mathbb{R}^2$, une ouvert étoilé, et ω, une 1-forme fermée de classe C^1. Alors ω est exacte.

Démonstration. Soit $D \subset \mathbb{R}^2$, un ouvert étoilé par rapport à l'origine. Soient $f\colon D \to \mathbb{R}$, $g\colon D \to \mathbb{R}$, des fonctions de classe C^1 telles que

$$\frac{\partial f}{\partial y} = \frac{\partial g}{\partial x} \tag{19.287}$$

sur D, et

$$F(x,y) = \int_0^1 \big[f(tx,ty)x + g(tx,ty)y\big] dt \tag{19.288}$$

pour tout $(x,y) \in D$.

Étant donné que nous ne définissons $F(x,y)$ que pour des $(x,y)\in D$, la fonction $t\mapsto f(tx,ty)$ est C^1 sur tout le compact $[0,1]$ et aucune divergence de l'intégrale n'est à craindre. Nous sommes donc dans le cadre de la proposition 16.25, et nous pouvons dériver sous le signe intégral.

Nous calculons, en utilisant la règle de dérivation de fonctions composées

$$\begin{aligned}\frac{\partial F}{\partial x}(x,t)&=\int_0^1\left[f\frac{\partial f}{\partial x}(tx,ty)x+f(tx,ty)+t\frac{\partial g}{\partial x}(tx,ty)y\right]dt\\&=\int_0^1\left[t\Big(x\frac{\partial f}{\partial x}(tx,ty)+y\frac{\partial f}{\partial y}(tx,ty)\Big)+f(tx,ty)\right]dt\end{aligned}\tag{19.289}$$

où nous avons utilisé l'hypothèse $\partial_y f=\partial_x g$. Ce qui se trouve dans la parenthèse n'est autre que $\partial_t\big(f(tx,ty)\big)$, plus précisément, si nous posons $\mathcal{F}(x,y,t)=f(tx,ty)$, nous avons

$$\frac{\partial\mathcal{F}}{\partial t}(x,y,t)=x\frac{\partial f}{\partial x}(tx,ty)+y\frac{\partial f}{\partial y}(tx,ty).\tag{19.290}$$

En recopiant le résultat (19.289) en termes de $\mathcal{F}$, nous avons

$$\begin{aligned}\frac{\partial F}{\partial x}(x,t)&=\int_0^1\left(t\frac{\partial\mathcal{F}}{\partial t}(x,y,t)+\mathcal{F}(x,y,t)\right)dt\\&=\int_0^1\partial_t\big(t\mathcal{F}(x,y,t)\big)dt\\&=\big[f\mathcal{F}(x,y,t)\big]_0^1\\&=\mathcal{F}(x,y,1)\\&=f(x,y).\end{aligned}\tag{19.291}$$

Le résultat correspondant pour $\frac{\partial F}{\partial y}(x,y)=g(x,y)$ s'obtient de la même manière. Nous avons donc obtenu que

$$\frac{\partial F}{\partial x}=f,\quad\text{et}\quad\frac{\partial F}{\partial y}=g.\tag{19.292}$$

En ayant prouvé cela, nous avons prouvé que si $\omega=fdx+gdy$ avec $\partial_y f=\partial_x g$, alors $\omega=dF$ où F est définie par (19.288). □

Démonstration alternative du théorème 19.83. Nous posons $u=tx$ et $v=ty$, ainsi que $\mathcal{F}(x,y,t)=f(u,v)$ et $\mathcal{G}(x,y,t)=g(u,v)$. Avec cette notation, nous avons

$$F(x,y)=\int_0^1\big(x\mathcal{F}(x,y,t)+y\mathcal{G}(x,y,t)\big)dt,\tag{19.293}$$

et

$$\begin{aligned}\frac{\partial\mathcal{F}}{\partial x}&=\frac{\partial f}{\partial u}\frac{\partial u}{\partial x}+\frac{\partial f}{\partial v}\frac{\partial v}{\partial x}=t\frac{\partial f}{\partial u},\\\frac{\partial\mathcal{G}}{\partial x}&=t\frac{\partial g}{\partial u}.\end{aligned}\tag{19.294}$$

Ainsi,

$$\begin{aligned}\frac{\partial F}{\partial x}&=\int_0^1\left(x\frac{\partial\mathcal{F}}{\partial x}+\mathcal{F}+y\frac{\partial G}{\partial x}\right)dt\\&=\int_0^1\left(xt\frac{\partial f}{\partial u}+\mathcal{F}+yt\frac{\partial g}{\partial u}\right)dt\\&=\int_0^1\left[t\left(x\frac{\partial f}{\partial u}+y\frac{\partial f}{\partial v}\right)+\mathcal{F}\right]dt.\end{aligned}\tag{19.295}$$

où nous avons utilisé le fait que, par hypothèse, $\frac{\partial g}{\partial u}=\frac{\partial f}{\partial v}$. Nous calculons par ailleurs que

$$\frac{\partial F}{\partial t}=\frac{\partial f}{\partial u}\frac{\partial u}{\partial t}+\frac{\partial f}{\partial v}\frac{\partial v}{\partial t}=x\frac{\partial f}{\partial u}+y\frac{\partial f}{\partial v}.\tag{19.296}$$

Donc, nous avons

$$\frac{\partial F}{\partial x} = \int_0^1 \left(t\frac{\partial \mathcal{F}}{\partial t} + \mathcal{F} \right) dt = \int_0^1 \frac{\partial}{\partial t}(t\mathcal{F})dt. \tag{19.297}$$

Par conséquent,

$$\frac{\partial F}{\partial x} = [t\mathcal{F}]_0^1 = \mathcal{F}(x, y, 1) = f(x, y). \tag{19.298}$$

Le même genre de calculs fournit $\frac{\partial F}{\partial y} = g(x, y)$. □

19.14 Théorème d'Abel angulaire

Théorème 19.84 (Abel angulaire[54]).
Soit $\sum_n a_n z^n$ une série entière de rayon de convergence plus grand ou égal à 1 et de somme f. Soit $\theta_0 \in [0, \frac{\pi}{2}[$. Nous posons

$$\Delta_{\theta_0} = \{z = 1 - \rho e^{i\varphi} \text{ tel que } \rho > 0, \varphi \in [\theta_0, \theta_0], |z| < 1\}. \tag{19.299}$$

Nous supposons de plus que $\sum_n a_n$ converge. Alors

$$\lim_{\substack{z \to 1 \\ z \in \Delta_0}} f(z) = \sum_{k=0}^{\infty} a_k. \tag{19.300}$$

Démonstration. Le résultat de ce théorème est que l'on peut calculer la limite $z \to 1$ avec des chemins contenus dans un domaine de la forme de celui dessiné à la figure 19.4.

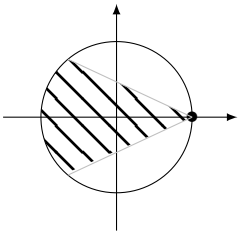


FIGURE 19.4 – La zone dans laquelle peut être le chemin qui va vers $z = 1$.

De façon très classique nous posons

$$S = \sum_{k=0}^{\infty} a_k \quad S_n = \sum_{k=0}^{n} a_k, \tag{19.301}$$

et $R_n = S - S_n$. En particulier $a_n = R_{n-1} - R_n$.

Le but du théorème est de montrer que $\sum a_n z^n$ converge vers S lorsque z converge vers 1 à l'intérieur de Δ_{θ_0}. Pour cela nous calculons pour un N donné la différence $\sum_{n=0}^{N} a_n z^n - S_N$ en triant les termes par ordre de R_n, en isolant le terme R_0 et le terme R_N :

$$\sum_{n=0}^{N} a_n z^n - S_N = \sum_{n=1}^{N} a_n(z^n - 1) \tag{19.302a}$$

$$= \sum_{n=1}^{N} (R_{n-1} - R_n)(z^n - 1) \tag{19.302b}$$

$$= R_0(z - 1) + \sum_{n=1}^{N-1} R_n(z^{n+1} - 1 - z^n + 1) + R_N(z^N - 1) \tag{19.302c}$$

$$= R_0(z - 1) + \sum_{n=1}^{N-1} R_n z^n(z - 1) + R_N(z^N - 1) \tag{19.302d}$$

$$= (z - 1) \sum_{n=0}^{N-1} R_n z^n + R_N(z^N - 1). \tag{19.302e}$$

Cela est valable pour tout N et $|z| < 1$. Nous avons donc

$$\sum_{n=0}^{N} a_n z^n - S_N = (z-1)\sum_{n=0}^{N-1} R_n z^n + R_N(z^N - 1). \tag{19.303}$$

Par hypothèse nous avons $\lim_{N\to\infty} R_N = 0$. Et de plus le membre de gauche converge parce que chacun des deux termes converge séparément. En passant à la limite nous avons pour tout $|z| < 1$:

$$f(z) - S = (z-1)\sum_{n=0}^{\infty} R_n z^n. \tag{19.304}$$

Nous voudrions étudier le comportement de la différence $f(z) - S$ lorsque z tend vers 1. Pour cela nous nous fixons $\epsilon > 0$ et $N \geqslant 1$ tel que $|R_n| < \epsilon$ dès que $n \geqslant N$. Alors pour tout $|z| < 1$ nous avons

$$|f(z) - S| \leqslant |z-1| \left(\sum_{n=0}^{N} |R_n| \underbrace{|z^n|}_{\leqslant 1} + \sum_{n=N+1}^{\infty} \underbrace{|R_n|}_{\leqslant \epsilon} |z^n| \right) \tag{19.305a}$$

$$\leqslant |z-1| \sum_{n=0}^{N} |R_n| + \epsilon \frac{|z-1|}{1-|z|} \tag{19.305b}$$

où nous avons utilisé la somme de la série géométrique (10.125) et l'égalité $|z^n| = |z|^n$. Avant de nous particulariser à $z \in \Delta_{\theta_0}$ nous devons anticiper un problème au dénominateur en multipliant par le binôme conjugué :

$$\frac{|z-1|}{1-|z|} = \frac{|z-1|(1+|z|)}{1-|z|^2}. \tag{19.306}$$

C'est maintenant que nous nous particularisons à $z \in \Delta_{\theta_0}$ en posant $z = \rho e^{i\varphi}$ et en remarquant que $|z|^2 = 1 - 2\rho\cos(\varphi) + \rho^2$. Nous avons le calcul suivant :

$$\frac{|z-1|}{1-|z|} = \frac{\rho(1+|z|)}{2\rho\cos(\varphi) - \rho^2} \tag{19.307a}$$

$$= \frac{1+|z|}{2\cos(\varphi) - \rho} \tag{19.307b}$$

$$\leqslant \frac{2}{2\cos(\varphi) - \rho} \tag{19.307c}$$

$$\leqslant \frac{2}{2\cos(\varphi) - \cos(\theta_0)} \tag{19.307d}$$

$$\leqslant \frac{2}{2\cos(\theta_0) - \cos(\theta_0)} \tag{19.307e}$$

$$= \frac{2}{\cos(\theta_0)}. \tag{19.307f}$$

Quelques justifications.

— Vu que nous avons dans l'idée de faire $\rho \to 0$ nous supposons que $\rho < \cos(\theta_0)$.

— Nous avons $\cos(\varphi) > \cos(\theta_0)$ parce que z est dans Δ_{θ_0}.

Nous avons donc, pour tout $z \in \Delta_{\theta_0}$ que

$$|f(z) - S| \leqslant |z-1| \sum_{n=0}^{N} |R_n| + \epsilon \frac{2}{\cos(\theta_0)}. \tag{19.308}$$

Il suffit de prendre ρ assez petit pour que

$$|z-1| \sum_{n=0}^{N} |R_n| < \epsilon \tag{19.309}$$

et nous avons

$$|f(z) - S| \leqslant \epsilon \left(1 + \frac{2}{\cos(\theta_0)}\right). \tag{19.310}$$

Nous avons donc bien $\lim_{\substack{z\to 1\\ z\in\Delta_0}} f(z) = S$, comme nous le voulions. □

La réciproque du théorème d'Abel angulaire est que si $f(z) = \sum_n a_n z^n$ sur $B(0,1)$ se prolonge par continuité en $z = 1$ alors cette prolongation se fait par $f(1) = \sum_n a_n$. Cela est faux comme le montre l'exemple suivant.

Exemple 19.85.
Nous considérons la série entière $\sum_{n=0}^{\infty}(-1)^n z^n$ qui converge[14] vers

$$f(z) = \frac{1}{1+z} \tag{19.311}$$

sur $B(0,1)$. De plus nous avons

$$\lim_{\substack{z\to 1\\ |z|<1}} \frac{1}{1+z} = \frac{1}{2}. \tag{19.312}$$

Donc la fonction converge bien vers quelque chose lorsque z tend vers 1. La fonction f se prolonge par continuité en 1. Pourtant la série es coefficients $\sum_n(-1)^n$ ne converge pas. △

Le théorème suivant donne une espèce d'inverse au théorème d'Abel angulaire. En effet il dit que si la série converge en allant vers 1 le long de l'axe réel, alors ça converge vers la somme des coefficients. Il faut cependant une hypothèse en plus sur les a_n.

Théorème 19.86 (Théorème taubérien faible[54])**.**
Soit $\sum_n a_n z^n$ une série entière de rayon de convergence 1 et de somme f. Nous supposons

(1) Il existe $S \in \mathbb{C}$ tel que $\lim_{\substack{x\to 1\\ x\in]-1,1[}} f(x) = S$.

(2) $\lim_{n\to\infty} na_n = 0$.

Alors la série $\sum_{n=0}^{\infty} a_n$ converge et vaut S.

Démonstration. Nous notons $S_n = \sum_{k=0} a_k$ et $M = \sup_{k\geqslant 1} k|a_k|$, qui est fini par hypothèse. Pour $x \in]0,1[$ et $n \geqslant 0$ nous avons

$$S_n - f(x) = \sum_{k=1}^{n} a_k - \sum_{k=1}^{n} a_k x^k - \sum_{k=n+1}^{\infty} a_k x^k = \sum_{k=1}^{n} a_k(1-x^k) - \sum_{k=n+1}^{\infty} a_k x^k. \tag{19.313}$$

Nous utilisons la série géométrique sous la forme $1 - x^k = (1-x)\sum_{i=0}^{n} x^i$ pour écrire

$$S_n - f(x) = \sum_{k=1}^{n} a_k(1-x)\underbrace{\sum_{i=0}^{k-1} x^i}_{\leqslant k} - \sum_{k=n+1}^{\infty} a_k x^k \tag{19.314a}$$

$$\leqslant \sum_{k=1}^{n} k a_k(1-x) - \sum_{k=n+1}^{\infty} a_k x^k, \tag{19.314b}$$

14. C'est la série géométrique de raison $-z$.

donc en passant à la norme

$$\begin{aligned}
\big|S_n - f(x)\big| &\leqslant (1-x)Mn + \sum_{k=n+1} |a_k| x^k && (19.315a)\\
&\leqslant (1-x)Mn + \sum_{k=n+1}^{\infty} \underbrace{\frac{k}{n}|a_k|}_{\leqslant M/n} x^k && (19.315b)\\
&\leqslant (1-x)Mn + \frac{M}{n} \sum_{k=n+1}^{\infty} x^k && (19.315c)\\
&\leqslant (1-x)Mn + \frac{M}{n}\frac{1}{1-x}. && (19.315d)
\end{aligned}$$

Ce que nous cherchons à étudier est le comportement $x \to 1$ et montrer que $S_n \to S$, ce qui nous incite à calculer $|S_n - f(1-\frac{\epsilon}{n})|$ avec $0 < \epsilon < 1$:

$$\big|S_n - f\big(1 - \frac{\epsilon}{n}\big)\big| \leqslant \epsilon M + \epsilon. \tag{19.316}$$

Nous choisissons N_1 tel que $\frac{M}{n} \leqslant \epsilon^2$ dès que $n \geqslant N_1$. En sus nous savons que

$$\lim_{\epsilon \to 0} f(1-\epsilon) = S. \tag{19.317}$$

Nous choisissons N_2 de telle sorte à avoir

$$\left|f\left(1 - \frac{\epsilon}{n}\right) - S\right| < \epsilon, \tag{19.318}$$

et en prenant $n \geqslant \max(N_1, N_2)$ nous avons

$$|S_n - S| \leqslant \left|S_n - f\left(1 - \frac{\epsilon}{n}\right)\right| + \left|f\left(1 - \frac{\epsilon}{n}\right) - S\right| \leqslant \epsilon M + 2\epsilon. \tag{19.319}$$

Il suffit de choisir ϵ suffisamment petit (en particulier pour que ϵM soit petit) pour montrer que $|S_n - S|$ est borné par un nombre arbitrairement petit. □

19.14.1 Passage à la limite sous le signe intégral

Un autre résultat très important pour l'étude de l'intégrabilité est le théorème de la **convergence dominée de Lebesgue** :

Théorème 19.87.
Soit $E \subset \mathbb{R}^n$ un ensemble mesurable et $\{f_k\}$, une suite de fonctions intégrables sur E qui converge simplement vers une fonction $f\colon E \to \overline{\mathbb{R}}$. Supposons qu'il existe une fonction g intégrable sur E telle que pour tout k,

$$|f(x)| \leqslant g(x) \tag{19.320}$$

pour tout $x \in E$. Alors f est intégrable sur E et

$$\int_E f = \lim_{k \to \infty} \int_E f_k. \tag{19.321}$$

19.14.2 Intégrale en dimension un

Proposition 19.88 (Critère de comparaison)**.**
Soit f mesurable sur $]a, \infty[$ et bornée sur tout $]a, b]$, et supposons qu'il existe un $X_0 \geqslant a$, tel que sur $]X_0, \infty[$,

$$|f(x)| \leqslant g(x) \tag{19.322}$$

où $g(x)$ est intégrable. Alors $f(x)$ est intégrable sur $]a, \infty[$.

Corolaire 19.89 (Critère d'équivalence)**.**
Soient f et g des fonctions mesurables et positives ou nulles sur $]a,\infty[$, bornées sur tout $]a,b]$, telles que

$$\lim_{x\to\infty}\frac{f(x)}{g(x)}=L \tag{19.323}$$

existe dans $\bar{\mathbb{R}}$.

(1) Si $L\neq\infty$ et $\int_a^\infty g(x)$ existe, alors $\int_a^\infty f(x)dx$ existe,
(2) Si $L\neq 0$ et si $\int_a^\infty f(x)dx$ existe, alors $\int_a^\infty g(x)dx$ existe,

Corolaire 19.90 (Critère des fonctions test)**.**
Soit $f(x)$ une fonction mesurable et positive ou nulle sur $]a,\infty[$ et bornée pour tout $]a,b]$. Nous posons

$$L(\alpha)=\lim_{x\to\infty}x^\alpha f(x), \tag{19.324}$$

et nous supposons qu'elle existe.

(1) Si il existe $\alpha>1$ tel que $L(\alpha)\neq\infty$, alors $\int_a^\infty f(x)dx$ existe,
(2) Si il existe $\alpha\leqslant 1$ et $L(\alpha)\neq 0$, alors $\int_a^\infty f(x)dx$ n'existe pas.

Corolaire 19.91.
Soit $f\colon]a,b]\to\mathbb{R}$ une fonction mesurable, positive ou nulle, et bornée sur $[a+\epsilon,b]$ $\forall\epsilon>0$. Si $\lim_{x\to a}(x-a)^\alpha f(x)=L$ existe, alors

(1) Si $\alpha<1$ et $L\neq\infty$, alors $\int_a^b f(x)dx$ existe,
(2) Si $\alpha\geqslant 1$ et $L\neq 0$, alors $\int_a^b f(x)dx$ n'existe pas.

19.14.3 Intégrales convergentes

Définition 19.92.
Soit f, une fonction mesurable sur $[a,\infty[$, bornée sur tout intervalle $[a,b]$. On dit que l'intégrale

$$\int_a^\infty f(x)dx \tag{19.325}$$

converge *si la limite*

$$\lim_{X\to\infty}\int_a^X f \tag{19.326}$$

existe et est finie.

19.14.4 La méthode de Rothstein-Trager

Mes sources pour parler d'intégration de fractions rationnelles : [392].

Théorème 19.93 (Rothstein-Trager[393])**.**
Soient $P,Q\in\mathbb{Q}[X]$ premiers entre eux avec $\operatorname{pgcd}(P,Q)=1$ et $\deg(P)<\deg(Q)$. Nous supposons que Q est unitaire et sans facteurs carrés. Supposons que nous puissions écrire, dans un extension $\mathbb{K}$ de $\mathbb{Q}$ la primitive de P/Q de la façon suivante :

$$\int\frac{P}{Q}=\sum_{i=1}^n c_i\ln(P_i) \tag{19.327}$$

où les c_i sont des constantes non nulles et deux à deux distinctes et où les P_i sont des polynômes unitaires non constants sans facteurs carrés et premiers deux à deux entre eux dans $\mathbb{K}[X]$.

Alors les c_i sont les racines distinctes du polynôme

$$R(Y)=\operatorname{res}_X(P-YQ',Q)\in\mathbb{K}[Y] \tag{19.328}$$

et

$$P_i=\operatorname{pgcd}(P-c_iQ',Q). \tag{19.329}$$

Démonstration. Nous posons

$$U_i = \prod_{j\neq i} P_j. \tag{19.330}$$

Question de division Ensuite nous dérivons formellement l'équation (19.327) et nous multiplions les deux côtés du résultat par $\prod_{j=1}^n P_j$:

$$P\prod_{j=1}^n P_j = Q\sum_{i=1}^n c_i\frac{P'i}{P_i}\prod_{j=1}^n P_j = Q\sum_{i=1}^n c_iP_i'U_i. \tag{19.331}$$

Une première chose que nous en tirons est que Q divise le produit $P\prod_{j=1}^n P_j$; mais P et Q étant premiers entre eux,

$$Q \mid \prod_{j=1}^n P_j \tag{19.332}$$

par le théorème de Gauss 6.44.

Une seconde chose que nous tirons de (19.331) est que P_j divise $Q\sum_{i=1}^n c_iP_i'U_i$. De cette somme, à cause du U_i qui est divisé par P_j pour tout i sauf $i=j$, le polynôme P_j divise tous les termes sauf peut-être un. Donc il les divise tous et en particulier

$$P_j \mid Qc_jP_J'U_j \tag{19.333}$$

En nous souvenant que les P_k sont premiers entre eux, P_j ne divise pas U_j. De plus P_j étant sans facteurs carrés, les polynômes P_j et P_j' sont premiers entre eux. Il ne reste que Q. Nous en déduisons que

$$P_j \mid Q \tag{19.334}$$

pour tout $1\leqslant j\leqslant n$. Et vu que les P_i sont premiers entre eux, le fait que chacun divise Q implique que leur produit divise Q, c'est-à-dire

$$\prod_{j=1}^n P_j \mid Q. \tag{19.335}$$

Or nous avions déjà prouvé la division contraire. Du fait que les deux polynômes sont unitaires nous en déduisons qu'ils sont en réalité égaux :

$$Q = \prod_{j=1}^n P_j. \tag{19.336}$$

Nous pouvons simplifier les deux membres de (19.331) par cela :

$$P = \sum_{i=1}^n c_iP_i'U_i. \tag{19.337}$$

Encore un peu de division En dérivant (19.336) nous trouvons

$$Q' = \sum_{j=1}^n P_j'U_j, \tag{19.338}$$

et en écrivant P sous sa forme (19.337),

$$P - c_iQ' = \sum_{j=1}^n c_jP_j'U_j - \sum_{j=1}^n c_iP_j'U_j = \sum_{j=1}^n (c_j-c_i)P_j'U_j. \tag{19.339}$$

Le terme $i=j$ de la somme est nul ; en ce qui concerne les autres termes, ils sont divisés par P_i parce que $P_i \mid U_j$. Donc P_i divise tous les termes de la somme et nous avons

$$P_i \mid P - c_iQ'. \tag{19.340}$$

Un pgcd pour continuer Nous montrons à présent que $P_i = \text{pgcd}(P - c_iQ', Q)$. Pour cela nous utilisons la multiplicativité du PGCD lorsque les facteurs sont premiers entre eux :

$$\text{pgcd}(P - c_iQ', Q) = \text{pgcd}(P - c_iQ', \prod_{j=1}^{n} P_j) = \prod_{j=1}^{n} \text{pgcd}(P - c_iQ', P_j). \tag{19.341}$$

Nous remplaçons $P - c_iQ'$ par son expression (19.339) et nous écrivons un des facteurs du produit :

$$\text{pgcd}(P - c_iQ', P_j) = \text{pgcd}(\sum_{k=1}^{n} (c_k - c_i)P'_kU_k, P_j) \tag{19.342}$$

Le polynôme P_j divise tous les U_k sauf celui avec $k = j$. Donc le lemme 6.50(1) nous permet de dire

$$\text{pgcd}(P - c_iQ', P_j) = \text{pgcd}\left((c_j - c_i)P'_jU_j, P_j\right) = \begin{cases} 1 & \text{si } i \neq j \\ P_j & \text{si } i = j. \end{cases} \tag{19.343}$$

La seconde ligne provient du fait que nous ayons déjà montré que $P_j \mid P - c_jQ'$. En fin de compte,

$$\text{pgcd}(P - c_iQ', Q) = P_i. \tag{19.344}$$

Une histoire de résultant Les nombres c_i sont tels que les polynômes $P - c_iQ'$ et Q ne sont pas premiers entre eux. Vu que les P_i sont non nuls, la proposition 9.18 nous dit que le résultant

$$\text{res}_X(P - c_iQ', Q) = 0. \tag{19.345}$$

Donc les c_i sont des racines du polynôme (en Y)

$$R(Y) = \text{res}_X(P - YQ', Q). \tag{19.346}$$

Nous n'avons pas prouvé qu'ils étaient *toutes* les racines[15].

Toutes les racines Nous allons maintenant montrer que les c_i étaient toutes les racines imaginables du polynôme (19.346) dans toutes les extensions de $\mathbb{Q}$. Soit donc c une racine de R dans une extension $\hat{\mathbb{K}}$ de $\mathbb{K}$ qui ne soit pas parmi les c_i de la formule (19.327). Étant donné que c est racine du résultat, les polynômes $P - cQ'$ et Q ont un PGCD non trivial, c'est-à-dire non constant. Donc

$$\text{pgcd}(P - cQ', Q) = s \in \hat{\mathbb{K}}[X] \tag{19.347}$$

est un polynôme non constant. Si T un facteur irréductible de S, alors T divise $P - cQ'$ et Q, mais $Q = \prod_{i=1}^{n} P_i$ avec les P_i premiers entre eux. Donc T ne peut diviser que l'un (et exactement un) d'entre eux[16]. Soit P_{i_0} celui qui est divisé par T. La relation (19.339) dans ce contexte donne :

$$P - cQ' = \sum_{j=1}^{n} (c_j - c)P'_jU_j \tag{19.348}$$

Le polynôme T divise tous les U_j avec $j \neq i_0$, mais comme en plus il divise $P - cQ'$, il divise aussi le dernier terme de la somme :

$$T \mid (c_{i_0} - c)P'_{i_0}U_{i_0}. \tag{19.349}$$

Le polynôme T ne divisant pas U_{i_0} et $(c_{i_0} - c)$ étant non nul, nous concluons que T divise P'_{i_0}. Mais cela n'est pas possible parce que nous avons supposé que P_{i_0} était sans facteur carré, ce qui voulait entre autres dire que P_{i_0} et P'_{i_0} n'ont pas de facteurs communs.

□

15. De plus, nous n'avons pas de garanties que ces racines soient dans $\mathbb{Q}$, et en fait il y a des cas dans lesquels les c_i n'y sont pas.

16. On ne peut pas diviser deux trucs qui sont premiers entre eux ; c'est une question de cohérence, madame !

Ce théorème suggère la méthode suivante pour trouver la primitive de la fraction rationnelle P/Q (si elle vérifie les hypothèses)

(1) Écrire le résultant $R(y) = \text{res}_X(P - yQ', Q)$ et en trouver les racines $\{c_i\}_{i=1,\dots,n}$.
(2) Calculer les $P_p = \text{pgcd}(P - c_iQ', Q)$.
(3) Écrire la réponse :

$$\int \frac{P}{Q} = \sum_{i=1}^{n} c_i \ln(P_i). \tag{19.350}$$

Notons que le polynôme $R(Y)$ est de degré $\deg(Q)$ (pour le voir, faire un peu de comptage de lignes et colonnes dans la matrice de Sylvester), donc il n'est à priori pas pire à factoriser que Q lui-même [17]. Mais il se peut que nous ayons de la chance et que R soit plus facile que Q.

À part qu'on a peut-être plus de chance avec R qu'avec Q, l'avantage de la méthode est qu'elle permet d'éviter de passer par des extensions de $\mathbb{Q}$ non nécessaires [18].

Exemple 19.94 ([393]).
Prenons la fraction rationnelle $\frac{x}{x^2-3}$. L'intégration via les fractions simples est :

$$\int \frac{x}{x^2-3}dx = \frac{1}{2}\int \frac{1}{x-\sqrt{3}} + \frac{1}{2}\int \frac{1}{x+\sqrt{3}} = \frac{1}{2}\ln(x-\sqrt{3}) + \frac{1}{2}\ln(x+\sqrt{3}) = \frac{1}{2}\ln(x^2-3). \tag{19.351}$$

Nous voyons que dans la réponse, il n'y a pas de racines. Passer par l'extension $\mathbb{Q}[\sqrt{3}]$ est par conséquent peut-être un effort inutile. Voyons comment les choses se mettent avec la méthode Rothstein-Trager.

D'abord

$$\begin{aligned}
R(Y) &= \text{res}_X(X - 2YX, X^2 - 3) && \text{(19.352a)}\\
&= \text{res}_X\big((1-2Y)X, X^2-3\big) && \text{(19.352b)}\\
&= \det\begin{pmatrix} 1-2Y & 0 & 0\\ 0 & 1-2Y & 0\\ 1 & 0 & -3\end{pmatrix} && \text{(19.352c)}\\
&= (1-2Y)\big(-3(1-2Y)\big) && \text{(19.352d)}\\
&= -3(1-2Y)^2, && \text{(19.352e)}
\end{aligned}$$

dont les solutions sont faciles : il n'y a que la racine double $y = \frac{1}{2}$. La somme (19.327) sera donc réduite à un seul terme avec $c_1 = \frac{1}{2}$. Nous calculons P_1 :

$$P_1 = \text{pgcd}(X - \frac{1}{2}2X, X^2 - 3) = \text{pgcd}(0, X^2 - 3) = X^2 - 3, \tag{19.353}$$

et par conséquent

$$\int \frac{X}{X^2-3} = \frac{1}{2}\ln(X^2-3). \tag{19.354}$$

À aucun moment nous ne sommes sortis de $\mathbb{Q}$. △

Comme vu sur cet exemple, l'intérêt du théorème de Rothstein-Trager est de permettre, lorsqu'on a de la chance, d'en profiter, et non de nous en rendre compte à la fin en remarquant bêtement que la réponse pouvait s'écrire dans $\mathbb{Q}[X]$.

Remarque 19.95.
Afin d'utiliser cette méthode, il faut s'assurer que Q soit sans facteurs carrés. Si nous devons intégrer un $\frac{P}{Q}$ quelconque, nous devons commencer par écrire

$$Q = Q_1Q_2^2Q_3^3\dots Q_r^r, \tag{19.355}$$

17. C'est de la factorisation de Q qu'on a besoin pour utiliser la méthode de décomposition en fractions simples.
18. J'imagine que pour un ordinateur, c'est plus facile d'éviter les extensions.

et ensuite il y a moyen de ramener l'intégrale de P/Q à des intégrales de $\frac{P}{Q_1 \dots Q_r}$. Cela ne demande pas de factoriser complètement Q, mais seulement de trouver ses facteurs irréductibles Q_i dans $\mathbb{Q}[X]$.

Dans l'exemple donné plus haut, $Q = X^2 - 3$ a des facteurs irréductibles autres que Q lui-même dans $\mathbb{R}[X]$, mais nous n'en avons pas besoin.

Voici une exemple où nous évitons de passer par les complexes.

Exemple 19.96 ([394]).
À calculer : $\int \frac{1}{x^3+x}$. La décomposition en fractions simples donne :

$$\frac{1}{x^3+x} = \frac{1}{x} - \frac{1/2}{x-i} - \frac{1/2}{x+i}. \tag{19.356}$$

Déjà cette décomposition passe par l'extension $\mathbb{Q}[i]$, et le calcul de la primitive de $\frac{1}{x+i}$ demande le logarithme complexe qui ne sera vu que dans la proposition 25.63 au prix d'un peu de sang. Nous verrons dans l'exemple 25.64 comment ça se passe en passant par les complexes.

En ce qui concerne la méthode de Rothstein-Trager, nous commençons par calculer le résultant (qui est tout de même un peu de calcul) :

$$P(y) = \operatorname{res}_X\big(-3yX^2 - y + 1, X^3 - X\big) \tag{19.357a}$$

$$= \det \begin{pmatrix} -3y & 0 & 1-y & 0 & 0 \\ 0 & -3y & 0 & 1-y & 0 \\ 0 & 0 & -3y & 0 & 1-y \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix} \tag{19.357b}$$

$$= -(y-1)^2(2y+1)^2 \tag{19.357c}$$

```
-----------------------------------------------------------------------
| Sage Version 5.7, Release Date: 2013-02-19                          |
| Type "notebook()" for the browser-based notebook interface.         |
| Type "help()" for help.                                             |
-----------------------------------------------------------------------
sage: y=var('y')
sage: R=matrix(5,5,[-3*y,0,1-y,0,0,0,-3*y,0,1-y,0,0,0,
                    -3*y,0,1-y,1,0,1,0,0,0,1,0,1,0])
sage: R.determinant().factor()
-(y - 1)*(2*y + 1)^2
```

Les solutions sont $c_1 = 1$ et $c_2 = -\frac{1}{2}$. Nous pouvons alors calculer les P_i :

$$P_1 = \operatorname{pgcd}(-3X^2, X^3 + X) = X \tag{19.358}$$

et

$$P_2 = \operatorname{pgcd}(\frac{3}{2}X^2 + \frac{3}{2}, X^3 + X) = X^2 + 1, \tag{19.359}$$

et finalement

$$\int \frac{1}{X^3 + X} = \ln(X) - \frac{1}{2}\ln(X^2 + 1). \tag{19.360}$$

△

Notons qu'il n'y a pas de miracles : lorsque la réponse contient des racines, nous ne pouvons pas couper à passer par des extensions et factoriser un peu à la dure.

Exemple 19.97 ([394]).
Nous voulons calculer

$$\int \frac{1}{X^2+1}. \tag{19.361}$$

Nous posons donc $P = 1$ et $Q = X^2 + 1$. Le résultant à calculer est

$$P(y) = \mathrm{res}_X(-2yX + 1, X^2 + 1) = \det \begin{pmatrix} -2y & 1 & 0 \\ 0 & -2y & 1 \\ 1 & 0 & 1 \end{pmatrix} = 4y^2 + 1. \tag{19.362a}$$

Les racines de cela sont complexes et il n'y a donc pas d'échappatoires : $c_1 = \frac{i}{2}$, $c_2 = -\frac{i}{2}$. Ensuite, étant donné que $X^2 + 1 = (X + i)(X - i) = i(-iX + 1)(X - 1)$ nous avons

$$P_1 = \mathrm{pgcd}(-iX + 1, X^2 + 1) = X + i. \tag{19.363}$$

Notons que de façon naturelle, nous aurions écrit $P_1 = 1-iX$, mais par convention nous considérons le PGCD unitaire. Cela ne change rien à la réponse parce que changer P_i en kP_i ne fait que rajouter une constante $\ln(k)$ à la primitive trouvée.

De la même façon,

$$P_2 = \mathrm{pgcd}(1 + iX, X^2 + 1) = X - i. \tag{19.364}$$

Au final nous écrivons

$$\int \frac{1}{X^2+1} = \frac{i}{2}\ln(X + i) - \frac{i}{2}\ln(X - i). \tag{19.365}$$

△

Remarque 19.98.
Tout cela est si nous voulons absolument écrire la primitive avec des logarithmes de polynômes. Pour celui de l'exemple 19.97, nous avons trouvé

$$\int \frac{1}{X^2+1} = \frac{i}{2}\ln(X + i) - \frac{i}{2}\ln(X - i). \tag{19.366}$$

Mais

```
sage: f(x)=1/(x**2+1)
sage: f.integrate(x)
x |--> arctan(x)
```

Si nous acceptons de passer aux fonctions trigonométriques (inverses), la primitive prend un tour très différent et bien réel. Ces deux visions de l'univers sont bien entendu [19] compatibles. En effet, afin de tomber juste, nous allons prendre la primitive

$$f(x) = \frac{i}{2}\ln(ix - 1) - \frac{i}{2}\ln(ix + 1) \tag{19.367}$$

au lieu de (19.366). Il s'agit seulement de multiplier l'intérieur des logarithmes, ce qui ne donne qu'une constante de différence. Ensuite nous passons à la forme trigonométrique des nombres complexes : $ix - 1 = \sqrt{x^2+1}e^{i\arctan(-x)}$ et $ix + 1 = \sqrt{x^2+1}e^{i\arctan(x)}$. Avec un peu de calcul,

$$f(x) = -\frac{1}{2}\Big(\arctan(-x) - \arctan(x)\Big) = \arctan(x). \tag{19.368}$$

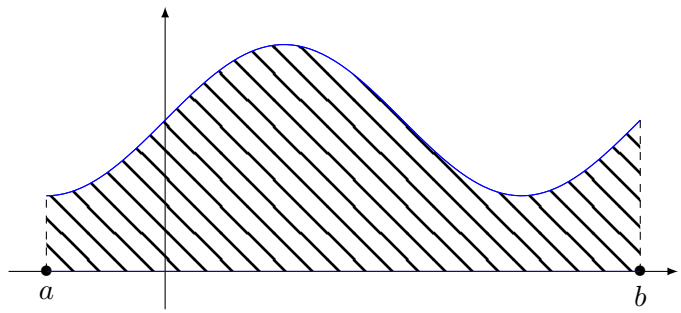


FIGURE 19.5 – L'intégrale de f entre a et b représente la surface sous la fonction.

19.15 Rappel sur les intégrales usuelles

Soit une fonction

$$\begin{aligned} f\colon [a,b] \subset \mathbb{R} &\to \mathbb{R}^+ \\ x &\mapsto f(x). \end{aligned} \tag{19.369}$$

L'intégrale de f sur le segment $[a,b]$, notée $\int_a^b f(x)dx$ est le nombre égal à l'aire de la surface située entre le graphe de f et l'axe des x, comme indiqué à la figure 19.5.

Définition 19.99.
*Si f est une fonction de une variable à valeurs réelles, une **primitive** de f est une fonction F telle que $F' = f$.*

Toute fonction continue admet une primitive.

Théorème 19.100 (Théorème fondamental du caclul intégral)**.**
Si f est une fonction positive et continue, et si F est une primitive de f, alors

$$\int_a^b f(x)dx = F(b) - F(a). \tag{19.370}$$

Remarque 19.101.
Si f est une fonction continue par morceaux, l'intégrale de f se calcule comme la somme des intégrales de ses morceaux. Plus précisément si nous avons $a = x_0 < x_1 < \ldots < x_n = b$ et si f est continue sur $]x_i, x_{i+1}[$ pour tout i, alors nous posons

$$\int_a^b f(x)dx = \int_{x_0}^{x_1} f(x)dx + \int_{x_1}^{x_2} f(x)dx + \cdots + \int_{x_{n-1}}^{n_n} f(x)dx. \tag{19.371}$$

Sur chacun des morceaux, l'intégrale se calcule normalement en passant par une primitive.

19.16 Intégrales le long de chemins

19.16.1 Circulation d'un champ de vecteur

Définition 19.102.
*Soit $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ un champ de vecteurs et un chemin $\sigma\colon [a,b] \to \mathbb{R}^3$. On appelle **circulation** de F le long du chemin σ le scalaire*

$$\int_a^b F\big(\sigma(t)\big) \cdot \sigma'(t)dt. \tag{19.372}$$

Il existe de nombreuses notations pour cela ; entre autres :

$$\int_\sigma F = \int_\sigma F \cdot ds. \tag{19.373}$$

19. Si on croit que la mathématique est cohérente.

En physique, la circulation de la force le long d'un chemin est la travail de la force.

Exemple 19.103.
À la surface de la Terre, le champ de gravitation est donné par

$$G(x,y,z) = -mg \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{19.374}$$

Si nous considérons un mobile qui monte à vitesse constante jusqu'à la hauteur h, c'est-à-dire le chemin

$$\sigma(t) = \begin{pmatrix} 0 \\ 0 \\ t \end{pmatrix} \tag{19.375}$$

avec $t \in [0,h]$. Le travail de la gravitation est alors donné par

$$W = \int_0^h G\big(\sigma(t)\big) \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -mg \int_0^h \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = -mgh. \tag{19.376}$$

Cela est bien le résultat usuel de l'énergie potentielle. Nous allons voir bientôt que nous nommons la fonction mgh énergie *potentielle* précisément parce que la force dérive de ce potentiel. △

Exemple 19.104.
Soit le chemin

$$\begin{aligned} \sigma\colon [0,2\pi] &\to \mathbb{R}^3 \\ t &\mapsto \begin{pmatrix} \sin(t) \\ \cos(t) \\ t \end{pmatrix}. \end{aligned} \tag{19.377}$$

et le champ de vecteurs

$$F \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{19.378}$$

La circulation de ce champ de vecteur le long de l'hélice σ est

$$\begin{aligned} \int_\sigma F \cdot ds &= \int_0^{2\pi} (F \circ \sigma)(t) \cdot \sigma'(t) dt \\ &= \int_0^{2\pi} \begin{pmatrix} \sin(t) \\ \cos(t) \\ t \end{pmatrix} \cdot \begin{pmatrix} \cos(t) \\ \sin(t) \\ 1 \end{pmatrix} dt \\ &= \int_0^{2\pi} t dt \\ &= \left[\frac{t^2}{2} \right]_0^{2\pi} \\ &= 2\pi^2. \end{aligned} \tag{19.379}$$

△

Proposition 19.105.
La circulation d'un champ de vecteurs le long d'un chemin ne dépend pas de le paramétrage. En d'autres termes, si σ_1 et σ_2 sont deux chemins équivalents, alors

$$\int_{\sigma_1} F = \int_{\sigma_2} F. \tag{19.380}$$

Démonstration. Soient deux chemins $\sigma_1\colon [a,b] \to \mathbb{R}^3$ et $\sigma_2\colon [c,d] \to \mathbb{R}^3$ équivalents, c'est-à-dire tels que

$$\sigma_1(t) = \sigma_2\big(\varphi(t)\big) \tag{19.381}$$

où $\varphi\colon [a,b] \to [c,d]$ strictement croissante. En utilisant le fait que $\sigma_1(t) = \varphi'(t)\sigma_2'\big(\varphi(t)\big)$, nous avons

$$\begin{aligned}
\int_{\sigma_1} F \cdot ds &= \int_a^b F\big(\sigma_1(t)\big) \cdot \sigma_1'(t)dt \\
&= \int_a^b F\Big(\sigma_2\big(\varphi(t)\big)\Big) \cdot \sigma_2'\big(\varphi(t)\big)\varphi'(t)dt \\
&= \int_{\varphi(a)}^{\varphi(b)} F\big(\sigma_2(s)\big) \cdot \sigma_2(s)ds \\
&= \int_c^d F\big(\sigma_2(s)\big) \cdot \sigma_2'(s)ds \\
&= \int_{\sigma_2} F \cdot ds.
\end{aligned} \tag{19.382}$$

où nous avons effectué le changement de variables $s = \varphi(t)$, $ds = \varphi'(t)dt$. □

Remarque 19.106.
Si σ_2 est le chemin opposé de σ, alors

$$\int_{\sigma_2} F = -\int_{\sigma_1} F. \tag{19.383}$$

19.17 Circulation d'un champ conservatif

Si nous avons une fonction scalaire $V\colon \mathbb{R}^3 \to \mathbb{R}$, nous pouvons construire un champ de vecteur en prenant le gradient :

$$F(x) = \nabla V(x). \tag{19.384}$$

On dit que le champ de vecteur F **dérive** de V, et on dit que V est le **potentiel** de F. Nous posons la définition suivante :

Définition 19.107.
*Un champ de vecteurs $F\colon \mathbb{R}^3 \to \mathbb{R}^3$ est un champ **conservatif** s'il existe une fonction $V\colon \mathbb{R}^3 \to \mathbb{R}$ telle que*

$$F(x) = \nabla V(x). \tag{19.385}$$

Nous disons aussi parfois que le champ V dérive d'un potentiel *ou bien qu'il s'agit d'un* champ de gradient.

Les champs de vecteurs conservatifs sont particulièrement importants parce que presque toutes les forces connues en physiques dérivent d'un potentiel. Nous verrons que la terminologie « conservatif » provient du fait que les forces de ce type conservent l'énergie associée.

Proposition 19.108.
Considérons une fonction $V\colon \mathbb{R}^3 \to \mathbb{R}$ (que nous appellerons potentiel*) et le champ de vecteur qui en dérive :*

$$F = \nabla V. \tag{19.386}$$

Alors

$$\int_\sigma F \cdot ds = V\big(\sigma(b)\big) - V\big(\sigma(a)\big). \tag{19.387}$$

Autrement dit, le travail nécessaires pour déplacer un objet d'un point à un autre dans un champ de force conservatif vaut la différence de potentiel entre le point de départ et le point d'arrivée.

Démonstration. Par définition,

$$\int_\sigma F \cdot ds = \int_a^b F\big(\sigma(t)\big) \cdot \sigma'(t)dt. \tag{19.388}$$

Nous pouvons transformer l'intégrante de la façon suivante :

$$\begin{aligned} F\big(\sigma(t)\big) \cdot \sigma'(t) &= \nabla V\big(\sigma(t)\big) \cdot \sigma'(t) \\ &= \frac{\partial V}{\partial x}\big(\sigma(t)\big)\sigma'_x(t) + \frac{\partial V}{\partial y}\big(\sigma(t)\big)\sigma'_y(t) + \frac{\partial V}{\partial z}\big(\sigma(t)\big)\sigma'_z(t) \\ &= \frac{d}{dt}\Big[V\big(\sigma(t)\big)\Big] \end{aligned} \tag{19.389}$$

où nous avons posé

$$\sigma(t) = \begin{pmatrix} \sigma_x(t) \\ \sigma_y(t) \\ \sigma_z(t) \end{pmatrix} \tag{19.390}$$

et utilisé à l'envers la formule de dérivation de fonction composée pour

$$\frac{d}{dt}\Big[V\big(\sigma(t)\big)\Big] = \Big((V \circ \sigma)(t)\Big)'. \tag{19.391}$$

En remettant ces expressions dans l'intégrale (19.388),

$$\int_\sigma F \cdot ds = \int_a^b \frac{d}{dt}\Big[V\big(\sigma(t)\big)\Big]dt = V\big(\sigma(b)\big) - V\big(\sigma(a)\big). \tag{19.392}$$

□

Exemple 19.109.
Nous savons que le champ de gravitation dérive d'un potentiel. À la surface de la Terre, le potentiel de gravitation vu par une masse m est donné par la fonction $V(x, y, z) = mgz$. Si nous voulons soulever cette masse d'une hauteur h, cela demandera toujours une énergie mgh, quel que soit le chemin suivit : en ligne droite vertical, en diagonal, en hélice, ... △

Exemple 19.110.
À plus grande échelle, le champ de gravitation est encore un champ qui dérive d'un potentiel. En coordonnées sphériques,

$$V(\rho, \theta, \varphi) = k\frac{m}{\rho} \tag{19.393}$$

Lorsqu'un satellite a une orbite de rayon R autour la Terre, il reste sur la sphère $\rho = R$. Donc il reste sur une surface sur laquelle V est constante. Il n'y a donc pas de travail de la force de gravitation ! C'est pour cela qu'un satellite peut tourner pendant des siècles sans apport énergétique. △

Exemple 19.111.
Soit le champ de vecteurs

$$F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix} \tag{19.394}$$

et le chemin

$$\sigma(t) = \begin{pmatrix} t^4/4 \\ \sin^3(t\frac{\pi}{2}). \end{pmatrix} \tag{19.395}$$

Nous voulons calculer la circulation de F le long du chemin σ entre $t = 0$ et $t = 1$.

La première chose à voir est que $F = \nabla V$ avec $V(x, y) = xy$. Donc la circulation sera donnée par

$$\int_\sigma F \cdot ds = V\big(\sigma(1)\big) - V\big(\sigma(0)\big) = V\big(\frac{1}{4}, 1\big) - V(0, 0) = \frac{1}{4} - 0 = \frac{1}{4}. \tag{19.396}$$

Nous n'avons pas réellement calculé l'intégrale. △

19.18 Intégration de fonction à deux variables

19.18.1 Intégration sur un domaine rectangulaire

Soit une fonction positive

$$\begin{aligned} f\colon [a,b]\times[c,d] &\to \mathbb{R}^+ \\ (x,y) &\mapsto f(x,y). \end{aligned} \tag{19.397}$$

L'intégrale de f sur le rectangle $[a,b]\times[c,d]$ est le volume sous le graphe de la fonction. C'est-à-dire le volume de l'ensemble

$$\{(x,y,z) \text{ tel que } (x,y)\in[a,b]\times[c,d], z\leqslant f(x,y)\}. \tag{19.398}$$

Théorème 19.112 (Théorème de Fubini)**.**
Soit une fonction $f\colon \mathbb{R}^2\to\mathbb{R}$ une fonction continue par morceaux sur $\mathcal{R}=[a,b]\times[c,d]$. Alors

$$\int_{\mathcal{R}} f(x,y)dxdy = \int_a^b\left[\int_c^d f(x,y)dy\right]dx = \int_c^d\left[\int_a^b f(x,y)dx\right]dy. \tag{19.399}$$

En pratique, nous utilisons le théorème de Fubini pour calculer les intégrales sur des rectangles.

Exemple 19.113.
Nous voudrions intégrer la fonction $f(x,y)-4+x^2+y^2$ sur le rectangle de la figure 19.6.

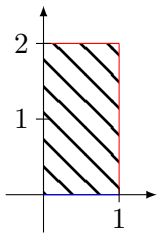


FIGURE 19.6 – Intégration sur un rectangle

L'ensemble sur lequel nous intégrons est donné par le produit cartésien d'intervalles $E=[0,1]\times[0,2]$. Le théorème de Fubini montre que nous pouvons intégrer séparément sur l'intervalle horizontal et vertical :

$$\int_{E=[0,1]\times[0,2]} f = \int_{[0,1]}\left(\int_{[0,2]}(4-x^2-y^2)dy\right)dx. \tag{19.400}$$

Ces intégrales sont maintenant des intégrales usuelles qui s'effectuent en calculant des primitives :

$$\begin{aligned} \int_0^1\int_0^2(4-x^2-y^2)dy\,dx &= \int_0^1\left[4y-x^2y-\frac{y^3}{3}\right]_0^2 dx \\ &= \int_0^1(8-2x^2-\frac{8}{3})dx \\ &= \left[\frac{16x}{3}-\frac{2x^3}{3}\right]_0^1 \\ &= \frac{14}{3}. \end{aligned} \tag{19.401}$$

Avec Sage, on peut faire comme ceci :

```
----------------------------------------------------------------------
| Sage Version 4.6.1, Release Date: 2011-01-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
```

```
sage: f(x,y)=4-x**2-y**2
sage: f.integrate(y,0,2).integrate(x,0,1)
(x, y) |--> 14/3
```

△

19.18.2 Intégration sur un domaine non rectangulaire

Nous voulons maintenant intégrer la fonction $f(x,y) = x^2 + y^2$ sur le triangle de la figure 19.7.

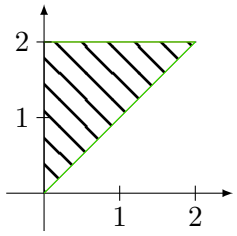
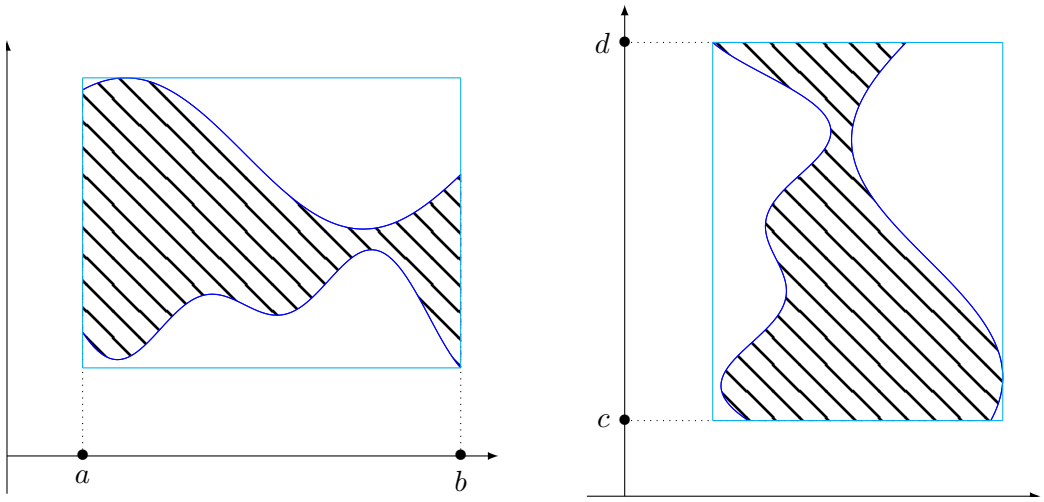


FIGURE 19.7 – Intégration sur un triangle

Étant donné que y varie de 0 à 2 et que *pour chaque* y, la variable x varie de 0 à y, nous écrivons l'intégrale sur le triangle sous la forme :

$$\int_{\text{triangle}} (x^2 + y^2)dxdy = \int_0^2 \left(\int_0^y (x^2 + y^2)dx \right) dy. \tag{19.402}$$

Il existe principalement deux types de domaines non rectangulaires : les « horizontaux » et les « verticaux », voir figure 19.8.



(a) Un domaine horizontal. (b) Un domaine vertical.

FIGURE 19.8 – Deux types de surfaces. Nous avons tracé un rectangle qui contient chacune des deux surfaces. L'intégrale sur un domaine sera l'intégrale sur le rectangle de la fonction qui vaut zéro en dehors du domaine.

Les surfaces horizontales sont de la forme

$$D = \{(x,y) \text{ tel que } x \in [a,b], \varphi_1(x) \leqslant y \leqslant \varphi_2(x)\} \tag{19.403}$$

où φ_1 et φ_2 sont les deux fonctions qui bornent le domaine. Le domaine D est la région comprise entre les graphes de φ_1 et φ_2. Pour un tel domaine nous avons

$$\int_D f(x,y)dxdy = \int_a^b dx \int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)dy. \tag{19.404}$$

Les surfaces verticales sont de la forme

$$D = \{(x,y) \text{ tel que } y \in [c,d], \psi_1(y) \leqslant x \leqslant \psi_2(y)\} \tag{19.405}$$

où φ_1 et φ_2 sont les deux fonctions qui bornent le domaine. Le domaine D est la région comprise entre les graphes de φ_1 et φ_2. Dans ces cas nous avons

$$\int_D f = \int_c^d dy \int_{\psi_1(y)}^{\psi_2(y)} f(x,y)dx. \tag{19.406}$$

Proposition 19.114.
L'aire du domaine D vaut l'intégrable de la fonction $f(x,y) = 1$ sur D :

$$Aire(D) = \int_D dxdy. \tag{19.407}$$

Démonstration. Supposons que le domaine soit du type « horizontal ». En utilisant le théorème de Fubini avec $f(x,y) = 1$ nous avons

$$\int_D dxdy = \int_a^b \left[\int_{\varphi_1(x)}^{\varphi_2(x)} dy \right] dx = \int_a^b \big[\varphi_2(x) - \varphi_1(x)\big]. \tag{19.408}$$

Cela représente l'aire sous φ_2 moins l'aire sous φ_1, et par conséquent l'aire contenue entre les deux. □

Exemple 19.115.
Cherchons la surface du disque de centre $(0,0)$ et de rayon 1 dessinée à la figure 19.9.

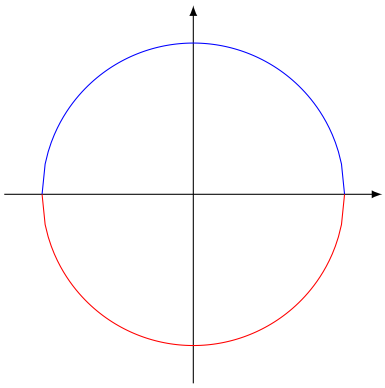


FIGURE 19.9 – En bleu, la fonction $\sqrt{r^2 - x^2}$ et en rouge, la fonction $-\sqrt{r^2 - x^2}$.

Le domaine est donné par $\varphi_1(x) \leqslant y \leqslant \varphi_2(x)$ et $x \in [-r, r]$ où $\varphi_1(x) = -\sqrt{r^2 - x^2}$ et $\varphi_2(x) = \sqrt{r^2 - x^2}$. L'aire est donc donnée par

$$A = \int_{-r}^{r} \big[\varphi_2(x) - \varphi_1(x)\big]dx = 2\int_{-r}^{r} \sqrt{r^2 - x^2}dx = 4\int_0^r \sqrt{r^2 - x^2}. \tag{19.409}$$

Nous effectuons le premier changement de variables $x = ru$, donc $dx = rdu$. En ce qui concerne les bornes, si $x = 0$, alors $u = 0$ et si $x = r$, alors $u = 1$. L'intégrale à calculer devient

$$A = 4\int_0^1 \sqrt{r^2 - r^2u^2}rdu = 4r^2\int_0^1 \sqrt{1 - u^2}du. \tag{19.410}$$

Cette dernière intégrale se calcule en posant

$$\begin{aligned} u &= \sin(t) & du &= \cos(t)dt \\ u &= 0 & t &= 0 \\ u &= 1 & t &= \pi/2. \end{aligned} \tag{19.411}$$

Nous avons

$$A = 4r^2 \int_0^{\pi/2} \sqrt{1-\sin^2(t)}\cos(t)dt = 4r^2 \int_0^{\pi/2} \cos^2(t)dt. \tag{19.412}$$

En utilisant la formule $2\cos^2(x) = 1 + \cos(2x)$, nous avons

$$A = 4r^2 \int_0^{\pi/2} \frac{1+\cos(2t)}{2} dt = \pi r^2. \tag{19.413}$$

△

19.18.3 Changement de variables

Nous n'allons pas parler de changements de variables maintenant parce que les principaux exemples sont les coordonnées polaires, cylindriques et sphériques qui requièrent les fonctions trigonométriques. Ce sera pour la section 17.14.

19.19 Les intégrales triples

Les intégrales triples fonctionnent exactement de la même manière que les intégrales doubles. Il s'agit de déterminer sur quelle domaine les variables varient et d'intégrer successivement par rapport à x, y et z. Il est autorisé de permuter l'ordre d'intégration [20] à condition d'adapter les domaines d'intégration.

Exemple 19.116.
Soit le domaine parallélépipédique rectangle

$$R = [0,1] \times [1,2] \times [0,4]. \tag{19.414}$$

Pour intégrer la fonction $f(x,y,z) = x^2 y \sin(z)$ sur R, nous faisons

$$\begin{aligned} I &= \int_R x^2 y \sin(z)\, dxdydz \\ &= \int_0^1 dx \int_1^2 dy \int_0^4 x^2 y \sin(z) dz \\ &= \int_0^1 dx \int_1^2 x^2 y (1-\cos(4)) dy \\ &= \int_0^1 \frac{3}{2}(1-\cos(4)) x^2 dx \\ &= \frac{1}{2}\big(1-\cos(4)\big). \end{aligned} \tag{19.415}$$

```
----------------------------------------------------------------------
| Sage Version 4.6.1, Release Date: 2011-01-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: f(x,y,z)=x**2*y*sin(z)
sage: f.integrate(x,0,1).integrate(y,1,2).integrate(z,0,4)
(x, y, z) |--> -1/2*cos(4) + 1/2
```

20. En toute rigueur, cela n'est pas vrai, mais nous ne considérons seulement des cas où cela est autorisé.

△

Exemple 19.117.
Soit D la région délimitée par le plan $x = 0$, $y = 0$, $z = 2$ et la surface d'équation

$$z = x^2 + y^2. \tag{19.416}$$

Cherchons à calculer $\int_D x\,dx\,dy\,dz$. Ici, un dessin indique que le volume considéré est $z \geqslant x^2 + y^2$. Il y a plusieurs façons de décrire cet ensemble. Une est celle-ci :

$$\begin{aligned} z &\colon 0 \to 2 \\ x &\colon 0 \to \sqrt{z} \\ y &\colon 0 \to \sqrt{z - x^2}. \end{aligned} \tag{19.417}$$

Cela revient à dire que z peut prendre toutes les valeurs de 0 à 2, puis que pour chaque z, la variable x peut aller de 0 à $\sqrt{z}$, mais que pour chaque z et x fixés, la variable y ne peut pas dépasser $\sqrt{z - x^2}$. En suivant cette méthode, l'intégrale à calculer est

$$\int_0^2 dz \int_0^{\sqrt{z}} dx \int_0^{\sqrt{z-x^2}} f(x, y, z)dy. \tag{19.418}$$

```
----------------------------------------------------------------------
| Sage Version 4.6.1, Release Date: 2011-01-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: f(x,y,z)=x
sage: assume(z>0)
sage: assume(z-x**2>0)
sage: f.integrate(y,0,sqrt(z-x**2)).integrate(x,0,sqrt(z)).integrate(z,0,2)
(x, y, z) |--> 8/15*sqrt(2)
```

Notez qu'il a fallu aider Sage en lui indiquant que $z > 0$ et $z - x^2 > 0$.

Un autre paramétrage serait

$$\begin{aligned} x &\colon 0 \to \sqrt{2} \\ y &\colon 0 \to \sqrt{2 - x^2} \\ z &\colon x^2 + y^2 \to 2. \end{aligned} \tag{19.419}$$

```
----------------------------------------------------------------------
| Sage Version 4.6.1, Release Date: 2011-01-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: f(x,y,z)=x
sage: assume(2-x**2>0)
sage: f.integrate(y,0,sqrt(z-x**2)).integrate(x,0,sqrt(z)).integrate(z,0,2)
(x, y, z) |--> 8/15*sqrt(2)
```

Écrivons le détail de cette dernière intégrale :

$$\begin{aligned} I &= \int_0^{\sqrt{2}} dx \int_0^{\sqrt{2-x^2}} dy \int_{x^2+y^2}^2 x dz \\ &= \int_0^{\sqrt{2}} dx \int_0^{\sqrt{2-x^2}} x(2 - x^2 - y^2)dy \\ &= \int_0^{\sqrt{2}} dx\, x \left[(2 - x^2)y - \frac{y^3}{3} \right]_0^{\sqrt{2-x^2}} \\ &= \int_0^{\sqrt{2}} \frac{2}{3} x(2 - x^2)^{3/2} dx. \end{aligned} \tag{19.420}$$

Ici nous effectuons le changement de variable $u = x^2$, $du = 2xdx$. Ne pas oublier de changer les bornes de l'intégrale :

$$I = \frac{1}{3}\int_0^2 (2-u)^{3/2}du. \tag{19.421}$$

Le changement de variable $t = 2 - u$, $dt = -du$ fait venir (attention aux bornes!!)

$$I = -\frac{1}{3}\int_2^0 t^{3/2}dt = \frac{1}{3}\left[\frac{t^{5/2}}{5/2}\right]_0^2 = \frac{8}{15}\sqrt{2}. \tag{19.422}$$

△

19.19.1 Volume

Parmi le nombreuses interprétations géométriques de l'intégrale triple, notons celle-ci :

Proposition 19.118.
Soit $D \subset \mathbb{R}^3$. Le volume de D est donné par

$$Vol(D) = \int_D dxdydz. \tag{19.423}$$

C'est-à-dire l'intégrale de la fonction $f(x, y, z) = 1$ sur D.

Suivant les points de vue, cette proposition peut être considérée comme une *définition*] du volume.

Exemple 19.119.
Calculons le volume de la sphère de rayon R. Le domaine de variation des variables x, y et z pour la sphère est

$$\begin{aligned} x&\colon -R \to R \\ y&\colon -\sqrt{R^2-x^2} \to \sqrt{R^2-x^2} \\ z&\colon -\sqrt{R^2-x^2-y^2} \to \sqrt{R^2-x^2-y^2}. \end{aligned} \tag{19.424}$$

Par conséquent nous devons calculer l'intégrale

$$V = \int_{-R}^{R} dx \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy \int_{-\sqrt{R^2-x^2-y^2}}^{\sqrt{R^2-x^2-y^2}} dz. \tag{19.425}$$

La première intégrale est simple :

$$V = 2\int_{-R}^{R} dx \int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} \sqrt{R^2-x^2-y^2}dy. \tag{19.426}$$

Afin de simplifier la notation, nous posons $a = R^2 - x^2$. Ceci n'est pas un changement de variables : juste une notation provisoire le temps d'effectuer l'intégration sur y. Étudions donc

$$I = \int_{-\sqrt{a}}^{\sqrt{a}} \sqrt{a-y^2}dy, \tag{19.427}$$

ce qui est la surface du demi-disque de rayon $\sqrt{a}$. Nous avons donc

$$I = \frac{\pi a}{2} = \frac{\pi}{2}(R^2 - x^2), \tag{19.428}$$

et

$$V = 2\int_{-R}^{R} \frac{\pi}{2}(R^2-x^2)dx = \pi\left[R^2x - \frac{x^3}{3}\right]_{-R}^{R} = \frac{4}{3}\pi R^3. \tag{19.429}$$

△

Exemple 19.120.
Nous pouvons calculer le volume de la sphère en utilisant les coordonnées sphériques. Les bornes des variables pour la sphère de rayon R sont

$$\begin{aligned}\rho &: 0 \to R\\ \theta &: 0 \to \pi\\ \varphi &: 0 \to 2\pi.\end{aligned} \tag{19.430}$$

En n'oubliant pas le jacobien $\rho^2 \sin(\theta)$, l'intégrale à calculer est

$$V = \int_0^R d\rho \int_0^{2\pi} d\varphi \int_0^{\pi} \rho^2 \sin(\theta) d\theta \tag{19.431}$$

L'intégrale sur φ fait juste une multiplication par 2π. Celle sur ρ vaut

$$\int_0^R \rho^2 d\rho = \frac{R^3}{3}. \tag{19.432}$$

L'intégrale sur θ donne

$$\int_0^{\pi} \sin(\theta) d\theta = [-\cos(\theta)]_0^{\pi} = 2. \tag{19.433}$$

Le tout fait par conséquent

$$V = \frac{4}{3}\pi R^3. \tag{19.434}$$

Sans contestes, le passage aux coordonnées sphériques a considérablement simplifié le calcul par rapport à celui de l'exemple 19.119. △

19.20 Un petit peu plus formel

19.20.1 Intégration sur un domaine non rectangulaire

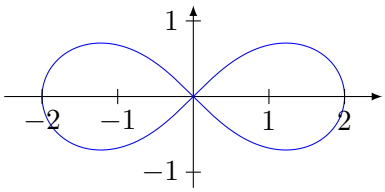


FIGURE 19.10 – Intégrer sur des domaines plus complexes.

La méthode de Fubini ne fonctionne plus sur un domaine non rectangulaire tel que celui de la figure 19.10. Nous allons donc utiliser une astuce. Considérons le domaine

$$E = \{(x, y) \in \mathbb{R}^2 \text{ tel que } a < x < b \text{ et } \alpha(x) < y < \beta(x)\} \tag{19.435}$$

représenté sur la figure 19.10. Nous considérons la fonction

$$\tilde{f}(x, y) = \begin{cases} f(x, y) & \text{si } (x, y) \in E \\ 0 & \text{sinon.} \end{cases} \tag{19.436}$$

Ensuite intégrons $\tilde{f}$ sur un rectangle qui englobe la surface à intégrer à l'aide de Fubini. Étant donné que $\tilde{f} = f$ sur la surface et que $\tilde{f}$ est nulle en dehors, nous avons

$$\int_E f = \int_E \tilde{f} = \int_{\text{rectangle}} \tilde{f} = \int_a^b \left(\int_{\alpha(x)}^{\beta(x)} f(x, y) dy \right) dx. \tag{19.437}$$

Dans le cas de l'intégrale de $f(x, y) = x^2 + y^2$ sur le triangle de la figure 19.7, nous avons

$$\int_{\text{triangle}} (x^2 + y^2) dx dy = \int_0^2 \left(\int_0^y (x^2 + y^2) dx \right) dy. \tag{19.438}$$

Remarque 19.121.
Le nombre $\int_D f(x,y)dxdy$ ne dépend pas du choix du rectangle englobant D.

En pratique, nous calculons l'intégrale en utilisant une extension du théorème de Fubini :

Théorème 19.122.
Soit $f\colon D \subset \mathbb{R}^2 \to \mathbb{R}$ une fonction continue où D est un domaine de type vertical ou horizontal.

(1) Si D est vertical, alors

$$\int_D f = \int_a^b \left[\int_{\varphi_1(x)}^{\varphi_2(x)} f(x,y)dy \right] dx. \tag{19.439}$$

(2) Si D est horizontal, alors

$$\int_D f = \int_c^d \left[\int_{\psi_1(y)}^{\psi_2(y)} f(x,y)dx \right] dy. \tag{19.440}$$

19.20.2 Changement de variables

Le théorème du changement de variable est le suivant.

Théorème 19.123.
Soit $g\colon A \to B$ un difféomorphisme. Soient $F \subset B$ un ensemble mesurable et borné et $f\colon F \to \mathbb{R}$ une fonction bornée et intégrable. Supposons que $g^{-1}(F)$ soit borné et que Jg soit borné sur $g^{-1}(F)$. Alors

$$\int_F f(x)dy = \int_{g^{-1}(F)} f\big(g(x)\big)|Jg(x)|dx \tag{19.441}$$

Pour rappel, Jg est le déterminant de la matrice jacobienne (aucun lien de parenté) donnée par

$$Jg = \det \begin{pmatrix} \partial_x g_1 & \partial_y g_1 \\ \partial_x g_2 & \partial_t g_2 \end{pmatrix}. \tag{19.442}$$

Un **difféomorphisme** est une application $g\colon A \to B$ telle que g et $g^{-1}\colon B \to A$ soient de classe C^1.

19.20.2.1 Coordonnées polaires

Les coordonnées polaires sont données par le difféomorphisme

$$\begin{aligned} g\colon]0,\infty[\times]0,2\pi[&\to \mathbb{R}^2 \backslash D \\ (r,\theta) &\mapsto \big(r\cos(\theta), r\sin(\theta)\big) \end{aligned} \tag{19.443}$$

où D est la demi-droite $y = 0$, $x \geqslant 0$. Le fait que les coordonnées polaires ne soient pas un difféomorphisme sur tout $\mathbb{R}^2$ n'est pas un problème pour l'intégration parce que le manque de difféomorphisme est de mesure nulle dans $\mathbb{R}^2$. Le jacobien est donné par

$$Jg = \det \begin{pmatrix} \partial_r x & \partial_\theta x \\ \partial_r y & \partial_\theta y \end{pmatrix} = \det \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix} = r. \tag{19.444}$$

19.20.2.2 Coordonnées sphériques

Les coordonnées sphériques sont données par

$$\begin{cases} x = r\cos\theta\sin\varphi \\ y = r\sin\theta\sin\varphi \\ z = r\cos\varphi \end{cases} \quad \text{avec} \quad \begin{aligned} r &\in]0,\infty[\\ \theta &\in]0,2\pi[\\ \phi &\in]0,\pi[. \end{aligned} \tag{19.445}$$

Le jacobien associé est $Jg(r,\theta,\varphi) = -r^2\sin\varphi$. Rappelons que ce qui rentre dans l'intégrale est la valeur absolue du jacobien.

Si nous voulons calculer le volume de la sphère de rayon R, nous écrivons donc

$$\int_0^R dr \int_0^{2\pi} d\theta \int_0^{\pi} r^2 \sin(\phi) d\phi = 4\pi R = \frac{4}{3}\pi R^3. \tag{19.446}$$

Ici, la valeur absolue n'est pas importante parce que lorsque $\phi \in]0, \pi, [$, le sinus de ϕ est positif.

Des petits malins pourraient remarquer que le changement de variable (19.445) est encore un paramétrage de $\mathbb{R}^3$ si on intervertit le domaine des angles :

$$\begin{aligned} \theta &: 0 \to \pi \\ \phi &: 0 \to 2\pi, \end{aligned} \tag{19.447}$$

alors nous paramétrons encore parfaitement bien la sphère, mais hélas

$$\int_0^R dr \int_0^{\pi} d\theta \int_0^{2\pi} r^2 \sin(\phi) d\phi = 0. \tag{19.448}$$

Pourquoi ces « nouvelles » coordonnées sphériques sont-elles mauvaises ? Il y a que quand l'angle ϕ parcours $]0, 2\pi[$, son sinus n'est plus toujours positif, donc la *valeur absolue* du jacobien n'est plus $r^2 \sin(\phi)$, mais $r^2 \sin(\phi)$ pour les ϕ entre 0 et π, puis $-r^2 \sin(\phi)$ pour ϕ entre π et 2π. Donc l'intégrale (19.448) n'est pas correcte. Il faut la remplacer par

$$\int_0^R dr \int_0^{\pi} d\theta \int_0^{\pi} r^2 \sin(\phi) d\phi - \int_0^R dr \int_0^{\pi} d\theta \int_{\pi}^{2\pi} r^2 \sin(\phi) d\phi = \frac{4}{3}\pi R^3 \tag{19.449}$$

19.21 Aire et primitive

Soient $f : \mathbb{R} \to \mathbb{R}$ une fonction continue et $x \in \mathbb{R}$. Pour chaque $x \in \mathbb{R}$, nous pouvons considérer le nombre $F(x)$ défini par

$$F(x) = \int_a^x f(t) dt. \tag{19.450}$$

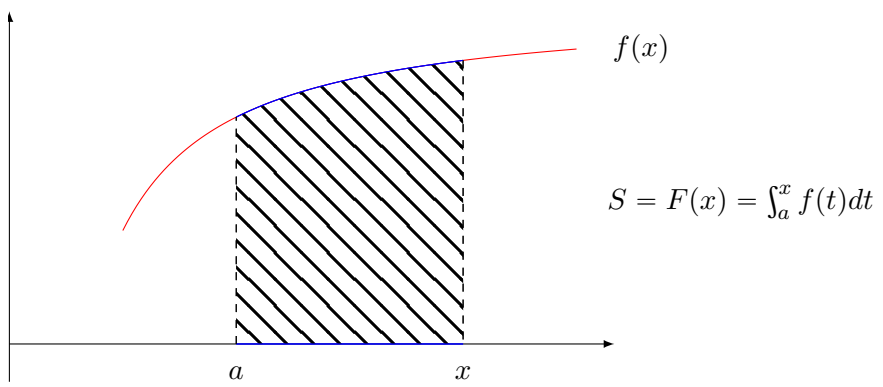


FIGURE 19.11 – Surface sous une courbe

La fonction F ainsi définie a deux importantes propriétés :

(1) C'est une primitive de f,

(2) Elle donne la surface en dessous de f entre les points a et x, voir la figure 19.11.

Notons que tant que f est positive, la surface est croissante.

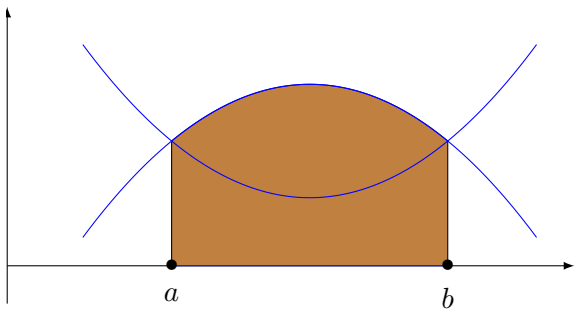
La manière de calculer la surface comprise entre deux fonctions est dessinée à la figure 19.12.

La surface entre les deux fonctions $y_1(x)$ et $y_2(x)$ se calcule comme suit.

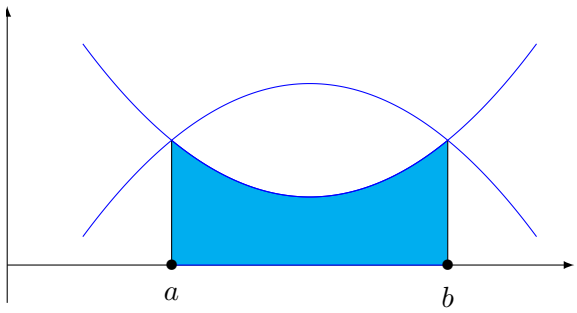
(1) On calcule les intersections entre $y1$ et y_2. Notons a et b les ordonnées obtenues.

(2) La surface demandée est la différence entre la surface sous la fonction y_1 (la plus grande) et la surface sous la fonction y_2 (la plus petite), donc

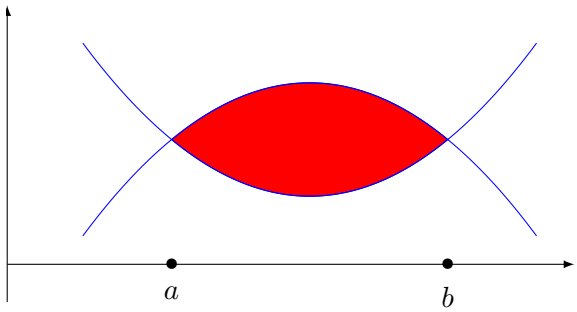
$$S = \int_a^b y_1 - \int_a^b y_1. \tag{19.451}$$



(a) La grande surface.



(b) La surface à soustraire.



(c) La surface entre les deux fonctions.

FIGURE 19.12 – Le calcul de la surface comprise entre deux fonctions.

19.21.1 Longueur d'arc de courbe

La longueur de l'arc de courbe de la fonction $y = f(x)$ entre les abscisses x_0 et x_1 est donné par la formule

$$l(x_0, x_1) = \int_{x_0}^{x_1} \sqrt{1 + y'(t)^2} dt. \tag{19.452}$$

Lorsque la courbe est donnée sous forme paramétrique

$$\begin{cases} x = x(t) & (19.453a) \\ y = y(t), & (19.453b) \end{cases}$$

alors la formule devient

$$l(t_1, t_2) = \int_{t_1}^{t_2} \sqrt{\dot{x}(t)^2 + \dot{y}(t)^2} dt, \tag{19.454}$$

où $\dot{x}(t) = x'(t)$.

19.21.2 Aire de révolution

Pour savoir l'aire engendrée par la ligne $y = f(x)$ entre a et b autour de l'axe Ox, on utilise la formule

$$S = 2\pi \int_a^b \sqrt{1 + f'(x)^2} f(x) dx. \tag{19.455}$$

19.22 L'aire en dessous d'une courbe

Soit f une fonction à valeurs dans $\mathbb{R}^+$.

Nous voudrions pouvoir calculer l'aire au-dessous du graphe de la fonction f. Nous notons $S_f(x)$ l'aire là-dessous de la fonction f entre l'abscisse 0 et x, c'est-à-dire l'aire bleue de la figure 19.13.

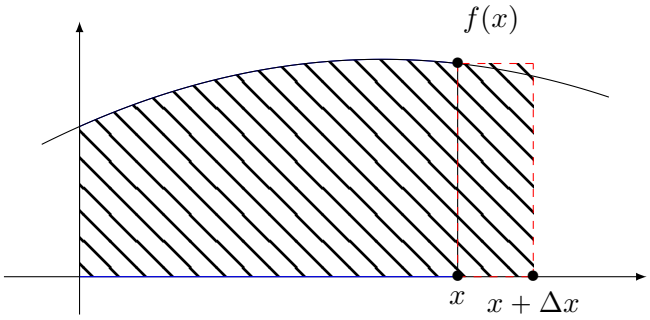


FIGURE 19.13 – L'aire en dessous d'une courbe. Le rectangle rouge d'aire $f(x)\Delta x$ approxime de combien la surface augmente lorsqu'on passe de x à $x + \Delta x$.

Si la fonction f est continue et que Δx est assez petit, la fonction ne varie pas beaucoup entre x et $x+\Delta x$. L'augmentation de surface entre x et $x+\Delta x$ peut donc être approximée par le rectangle de surface $f(x)\Delta x$. Ce que nous avons donc, c'est que quand Δx est très petit,

$$S_f(x + \Delta x) - S_f(x) = f(x)\Delta x, \tag{19.456}$$

ou encore

$$f(x) = \frac{S_f(x + \Delta x) - S_f(x)}{\Delta x}. \tag{19.457}$$

Nous formalisons la notion de « lorsque Δx est très petit » par une limite :

$$f(x) = \lim_{\Delta x \to 0} \frac{S_f(x + \Delta x) - S_f(x)}{\Delta x}. \tag{19.458}$$

Donc, la fonction f est la dérivée de la fonction qui représente l'aire là-dessous de f. Calculer des surfaces revient donc au travail inverse de calculer des dérivées.

19.23 Propriétés des intégrales

Lemme 19.124.
Pour toute fonction $u\colon [a, b] \to \mathbb{R}^n$, nous avons

$$\| \int_a^b u(t)dt\| \leqslant \int_a^b \|u(t)\|dt \tag{19.459}$$

pourvu que le membre de gauche ait un sens.

Démonstration. Étant donné que $\int_a^b u(t)dt$ est un élément de $\mathbb{R}^n$, par la proposition 9.273, il existe un $\xi \in \mathbb{R}^n$ de norme 1 tel que

$$\| \int_a^b u(t)dt\| = \xi \cdot \int_a^b u(t)dt = \int_a^b u(t) \cdot \xi dt \leqslant \int_a^b \|u(t)\|\|\xi\| = \int_a^b \|u(t)\|dt. \tag{19.460}$$

□

Proposition 19.125 (Relations de Chasles)**.**
Soit f une fonction continue sur l'intervalle I. Si $a, b, c \in I$ nous avons

$$\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx. \tag{19.461}$$

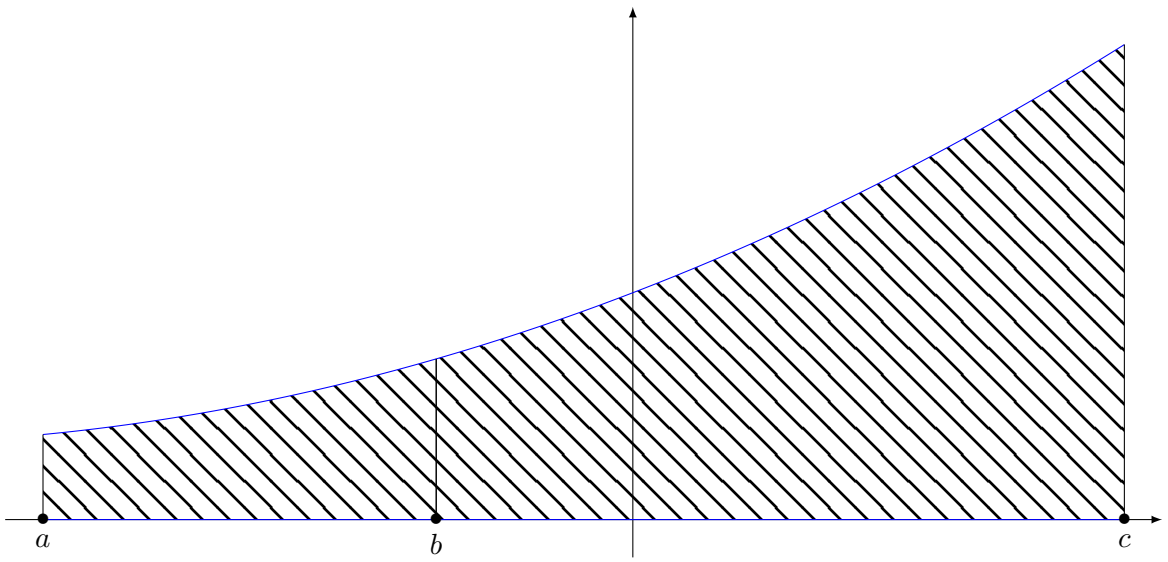


FIGURE 19.14 – Illustration pour les relations de Chasles.

Sur la figure 19.14, la surface de a à c est évidemment égale à la somme des surfaces de a à b et de b à c.

Corolaire 19.126.

$$\int_a^b f(x)dx = -\int_b^a f(x)dx. \tag{19.462}$$

Proposition 19.127 (Linéarité de l'intégrale)**.**
Si f et g sont deux fonctions continues sur $I \subset \mathbb{R}$, a, $b \in I$ et $\lambda \in \mathbb{R}$ nous avons

$$\int_a^b \big(f(x) + g(x)\big)dx = \int_a^b f(x)dx + \int_a^b g(x)dx, \tag{19.463}$$

et

$$\int_a^b \lambda f(x)dx = \lambda \int_a^b f(x)dx. \tag{19.464}$$

Proposition 19.128 (L'intégrale est monotone)**.**
Soient $a, b \in I$ avec $a < b$. Si $f \geqslant g$ sur $[a, b]$ alors

$$\int_a^b f(x)dx \geqslant \int_a^b g(x)dx. \tag{19.465}$$

Corolaire 19.129 (Positivité)**.**
Si $a < b$ et $f \geqslant 0$ sur $[a, b]$ alors

$$\int_a^b f(x)dx \geqslant 0. \tag{19.466}$$

Ce résultat n'est qu'une application de la proposition 19.128 car il consiste à prendre comme fonction g la fonction nulle.

19.24 Techniques d'intégration

Par le théorème 13.236, la calcul d'une intégrale consiste essentiellement à trouver une primitive de la fonction à intégrer. Il est donc indispensable de bien connaitre les dérivées des fonctions usuelles.

Voici un tableau des primitives à connaitre.

Fonction $f(x)$	Primitive $\int f(x)\,dx$	Ensemble de définition de f	Remarques
x^α	$\frac{x^{\alpha+1}}{\alpha+1}+C$	dépend de α	$\alpha \in \mathbb{R}\backslash\{-1\}$
$\frac{1}{x}$	$\ln\big(\lvert x\rvert\big)+C$	$x \neq 0$	
$\frac{1}{1+x^2}$	$\arctan(x)+C$	$\mathbb{R}$	
$\frac{1}{\sqrt{1-x^2}}$	$\arcsin(x)+C$	$]-1,1[$	
$\frac{-1}{\sqrt{1-x^2}}$	$\arccos(x)+C$	$]-1,1[$	
e^x	e^x+C	$\mathbb{R}$	
$\sin(x)$	$-\cos(x)+C$	$\mathbb{R}$	
$\cos(x)$	$\sin(x)+C$	$\mathbb{R}$	
$1+\tan^2(x)$	$\tan(x)+C$	in intervalle de la forme $]-\frac{\pi}{2},\frac{\pi}{2}[+k\pi$	

Notez que au signe près, les fonctions arcsin et arccos ont la même dérivée.

Si la fonction à intégrer est une combinaison linéaire de fonctions usuelles alors sa primitive peut être calculée en utilisant la proposition 19.127. Dans les sections suivantes on abordera deux autres cas où la fonction à intégrer peut s'écrire en termes de fonctions dont on connaît une primitive.

19.24.1 Intégration par parties

Proposition 19.130.
Si u et v sont deux fonctions dérivables de dérivées continues sur l'intervalle $[a,b]$ alors

$$\int_a^b u(x)v'(x)dx = \big[u(x)v(x)\big]_a^b - \int_a^b u'(x)v(x)dx. \tag{19.467}$$

Démonstration. Il s'agit d'utiliser à l'envers la formule de dérivation d'un produit :

$$uv' = (uv)' - u'v. \tag{19.468}$$

Les fonctions à gauche et à droite étant égales, elles ont même intégrale sur $[a,b]$ et par linéarité, voir proposition 19.127, on a :

$$\int_a^b u(x)v'(x)dx = \int_a^b (uv)'(x) - \int_a^b u'(x)v(x)dx. \tag{19.469}$$

La fonction uv est évidemment une primitive de $(uv)'$, de telle sorte que l'on puisse un peu simplifier cette expression :

$$\int_a^b u(x)v'(x)dx = \Big[u(x)v(x)\Big]_a^b - \int_a^b u'(x)v(x)dx, \tag{19.470}$$

ce qu'il fallait démontrer. □

Exemple 19.131.
Un cas typique d'utilisation de l'intégrale par parties est le suivant. Soit à calculer

$$\int_0^\pi x\cos(x)dx. \tag{19.471}$$

Nous devons écrire $x\cos(x)$ comme un produit $u(x)v'(x)$. Il y a (au moins) deux moyens de le faire :

$$\begin{cases} u = x & (19.472a)\\ v' = \cos(x). & (19.472b)\end{cases}$$

ou

$$\begin{cases} u = \cos(x) & (19.473a)\\ v' = x. & (19.473b)\end{cases}$$

Nous allons choisir le premier [21]. Nous avons donc

$$\begin{aligned} u &= x, \quad & v' &= \cos(x) \\ u' &= 1 & v &= \sin(x). \end{aligned} \tag{19.474}$$

En utilisant la formule d'intégration par parties,

$$\int_0^\pi x\cos(x)dx = \Big[x\sin(x)\Big]_0^\pi - \int_0^\pi 1\times\sin(x)dx = \pi\sin(\pi) - \Big[-\cos(x)\Big]_0^\pi = -2. \tag{19.475}$$

△

Le plus souvent, pour alléger les notations, il est plus pratique d'utiliser l'intégration par parties pour déterminer une primitive. Nous utilisons pour cela la formule (sans doute plus simple à retenir)

$$\int uv' = uv - \int u'v. \tag{19.476}$$

Exemple 19.132.
Nous reprenons l'exemple 19.131 en déterminant cette fois une primitive de $x\cos(x)$:

$$\int x\cos(x)dx = x\sin(x) - \int \sin(x)dx = x\sin(x) + \cos(x) + C, \qquad C\in\mathbb{R}. \tag{19.477}$$

Nous retrouvons le résultat numérique de l'exemple précédent en ajoutant les extrêmes d'intégration

$$\int_0^\pi x\cos(x)dx = \Big[x\sin(x) + \cos(x)\Big]_0^\pi = -2. \tag{19.478}$$

△

Remarque 19.133.
Lorsqu'on calcule des intégrales, il est bon de passer par la primitive (c'est-à-dire en suivant l'exemple 19.132 et non 19.131) parce qu'il est alors facile de vérifier le résultat en calculant la dérivée de la primitive trouvée.

Par exemple pour vérifier si (19.477) est correct, il suffit de dériver $x\sin(x)+\cos(x)$:

$$\big(x\sin(x)+\cos(x)\big)' = \sin(x) + x\cos(x) - \sin(x) = x\cos(x). \tag{19.479}$$

La fonction $x\sin(x)+\cos(x)$ est donc bien une primitive de $x\cos(x)$.

Exemple 19.134 (Primitive du logarithme)**.**
La primitive de la fonction logarithme définie en 14.58 nous offre un bon moment d'intégration par partie.

Trouver la primitive de la fonction $x\mapsto\ln(x)$. Pour calculer

$$\int \ln(x)dx \tag{19.480}$$

nous écrivons $\ln(x) = 1\times\ln(x)$ et nous posons $u'=1$ et $v=\ln(x)$, c'est-à-dire

$$\begin{aligned} u' &= 1 \quad & v &= \ln(x) \\ u &= x & v' &= \frac{1}{x}. \end{aligned} \tag{19.481}$$

La formule d'intégration par parties (19.467) donne donc

$$\int \ln(x) = x\ln(x) - \int x\times\frac{1}{x} = x\ln(x) - \int 1 = x\ln(x) - x + C, \qquad C\in\mathbb{R}. \tag{19.482}$$

Il est facile de vérifier par un petit calcul que

$$\big(x\ln(x) - x\big)' = \ln(x). \tag{19.483}$$

△

21. Mais nous conseillons vivement au lecteur d'essayer le deuxième pour se rendre compte qu'il ne fonctionne pas.

19.24.2 Changement de variables – pour trouver des primitives

De la même manière que l'utilisation « à l'envers » de la formule de dérivation du produit avait donné la méthode d'intégration par parties, nous allons voir que que l'utilisation « à l'envers » de la formule de dérivation d'une fonction composée donne lieu à la méthode d'intégration par changement de variables.

Proposition 19.135.
Soient I et J des intervalles de $\mathbb{R}$, $u\colon I \to J$ une fonction qui est dérivable de dérivée continue et $f\colon J \to \mathbb{R}$ une fonction admettant une primitive F. Alors la fonction

$$x \mapsto F\big(u(x)\big) \tag{19.484}$$

est une primitive de

$$f\big(u(x)\big)u'(x). \tag{19.485}$$

Démonstration. Cela est une utilisation immédiate de la formule de dérivée des fonctions composées. □

Exemple 19.136.
Soit à calculer

$$\int x\sqrt{1-x^2}dx. \tag{19.486}$$

La fonction $g(x) = x\sqrt{1-x^2}$ est le produit de x et de $\sqrt{1-x^2}$. On remarque que la dérivée de $1-x^2$ est $-2x$: nous avons alors, à un facteur -2 près, une expression de la forme (19.485) où la racine carrée joue le rôle de f, $f(t) = \sqrt{t}$, et $1-x^2$ le rôle de u. Une primitive de la fonction $f(t) = \sqrt{t}$ est $F(t) = 2t^{3/2}/3$.

Donc la fonction $\frac{2u(x)^{3/2}}{3} = \frac{2}{3}(1-x^2)^{3/2}$ est primitive de $-2x\sqrt{1-x^2} = -2g(x)$. Autrement dit,

$$\int -2x\sqrt{1-x^2}\,dx = \frac{2(1-x^2)^{3/2}}{3} + C, \tag{19.487}$$

et en divisant par -2 nous trouvons la primitive demandée :

$$\int x\sqrt{1-x^2}\,dx = -\frac{(1-x^2)^{3/2}}{3} + C. \tag{19.488}$$

△

L'exemple suivant donne une façon plus économe de retenir la méthode du changement de variables.

Exemple 19.137.
Soit à calculer

$$\int \cos(x)e^{\sin(x)}dx. \tag{19.489}$$

Vu qu'il y a beaucoup de fonctions trigonométriques dans la fonction à intégrer, nous allons poser $u(x) = \sin(x)$, et remplacer élément par élément tout ce qui contient du « x » dans l'intégrale demandée par la quantité correspondante en termes de u.

La difficulté est de savoir ce que nous allons faire du « dx » dans l'intégrale. Ce dx marque une variation (infinitésimale) de x. La formule des accroissements finis dit que si x augmente de la valeur dx, alors $u(x)$ augmente de $u'(x)dx$, c'est-à-dire que

$$du = \cos(x)dx. \tag{19.490}$$

Nous avons donc les substitutions suivantes à faire :

$$\sin(x) = u \tag{19.491a}$$
$$du = \cos(x)dx \tag{19.491b}$$
$$dx = \frac{du}{\cos(x)}. \tag{19.491c}$$

La chose « magique » est que le $\cos(x)$ se trouvant dans la fonction se simplifie avec le cosinus qui arrive lorsqu'on remplace dx par $\frac{du}{\cos(x)}$. Les substitutions faites nous restons avec

$$\int \cos(x)e^{\sin(x)}dx = \int e^u du = e^u + C, \qquad \text{où } u = \sin(x). \tag{19.492}$$

Attention : la réponse doit être impérativement donnée en termes de x et non de u. Nous écrivons donc

$$\int \cos(x)e^{\sin(x)} = e^{\sin(x)} + C. \tag{19.493}$$

△

19.24.3 Changement de variables – pour calculer des intégrales

Le théorème 13.236 fixe la relation entre la recherche des primitives de f et la calcul de l'intégrale de f sur l'intervalle d'extrêmes a et b. On a vu dans la section précédente comment utiliser le changement de variable pour trouver une primitive de f. Il faut maintenant comprendre comment appliquer ce qu'on a vu dans le calcul d'une intégrale.

En effet nous avons le choix entre

— trouver une primitive de f comme dans la section précédente et appliquer ensuite la formule du corolaire 13.236 ;

— écrire une intégrale pour la nouvelle variable $u = u(x)$ sur l'intervalle entre $u(a)$ et $u(b)$.

Nous allons voir ce deux méthodes dans des exemples.

Exemple 19.138.
Soit à calculer

$$\int_{1/3}^{1/2} x\sqrt{1-x^2}dx. \tag{19.494}$$

Les primitives $\int x\sqrt{1-x^2}dx$ ont été trouvé dans l'exemple 19.137. Une primitive est

$$F(x) = \int x\sqrt{1-x^2}dx = -\frac{(1-x^2)^{3/2}}{3}. \tag{19.495}$$

Nous pouvons maintenant calculer l'intégrale de $x\sqrt{1-x^2}$ sur l'intervalle $[1/3, 1/2]$ par la définition

$$\int_{1/3}^{1/2} x\sqrt{1-x^2}dx = F\left(\frac{1}{2}\right) - F\left(\frac{1}{3}\right) = -\frac{\sqrt{3}}{8} + \frac{16\sqrt{2}}{81}. \tag{19.496}$$

△

Remarque 19.139.
Pour que le calcul d'intégrale donne quelque chose de sensé il faut absolument que la primitive soit écrite en tant que fonction de x et non comme fonction de u. La méthode que nous allons voir dans l'exemple suivant réduit grandement la probabilité d'oublier ce détail, d'où le fait qu'elle soit de loin la plus utilisée.

Exemple 19.140.
Calculons à nouveau

$$\int_{1/3}^{1/2} x\sqrt{1-x^2}dx. \tag{19.497}$$

Cette fois nous allons toucher à l'intervalle d'intégration en même temps que faire le changement de variables. Nous savons déjà les substitutions

$$\begin{cases} u = 1-x^2 & (19.498a)\\ du = -2xdx & (19.498b)\\ dx = \dfrac{du}{-2x}. & (19.498c)\end{cases}$$

En ce qui concerne les extrêmes d'intégration, si $x = 1/3$ alors $u = 1 - \frac{1}{9} = \frac{8}{9}$ et si $x = \frac{1}{2}$ alors $u = \frac{3}{4}$. Nous avons donc encore les substitutions suivantes :

$$\begin{cases} x = 1/3 \to u = 8/9 & (19.499a)\\ x = 1/2 \to u = 3/4 & (19.499b)\end{cases}$$

Le calcul est alors

$$\int_{1/3}^{1/2} x\sqrt{1-x^2}dx = -\frac{1}{2}\int_{8/9}^{3/4}\sqrt{u}du = -\frac{1}{2}\left[\frac{u^{3/2}}{3/2}\right]_{8/9}^{3/4} = -\frac{\sqrt{3}}{8} + \frac{16\sqrt{2}}{81}. \tag{19.500}$$

Attention : la dernière égalité n'est pas immédiate ; elle demande quelques calculs et une bonne utilisation des règles de puissances. △

La deuxième méthode est plus utilisée et, avec un peu d'exercice, plus rapide à mettre en place que la première.

Jusqu'à présent nous avons utilisé des changements de variables dans lesquels nous exprimions u en termes de x. Comme le montre l'exemple suivant, il est parfois fructueux d'utiliser le changement de variable dans le sens inverse : avec x exprimé en termes d'un paramètre.

Exemple 19.141.
À calculer :

$$\int_{1/2}^{\sqrt{3}/2} \sqrt{1-x^2}dx. \tag{19.501}$$

Nous posons $x = \sin(\theta)$ parce que nous savons que $1 - \sin^2(\theta) = \cos^2(\theta)$; nous espérons que le changement de variables simplifie l'expression [22]. Les substitutions à faire dans l'intégrale sont :

$$\begin{cases} x = \sin(\theta) & (19.502a)\\ dx = \cos(\theta)d\theta, & (19.502b)\end{cases}$$

et en ce qui concerne les bornes, si $x = 1/2$ alors $\sin(\theta) = \frac{1}{2}$, c'est-à-dire $\theta = \frac{\pi}{6}$. Si $x = \sqrt{3}/2$ alors $\theta = \frac{\pi}{3}$. Donc

$$\int_{1/2}^{\sqrt{3}/2} \sqrt{1-x^2}dx = \int_{\pi/6}^{\pi/3} \sqrt{1-\sin^2(\theta)}\cos(\theta)dt. \tag{19.503}$$

Nous avons $1-\sin^2(\theta) = \cos^2(\theta)$ et vu que $\theta \in [\frac{\pi}{6}, \frac{\pi}{3}]$ nous avons toujours $\cos(\theta) > 0$, ce qui donne $\sqrt{\cos^2(\theta)} = \cos(\theta)$. Nous devons donc calculer

$$\int_{\pi/6}^{\pi/3} \cos^2(\theta)d\theta. \tag{19.504}$$

22. Lorsqu'on fait un changement de variables, il s'agit toujours d'*espérer* que l'expression se simplifie. Il n'y a pas moyen de savoir à priori si tel changement de variable va être utile. Il faut essayer.

Pour celle-là, il faut utiliser une formule de trigonométrie [23] :

$$\cos^2(\theta) = \frac{1+\cos(2\theta)}{2}. \tag{19.505}$$

Donc

$$\int_{\pi/6}^{\pi/3} \cos^2(\theta)d\theta = \int_{\pi/6}^{\pi/3} \frac{1+\cos(2\theta)}{2}d\theta = \left[\frac{\theta}{2}\right]_{\pi/6}^{\pi/3} + \int_{\pi/6}^{\pi/3} \frac{\cos(2\theta)}{2}d\theta, \tag{19.506}$$

Pour calculer proprement la dernière intégrale nous effectuons un autre changement de variable (facile) en posant $t = 2\theta$, $dt = 2d\theta$, $t(\pi/6) = \pi/3$ et $t(\pi/3) = 2\pi/3$, nous avons alors

$$\int_{\pi/6}^{\pi/3} \cos^2(\theta)d\theta = \left[\frac{\theta}{2}\right]_{\pi/6}^{\pi/3} + \int_{\pi/3}^{2\pi/3} \frac{\cos(t)}{4}dt = \left[\frac{\theta}{2}\right]_{\pi/6}^{\pi/3} = \frac{\pi}{6} - \frac{\pi}{12} = \frac{\pi}{12}, \tag{19.507}$$

parce que $\sin\left(\frac{2\pi}{3}\right) = \sin\left(\frac{\pi}{3}\right)$. Au final,

$$\int_{1/2}^{\sqrt{3}/2} \sqrt{1-x^2}dx = \frac{\pi}{12}. \tag{19.508}$$

△

19.24.4 Intégrations des fractions rationnelles réduites

Définition 19.142.
*Une **fraction rationnelle** est un quotient de deux polynômes à coefficients réels ou complexes.*

Par exemple

$$\frac{x^5 + 7x^4 - \frac{x^3}{2} + x}{x^2 - 1} \tag{19.509}$$

est une fraction rationnelle.

Il sera expliqué dans le cours d'algèbre que toute fraction rationnelle peut être écrite sous forme d'une somme d'éléments simples, c'est-à-dire de fractions rationnelles d'un des deux types suivants :

$$\frac{\alpha}{(x-a)^m}, \qquad \alpha, a \in \mathbb{R}, m \in \mathbb{N} \tag{19.510a}$$

$$\frac{\alpha x + \beta}{(x^2+ax+b)^m}; \qquad \alpha, \beta, a, b \in \mathbb{R}, m \in \mathbb{N}, a^2 - 4b < 0. \tag{19.510b}$$

Nous allons nous contenter de donner un exemple de chaque type.

(1) En ce qui concerne le cas (19.510a) avec $m = 1$, nous avons par exemple

$$\int \frac{1}{x-3}dx = \ln\left(|x-3|\right) + C. \tag{19.511}$$

Si vous voulez en être tout à fait sûr, effectuez d'abord le changement de variables $u = x - 3$ qui donne $dx = du$.

(2) En ce qui concerne le cas (19.510a) avec $m \neq 1$, nous avons par exemple

$$\int \frac{1}{(x-1)^4}dx = -\frac{1}{3(x-1)^3} + C. \tag{19.512}$$

Encore une fois, pour s'en convaincre, utiliser le changement de variables $u = x-1$, $dx = du$:

$$\int \frac{1}{(x-1)^4}dx = \int \frac{1}{u^4}du = \int u^{-4}du = -\frac{u^{-3}}{3} + C = -\frac{1}{3}\frac{1}{(x-1)^3} + C. \tag{19.513}$$

23. En fait, il y a moyen de terminer le calcul en intégrant deux fois par parties, mais c'est plus compliqué.

(3) En ce qui concerne le cas (19.510b) avec $\alpha \neq 0$, nous avons par exemple

$$\int \frac{x}{x^2+4}dx = \frac{1}{2}\ln(x^2+4) + C. \tag{19.514}$$

Pour ce faire, il faut faire le changement de variables $u = x^2+4$, $du = 2xdx$, $dx = \frac{du}{2x}$ qui donne

$$\int \frac{x}{x^2+4}dx = \frac{1}{2}\int \frac{du}{u} = \frac{1}{2}\ln(|u|) + C = \frac{1}{2}\ln(|x^2+4|) + C. \tag{19.515}$$

Dans ce cas nous pouvons oublier d'écrire la valeur absolue dans le logarithme parce que de toutes façons, x^2+4 est toujours positif.

(4) En ce qui concerne le cas (19.510b) avec $\alpha = 0$, nous avons par exemple

$$\int \frac{dx}{x^2+4} = \frac{1}{4}\int \frac{dx}{(\frac{x}{2})^2+1} = \frac{1}{2}\arctan(\frac{x}{2}) + C. \tag{19.516}$$

où nous avons utilisé la primitive $\int \frac{dx}{x^2+1}dx = \arctan(x)$ du tableau de la page 1475. Pour vous en convaincre vous pouvez faire la dernière étape avec le changement de variables $u = x/2$, $dx = 2du$.

19.24.5 Quelques formules à connaitre

À retenir 19.143

$$\int (\alpha f(x) + \beta g(x))\, dx = \alpha \int f(x)\, dx + \beta \int g(x)\, dx. \tag{19.517a}$$

$$\int f(x)g'(x)\, dx = f(x)g(x) - \int f'(x)g(x)\, dx. \tag{19.517b}$$

$$\int f'(u(x))u'(x)\, dx = \int f(t)\, dt, \qquad \text{avec } t = u(x). \tag{19.517c}$$

$$\int \frac{f'(x)}{f(x)}\, dx = \log|f(x)| + C, \qquad \text{c'est un cas particulier de la formule précédente.} \tag{19.517d}$$

19.24.6 Approximation de $\ln(2)$

Théorème 19.144.
Soit une fonction $f\colon [a,b] \to \mathbb{R}$ de classe C^{n+1}. Alors pour tout $N \leqslant n$ nous avons

$$f(b) = f(a) + \sum_{k=1}^{N} \frac{f^{(k)}(a)}{k!}(b-a)^k + \frac{1}{N!}\int_a^b (b-t)^N f^{(N+1)}(t)dt. \tag{19.518}$$

Démonstration. Notons que dans l'énoncé, n est fixé ; nous faisons une récurrence sur N. Ça ne change pas grand chose, mais il faut être conscient de ce qui est exactement dans l'hypothèse du théorème et ce qui est dans l'hypothèse de récurrence.

Bref, n est fixé, la fonction f est de classe C^{n+1} et nous vérifions d'abord la formule avec $N = 1$. À droite dans (19.518) nous avons

$$f(a) + f'(a)(b-a)\int_a^b (b-t)f''(t)dt. \tag{19.519}$$

Nous évaluons l'intégrale à part en faisant une intégration par parties [24]. Il s'agit de poser

$$u = b - t \tag{19.520a}$$

$$v' = f'', \tag{19.520b}$$

24. Proposition 19.130.

de déduire

$$u' = -1 \tag{19.521a}$$
$$v = f' \tag{19.521b}$$

et d'écrire

$$\int_a^b (b-t)f''(t)dt = \big[(b-t)f'(t)\big]_a^b - \int_a^b (-)f'(t)dt \tag{19.522a}$$
$$= -(b-a)f'(a) + \int_a^b f'(t)dt \tag{19.522b}$$
$$= -(b-a)f'(a) + f(b) - f(a). \tag{19.522c}$$

Dans le calcul nous avons utilisé le théorème fondamental du calcul intégral 13.236. En remettant ça dans (19.519) nous trouvons $f(b)$ comme il se doit.

En ce qui concerne la récurrence, nous devons calculer

$$f(a) + \sum_{k=1}^{N+1} \frac{f^{(k)(a)}}{k!}(b-a)^k + \frac{1}{(N+1)!}\int_a^b (b-t)^{N+1} f^{(N+2)}(t)dt. \tag{19.523}$$

Ici encore, il s'agit de faire une intégration par partie, et sortir de la somme le terme $k = N+1$. L'intégration par partie donne

$$\int_a^b (b-t)^{N+1} f^{(N+2)}(t)dt = -(b-a)^{N+1} f^{(N+1)}(a) + (N+1)\int_a^b (b-t)^N f^{(N+1)}(t)dt. \tag{19.524}$$

En remettant tout ensemble, il y a encore deux termes qui se simplifient, et des termes qui se remettent pour former la formule de récurrence. Bref, on obtient que (19.523) se réduit bien à $f(b)$. □

Cette formule avec reste intégral sert par exemple à prouver un encadrement pour $\ln(2)$, voir la proposition 19.145.

Proposition 19.145 (Approximation de $\ln(2)$[395])**.**
Pour tout n nous avons

$$\left|\ln(2) - \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k}\right| \leqslant \frac{1}{n+1}. \tag{19.525}$$

Démonstration. Nous écrivons la formule de Taylor avec reste intégral du théorème 19.144 pour la fonction $f = \ln$ et pour $a = 1$ et $b = x$. Cela donne :

$$\ln(x) = \ln(1) + \sum_{k=1}^{N} \frac{\ln^{(k)}(1)}{k!}(x-1)^k + \frac{1}{N!}(x-t)^N \ln^{(N+1)}(t)dt. \tag{19.526}$$

Sachant que la dérivée du logarithme[25] est $1/x$ et faisant une petite récurrence, pour $k \geqslant 1$ nous avons

$$\ln^{(k)}(x) = \frac{(-1)^{k+1}(k-1)!}{x^k}. \tag{19.527}$$

En remplaçant,

$$\ln(x) = \sum_{k=1}^{N} \frac{(-1)^{k+1}}{k}(x-1)^k + \int_1^x \frac{(-1)^N (x-t)^N}{t^{N+1}}dt. \tag{19.528}$$

C'est le moment de poser $x = 2$ et de faire les simplifications qui s'imposent,

$$\ln(2) = \sum_{k=1}^{N} \frac{(-1)^{k+1}}{k} + \int_1^2 \frac{(-1)^N (2-t)^N}{t^{N+1}}dt. \tag{19.529}$$

25. Voir la proposition 14.64.

Nous déplaçons la somme à gauche, et nous prenons la valeur absolue des deux côtés :

$$|\ln(2) - \sum_{k=1}^{N} \frac{(-1)^{k+1}}{k}| = |\int_1^2 \frac{(-1)^N (2-t)^N}{t^{N+1}} dt| \tag{19.530a}$$

$$\leqslant \int_1^2 \int_1^2 \frac{(2-t)^N}{t^{N+1}} dt \tag{19.530b}$$

$$\leqslant \int_1^2 (2-t)^N. \tag{19.530c}$$

Justifications :

— Pour 19.530b. Majoration en rentrant la valeur absolue dans l'intégrale, suppression de $(-1)^N$, et le fait que pour $t \in [1,2]$, $2-t \geqslant 0$.

— Pour 19.530c. Majoration en supprimant purement et simplement le dénominateur $t^N + 1 \geqslant 1$.

Ais-je vraiment besoin de vous dire que la dernière intégrale se calcule en posant le changement de variables [26] $u = 2 - t$? Le résultat est que

$$\int_1^2 (2-t)^N dt = \frac{1}{N+1}. \tag{19.531}$$

□

Exemple 19.146 ([1]).
La convergence de l'encadrement (19.145) n'est pas terrible. Pour avoir une erreur de $\frac{1}{10}$, il faut

$$\frac{1}{10} = \frac{1}{n+1}, \tag{19.532}$$

ce qui demande $n = 9$. Ça reste jouable, même pour les jeunes d'aujourd'hui. Écrivons 9 termes :

$$|\ln(2) - 1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \frac{1}{5} + \frac{1}{6} - \frac{1}{7} + \frac{1}{8} - \frac{1}{9}| \leqslant \frac{1}{10}. \tag{19.533}$$

En calculant [27],

$$|\ln(2) - \frac{1879}{2520}| \leqslant \frac{1}{10}. \tag{19.534}$$

Voici donc un bel encadrement

$$\frac{1879}{2520} - \frac{1}{10} \leqslant \ln(2) \leqslant \frac{1879}{2520} + \frac{1}{10}. \tag{19.535}$$

Pour avoir quelque chose avec des virgules, d'abord un peu de calcul mental donne

$$\frac{1879}{2520} \simeq 0.745634920634921. \tag{19.536}$$

Donc en majorant et minorant, disons, la troisième décimale [28], on n'est pas moins précis que le $\frac{1}{10}$. On a

$$0.744 - \frac{1}{10} \leqslant \ln(2) \leqslant 0.746 + \frac{1}{10}. \tag{19.537}$$

Bref, on retient l'approximation

$$0.644 \leqslant \ln(2) \leqslant 0.846. \tag{19.538}$$

Pour la quantité de travail, avouez que ce n'est pas terrible comme résultat. Eh oui ; le calcul numérique c'est tout un métier ; il existe des méthodes nettement plus efficaces que ce que nous venons de faire. △

26. Proposition 19.135.
27. Moi j'ai utilisé Sage, mais si tu es au tableau, débrouilles-toi.
28. Notez ici que nous utilisons le fait que la division euclidienne, elle, donne un encadrement pour les fractions. Pensez-y.

Proposition 19.147 (Formule de Taylor avec reste intégral[396, 257]).
Soient X et Y des espaces normés et un ouvert $\mathcal{O} \subset X$. Si $f \in C^m(\mathcal{O}, Y)$ et si $[p, x] \subset \mathcal{O}$ alors

$$\begin{aligned} f(x) = f(p) &+ \sum_{k=1}^{m-1} \frac{1}{k!}(d^k f)_p (x-p)^k \\ &+ \frac{1}{(m-1)!} \int_0^1 (1-t)^{m-1} (d^m f)_{p+t(x-p)} (x-p)^m \, dt \end{aligned} \tag{19.539}$$

où $\omega_p u^k$ signifie $\omega_p(u, \dots, u)$ lorsque $\omega \in \Omega^k$.

Notez que l'intégrale n'est pas une intégrale faisant intervenir les espaces X ou Y. Elle est une simple intégrale d'une fonction $\mathbb{R} \to \mathbb{R}$, comme définie par la mesure de Lebesgue de la définition 13.127.

Comme expliqué dans l'exemple 11.285, toute ces applications de différentielles se réduisent à des termes de la forme

$$f^{(k)}(p)(x-p)^k \tag{19.540}$$

dans le cas d'une fonction $\mathbb{R} \to \mathbb{R}$.

19.25 Constructions plus naïves de l'intégrale dans le cas réel

Les sections 13.2 et 13.8 ont donné une construction très complète de la mesure de Lebesgue, et nous avons définit la théorie de l'intégration sur un espace mesuré quelconque dans la définition 13.151.

Dans cette section nous allons donner différentes choses plus rapides qui servent souvent de définition dans les cours moins avancés.

19.25.1 Mesure de Lebesgue, version rapide

Nous construisons à présent la mesure de Lebesgue sur $\mathbb{R}^n$. Un **pavé** dans $\mathbb{R}^n$ est un ensemble de la forme

$$B = \prod_{i=1}^{n} [a_i, b_i]; \tag{19.541}$$

le volume d'un tel pavé est défini par $\mathrm{Vol}(B) = \prod_i (b_i - a_i)$. Soit maintenant $A \subset \mathbb{R}^n$. La **mesure externe** de A est le nombre

$$m^*(A) = \inf\{ \sum_{B \in \mathcal{F}} \mathrm{Vol}(B) \text{ où } \mathcal{F} \text{ est un ensemble dénombrable de pavés dont l'union recouvre } A.\} \tag{19.542}$$

Définition 19.148.
*Nous disons que A est **mesurable** au sens de Lebesgue si pour tout ensemble $S \subset \mathbb{R}^n$ nous avons l'égalité*

$$m^*(S) = m^*(A \cap S) + m^*(S \backslash A). \tag{19.543}$$

Dans ce cas nous disons que la mesure de Lebesgue de A est $m(A) = m^*(A)$.

Proposition 19.149.
Deux fonctions continues égales presque partout pour la mesure de Lebesgue[29] *sont égales.*

Démonstration. Soient f et g deux fonctions continues telles que $f(x) = g(x)$ pour presque tout $x \in D$. La fonction $h = f - g$ est alors presque partout nulle et nous devons prouver qu'elle est nulle sur tout D. La fonction h est continue ; si $h(a) \neq 0$ pour un certain $a \in D$ alors h est non nulle sur un ouvert autour de a par continuité et donc est non nulle sur un ensemble de mesure non nulle. □

29. Définition 19.148.

19.25.2 Pavés et subdivisions

Définition 19.150.
*Nous appelons **pavé** de $\mathbb{R}^p$ toute partie de $\mathbb{R}^p$ obtenue comme produit de p intervalles de $\mathbb{R}$. Plus explicitement, une partie R est un pavé de $\mathbb{R}^p$ s'il s'écrit sous la forme*

$$R = \left\{(x_1, \ldots, x_p) \in \mathbb{R}^p \,\middle|\, x_i \in \mathcal{I}_i, i = 1, \ldots, p\right\},$$

où $\mathcal{I}_i$ est un intervalle de $\mathbb{R}$ pour tout $i = 1, \ldots, p$.

On appelle pavé fermé de $\mathbb{R}^p$ le produit de p intervalles fermés

$$R = \prod_{i=1}^{p} [a_i, b_i].$$

On définit de même le pavé ouvert

$$S = \prod_{i=1}^{p}]a_i, b_i[.$$

Un pavé $R = \prod_{i=1}^{p} \mathcal{I}_i$ est dit borné si tous les intervalles $\mathcal{I}_i$ sont bornés dans $\mathbb{R}$. Les pavés non bornés sont des produits d'intervalles où un (ou plusieurs) des intervalles n'est pas borné. Par exemple,

$$N =]-\infty, 5] \times [0, 13].$$

L'espace $\mathbb{R}^p$, lui-même, est un pavé de $\mathbb{R}^p$.

Définition 19.151.
*Une partie A de $\mathbb{R}^p$ est dite **pavable** s'il existe une famille finie de pavés bornés R_j, $j = 1, \ldots, n$, et deux à deux disjoints tels que*

$$A = \bigcup_{j=1}^{n} R_j.$$

Un exemple d'ensemble pavable dans $\mathbb{R}^2$ est donné à la figure 19.15. Il existe beaucoup d'ensembles dans $\mathbb{R}^2$ qui ne sont pas pavables, par exemple les ellipses.

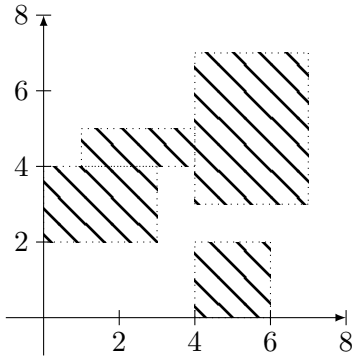


FIGURE 19.15 – Un ensemble pavable.

Le complémentaire d'un pavé est un ensemble pavable et, en particulier, tout complémentaire d'un pavé borné est une réunion de pavés non bornés. Toute union finie et toute intersection d'ensemble pavables est pavable.

Définition 19.152.
*Soit R un pavé borné de $\mathbb{R}^p$, pour fixer les idées on peut penser $R = \prod_{i=1}^{p} [a_i, b_i]$. On appelle **longueur** de l'i-ème arrête de R le nombre $b_i - a_i$. La **mesure p-dimensionnelle de** R, $m(R)$, est le produit des longueurs*

$$m(R) = \prod_{i=1}^{p} (b_i - a_i).$$

Exemple 19.153.
Dans $\mathbb{R}^3$, l'ensemble $R = [-1,1] \times [3,4] \times [0,2]$ est un pavé fermé de mesure

$$m(R) = (1+1) \cdot (4-3) \cdot (2-0) = 4.$$

Il s'agit du volume usuel du parallélépipède rectangle. △

Exemple 19.154.
L'ensemble $R =]-1,1[\times [3,4] \times [0,2]$ est un pavé de $\mathbb{R}^3$. Il n'est ni fermé ni ouvert, sa mesure est encore 4. △

Si R est un pavé non borné on peut encore définir sa mesure. La notion de mesure se généralise en deux étapes. D'abord on dit que la longueur d'une arête non bornée est ∞. Ensuite, on adopte la convention $0 \cdot \infty = 0$. Il faut remarquer que avec cette généralisation tout point et toute droite dans $\mathbb{R}^2$ ont mesure nulle.

Afin de définir les intégrales, nous allons intensivement faire appel à la notion de subdivision d'intervalles, voir définition 20.3 et la discussion qui suit.

Lorsqu'on considère un pavé borné $R = \prod_{i=1}^p \mathcal{I}_i$ de $\mathbb{R}^p$, on note $\mathcal{S}_i$ l'ensemble des subdivisions de l'intervalle $\mathcal{I}_i$. La notion de subdivision de généralise au cas des pavés.

Définition 19.155.
Soir R un pavé fermé borné de $\mathbb{R}^p$, pour fixer les idées on peut penser à $R = \prod_{i=1}^p [a_i, b_i]$. On appelle ***subdivision*** *finie de R les éléments de l'ensemble $\mathcal{S} = \prod_{i=1}^p \mathcal{S}_i$,*

$$\mathcal{S} = \left\{ (Y_1, \dots, Y_p) \,\middle|\, Y_i = (y_{i,j})_{j=1}^{n_i} \in \mathcal{S}_i,\ i = 1, \dots, p \right\}.$$

On peut définir de même l'ensemble des subdivisions d'un pavé non borné.

Souvent, une subdivision d'un pavé $R = \prod_{i=1}^p \mathcal{I}_i$ sera noté $\sigma = (y_{i,j})_{j=1}^{n_i}$. Dans cette notation, on sous-entend que pour chaque i fixé, les nombres $y_{i,j}$ (il y en a n_i) forment une subdivision de l'intervalle $\mathcal{I}_i$. Afin de vous familiariser avec ces notations, repérez bien tous les éléments de la figure 19.16.

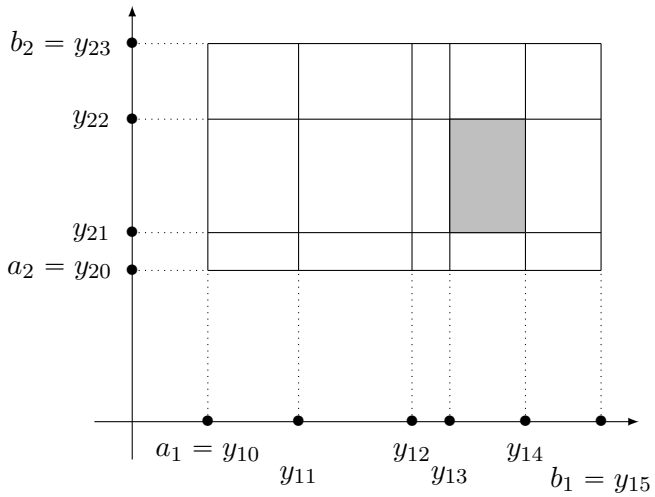


FIGURE 19.16 – Une cellule d'une subdivision d'un pavé de $\mathbb{R}^2$. La cellule grisée est $R_{(4,2)}$.

Définition 19.156.
Si σ est une subdivision d'un pavé R, un ***raffinement*** *de σ est une subdivision de R obtenue en fixant plus de points dans chaque intervalle.*

La subdivision σ de R détermine $n_1 n_2 \dots n_p$ pavés fermés de la forme

$$R_{(k_1,\dots,k_p)} = \{(x_1, \dots, x_p) \in \mathbb{R}^p \,|\, y_{i,k_{i-1}} \leqslant x_i \leqslant y_{i,k_i}\},$$

où k_i est dans $\{1, \ldots, n_i\}$ et i dans $\{1, \ldots, p\}$. On les appelles **cellules** de σ. On remarque que les cellules de σ sont toujours deux à deux disjointes (sauf au plus sur leurs bords).

Lemme 19.157.
Soit R un pavé borné de $\mathbb{R}^p$ et soit $\sigma = (y_{i,j})_{j=1}^{n_i}$ une subdivision de R. On a

$$m(R) = \sum_{(k_1,\ldots,k_p)\in K} m(R_{(k_1,\ldots,k_p)}),$$

où $K = \{1, \ldots, n_1\} \times \{1, \ldots, n_2\} \times \ldots \times \{1, \ldots, n_p\}$.

Le lemme 19.157 suggère de définir la mesure d'un ensemble borné pavable $P = \bigcup_{j=1}^{n} R_j$ comme la somme des mesures des pavés disjoints R_j, $j = 1, \ldots, n$.

Définition 19.158.
*Une application $f : \mathbb{R}^p \to \mathbb{R}$ est dite **application en escalier** sur $\mathbb{R}^m$ si*

— *f est une application bornée,*

— *il existe une subdivision σ de $\mathbb{R}^p$ telle que la restriction de f est une application constante sur toute cellule R_k de σ*

$$f_{|_{R_k}} = C_k, \qquad C_k \in \mathbb{R},$$

*Une telle subdivision σ est dite **associée** à f.*

Exemple 19.159.
La fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par

$$f(x, y) = \begin{cases} 1 & \text{si } (x, y) \in [0, 3] \times [-1, 2], \\ 2 & \text{sinon.} \end{cases} \tag{19.544}$$

est une application en escalier. Exercice : donner une subdivision de $\mathbb{R}^2$ associée à cette fonction. △

Exemple 19.160.
La fonction f de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par

$$f(x, y) = \begin{cases} \frac{1}{m^2+n^2}, & \text{si } (x, y) \in [m, m+1] \times [n, n+1], \quad m,\, n \in \mathbb{N}_0, \\ 0, & \text{sinon} \end{cases} \tag{19.545}$$

est une application en escalier. Observez que, dans ce cas, il n'existe pas une subdivision finie de $\mathbb{R}^2$ associée à f. △

Remarque 19.161.
Si la subdivision σ est associée à f alors tout raffinement de σ (c'est-à-dire, toute subdivision obtenue en fixant plus de points dans chaque intervalle) a la même propriété.

Si f et g sont deux applications en escalier sur R et σ_f et σ_g sont des subdivisions de R associées respectivement à f et g, alors on peut construire une troisième subdivision de R qui est associée à f et à g en même temps. Soient $\sigma_f = (Y_1, \ldots, Y_p)$ et $\sigma_g = (Z_1, \ldots, Z_p)$, où $Y_i = (y_{i,j})_{j=1}^{m_i}$ et $Z_i = (z_{i,j})_{j=1}^{n_i}$ sont des subdivisions de l'intervalle $[a_i, b_i]$, pour $i = 1, \ldots, p$. La subdivision de $[a_i, b_i]$ obtenue par l'union de Y_i et Z_i est encore une subdivision finie, qu'on appellera $\bar{Y}_i$. La subdivision $\bar{\sigma} = (\bar{Y}_1, \ldots, \bar{Y}_p)$ de R est un raffinement de σ_f et de σ_g, donc elle est associée à la fois à f et à g.

Cela nous permet de prouver que si f et g sont des applications en escalier, alors $f + g$, fg, $\min\{f, g\}$, $\max\{f, g\}$ et $|f|$ sont des applications en escalier.

19.25.3 Intégrale d'une fonction en escalier

Définition 19.162.
*Soit f une fonction de $\mathbb{R}^m$ dans $\mathbb{R}^n$. Le **support** de f est la fermeture de l'ensemble des points x tels que $f(x) \neq 0$.*

Définition 19.163.
*Une application en escalier f est dite **intégrable** si son support est compact.*

Soit f une application en escalier sur $\mathbb{R}^p$. Soit σ une subdivision de $\mathbb{R}^p$ associée à f et appelons R_k les cellules de σ, avec $k = (k_1, \ldots, k_p)$ dans $K = \{1, \ldots, n_1\} \times \{1, \ldots, n_2\} \times \ldots \times \{1, \ldots, n_p\}$. Alors

$$f_{|R_k} = C_k, \qquad C_k \in \mathbb{R}.$$

Définition 19.164.
*On définit l'**intégrale** de f sur $\mathbb{R}^p$ par*

$$\int_{\mathbb{R}^p} f\, dV = \sum_{k \in K} C_k m(R_k).$$

L'intégrale ainsi définie est un nombre réel. La proposition suivante nous dit que l'intégrale est « bien définie », au sens que sa valeur ne dépend pas de la subdivision associée à f qu'on utilise dans le calcul.

Proposition 19.165.
Soit f une application en escalier intégrable sur $\mathbb{R}^p$. Soient σ_1 et σ_2 deux subdivisions de $\mathbb{R}^p$ associées à f. L'intégrale de f ne dépend pas de la subdivision choisie.

On ne donne pas une preuve complète de cette proposition. En fait elle est une conséquence de la formule de réduction introduite dans la suite de ce chapitre.

19.25.4 Intégrales partielles

Soit f de $\mathbb{R}^p$ dans $\mathbb{R}$ une fonction continue, nulle hors du pavé borné R. Posons $R = \prod_{i=1}^p [a_i, b_i]$, pour fixer les idées. Pour chaque i dans $\{1, \ldots, p\}$ fixé, on peut associer à f la fonction F_i de $p-1$ variables définie par

$$F_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p) = \int_{a_i}^{b_i} f(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_p)\, dy.$$

La fonction F_i est l'intégrale partielle de f par rapport à la i-ème variable. En particulier, si $f(x_1, \ldots, x_p) = g(x_i)h(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)$ on obtient

$$F_i = \int_{a_i}^{b_i} g(y)h(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p)\, dy = h \cdot \int_{a_i}^{b_i} g\, dy.$$

La fonction d'une seule variable qu'on obtient à partir de f en fixant $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p$ et qui associe à x_i la valeur $f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_p)$, est appelée x_i-ème section de f en $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_p$.

Exemple 19.166.
Soit f la fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ définie par

$$f(x, y) = \begin{cases} x + 3y & \text{si } (x, y) \in [9, 10] \times]\pi, 5] \\ 0 & \text{sinon.} \end{cases} \tag{19.546}$$

Les intégrales partielles de f sont

$$F_1(y) = \int_9^{10} x + 3y\, dx = \left[\frac{x^2}{2} + 3xy\right]_{x=9}^{x=10} = \frac{19}{2} + 3y,$$

$$F_2(x) = \int_\pi^5 x + 3y\,dy = \left[xy + \frac{3y^2}{2}\right]_{y=\pi}^{y=5} = x(5-\pi) + \frac{3}{2}(25-\pi^2).$$

△

19.25.5 Réduction d'une intégrale multiple

Soit $R = [a,b] \times [c,d]$ un pavé fermé et borné de $\mathbb{R}^2$ et soit f une application en escalier intégrable sur $\mathbb{R}^2$ telle que le support de f soit contenu dans R. On considère la subdivision σ de R définie par les subdivisions

$$a = x_0 \leqslant x_1 \leqslant \ldots \leqslant x_m = b,$$

$$c = y_0 \leqslant y_1 \leqslant \ldots \leqslant y_n = d.$$

Les cellules de σ sont

$$R_{i,j} = [x_i, x_{i+1}] \times [y_j, y_{j+1}], \qquad i = 0, \ldots, m-1, \quad j = 0, \ldots, n-1.$$

La mesure de R est la somme des mesures des $R_{i,j}$

$$\begin{aligned} m(R) &= \sum_{(i,j)\in\{0,\ldots,m-1\}\times\{0,\ldots,n-1\}} m(R_{i,j}) = \\ &= \sum_{j=0}^{n-1}\sum_{i=0}^{m-1} (x_{i+1}-x_i)\cdot(y_{i+1}-y_i) = \\ &= \sum_{i=0}^{m-1} (x_{i+1}-x_i)\cdot\sum_{j=0}^{n-1}(y_{i+1}-y_i) = \\ &= (b-a)\cdot(d-c). \end{aligned} \tag{19.547}$$

Si f est constante sur chaque cellule de σ on peut écrire f de la forme suivante

$$f(x,y) = \sum_{j=0}^{n-1}\sum_{i=0}^{m-1} C_{i,j}\,\chi_{R_{i,j}}$$

où les $C_{i,j}$ sont des constantes réelles et $\chi_{R_{i,j}}$ est la **fonction caractéristique** de $R_{i,j}$

$$\chi_{R_{i,j}}(x,y) = \begin{cases} 1, & \text{si } (x,y) \in R_{i,j}, \\ 0, & \text{sinon.} \end{cases} \tag{19.548}$$

Comme (x,y) est dans $R_{i,j}$ si et seulement si $x \in [x_i, x_{i+1}]$ et $y \in [y_j, y_{j+1}]$, on vérifie que la fonction $\chi_{R_{i,j}}$ est égal au produit des fonctions caractéristiques des intervalles $[x_i, x_{i+1}]$ et $[y_j, y_{j+1}]$

$$\chi_{R_{i,j}}(x,y) = \chi_{[x_i,x_{i+1}]}(x)\cdot\chi_{[y_j,y_{j+1}]}(y).$$

On peut donc écrire la fonction f de la façon suivante

$$f(x,y) = \sum_{j=0}^{n-1}\sum_{i=0}^{m-1} C_{i,j}\,\chi_{[x_i,x_{i+1}]}(x)\cdot\chi_{[y_j,y_{j+1}]}(y).$$

Comme on suppose que le support de f est une partie de R, l'intégrale de f sur $\mathbb{R}^2$ est

$$\int_{\mathbb{R}^2} f\,dV = \sum_{j=0}^{n-1}\sum_{i=0}^{m-1} C_{i,j}\,m(R_{i,j}) = \sum_{j=0}^{n-1}\sum_{i=0}^{m-1} C_{i,j}\,(x_{i+1}-x_i)\cdot(y_{j+1}-y_j). \tag{19.549}$$

Cette intégrale peut être réduite à la composition de deux intégrales partielles. Il suffit de remarquer que la valeur de l'intégrale de la fonction caractéristique d'un intervalle est la longueur de l'intervalle,

$$\begin{aligned} C_{i,j}(x_{i+1}-x_i)\cdot(y_{j+1}-y_j) &= \\ &= C_{i,j}\left(\int_{x_i}^{x_{i+1}}\chi_{[x_i,x_{i+1}]}(x)\,dx\right)\cdot\left(\int_{y_j}^{y_{j+1}}\chi_{[y_j,y_{j+1}]}(y)\,dy\right)= \\ &= C_{i,j}\left(\int_a^b\chi_{[x_i,x_{i+1}]}(x)\,dx\right)\cdot\left(\int_c^d\chi_{[y_j,y_{j+1}]}(y)\,dy\right), \end{aligned} \tag{19.550}$$

et utiliser les propriétés de linéarité de l'intégrale

$$\begin{aligned} \int_{\mathbb{R}^2} f\,dV &= \sum_{j=0}^{n-1}\sum_{i=0}^{m-1} C_{i,j}\left(\int_a^b\chi_{[x_i,x_{i+1}]}(x)\,dx\right)\cdot\left(\int_c^d\chi_{[y_j,y_{j+1}]}(y)\,dy\right)= \\ &= \int_c^d\int_a^b\sum_{j=0}^{n-1}\sum_{i=0}^{m-1} C_{i,j}\,\chi_{[x_i,x_{i+1}]}(x)\cdot\chi_{[y_j,y_{j+1}]}(y)\,dxdy= \\ &= \int_c^d\int_a^b f\,dxdy. \end{aligned} \tag{19.551}$$

De même on obtient

$$\begin{aligned} \int_{\mathbb{R}^2} f\,dV &= \int_a^b\int_c^d\sum_{j=0}^{n-1}\sum_{i=0}^{m-1} C_{i,j}\,\chi_{[x_i,x_{i+1}]}(x)\cdot\chi_{[y_j,y_{j+1}]}(y)\,dxdy= \\ &= \int_a^b\int_c^d f\,dxdy. \end{aligned} \tag{19.552}$$

En général, on prouve la proposition suivante

Proposition 19.167.
Soit f une application en escalier intégrable sur $\mathbb{R}^p$ et soit R un pavé borné dans $\mathbb{R}^p$ qui contient le support de f. Comme d'habitude, pour fixer les idées nous écrivons $= \prod_{i=1}^p[a_i,b_i]$. Alors

$$\begin{aligned} \int_{\mathbb{R}^p} f(x_1,\dots,x_p)\,dV &= \int_{a_p}^{b_p}\int_{a_{p-1}}^{b_{p-1}}\cdots\int_{a_1}^{b_1} f(x_1,\dots,x_p)\,dx_1\cdots dx_p= \\ &= \int_{a_{s_p}}^{b_{s_p}}\int_{a_{s_{p-1}}}^{b_{s_{p-1}}}\cdots\int_{a_{s_1}}^{b_{s_1}} f(x_1,\dots,x_p)\,dx_1\cdots dx_p, \end{aligned} \tag{19.553}$$

pour toute permutation $(s_1,\dots,s_p)$ de l'ensemble $\{1,\dots p\}$.

19.25.6 Propriétés de l'intégrale

Soient f et g deux fonctions en escalier intégrables de $\mathbb{R}^p$ dans $\mathbb{R}$, et soient a et b dans $\mathbb{R}$.

Linéarité de l'intégrale :

— Additivité : $f+g$ est intégrable et

$$\int_{\mathbb{R}^p}(f+g)\,dV = \int_{\mathbb{R}^p} f\,dV + \int_{\mathbb{R}^p} g\,dV,$$

— Homogénéité : λf est intégrable pour tout réel λ

$$\int_{\mathbb{R}^p}\lambda f\,dV = \lambda\int_{\mathbb{R}^p} f\,dV,$$

Monotonie Si $f \leqslant g$ alors

$$\int_{\mathbb{R}^p} f\,dV \leqslant \int_{\mathbb{R}^p} g\,dV,$$

Inégalité fondamentale

$$|\int_{\mathbb{R}^p} f\,dV| \leqslant \int_{\mathbb{R}^p} |f|\,dV.$$

Cette dernière inégalité s'obtient de la façon suivante :

$$|\int_{\mathbb{R}^p} f\,dV| = |\sum_{k\in K} C_k m(R_k)| \leqslant \sum_{k\in K} |C_k| m(R_k) = \int_{\mathbb{R}^p} |f|\,dV.$$

Inégalité de Čebičeff Si f est une application en escalier alors pour tout $a > 0$ dans $\mathbb{R}$ l'ensemble $\{x \in \mathbb{R}^p \,:\, |f(x)| \geqslant a\}$ est pavable et borné, et l'inégalité suivante est satisfaite

$$m\left(\{x \in \mathbb{R}^p \,:\, |f(x)| \geqslant a\}\right) \leqslant \frac{1}{a}\int_{\mathbb{R}^p} |f|\,dV.$$

19.25.7 Intégrales multiples, cas général

Nous voulons généraliser la définition d'intégrale multiple au cas des domaines non pavables et de fonctions qui ne sont pas en escalier. Il y a plusieurs méthodes de le faire et ici on ne considère qu'une seule, introduite par Riemann.

Définition 19.168.
Soit $f : \mathbb{R}^p \to \mathbb{R}$ une fonction.

— *Pour toute application en escalier intégrable f_* telle que $f_* \leqslant f$, l'intégrale de f_* est dit une* ***somme inférieure*** *de f.*

— *Pour toute application en escalier intégrable f^* telle que $f_* \geqslant f$, l'intégrale de f^* est dit une* ***somme supérieure*** *de f.*

Soient $\sum_* f$ et $\sum^* f$ les ensembles des sommes inférieures et supérieures de f. Grâce à la propriété de monotonie de l'intégrale on sait que si a est dans $\sum_* f$ et b est dans $\sum^* f$ alors $a \leqslant b$.

Définition 19.169.
La fonction f est intégrable (au sens de Riemann) si $\sum_* f$ et $\sum^* f$ ne sont pas vides et

$$\inf \Sigma^* f = I = \sup \Sigma_* f.$$

Dans ce cas, la valeur I est appelée intégrale de f sur $\mathbb{R}^p$.

Remarque 19.170.
Toute fonction intégrable est bornée et à support compact. En effet, si le support de la fonction n'est pas compact alors soit $\sum_* f$ soit $\sum^* f$ doit être vide !

L'intégrale qu'on vient de définir possède toutes les propriétés de l'intégrale pour les fonctions en escalier. Le produit de deux fonctions intégrables est intégrable.

Il y a des cas où l'intégrabilité d'une fonction n'est pas évidente. Cependant, dans la plupart des exercices et des exemples de ce cours, nous nous aidons avec le critère suivant

Proposition 19.171.
Toute fonction continue à support compact est intégrable.

Cette proposition n'est à priori pas étonnante, vu qu'une fonction continue sur un support compact est bornée (théorème de Weierstrass 7.105).

19.25.8 Réduction d'une intégrale multiple

On n'utilise jamais la définition pour calculer la valeur d'une intégrale multiple. La méthode plus efficace, en pratique, est de réduire l'intégrale à la composition de plusieurs intégrales d'une variable.

Théorème 19.172 (de Fubini)**.**
Soit f une fonction intégrable de $\mathbb{R}^2$ dans $\mathbb{R}$. Si pour tout x dans $\mathbb{R}$ la section $f(x, \cdot)$ est intégrable par rapport à y, alors

$$\int_{\mathbb{R}^2} f(x,y)\,dV = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)\,dx \right) dy.$$

De même, si pour tout y dans $\mathbb{R}$ la section $f(\cdot, y)$ est intégrable par rapport à x, alors

$$\int_{\mathbb{R}^2} f(x,y)\,dV = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)\,dy \right) dx.$$

En général, on ne peut pas dire que les sections d'une fonction intégrable sont intégrables, donc il faut vraiment se souvenir des hypothèses du théorème 19.172. En dimension plus haute, on a le même résultat

Théorème 19.173.
Soit f une fonction intégrable de $\mathbb{R}^p$ dans $\mathbb{R}$. Si pour tout $(p-1)$-uple $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_p)$ dans $\mathbb{R}^{p-1}$ la section $f(x_1, \dots, x_{i-1}, \cdot, x_{i+1}, \dots, x_p)$ est intégrable par rapport à x_i, alors

$$\int_{\mathbb{R}^p} f\,dV = \int_{\mathbb{R}} \left(\int_{\mathbb{R}^{p-1}} f\,dV \right) dx_i.$$

Si f est une fonction positive et intégrable de $\mathbb{R}^2$ dans $\mathbb{R}$ on peut interpréter l'intégrale de f comme le volume du solide au-dessous du graphe de f. Avec cette interprétation, l'intégrale partielle par rapport à x pour $y = y_0$ fixé est l'aire de la tranche qu'on obtient en coupant le solide par le plan $y = y_0$.

Exemple 19.174.
Le premier exemple à faire est celui d'une fonction en escalier intégrable et positive. Soit $f\colon \mathbb{R}^2 \to \mathbb{R}$ la fonction

$$f(x,y) = \begin{cases} 1 & \text{si } (x,y) \in R_1 =]-1,3] \times [4,5] \\ 3 & \text{si } (x,y) \in R_2 =]13,15[\times [0,2[\\ 0 & \text{dans les autres cas.} \end{cases} \tag{19.554}$$

L'intégrale de f sur $\mathbb{R}^2$ est $1 \cdot m(R_1) + 3 \cdot m(R_2) = 16$. On voit tout de suite qu'il s'agit de la somme du volume des deux parallélépipèdes de hauteurs respectives 1 et 3 et bases R_1 et R_2. △

Exemple 19.175.
On veut calculer le volume du solide S, borné par le paraboloïde elliptique $x^2 + 2y^2 + z = 16$ et le plans $x = 2$, $x = 0$, $y = 2$ $y = 0$, $z = 0$. On observe que la portion de paraboloïde elliptique qui nous intéresse est le graphe de la fonction $f(x,y) = 16 - x^2 - 2y^2$ pour (x,y) dans $R = [0,2] \times [0,2]$. La fonction f est continue ainsi que ses sections, donc on peut appliquer le théorème 19.172 et décomposer l'intégrale double en deux intégrales simples :

$$\begin{aligned} \int_R 16 - x^2 - 2y^2\,dV &= \int_0^2 \int_0^2 f(x,y)\,dxdy = \\ &= \int_0^2 \left[(16 - 2y^2)x - \frac{x^3}{3} \right]_{x=0}^{x=2} dy = \\ &= \left[\left(32 - \frac{8}{3} \right) y - \frac{4y^3}{3} \right]_{x=0}^{x=2} = 64 - \frac{16+32}{3} = 48. \end{aligned} \tag{19.555}$$

Vérifiez, comme exercice, qu'on obtient le même résultat en intégrant d'abord par rapport à y et puis par rapport à x. △

Exemple 19.176.
Dans les hypothèses du théorème 19.172 l'ordre des intégrations partielles ne change pas la valeur de l'intégrale. En fait, si les calculs sont faits par des êtres humains l'ordre d'intégration peut faire une certaine différence comme dans cet exemple. On veut évaluer la valeur de l'intégrale

$$\int_{\mathbb{R}^2} f(x,y)\,dV$$

où

$$f(x,y) = \begin{cases} y\sin(x,y) & \text{si } (x,y) \in [1,2,] \times [0,\pi] \\ 0 & \text{sinon.} \end{cases} \tag{19.556}$$

Les deux sections de $f(x,y) = y\sin(xy)$ sont continues. Si on intègre d'abord par rapport à y on obtient

$$-\int_1^2 \frac{\pi\cos(\pi x)}{x}dx + \int_1^2 \frac{\sin(\pi x)}{x^2}dx,$$

qui n'est pas du tout immédiat, alors que, si on intègre d'abord par rapport à x on obtient

$$\int_0^\pi \cos y - \cos(2y)\,dy.$$

△

19.25.9 Intégrales sur des parties de $\mathbb{R}^2$

On veut évaluer l'intégrale de la fonction $f(x,y) = \sqrt{1-x^2}$ sur son domaine, la boule unité $B((0,0),1)$. La théorie introduite jusqu'ici n'est pas suffisante pour résoudre ce problème, parce que $B((0,0),1)$ n'est pas pavable. Les parties bornées de $\mathbb{R}^p$ sur lesquelles on peut intégrer des fonctions sont dites mesurables (au sens de Riemann) parce que, comme on verra dans la suite, la mesure d'une partie de $\mathbb{R}^p$ est l'intégrale (s'il existe) de sa fonction caractéristique.

On peut dire qu'une partie de $\mathbb{R}^p$ est mesurable si son bord est «assez régulier». Dans $\mathbb{R}^2$ il est suffisant que le bord de A soit une réunion finie de courbes paramétrées continues. En particulier, on est très souvent dans un des deux cas suivants

Régions du premier type A est borné et contenu entre les graphes de deux fonctions continues de x

$$A = \{(x,y) \in \mathbb{R}^2 \,:\, a \leqslant x \leqslant b,\ g_1(x) \leqslant y \leqslant g_2(x)\},$$

avec g_1 et g_2 continues.

Régions du deuxième type A est borné et contenu entre les graphes de deux fonctions continues de y

$$A = \{(x,y) \in \mathbb{R}^2 \,:\, c \leqslant y \leqslant d,\ h_1(y) \leqslant x \leqslant h_2(y)\},$$

avec h_1 et h_2 continues.

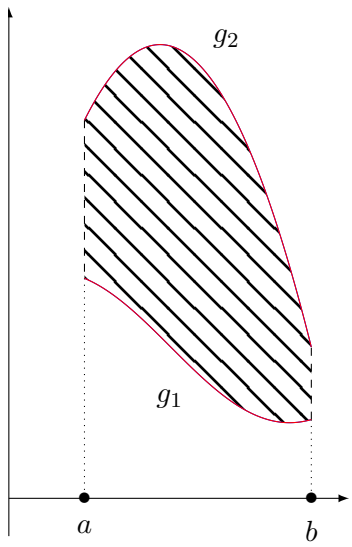
Exemple 19.177.
Il y a des régions qui sont des deux types au même temps, comme les boules centrées à l'origine, le triangle de sommets $(0,0)$, $(0,a)$ et $(b,0)$, ou la région C délimité par les courbes $y = 2x$ et $y = x^2$. Cette dernière admet les représentations suivantes

$$C = \{(x,y) \in \mathbb{R}^2 \,:\, 0 \leqslant x \leqslant 1,\ x^2 \leqslant y \leqslant 2x\},$$

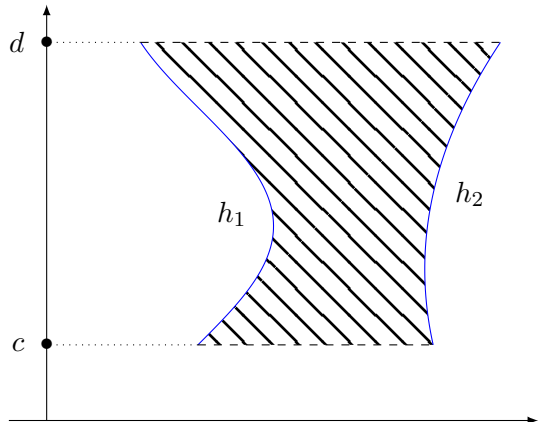
et

$$C = \{(x,y) \in \mathbb{R}^2 \,:\, 0 \leqslant y \leqslant 1,\ y/2 \leqslant x \leqslant \sqrt{y}\}.$$

△



(a) Une region du premier type



(b) Une region du deuxième type

FIGURE 19.17 – Régions du premier et du deuxième type

Définition 19.178.
Soit f une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$ dont le support A est une région du premier ou du deuxième type. On définit la fonction $\bar{f}$ comme

$$\bar{f}(x,y) = \begin{cases} f(x,y), & si\ (x,y) \in A, \\ 0, & sinon. \end{cases} \tag{19.557}$$

*La fonction f est dite **intégrable** si $\bar{f}$ est intégrable, et la valeur de son intégrale est*

$$\int_A f\,dV = \int_{\mathbb{R}^2} \bar{f}\,dV.$$

Une fonction continue définie sur une région du premier ou du deuxième type est toujours intégrable.

Pour fixer les idées on suppose ici que A est du premier type et contenue dans le pavé borné

$R = [a,b] \times [c,d]$. En suivant la définition on obtient

$$\begin{aligned}\int_A f\,dV &= \int_{\mathbb{R}^2} \bar{f}\,dV = \\ &= \int_a^b \int_c^d \bar{f}\,dydx = \\ &= \int_a^b \left(\int_c^{g_1(x)} \bar{f}\,dy + \int_{g_1(x)}^{g_2(x)} \bar{f}\,dy + \int_{g_2(x)}^d \bar{f}\,dy \right) dx = \\ &= \int_a^b \int_{g_1(x)}^{g_2(x)} f\,dydx.\end{aligned} \tag{19.558}$$

De même, si A est du deuxième type on obtient

$$\int_A f\,dV = \int_c^d \int_{h_1(y)}^{h_2(y)} f\,dxdy. \tag{19.559}$$

Exemple 19.179.
On peut maintenant résoudre notre problème de départ, évaluer l'intégrale de la fonction $f(x,y) = \sqrt{1-x^2}$ sur $B((0,0),1)$. Nous choisissons de décrire la boule unité de $\mathbb{R}^2$ comme une région du premier type : $B((0,0),1) = \{(x,y) \,:\, x \in [-1,1],\, -\sqrt{1-x^2} \leqslant y \leqslant \sqrt{1-x^2}\}$.

$$I = \int_B \sqrt{1-x^2}\,dV = \int_{-1}^1 \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1-x^2}dydx \tag{19.560}$$

La première intégrale à effectuer, par rapport à y, est l'intégrale d'une fonction constante. Ne pas oublier que l'on intègre $\sqrt{1-x^2}$ par rapport à y ; c'est bien une constante et l'intégrale consiste seulement à multiplier par y :

$$I = \int_{-1}^1 \left[y\sqrt{1-x^2} \right]_{y=-\sqrt{1-x^2}}^{y=\sqrt{1-x^2}} dx = 2\int_{-1}^1 (1-x^2)dx. \tag{19.561}$$

Cela est à nouveau une intégrale simple à effectuer. Le résultat est

$$2\int_{-1}^1 (1-x^2)dx = 2\left[x - \frac{x^3}{3} \right]_{x=-1}^{x=1} = \frac{8}{3}. \tag{19.562}$$

△

Remarque 19.180.
Toutes les techniques d'intégration à une variable restent valables. Par exemple, lorsqu'une des intégrales est l'intégrale d'une fonction impaire sur un intervalle symétrique par rapport à zéro, l'intégrale vaut zéro.

19.181.
Par le lemme 13.157 nous savons que la mesure d'une région bornée de $\mathbb{R}^2$ est l'intégrale de sa fonction caractéristique, si elle existe.

La mesure d'une région bornée de $\mathbb{R}^2$ est dite son **aire**, et celle d'une région bornée de $\mathbb{R}^3$ est son **volume**. Voir aussi la remarque 20.11.

Exemple 19.182.
On veut calculer l'aire de la région de la figure 19.18 définie par

$$A = \{(x,y) \in \mathbb{R}^2 \,|\, 0 \leqslant x \leqslant 1, x^3 - 1 \leqslant y \leqslant x\}.$$

On considère l'intégrale

$$\int_{\mathbb{R}^2} \chi_A\,dV = \int_0^1 \int_{x^3+1}^x 1\,dy\,dx = \int_0^1 -x^3 + x + 1\,dx = -\frac{1}{4} + \frac{1}{2} + 1 = \frac{5}{4}.$$

△

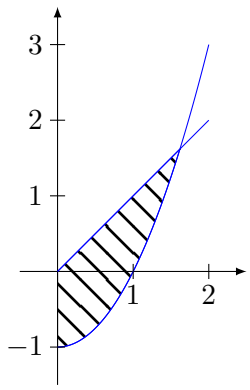


FIGURE 19.18 – La région A de l'exemple 19.182

Exemple 19.183.
Parfois la région sur laquelle on veut intégrer peut être décrite indifféremment de deux façons, mais la fonction à intégrer nous force a choisir un ordre particulier. Vérifiez que la fonction $f(x, y) = \sin(y^2)$ sur la région triangulaire de sommets $(0, 0)$, $(0, 2)$, $(2, 2)$ doit être intégrée d'abord par rapport à x. △

Si une région bornée n'est pas de premier ou de deuxième type on peut normalement la découper en morceaux plus faciles à décrire. On utilise alors la propriété suivante.

Lemme 19.184.
Soit A un sous-ensemble borné de $\mathbb{R}^2$ et soient B_1 et B_2 deux parties de A telles que $B_1 \cap B_2 = \varnothing$ et $B_1 \cup B_2 = A$. Alors, pour toute fonction f intégrable sur A (et en particulier pour sa fonction caractéristique) on a

$$\int_A f\,dV = \int_{B_1} f\,dV + \int_{B_2} f\,dV.$$

Exemple 19.185.
La région D que nous voyons sur la figure 19.19 est bornée par la parabole $y^2 = 2x + 6$ et la droite $y = x - 1$. La région D est une région du deuxième type. Nous pouvons aussi la décrire comme l'union de deux régions du premier type D_1 et D_2,

$$D_1 = \{(x, y) \,:\, -3 \leqslant x \leqslant -1,\ -\sqrt{2x+6} \leqslant y \leqslant \sqrt{2x+6}\},$$

et

$$D_2 = \{(x, y) \,:\, -3 \leqslant x \leqslant -1,\ x - 1 \leqslant y \leqslant \sqrt{2x+6}\}.$$

△

19.25.10 Intégrales sur des parties de $\mathbb{R}^3$

Dans ces notes nous n'avons pas l'ambition de traiter d'une façon rigoureuse l'étude des ensembles mesurables de $\mathbb{R}^3$. Comme dans la section précédente on se limitera à considérer des cas particuliers.

Définition 19.186.
Soit E une région de $\mathbb{R}^3$. On dit que E est une ***région solide de premier type*** *si E est contenue entre les graphes de deux fonctions continues de x et y.*

$$E = \{(x, y, z) \in \mathbb{R}^3 \,|\, (x, y) \in A \subset \mathbb{R}^2, u_1(x, y) \leqslant z \leqslant u_2(x, y)\}.$$

Le sous-ensemble de A de $\mathbb{R}^2$ qui apparaît dans la définition 19.186 est la projection (ou l'ombre) de E sur le plan x-y.

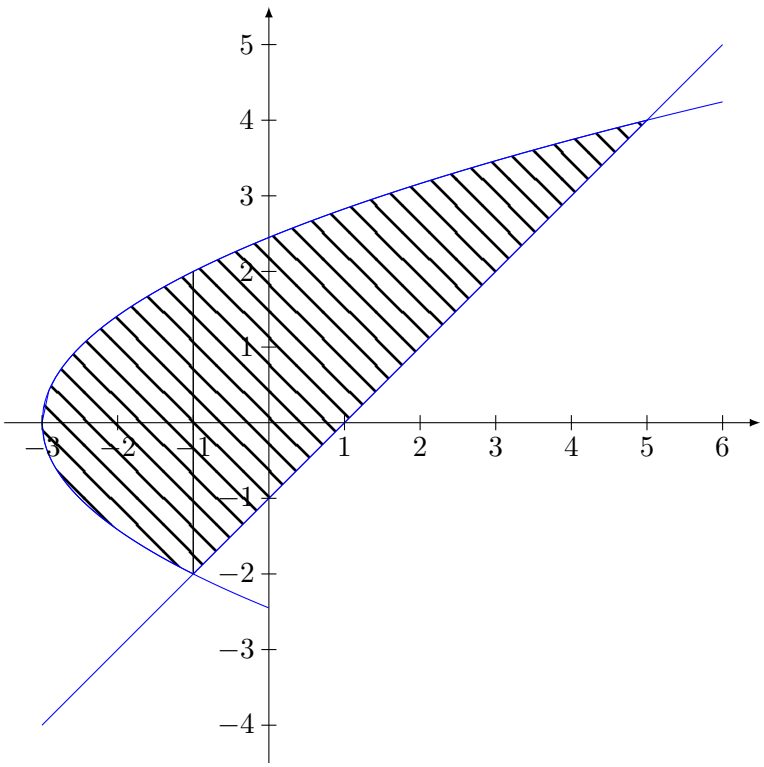


FIGURE 19.19 – La région D de l'exemple 19.185

Exemple 19.187.
La région E donnée par une portion de sphère collée à un cône est une région solide de premier type

$$E = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y) \in \overline{B\big((0, 0), 1\big)}, \sqrt{x^2 + y^2} \leqslant z \leqslant \sqrt{1 - x^2 - y^2}\}. \tag{19.563}$$

L'ombre de E est la boule unité de $\mathbb{R}^2$. L'ensemble $\sqrt{x^2 + y^2} \leqslant z$ est un cône posé sur sa pointe tandis que l'ensemble $z \leqslant \sqrt{1 - x^2 - y^2}$ est la demi-sphère. L'ensemble E contient les points entre les deux, voir la figure 19.20.

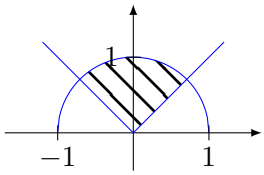


FIGURE 19.20 – Il faut voir ça en trois dimensions.

△

Si la fonction f, à intégrer sur E, et ses sections sont intégrables alors on peut réduire l'intégrale

$$\begin{aligned} \int_E f(x, y, z)\, dV &= \int_A \left(\int_{u_1(x,y)}^{u_2(x,y)} f(x, y, z)\, dz \right) dV = \\ &= \int_A \left(F(x, y, u_2(x, y)) - F(x, y, u_1(x, y)) \right) dV, \end{aligned} \tag{19.564}$$

où F est une primitive de f par rapport à la variable z, c'est-à-dire en considérant x et y comme des constantes. Il faut ensuite évaluer la partie qui reste comme dans la section précédente. Comme le calcul des aires dans $\mathbb{R}^2$, le calcul des volumes dans $\mathbb{R}^3$ est fait par des intégrales. En fait le **volume** d'une région solide dans $\mathbb{R}^3$ est sa mesure.

Définition 19.188.
La mesure d'une région de $\mathbb{R}^3$ est l'intégrale de sa fonction caractéristique.

Soit E une région solide du premier type, nous pouvons évaluer son volume par l'intégrale

$$\int_A (u_2(x,y) - u_1(x,y))\, dV.$$

Parfois c'est plus intéressant de calculer le volume avec la formule de réduction contraire : l'intégrale double d'abord et puis l'intégrale simple par rapport à z. On parle alors de calcul de volume « par tranche ».

Exemple 19.189.
On veut calculer le volume de la boule de rayon a, centrée à l'origine $B = \{(x,y,z) \in \mathbb{R}^3 \,|\, x^2 + y^2 + z^2 \leqslant a^2\}$. On peut décrire B par

$$B = \left\{(x,y,z) \in \mathbb{R}^3 \,|\, (x,y) \in D_a, -\sqrt{a^2 - x^2 - y^2} \leqslant z \leqslant \sqrt{a^2 - x^2 - y^2}\right\},$$

où D_a est le disque de rayon a centré en $(0,0)$, donc le volume B sera

$$2\int_{D_a} \sqrt{a^2 - x^2 - y^2} dV.$$

Cette intégrale est un peu ennuyeuse à calculer. On peut simplifier le calcul en observant que pour $\bar{z}$ fixé dans l'intervalle $[-a,a]$ la section de la boule au niveau $\bar{z}$ est un disque de rayon $\sqrt{a^2 - z^2}$. L'aire d'un tel disque est $\pi(a^2 + z^2)$. Si on réduit l'intégrale de volume de la façon

$$\int_B 1\, dV = \int_{-a}^{a} \sqrt{a^2 - z^2}\, dz,$$

on obtient tout de suite la valeur cherchée : le volume de B est $4/3\pi a^3$. △

Exemple 19.190.
On calcule l'intégrale de $f(x,y,z) = z$ sur la pyramide P bornée par le plans $x = 0$, $y = 0$, $x+y+z = 1$, $x+y+z/2 = 1$. On remarque tout de suite que le plans $x+y+z = 1$, $x+y+z/2 = 1$ se coupent en la droite $x + y = 1$, $z = 0$ (on se souvient qu'*une* droite dans $\mathbb{R}^3$, c'est *deux* équations). Cela veut dire que la projection de P sur le plan x-y est le triangle T borné par les droites $x = z = 0$, $y = z = 0$ et $x + y = 1$, $z = 0$. On décrit donc P par

$$P = \{(x,y,z) \in \mathbb{R}^3 \,|\, (x,y) \in T,\, 1 - 2x - 2y \leqslant z \leqslant 1 - x - y\}$$

et T par

$$T = \{(x,y) \in \mathbb{R}^2 \,|\, 0 \leqslant x \leqslant 1,\, 0 \leqslant y \leqslant 1 - x\},$$

donc l'intégrale de f sur P est

$$\int_p f(x,y,z)\, dV = \int_0^1 \int_0^{1-x} \int_{1-2x-2y}^{1-x-y} z\, dz\, dy\, dx = -\frac{1}{24}.$$

Notez que lorsque x et y sont entre 0 et 1, nous avons bien $1 - 2x - 2y < 1 - x - y$, d'où le fait que nous mettons $1 - 2x - 2y$ dans la borne inférieure de l'intégrale. △

De façon analogue on définit les régions solides du deuxième et du troisième type.

19.25.11 Intégrales de fonctions non bornées sur des ensembles non bornés

Soit $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, une fonction positive. On dit qu'elle est **intégrable** sur $E \subset \mathbb{R}^n$ si

(1) $\forall r > 0$, la fonction $f_r(x) = f(x)\mathbb{1}_{f<r}$ est intégrable sur E_r ;
(2) la limite $\lim_{r\to\infty} \int_{E_r} f_r$ est finie.

Dans ce cas, on pose

$$\int_E f = \lim_{r\to\infty} \int_{E_r} f_r. \tag{19.565}$$

Théorème 19.191.
Soit E mesurable dans $\mathbb{R}^n$ et $f\colon E \to \overline{\mathbb{R}}$. Si f est mesurable et s'il existe $g\colon E \to \overline{\mathbb{R}}$ intégrable sur E telle que $|f(x)| \leqslant g(x)$ pour tout $x \in E$, alors f est intégrable sur E.

Réciproquement, si f est intégrable sur E, alors f est mesurable.

Lemme 19.192.
Si f est une fonction sur $[a, \infty[$, alors nous avons la formule

$$\lim_{b\to\infty} \int_a^b f(x)dx = \int_a^\infty f(x)dx \tag{19.566}$$

au sens où si un des deux membres existe, alors l'autre existe et est égal.

Démonstration. Supposons que le membre de gauche existe. Cela signifie que la fonction

$$\psi(x) = \int_a^x f \tag{19.567}$$

est bornée. Soit M, un majorant. Pour toute fonction simple φ dominant f, on a $\int \varphi \leqslant M$, donc l'ensemble sur lequel on prend le supremum pour calculer $\int_a^\infty f$ est majoré par M et possède donc un supremum. Nous avons donc

$$\int_a^\infty f \leqslant \lim_{b\to\infty} \int_a^b f. \tag{19.568}$$

□

19.25.12 Lemme de Morse

Lemme 19.193 (Lemme de Morse)**.**
Soit $f \in C^3(\mathcal{U}, \mathbb{R})$ où $\mathcal{U}$ est un ouvert de $\mathbb{R}^n$ contenant 0. Nous supposons que $df_0 = 0$ et que $d^2 f_0$ est non dégénérée[30] *et de signature $(p, n-p)$. Alors il existe un C^1-difféomorphisme φ entre deux voisinages de 0 dans $\mathbb{R}^n$ tel que*

(1) $\varphi(0) = 0$,

(2) si $\varphi(x) = u$ alors

$$f(x) - f(0) = u_1^2 + \cdots + u_p^2 - u_{p+1}^2 - \ldots - u_n^2. \tag{19.569}$$

Une autre façon de dire est qu'il existe un C^1-difféomorphisme local ψ tel que

$$(f \circ \psi)(x) - f(0) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \ldots - x_n^2. \tag{19.570}$$

Démonstration. Nous allons noter Hf la matrice hessienne de f, c'est-à-dire $Hf_a = d^2 f_a \in \mathcal{L}^{(2)}(\mathbb{R}^n, \mathbb{R})$. Écrivons la formule de Taylor avec reste intégral (proposition 19.147 avec $p = 0$ et $m = 2$) :

$$f(x) - f(0) = \underbrace{df_0(x)}_{=0} + \int_0^1 (1-t) \underbrace{d^2 f_{tx}(x,x)}_{x^t (Hf)_{tx} x = \langle Hf_{tx} x, x\rangle} dt = x^t Q(x) x \tag{19.571}$$

avec

$$Q(x) = \int_0^1 (1-t)(Hf)_{tx} dt \tag{19.572}$$

30. En tant qu'application bilinéaire.

qui est une intégrale dans $\mathcal{L}^{(2)}(\mathbb{R}^n, \mathbb{R})$. Nous prouvons à présent que Q est de classe C^1 en utilisant le résultat de différentiabilité sous l'intégrale 16.26. Pour cela nous passons aux composantes (de la matrice) et nous considérons

$$\begin{aligned} h_{kl} \colon U \times [0,1] &\to \mathbb{R} \\ h_{kl}(x,t) &= (1-t)\frac{\partial^2 f}{\partial x_k \partial x_l}(tx). \end{aligned} \tag{19.573}$$

Étant donné que f est de classe C^3, la dérivée de h_{kl} par rapport à x_i ne pose pas de problèmes :

$$\frac{\partial h_{kl}}{\partial x_i} = t(t-1)\frac{\partial^3 f}{\partial x_i \partial x_k \partial x_l}(tx), \tag{19.574}$$

qui est encore continue à la fois en t et en x. La proposition 16.26 nous montre à présent que

$$Q_{kl}(x) = \int_0^1 (1-t)h_{kl}(tx)dt \tag{19.575}$$

est une fonction C^1. Étant donné que les composantes de Q sont C^1, la fonction Q est également C^1.

Nous avons $Q(0) = \frac{1}{2}(Hf)_0 \in S_n \cap \mathrm{GL}(n,\mathbb{R})$, d'abord parce que f est C^2 (et donc la matrice hessienne est symétrique), ensuite par hypothèse $d^2 f_0$ est non dégénérée.

À partir de là, le lemme 16.113 donne un voisinage V de $Q(0)$ dans S_n et une application ϕ de classe C^1

$$\phi \colon V \to \mathrm{GL}(n,\mathbb{R}) \tag{19.576}$$

telle que pour tout $A \in V$,

$$\phi(A)^t Q(0)\phi(A) = A. \tag{19.577}$$

Si on pose $M = \phi \circ Q$, et si x est dans un voisinage de zéro, Q étant continue nous avons $Q(x) \in V$ et donc

$$Q(x) = M(x)^t Q(0) M(x). \tag{19.578}$$

Notons que l'application $\mathbb{M} \colon \mathbb{R} \to \mathrm{GL}(n,\mathbb{R})$ est de classe C^1 parce que Q et ϕ le sont.

Nous avons

$$f(x) - f(0) = x^t Q(x) x = x^t M(x)^t Q(0) M(x) x = y(x)^t Q(0) y(x) \tag{19.579}$$

où $y(x) = M(x)x = (\phi \circ Q)(x)x$ est encore une fonction de classe C^1 parce que la multiplication est une application C^∞.

D'un autre côté le théorème de Sylvester 9.186 nous donne une matrice inversible P telle que

$$Q(0) = P^t \begin{pmatrix} \mathbb{1}_p & \\ & -\mathbb{1}_{n-p} \end{pmatrix} P. \tag{19.580}$$

Et nous posons enfin $u = \varphi(x) = Py(x)$ qui est toujours de classe C^1 et qui donne

$$f(x) - f(0) = y^t Q(0) y \tag{19.581a}$$

$$= y^t P^t \begin{pmatrix} \mathbb{1} & \\ & -\mathbb{1} \end{pmatrix} Py \tag{19.581b}$$

$$= u^t \begin{pmatrix} \mathbb{1} & \\ & -\mathbb{1} \end{pmatrix} u \tag{19.581c}$$

$$= u_1^2 + \cdots + u_p^2 - u_{p+1}^2 - \ldots - u_n^2. \tag{19.581d}$$

Nous devons maintenant montrer que, quitte à réduire son domaine à un ouvert plus petit, φ est un C^1-difféomorphisme. Dans la chaine qui donne φ, seule l'application

$$\begin{aligned} g \colon U \subset \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto M(x)x \end{aligned} \tag{19.582}$$

est sujette à caution. Nous allons appliquer le théorème d'inversion locale. Nous savons que g est de classe C^1 et donc différentiable ; calculons la différentielle en utilisant la formule (11.485) :

$$dg_0(x) = \frac{d}{dt}\Big[g(tx)\Big]_{t=0} = \frac{d}{dt}\Big[tM(tx)x\Big]_{t=0} = M(0)x. \tag{19.583}$$

Note que nous avons utilisé la règle de Leibnitz pour la dérivée d'un produit, mais le second terme s'est annulé. Donc $dg_0 = M(0) \in \mathrm{GL}(n, \mathbb{R})$ et g est localement un C^1-difféomorphisme.

Il suffit de restreindre φ au domaine sur lequel g est un C^1-difféomorphisme pour que φ devienne lui-même un C^1-difféomorphisme.

□

Définition 19.194.
*Un point a est un **point critique** de la fonction différentiable f si $df_a = 0$.*

Corolaire 19.195 ([397])**.**
Les points critiques non dégénérés d'une fonction C^3 sont isolés.

Démonstration. Soit a un point critique non dégénéré. Par le lemme de Morse 19.193, il existe un C^1-difféomorphisme ψ et un entier p tel que

$$(f \circ \psi)(x) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \ldots - x_n^2 + f(a) \tag{19.584}$$

sur un voisinage $\mathcal{U}$ de a. Vue la formule générale $df_x(u) = \nabla f(x) \cdot u$, si x est un point critique de f, alors $\nabla f(x) = 0$. Dans notre cas, les points critiques de $f \circ \psi$ dans $\mathcal{U}$ doivent vérifier $x_i = 0$ pour tout i, et donc $x = a$.

Nous devons nous assurer que la fonction f elle-même n'a pas de points critiques dans $\mathcal{U}$. Pour cela nous utilisons la formule générale de dérivation de fonction composée :

$$\nabla(f \circ \psi)(x) = \sum_k \frac{\partial f}{\partial y_k}\big(g(x)\big)\nabla g_k(x). \tag{19.585}$$

Si $\psi(x)$ est une point critique de f, alors le membre de droite est le vecteur nul parce que tous les $\partial_k f\big(\psi(x)\big)$ sont nuls. Par conséquent le membre de gauche est également nul, et x est un point critique de $f \circ \psi$. Or nous venons de voir que $f \circ \psi$ n'a pas de points critiques dans $\mathcal{U}$.

Donc f n'a pas de points critiques dans un voisinage d'un point critique non dégénéré. □

19.26 Autres intégrales sympathiques

19.26.1 Intégrale de Wallis

Lemme 19.196 ([398])**.**
Soit $n > 0$. En posant $I_n = \int \sin^n(x)dx$, nous avons :

$$I_n = \frac{\cos(x)\sin^{n-1}(x)}{n} + \frac{n-1}{n} I_{n-2}, \tag{19.586}$$

c'est-à-dire

$$\int \sin^n(x)dx = \frac{n-1}{n}\int \sin^{n-2}(t)dt - \frac{\sin^{n-1}(x)\cos(x)}{n}. \tag{19.587}$$

De la même manière,

$$\int \cos^n(x)dx = \frac{\cos^{n-1}(x)\sin(x)}{n} + \frac{n-1}{n}\int \cos^{n-2}(x)dx. \tag{19.588}$$

Démonstration. Nous posons $I_n = \int \sin^n(t)dt$, et nous y allons par récurrence. D'abord pour $n = 1$. Dans ce cas, la formule à démontrer se réduit à

$$\int \sin(x)dx = \cos(x). \tag{19.589}$$

Pas de problèmes.

Pour $n \geqslant 2$, nous évaluons l'intégrale I_n en utilisant une intégration par partie[31] en posant

$$\begin{cases} u = \sin^{n-1}(t) & (19.590a) \\ v' = \sin(t) & (19.590b) \end{cases}$$

et en déduisant

$$\begin{cases} u' = (n-1)\sin^{n-2}(t)\cos(t) & (19.591a) \\ v = -\cos(t). & (19.591b) \end{cases}$$

Nous avons alors le calcul

$$\begin{aligned} I_n &= -\cos(x)\sin^{n-1}(x) + \int (n-1)\sin^{n-2}(x)\cos^2(x)dx & (19.592a) \\ &= -\cos(x)\sin^{n-1}(x) + (n-1)\int \sin^{n-2}\underbrace{\cos^2(x)}_{=1-\sin^2(x)}\,dx & (19.592b) \\ &= \cos(x)\sin^{n-1}(x) + (n-1)\underbrace{\int \sin^{n-2}(x)dx}_{=I_{n-2}} - (n-1)\underbrace{\int \sin^n(x)\,dx}_{=I_n} & (19.592c) \end{aligned}$$

Nous avons donc déjà prouvé que

$$I_n = \cos(x)\sin^{n-1}(x) + (n-1)(I_{n-2} - I_n). \tag{19.593}$$

En isolant I_n,

$$I_n = \frac{\cos(x)\sin^{n-1}(x)}{n} + \frac{n-1}{n}I_{n-2}. \tag{19.594}$$

La formule (19.588) se démontre de la même façon. □

Lemme 19.197 ([399, 400])**.**
Nous posons

$$W_n = \int_0^{\pi/2} \cos^n(t)dt. \tag{19.595}$$

Alors :

(1) une formule de récurrence :

$$W_n = \frac{n-1}{n}W_{n-2}, \tag{19.596}$$

(2) et une formule un peu explicite :

$$W_{2n} = \frac{(2n)!}{(2^n n!)^2}\frac{\pi}{2}. \tag{19.597}$$

Démonstration. Le nombre W_n est seulement la seconde intégrale du lemme 19.196, évaluée entre 0 et $\pi/2$. En partant donc de (19.588), nous avons

$$W_n = \int_0^{\pi/2} \cos^n(x)dx = \left[\frac{\cos^{n-1}(x)\sin(x)}{n}\right]_0^{\pi/2} + \frac{n-1}{n}\int_0^{\pi/2} \cos^{n-2}(x)dx. \tag{19.598}$$

Le terme aux bords disparaît grâce aux valeurs trigonométriques remarquables[32]. Il reste immédiatement

$$W_n = \frac{n-1}{n}W_{n-2}. \tag{19.599}$$

31. Proposition 19.130.
32. Par exemple la proposition 17.17(10).

À partir de là, nous démontrons (19.597) par récurrence. D'abord pour $n = 0$, c'est l'égalité $W_0 = \pi/2$ qui est correcte parce que

$$\int_0^{\pi/2} \cos^0(x)dx = \int_0^{\pi/2} 1dx. \tag{19.600}$$

Pour la récurrence elle-même,

$$\begin{aligned} W_{2(n+1)} &= \frac{2n+1}{2(n+1)} W_{2n} && (19.601a)\\ &= \frac{2n+1}{2n+2} \frac{(2n)!}{(2^n n!)^2} \frac{\pi}{2} && (19.601b)\\ &\ldots \text{pas mal de petits calculs} \ldots && (19.601c)\\ &= \frac{(2n+2)!}{\big((n+1)!2^{n+1}\big)^2} \frac{\pi}{2}. && (19.601d) \end{aligned}$$

Voila. □

Maintenant, la suite (W_n) se divise en ses termes pairs et ses termes impairs. Pour les pairs, nous avons une formule assez explicite donnée par le lemme 19.197. Pour les termes impairs, nous n'avons rien. Dans tous les cas, nous avons la formule de récurrence

$$W_n = \frac{n-1}{n} W_{n-2} \tag{19.602}$$

qui ne sert à rien pour déduire des choses sur les termes impairs à partir de ce que l'on sait des termes pairs.

Sommes-nous perdus ? Non. La situation se débloque grâce au lemme suivant.

Lemme 19.198.
La suite donnée par

$$W_n = \int_0^{\pi/2} \cos^n(x)dx \tag{19.603}$$

est décroissante.

Démonstration. Vu que l'intégrale est sur $[0, \pi/2]$, le nombre $\cos(x)$ prend ses valeurs dans $[0, 1]$. Nous avons donc

$$\cos^{n+1}(x) \leqslant \cos^n(x). \tag{19.604}$$

Les intégrales suivent les mêmes inégalités. □

Ce lemme permet de relancer le jeu parce que les termes impairs sont coincés entre les termes pairs, qui décroissent. Les termes impairs doivent donc décroître à la même vitesse. Le lemme suivant met cela en musique.

Lemme 19.199.
Nous posons

$$W_n = \int_0^{\pi/2} \cos^n(x)dx \tag{19.605}$$

La fonction $\alpha\colon \mathbb{N} \to \mathbb{R}$ *définie par* $W_n = \alpha(n)W_{n-1}$ *vérifie* $\lim_{n\to\infty} \alpha(n) = 1$.

Démonstration. Parce que nous en aurons besoin, nous triturons d'abord un peu la formule de récurrence (19.596). D'abord nous l'inversons un peu pour avoir

$$W_n = \frac{n+2}{n} W_{n+2}, \tag{19.606}$$

et ensuite nous écrivons

$$W_{n-1} = \frac{n+1}{n} W_{n+1}. \tag{19.607}$$

Ne vous posez pas de questions, ça va être utile. Le fait que la suite soit décroissante (lemme 19.198) nous permet d'écrire

$$W_{n+2} \leqslant W_n \leqslant W_{n-1}. \tag{19.608}$$

En y remplaçant W_n par $\alpha(n)W_{n+1}$ et W_{n-1} par (19.607),

$$W_{n+1} \leqslant \alpha(n)W_{n+1} \leqslant \frac{n+1}{n}W_{n+1}. \tag{19.609}$$

Nous simplifions par W_{n+1} et nous trouvons l'encadrement, valable pour tout n :

$$1 \leqslant \alpha(n) \leqslant \frac{n+1}{n}. \tag{19.610}$$

Nous en déduisons par la règle de l'étau que $\alpha(n) \to 1$. □

Lemme 19.200.
Soit

$$W_n = \int_0^{\pi/2} \cos^n(x)dx. \tag{19.611}$$

Encore plusieurs choses à dire.

(1) Pour tout n nous avons

$$nW_nW_{n-1} = \frac{\pi}{2}. \tag{19.612}$$

(2) Nous avons l'équivalence de suites[33]

$$W_n \sim \sqrt{\frac{\pi}{2n}}. \tag{19.613}$$

Démonstration. En deux parties

La suite constante Nous posons $K_n = nW_nW_{n-1}$. Grâce aux formules de récurrence (19.596) que nous écrivons sous la forme

$$W_n = \frac{n-1}{n}W_{n-2} \tag{19.614a}$$

$$W_{n+1} = \frac{n}{n+1}W_{n-1}, \tag{19.614b}$$

nous avons

$$K_{n+1} = (n+1)W_{n+1}W_n = (n+1)\frac{n}{n+1}\frac{n-1}{n}W_{n-2}W_{n_2} = (n-1)W_{n-1}W_{n-2} = K_{n-1}. \tag{19.615}$$

Nous avons montré que $K_{n+1} = K_{n-1}$.
Il nous reste à prouver que $K_1 = K_2 = \frac{\pi}{2}$. Pour cela nous avons immédiatement $W_0 = \frac{\pi}{2}$ ainsi que

$$W_1 = \int_0^{\pi/2} \cos(t)dt = 1. \tag{19.616}$$

Pour W_2, il ne faut pas calculer d'intégrales, mais seulement utiliser la formule (19.597). Nous trouvons vite $W_2 = \frac{\pi}{4}$. Donc

$$K_1 = W_1W_0 = \frac{\pi}{2} \tag{19.617}$$

et

$$K_2 = 2W_2W_1 = \frac{\pi}{2}. \tag{19.618}$$

33. Définition 7.136.

L'équivalence de suites Soit la fonction α définie par $W_n = \alpha(n)W_{n-1}$. Le lemme 19.199 nous dit que $\alpha(n) \to 1$. Nous l'utilisons dans (19.612) :

$$\frac{\pi}{2} = nW_nW_{n-1} = n\alpha(n)W_{n-1}^2. \tag{19.619}$$

Donc

$$W_{n-1}^2 = \frac{\pi}{2n\alpha(n)}, \tag{19.620}$$

et

$$W_{n-1}\sqrt{\frac{1}{\alpha(n)}}\sqrt{\frac{\pi}{2n}}. \tag{19.621}$$

Vu que le coefficient $\sqrt{1/\alpha(n)}$ tend vers 1 pour $n \to 1$, nous avons l'équivalence demandée.

□

19.26.2 Formule de Stirling

Lemme 19.201 ([401]).
Si $n \in \mathbb{N}$ nous avons la formule

$$\ln(1+\frac{1}{n}) = 2\sum_{k=0}^{\infty}\frac{1}{2k+1}\left(\frac{1}{2n+1}\right)^{2k+1} \tag{19.622}$$

Démonstration. Il s'agit d'utiliser astucieusement le développement de la proposition 14.74. Nous avons d'une part

$$\ln(1+t) = \sum_{k=1}^{\infty}(-1)^{k+1}\frac{x^k}{k} \tag{19.623}$$

et d'autre part,

$$-\ln(1-t) = \sum_{k=1}^{\infty}\frac{x^k}{k}. \tag{19.624}$$

Cela permet de calculer, en utilisant l'associativité de la série [34]

$$\ln\left(\frac{1+t}{1-t}\right) = \ln(1+t) - \ln(1-t) \tag{19.625a}$$

$$= \sum_{k=1}^{\infty}\left(\frac{(-1)^{k+1}x^k}{k} + \frac{x^k}{k}\right) \tag{19.625b}$$

$$= 2\sum_{k=0}^{\infty}\frac{x^{2k+1}}{2k+1} \tag{19.625c}$$

parce que tous les termes pairs s'annulent, tandis que les termes impairs sont doublés. Notez que la somme dans (19.625c) commence à $k = 0$, contrairement aux autres qui commencent à 1.

Et là c'est l'astuce : on pose écrit l'égalité (19.625) avec le t qu'il faut pour que

$$\frac{1+t}{1-t} = 1 + \frac{1}{n}. \tag{19.626}$$

Nous posons donc $t = \frac{1}{2n+1}$ nous avons

$$\ln\left(1+\frac{1}{n}\right) = 2\sum_{k=0}^{\infty}\frac{1}{2k+1}\left(\frac{1}{2n+1}\right)^{2k+1}, \tag{19.627}$$

ce qu'il fallait. □

34. Proposition 10.60.

Lemme 19.202 (Formule de Stirling[402, 403, 399, 401, 404]).
Nous avons l'équivalence de suites[35]

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}. \tag{19.628}$$

Démonstration. Nous posons

$$a_n = \frac{n!}{\sqrt{2n}\left(\frac{n}{e}\right)^n} \tag{19.629}$$

et $b_n = \ln(a_n)$.

Une formule pour $b_n - b_{n+1}$ Nous faisons un beau calcul qui utilise les formules de la proposition 14.60 ainsi que $\ln(e) = 1$:

$$\begin{aligned}
b_n - b_{n+1} &= \ln\left(\frac{a_n}{a_{n+1}}\right) && (19.630a)\\
&= \ln\left(\frac{(n+1)^{n+1/2}}{n^{n+1/2}}\frac{1}{e}\right) && (19.630b)\\
&= \ln\left((\frac{n+1}{n})^{n+1/2}\right) - 1 && (19.630c)\\
&= \left(n+\frac{1}{2}\right)\ln(1+\frac{1}{n}) - 1. && (19.630d)
\end{aligned}$$

La suite (b_n) est décroissante Nous écrivons l'égalité (19.630) en utilisant le lemme 19.201 :

$$\begin{aligned}
b_n - b_{n+1} &= (n+\frac{1}{2})\ln(1+\frac{1}{n}) - 1 && (19.631a)\\
&= \frac{1}{2}(2n+1)2\sum_{k=0}^{\infty}\frac{1}{2k+1}\left(\frac{1}{2n+1}\right)^{2k+1} - 1 && (19.631b)\\
&= (2n+1)\sum_{k=1}^{\infty}\frac{1}{2k+1}\left(\frac{1}{2n+1}\right)^{2k+1} && (19.631c)\\
&= \sum_{k=1}^{\infty}\frac{1}{2k+1}\left(\frac{1}{2n+1}\right)^{2k}. && (19.631d)
\end{aligned}$$

Notez que le terme $k = 0$ s'est simplifié avec le -1. Vu que le tout est une somme de termes positifs, nous avons

$$b_n - b_{n+1} > 0 \tag{19.632}$$

et la suite est décroissante.

Majoration pour $b_n - b_{n+1}$ Vu que $\frac{1}{2k+1} < 1$, nous pouvons majorer :

$$b_n - b_{n+1} \leqslant \sum_{k=1}^{\infty}\left(\frac{1}{(2n+1)^2}\right)^k. \tag{19.633}$$

Nous remarquons que cela est une série géométrique déjà traitée dans l'exemple 10.67. Nous faisons un peu de calcul en partant de

$$\sum_{k=1}^{\infty} q^n = \frac{q}{1-q} \tag{19.634}$$

avec $q = \frac{1}{(2n+1)^2}$. Après quelques simplifications,

$$b_n - b_{n+1} \leqslant \frac{1}{4}\frac{1}{n(n+1)} \tag{19.635}$$

35. Définition 7.136.

Le coup de la somme télescopique Nous avons

$$b_1 - b_n = (b_1 - b_2) + (b_2 - b_3) + \ldots + (b_{n_1} - b_n). \tag{19.636}$$

Chacun de ces termes est majoré ; nous avons donc

$$b_1 - b_n < \frac{1}{4}\sum_{m=1}^{n-1}\frac{1}{m(m+1)} \leqslant \frac{1}{4}\sum_{m=1}^{\infty}\frac{1}{m(m+1)} = \frac{1}{4} \tag{19.637}$$

grâce au lemme 10.71 pour la dernière somme.

La suite b_n est bornée vers le bas Vu que $b_1 = 1 - \frac{\ln(2)}{2}$, nous avons

$$b_n > b_1 - \frac{1}{4} = \frac{3}{4} - \frac{\ln(2)}{2}. \tag{19.638}$$

En utilisant la majoration de l'exemple 19.146 nous trouvons

$$0.327 \leqslant b_n \leqslant 0.427. \tag{19.639}$$

Cet encadrement n'est pas très important. Le point est que la suite (b_n) soit bornée vers le bas ; savoir que la borne est strictement positive n'est pas indispensable.

Une limite pour (a_n) La suite (b_n) est décroissante et bornée vers le bas, donc elle est convergente par le lemme 7.137. Vu que l'exponentielle est une fonction continue [36], la suite $a_n = e^{b_n}$ est également convergente.

Vu que $\lim_{n\to\infty} b_n > 0$, nous avons $\lim_{n\to\infty} a_n = e^{\lim_{n\to\infty} b_n} > 1$.

L'important est que nous sachions que (a_n) est une suite convergente. Nous notons L sa limite :

$$\lim_{n\to\infty}\frac{n!}{\sqrt{2n}\left(\frac{n}{e}\right)^n} = L. \tag{19.640}$$

Ou encore :

$$\lim_{n\to\infty}\frac{n!}{L\sqrt{2n}\left(\frac{n}{e}\right)^n} = 1. \tag{19.641}$$

Autrement dit, en posant

$$n! = \alpha(n)L\sqrt{2n}\left(\frac{n}{e}\right)^n, \tag{19.642}$$

nous avons $\lim_{n\to\infty}\alpha(n) = 1$.

Introduction de Wallis Nous avons déjà parlé dans le lemme 19.197 du nombre

$$W_{2n} = \frac{(2n)!}{(2^n n!)^2}\frac{\pi}{2}. \tag{19.643}$$

Le lemme 19.200 implique que la suite $I_n = W_{2n}$ est équivalente à $\sqrt{\pi/4n}$.

Cela étant dit, nous faisons un gros calcul en remplaçant les factorielles dans W_{2n} par la formule (19.642). Après pas mal de calculs [37], nous trouvons

$$I_n = \frac{\alpha(2n)}{\alpha(n)^2}\frac{1}{\sqrt{n}}\frac{\pi}{2n}. \tag{19.644}$$

Nous avons donc

$$\sqrt{\frac{\pi}{4n}} \sim \frac{\alpha(2n)}{\alpha(n)^2}\frac{\pi}{\sqrt{n}}\frac{1}{2L}. \tag{19.645}$$

Il existe donc une fonction $\beta(n)$ avec $\beta(n) \to 1$ telle que

$$\frac{\sqrt{\pi}}{2}\frac{1}{\sqrt{n}} = \beta(n)\frac{\alpha(2n)}{\alpha(n)^2}\frac{\pi}{\sqrt{n}}\frac{1}{2L}. \tag{19.646}$$

36. Parce qu'elle est dérivable, voir par exemple le théorème 14.54.
37. Si vous êtes en manque de papier de brouillon, c'est le moment de vous inquiéter.

En simplifiant par $1/\sqrt{n}$,

$$\frac{\sqrt{\pi}}{2} = \beta(n)\frac{\alpha(2n)}{\alpha(n)^2}\frac{\pi}{2L}. \tag{19.647}$$

Nous prenons à présent la limite $n \to \infty$ en nous rappelant que α et β donnent 1. Après simplifications, nous trouvons

$$L = \sqrt{\pi}. \tag{19.648}$$

La fin Nous introduisons la valeur $L = \sqrt{\pi}$ dans l'expression (19.642) de la factorielle :

$$n! = \alpha(n)\sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \tag{19.649}$$

La dernière équation est exactement ce qui signifie l'équivalence de suite demandée. □

19.26.3 La fonction sinus cardinal, intégrale de Dirichlet

Définition 19.203.
*La fonction **sinus cardinal** est*

$$f(t) = \begin{cases} 1 & si\ t = 0 \\ \frac{\sin(t)}{t} & sinon. \end{cases} \tag{19.650}$$

Elle sert à plein de choses. Entre autres, le lemme 19.206 montrera que la fonction $x \mapsto |\sin(x)/x|$ a une intégrale sur $\mathbb{R}$ qui vaut ∞. Cela nous permettra de donner un exemple d'une fonction dans $L^1(\mathbb{R})$ dont la transformée de Fourier n'est pas dans $L^1(\mathbb{R})$ (lemme 28.8).

19.204.
Le but que nous nous fixons maintenant est de prouver que

$$\int_0^\infty \frac{\sin(t)}{t}dt = \frac{\pi}{2}. \tag{19.651}$$

Un adage dit que si un théorème est trop long, c'est qu'il n'a pas assez de lemmes. Nous allons faire plein de lemmes.

Lemme 19.205.
La fonction sinus cardinal est continue.

Démonstration. Elle est continue en zéro parce que le lemme 17.214 nous donne

$$\lim_{t\to 0}\frac{\sin(t)}{t} = 1. \tag{19.652}$$

□

Nous commençons par une mauvaise nouvelle.

Lemme 19.206 ([1])**.**
Nous avons

$$\int_{\mathbb{R}} |\frac{\sin(x)}{x}|dx = \infty. \tag{19.653}$$

Démonstration. Soit $\delta > 0$ tel que sur l'intervalle $[\frac{\pi}{2} - \delta, \frac{\pi}{2} + \delta]$, nous ayons $\sin(x) > 0.9$ [38] .

Les intervalles $I_k = [\frac{\pi}{2} - \delta + 2k\pi, \frac{\pi}{2} + \delta + 2k\pi]$ sont disjoints et la fonction que nous intégrons est partout positive. Nous découpons

$$\mathbb{R} = C + \bigcup_{k=0}^{\infty}[\frac{\pi}{2} - \delta + 2k\pi, \frac{\pi}{2} + \delta + 2k\pi] \tag{19.654}$$

38. Ça existe par une astucieuse combinaison du théorème 11.50 des valeurs intermédiaires, de la valeur remarquable $\sin(\pi/2) = 1$ (de (17.40)) et du fait que sin est continue (proposition 17.1).

où C est le complémentaire qu'il faut pour faire $\mathbb{R}$.

La σ-additivité de l'intégrale de Lebesgue (proposition 13.188) nous indique que

$$\int_{\mathbb{R}} |\frac{\sin(x)}{x}| = \int_C |\frac{\sin(x)}{x}| + \sum_{k=0}^{\infty} \int_{I_k} |\frac{\sin(x)}{x}| dx \tag{19.655}$$

Vu que tous les termes sont positifs, nous obtenons une majoration en en supprimant un. Allons-y :

$$\begin{aligned}
\int_{\mathbb{R}} |\frac{\sin(x)}{x}| dx &\geqslant \sum_{k=0}^{\infty} \int_{I_k} |\frac{\sin(x)}{x}| dx && (19.656a)\\
&\geqslant \sum_{k=0}^{\infty} 0.9 \int_{I_k} \frac{1}{x} dx && (19.656b)\\
&\geqslant 0.9 \sum_{k=0}^{\infty} 2\delta \frac{1}{\frac{\pi}{2} - \delta + 2k\pi} && (19.656c)\\
&\geqslant 1.8\delta \sum_{k=0}^{\infty} \frac{1}{2\pi(k+1)} && (19.656d)\\
&= \frac{1.8\delta}{2\pi} \sum_{k=0}^{\infty} \frac{1}{k+1}. && (19.656e)
\end{aligned}$$

Justifications :

— Pour (19.656c), nous avons majoré $\frac{1}{x}$ par $\frac{1}{\frac{\pi}{2}+\delta+2k\pi}$ sur I_k.

— Pour (19.656d), nous avons dit que $\frac{\pi}{2} + \delta < 2\pi$.

La dernière somme dans (19.656) diverge.

Donc la fonction sinus cardinal n'est pas dans $L^1(\mathbb{R})$. □

La mauvaise nouvelle suivante en est un corolaire immédiat.

Lemme 19.207.
La fonction $t \mapsto \frac{\sin(t)}{t}$ n'est pas intégrable sur $[0, \infty[$ au sens de Lebesgue.

Démonstration. Le lemme 19.206 nous dit que $\int_0^\infty |f| = \infty$. Dans ce cas, $\int_0^\infty f$ n'existe pas par le lemme 13.169. □

Donc l'intégrale $\int_0^\infty \frac{\sin(t)}{t} dt$ n'existe pas parce que la définition de l'intégrale de Lebesgue ne permet pas de profiter des compensations qui arrivent entre les valeurs positives et négatives.

Nous définissons donc

$$\int_0^\infty \frac{\sin(t)}{t} dt = \lim_{b\to\infty} \int_0^b \frac{\sin(t)}{t} dt. \tag{19.657}$$

Pour chaque b, l'intégrale existe sans problèmes (fonction continue sur le compact $[0, b]$), et les compensations se font. Il n'est pas pas sans espoir que la limite (19.657) existe et valle un nombre fini.

Lemme 19.208 ([405])**.**
La limite

$$\lim_{b\to\infty} \int_0^b \frac{\sin(t)}{t} dt \tag{19.658}$$

existe dans $\mathbb{R}$.

Démonstration. En plusieurs parties.

Découpage Nous découpons l'intervalle $[0,b]$ en morceaux du type $[k\pi,(k+1)\pi]$ et un morceau restant lorsque b n'est pas un multiple de π :

$$[0,b]=\bigcup_{k=0}^{N(b)-1}[k\pi,(k+1)\pi]\cup[N(b)\pi,b] \tag{19.659}$$

où $N(b)$ est un entier bien choisi [39]. En tout cas $\lim_{b\to\infty}N(b)=\infty$.

Majoration 1 Pour chaque $b\in\mathbb{R}^+$ nous avons

$$\int_0^b\frac{\sin(t)}{t}dt=\sum_{k=0}^{N(b)}\int_{[k\pi,(k+1)\pi]}\frac{\sin(t)}{t}dt+\int_{[(N(b)+1)\pi,b]}\frac{\sin(t)}{t} \tag{19.660}$$

Dans le dernier terme, nous majorons $|\sin(t)|\leqslant 1$ et $\frac{1}{t}\leqslant\big(N(b)+1\big)\pi$. Cela donne

$$\left|\int_{[(N(b)+1)\pi,b]}\frac{\sin(t)}{t}\right|\leqslant\frac{b-\big(N(b)+1\big)\pi}{\big(N(b)+1\big)\pi}\leqslant\frac{1}{N(b)+1} \tag{19.661}$$

où nous avons encore majoré $b\leqslant\big(N(b)+2\big)\pi$.

Majoration 2 En ce qui concerne les autres termes, sur l'intervalle $[k\pi,(k+1)\pi]$, nous avons $\sin(t)=(-1)^k|\sin(t)|$. Nous avons alors

$$\int_0^b\frac{\sin(t)}{t}dt=\sum_{k=0}^{N(b)}(-1)^k\int_{k\pi}^{(k+1)\pi}\frac{|\sin(t)|}{t}dt+\alpha(b) \tag{19.662}$$

où $|\alpha(b)|\leqslant\frac{1}{N(b)+1}$; l'important est que $\lim_{b\to\infty}\alpha(b)=0$.

Une suite alternée L'inégalité (19.662) nous incite à étudier la série $\sum_{k=0}^{\infty}(-1)^ka_k$ en ayant posé

$$a_k=\int_{k\pi}^{(k+1)\pi}\frac{|\sin(t)|}{t}dt. \tag{19.663}$$

Nous montrons à présent que la suite (a_k) vérifie les conditions du critère des séries alternées 10.72.

D'abord, $a_{k+1}\leqslant a_k$. En effet en utilisant le changement de variables [40] $u=t-\pi$,

$$a_{k+1}=\int_{(k+1)\pi}^{(k+2)\pi}\frac{|\sin(t)|}{t}dt=\int_{k\pi}^{(k+1)\pi}\frac{|\sin(u+\pi)|}{u+\pi}du=\int_{k\pi}^{(k+1)\pi}\frac{|\sin(u)|}{u+\pi}du<a_k. \tag{19.664}$$

Nous avons utilisé le fait que $|\sin(u+\pi)|=|\sin(u)|$ pour tout u.

De plus, vu que $|\sin(t)|\leqslant 1$ et que $t\in[k\pi,(k+1)\pi]$, nous avons

$$\int_{k\pi}^{(k+1)\pi}\frac{|\sin(t)|}{t}\leqslant\int_{k\pi}^{(k+1)\pi}\frac{1}{k\pi}=\frac{(k+1)\pi-k\pi}{k\pi}=\frac{1}{k}. \tag{19.665}$$

Donc $a_k\leqslant\frac{1}{k}\to 0$.

Le critère des séries alternées 10.72 nous dit que

$$\sum_{k=0}^{\infty}(-1)^ka_k<\infty. \tag{19.666}$$

39. Il me semble que le traitement de ce terme manque dans [405].
40. Le théorème 13.255 est toujours bon à citer.

Conclusion Nous repartons de (19.662) :

$$\int_0^b \frac{\sin(t)}{t}dt = \sum_{k=0}^{N(b)} (-1)^k \int_{k\pi}^{(k+1)\pi} \frac{|\sin(t)|}{t}dt + \alpha(b). \tag{19.667}$$

Cette égalité est valable pour tout $b \in \mathbb{R}^+$. Le passage à la limite $b \to 0$ à droite donne un nombre fini ; donc à gauche aussi, et nous avons prouvé que

$$\lim_{b\to\infty} \int_0^b \frac{\sin(t)}{t}dt < \infty. \tag{19.668}$$

□

Notre tache n'est donc pas sans espoir. Au moins l'intégrale que nous cherchons à évaluer est finie.

Lemme 19.209.
Soit $x \in]0, \infty[$. L'intégrale

$$F(x) = \int_0^\infty e^{-tx}\frac{\sin(t)}{t}dt \tag{19.669}$$

existe au sens de Lebesgue usuel.

Démonstration. Vu qu'en $t = 0$ nous avons $\frac{\sin(t)}{t} = 1$, il n'y a pas de problèmes de ce côté. Lorsque $t > 1$ nous avons la majoration

$$|e^{-tx}\frac{\sin(t)}{t}| \leqslant |e^{-tx}|. \tag{19.670}$$

Lorsque t est assez grand, le lemme 14.79 nous donne aussi la majoration

$$|e^{-tx}| \leqslant \frac{1}{t^2}. \tag{19.671}$$

La proposition 13.251(2) implique que $\int_1^\infty \frac{1}{t^2} < \infty$. Et les majorations font que la proposition 13.158 nous donne le résultat. □

Lemme 19.210 ([405])**.**
Il existe une constante $C \in \mathbb{R}$ telle que

$$I(x) = \int_0^\infty e^{-tx}\frac{\sin(t)}{t}dt = -\arctan(x) + C \tag{19.672}$$

pour tout $x > 0$.

Démonstration. En permutant dérivée et intégrale, nous allons prouver que $I'(x) = -\frac{1}{1+x^2}$.

Permuter Nous posons

$$f(x,t) = e^{-tx}\frac{\sin(t)}{t}, \tag{19.673}$$

et nous vérifions les hypothèses du théorème 16.18.

(1) Pour chaque $x > 0$ fixé, la fonction $t \mapsto f(x,t)$ est intégrable sur $[0, \infty[$, c'est le lemme 19.209.

(2) Pour $t > 0$ fixé, la fonction $x \mapsto f(x,t)$ est dérivable.

(3) Nous avons la dérivée partielle

$$\frac{\partial f}{\partial x}(x,t) = -e^{-tx}\sin(t) \tag{19.674}$$

qui vérifie

$$|\frac{\partial f}{\partial x}(x,t)| \leqslant e^{-tx}, \tag{19.675}$$

alors que la fonction $t \mapsto e^{-tx}$ est intégrable sur $[0, \infty[$.

Nous pouvons donc dériver sous l'intégrale et obtenir

$$F'(x) = -\int_0^\infty e^{-xt}\sin(t)dt. \tag{19.676}$$

Quelques intégrations par partie Nous posons

$$J(x) = \int_0^\infty e^{-xt}\sin(t)dt, \tag{19.677}$$

et nous allons la faire par parties [41], en deux fois.
D'abord en posant $u = e^{-xt}$ et $v' = \sin(t)$ nous avons

$$J(x) = \left[-e^{-xt}\cos(t)\right]_{t=0}^{t=\infty} - \int_0^\infty (-)xe^{-xt}(-)\cos(t)dt = 1 - x\int_0^\infty e^{-xt}\cos(t)dt. \tag{19.678}$$

Nous faisons l'intégrale encore par parties en posant $u = e^{-xt}$ et $v' = \cos(t)$:

$$\int_0^\infty e^{-xt}\cos(t)dt = \left[e^{-xt}\sin(t)\right]_0^\infty - \int_0^\infty (-)e^{-xt}\sin(t)dt = J(x). \tag{19.679}$$

Donc

$$\begin{aligned} J(x) &= 1 - x\int_0^\infty e^{-xt}\cos(t)dt && (19.680a)\\ &= 1 - x\int_0^\infty e^{-xt}\cos(t)dt && (19.680b)\\ &= 1 - x\Big(x\underbrace{\int_0^\infty e^{-xt}\sin(t)dt}_{J(x)}\Big) && (19.680c)\\ &= 1 - x^2J(x). && (19.680d) \end{aligned}$$

Voila qui prouve que $J(x) = \frac{1}{1+x^2}$, et donc que

$$F'(x) = -J(x) = -\frac{1}{1+x^2}. \tag{19.681}$$

Et enfin Le théorème 17.30(2) nous dit que la dérivée de la fonction arctan est précisément $1/(1+x)$. Donc F et arctan ont la même dérivée (au signe près). Donc il existe $C \in \mathbb{R}$ tel que

$$F(x) = -\arctan(x) + C \tag{19.682}$$

pour tout $x > 0$.

□

Lemme 19.211.
Nous avons

$$F(x) = \int_0^\infty e^{-xt}\frac{\sin(t)}{t}dt = -\arctan(x) + \frac{\pi}{2} \tag{19.683}$$

pour tout $x > 0$.

Démonstration. Le but de ce lemme est de fixer la constante laissée arbitraire dans le lemme 19.210. Nous savons qu'il existe $C \in \mathbb{R}$ tel que

$$F(x) = -\arctan(x) + C. \tag{19.684}$$

41. Proposition 19.130.

Le but est de prendre la limite $x \to \infty$ des deux côtés.

Par le lemme 17.215, nous avons $|\frac{\sin(t)}{t}| \leqslant 1$ sur $[0, \infty[$. Donc

$$F(x) \leqslant \int_0^\infty e^{-xt} dt = \left[-\frac{1}{x} e^{-xt} \right]_{t=0}^{t=\infty} = \frac{1}{x}. \tag{19.685}$$

Vu que $F(x) \leqslant \frac{1}{x}$ pour tout x, nous avons certainement $\lim_{x\to\infty} F(x) = 0$.

D'autre part,

$$\lim_{x\to\infty} \arctan(x) = \frac{\pi}{2}. \tag{19.686}$$

En passant à la limite dans (19.684), nous avons

$$0 = -\frac{\pi}{2} + C, \tag{19.687}$$

et donc $C = \pi/2$. □

Nous avons maintenant la formule

$$F(x) = \int_0^\infty e^{-tx} \frac{\sin(t)}{t} dt = -\arctan(x) + \frac{\pi}{2} \tag{19.688}$$

qui est valable pour tout $x > 0$.

Notre but sera de prendre la limite $x \to 0$ des deux côtés. Vu que arctan est continue, le membre de droite ne pose pas de problèmes et donne $\pi/2$. Pour le membre de gauche, il faut encore permuter une limite et une intégrale.

Pour la suite, nous allons étudier[405]

$$\int_0^\infty (1 - e^{-xt}) \frac{\sin(t)}{t} dt. \tag{19.689}$$

Cette intégrale n'existe pas au sens de Lebesgue et est définie par

$$L(x) = \lim_{b\to\infty} \int_0^b (1 - e^{-xt}) \frac{\sin(t)}{t} dt. \tag{19.690}$$

Rien n'indique cependant pour l'instant que cette limite existe.

Lemme 19.212 ([405]).
Soit $x > 0$. Nous posons

$$L_k(x) = \int_{k\pi}^{(k+1)\pi} (1 - e^{-tx}) \frac{|\sin(t)|}{t} dt. \tag{19.691}$$

La suite $(L_k(x))_{k\in\mathbb{N}}$ satisfait le critère des séries alternées[42]*, c'est-à-dire que cette suite est positive, décroissante à limite nulle.*

Démonstration. Notons que chacune des intégrales $L_k(x)$ est sans problèmes : fonction continue sur un compact. Trois éléments à prouver.

Positive Vu que dans toute notre histoire, $x, t > 0$, nous avons $1 - e^{-tx} > 0$ et donc toute la fonction intégrée est positive.

Tend vers zéro Vu que $1 - e^{-tx} < 1$, nous avons

$$L_k(x) \leqslant \int_{k\pi}^{(k+1)\pi} \frac{1}{t} dt \leqslant \pi \frac{1}{k\pi} = \frac{1}{k}. \tag{19.692}$$

Donc $\lim_{k\to\infty} L_k(x) = 0$.

42. Théorème 10.72.

Décroissante Nous devons à présent prouver que $L_k(x)$ est décroissante en k lorsque x est fixé.

Nous avons $L_{k+1}(0) = L_k(0)$ pour tout k. Nous allons montrer que $L'_{k+1}(x) < L'_k(x)$ pour tout $x > 0$. De cette façon nous aurons bien $L_{k+1}(x) < L_k(x)$ pour tout k et x.

En permutant (encore) intégrale et dérivée,

$$L'_{k+1}(x) = \int_{(k+1)\pi}^{(k+2)\pi} e^{-tx}|\sin(t)|dt \tag{19.693a}$$
$$= \int_{k\pi}^{(k+1)\pi} e^{-(u+\pi)x}|\sin(u+\pi)| \tag{19.693b}$$
$$= e^{-\pi x}\int_{k\pi}^{(k+1)\pi} e^{-ux}|\sin(u)|du \tag{19.693c}$$
$$= e^{-\pi x}L'_k(x). \tag{19.693d}$$

Justifications :

— Pour (19.693a), permuter dérivée et intégrale ; je ne donne pas tout le détail. Ça a déjà été fait.

— Pour (19.693b), nous avons fait le changement de variables $u = t - \pi$.

— Pour (19.693c), nous avons utilisé le fait que $|\sin(u+\pi)| = |\sin(u)|$ ainsi que $e^{-(u+\pi)x} = e^{-ux}e^{-\pi x}$ par (11.898).

Nous avons donc prouvé que

$$L'_{k+1}(x) = e^{-\pi x}L'_k(x) < L'_k(x). \tag{19.694}$$

□

Lemme 19.213.
Pour chaque $x > 0$, nous avons la limite

$$\lim_{b\to\infty}\int_0^b (1-e^{-tx})\frac{\sin(t)}{t}dt = \sum_{k=0}^{\infty}(-1)^k L_k(x) < \infty. \tag{19.695}$$

Démonstration. Nous fixons (provisoirement) b et nous découpons l'intervalle d'intégration comme

$$[0,b] = \bigcup_{k=1}^{N}[k\pi,(k+1)\pi] \cup \big[(N+1)\pi, b\big] \tag{19.696}$$

où N est une fonction de b ; quelque chose comme $N(b)$ est le plus grand entier tel que $\big(N(b)+1\big)\pi \leqslant \pi$. Sur chacun des intervalles nous avons $\sin(t) = (-1)^k|\sin(t)|$. Nous avons donc

$$\begin{aligned}\int_0^b (1-e^{-tx})\frac{\sin(t)}{t}dt = \sum_{k=0}^{N(b)}(-1)^k\int_{k\pi}^{(k+1)\pi}(1-e^{-tx})\frac{|\sin(t)|}{t}dt \\ + \int_{(N+1)\pi}^{b}(1-e^{-tx})\frac{\sin(t)}{t}dt\end{aligned} \tag{19.697}$$

Le premier terme est $\sum_{k=0}^{N(b)} L_k(x)$, dont nous savons que la limite $b \to \infty$ existe parce que $L_k(x)$ vérifie le critère des séries alternées (lemme 19.212). En ce qui concerne le second terme,

$$\Big|\int_{(N+1)\pi}^{b}(1-e^{-tx})\frac{\sin(t)}{t}\Big| < \frac{b-(N+1)\pi}{(N+1)\pi} < \frac{1}{N(b)+1}. \tag{19.698}$$

La dernière inégalité est le fait que $N(b)$ est choisi pour avoir $b - \big(N(b)+1\big)\pi < \pi$.

Les deux termes de (19.697) ont donc une limite lorsque $b \to \infty$. Nous pouvons donc passer à la limite en sommant les deux limites :

$$\int_0^\infty (1-e^{-tx})\frac{\sin(t)}{t}dt = \sum_{k=0}^\infty (-1)^k L_k(x). \tag{19.699}$$

Cette égalité est valable pour chaque $x > 0$.

Le fait que la limite soit finie est dans le critère des séries alternées. Pour chaque x, la suite $L_k(x)$ vérifie ce critère par le lemme 19.212. □

Lemme 19.214.
Nous avons

$$\lim_{x\to 0^+} \int_0^\infty (1-e^{-tx})\frac{\sin(t)}{t}dt = 0. \tag{19.700}$$

Démonstration. La définition de l'intégrale ainsi que le lemme 19.213 nous ont déjà donné

$$\int_0^\infty (1-e^{-tx})\frac{\sin(t)}{t}dt = \lim_{b\to\infty}\int_0^b (1-e^{-tx})\frac{\sin(t)}{t}dt = \sum_{k=0}^\infty (-1)^k L_k(x) \tag{19.701}$$

ainsi que l'assurance que le tout est un nombre réel fini [43].

Majoration pour la série alternée Nous majorons un peu. Pour $x > 0$ et $N \in \mathbb{N}$ nous avons

$$\begin{aligned}
|\sum_{k=0}^N (-1)^k L_k(x)| &\leqslant \sum_{k=0}^N |L_k(x)| && (19.702a)\\
&= \sum_{k=0}^N \Big|\int_{k\pi}^{(k+1)\pi} (1-e^{-tx})\frac{|\sin(t)|}{t}dt\Big| && (19.702b)\\
&\leqslant \sum_{k=0}^N \int_{k\pi}^{(N+1)\pi} (1-e^{-tx})\frac{|\sin(t)|}{t}dt && (19.702c)\\
&= \int_0^{(N+1)\pi} (1-e^{-tx})\frac{|\sin(t)|}{t}dt && (19.702d)\\
&\leqslant \int_0^{(N+1)\pi} tx\frac{|\sin(t)|}{t}dt && (19.702e)\\
&\leqslant \int_0^{(N+1)\pi} xdt && (19.702f)\\
&= x(N+1)\pi. && (19.702g)
\end{aligned}$$

Justifications :

— Pour (19.702b) c'est la définition (19.691).

— Pour (19.702e), c'est le fait que $0 \leqslant 1-e^{-u} \leqslant u$ pour tout $u \geqslant 0$ ainsi que la sous-additivité de l'intégrale de la proposition 13.174.

43. De toutes façons, il n'existe pas de nombres réels infinis, mais vous voyez ce que je veux dire.

Majoration pour l'intégrale Nous fixons $N \in \mathbb{N}$, et nous avons :

$$\begin{aligned}
\Big|\lim_{b\to\infty}\int_0^b (1-e^{-tx})\frac{\sin(t)}{t}dt\Big| &= \Big|\sum_{k=0}^{\infty}(-1)^k L_k(x)\Big| && (19.703a)\\
&= \Big|\sum_{k=0}^{N}(-1)^k L_k(x) + \sum_{k=N+1}^{\infty}(-1)^k L_k(x)\Big| && (19.703b)\\
&\leqslant \Big|\sum_{k=0}^{N}(-1)^k L_k(x)\Big| + \Big|\sum_{k=N+1}^{\infty}(-1)^k L_k(x)\Big| && (19.703c)\\
&\leqslant \Big|\sum_{k=0}^{N}(-1)^k L_k(x)\Big| + L_{N+1}(x) && (19.703d)\\
&\leqslant \Big|\sum_{k=0}^{N}(-1)^k L_k(x)\Big| + \frac{1}{N+1} && (19.703e)\\
& && (19.703f)
\end{aligned}$$

Justifications :

— Pour (19.703d) c'est le reste du critère des séries alternées, théorème 10.72(3).

— Pour (19.703e) c'est la majoration (19.692) déjà faite.

Les deux ensemble Pour chaque N et pour chaque $x > 0$ nous avons, en mettant (19.702) au bout de (19.703) :

$$\Big|\int_0^{\infty}(1-e^{-tx})\frac{\sin(t)}{t}dt\Big| \leqslant x(N+1)\pi + \frac{1}{N+1}. \tag{19.704}$$

En prenant la limite $x \to 0$ nous trouvons

$$\lim_{x\to 0}\int_0^{\infty}(1-e^{-tx})\frac{\sin(t)}{t}dt \leqslant \frac{1}{N+1} \tag{19.705}$$

pour tout N. Donc cette limite est nulle :

$$\lim_{x\to 0}\int_0^{\infty}(1-e^{-tx})\frac{\sin(t)}{t}dt = 0. \tag{19.706}$$

□

Maintenant que nous avons fait plein de lemmes, nous pouvons énoncer notre résultat principal, et le démontrer facilement.

Théorème 19.215 (Intégrale de Dirichlet[405])**.**
Nous avons

$$\lim_{b\to\infty}\int_0^b \frac{\sin(t)}{t}dt = \frac{\pi}{2}. \tag{19.707}$$

Nous avons écrit $\lim_{b\to\infty}\int_0^b \frac{\sin(t)}{t}dt$ et non $\int_0^{\infty}\frac{\sin(t)}{t}dt$ parce que cette dernière intégrale n'existe pas vraiment au sens de Lebesgue, voir le lemme 19.207. Dans la suite nous écrirons cependant $\int_0^{\infty}\frac{\sin(t)}{t}dt$, en gardant en tête que cela n'est défini que via la limite.

Démonstration. Nous nommons D la valeur que nous cherchons. Le lemme 19.208 nous assure que

$$D = \lim_{b\to\infty}\int_0^b \frac{\sin(t)}{t} < \infty. \tag{19.708}$$

Le lemme 19.211 nous donne, quant à lui,

$$\lim_{b\to\infty}\int_0^b e^{-tx}\frac{\sin(t)}{t}dt = -\arctan(x) + \frac{\pi}{2}. \tag{19.709}$$

Vu que les deux limites existent, on peut permuter somme et limite [44] :

$$D + \arctan(x) - \frac{\pi}{2} = \lim_{b\to\infty} \Big(\int_0^b \frac{\sin(t)}{t} + \int_0^b e^{-tx} \frac{\sin(t)}{t} dt \Big) \tag{19.710a}$$

$$= \lim_{b\to\infty} \int_0^b (1 - e^{-xt}) \frac{\sin(t)}{t} dt. \tag{19.710b}$$

Pour (19.710b), nous avons des fonctions bornées sur un intervalle borné $(]0, b[)$, donc il n'y a pas de mal à sommer les intégrales.

Donc pour tout $x > 0$, nous avons

$$D + \arctan(x) - \frac{\pi}{2} = \int_0^\infty (1 - e^{-xt}) \frac{\sin(t)}{t} dt. \tag{19.711}$$

Nous passons à la limite $x \to 0$ en utilisant le lemme 19.214 et le fait que $\arctan(0) = 0$ (lemme 17.33) :

$$D - \frac{\pi}{2} = 0, \tag{19.712}$$

c'est à dire que résultat annoncé. □

44. C'est une phrase un peu grandiloquente pour dire que $\lim_{b\to a} f(b) - \lim_{b\to a} g(b) = \lim_{b\to a}(f(b) - g(b))$. Ici nous avons $a = \infty$ et les fonctions f et g sont celles définies par les intégrales.

Chapitre 20

Arcs paramétrés

La structure de ce chapitre, comme beaucoup de choses dans le Frido, est fortement liée au choix de présenter toutes les matières dans l'ordre mathématiquement logique. Nous devons donc le placer après la trigonométrie ; les propriétés principales des fonctions trigonométriques étant dans la proposition 17.17, et c'est la proposition 17.46 qui nous permet de dire que $\big(\cos(t), \sin(t)\big)$ décrit le cercle.

Et enfin nous n'avons pas encore calculé la circonférence du cercle, et pour cause : nous n'avons pas encore donné de définition à la longueur d'un chemin dans $\mathbb{R}^2$. C'est pourquoi ce chapitre va aller droit à la longueur avant de donner des exemples.

20.1 Définitions

Définition 20.1.
*Un **arc paramétré** dans $\mathbb{R}^p$ est un couple (I, γ) où I est un intervalle de $\mathbb{R}$ et γ est une application continue de I dans $\mathbb{R}^p$. Nous disons que (I, γ) est un arc paramétré **compact** (ou un **chemin** dans $\mathbb{R}^p$) lorsque I est compact dans $\mathbb{R}$.*

L'intervalle I d'un arc paramétré compact est toujours de la forme $[a, b]$, étant donné que tous les intervalles compacts de $\mathbb{R}$ sont de cette forme. Un **sous arc** de (I, γ) est un arc de la forme (I_0, γ) avec $I_0 \subset I$.

Définition 20.2.
*Un **chemin** dans $\mathbb{R}$ est une application continue*

$$\begin{aligned} \sigma\colon [a, b] &\to \mathbb{R}^3 \\ t &\mapsto \sigma(t). \end{aligned} \tag{20.1}$$

La fonction $\sigma'(t)$ est la **vitesse** du chemin σ. Si la fonction $t \mapsto \sigma(t)$ est dérivable, on dit que $\sigma''(t)$ est l'**accélération**. Les points $\sigma(a)$ et $\sigma(b)$ sont les extrémités du chemin. L'ensemble

$$\{\sigma(t) \text{ tel que } t \in [a, b]\} \tag{20.2}$$

est la **courbe** σ.

20.2 Longueur d'arc

Nous voulons définir et étudier la notion de longueur d'un arc paramétré. Pour cela, le plus raisonnable est d'approcher l'arc par des petits segments de droites (dont les longueurs sont évidentes), et d'extraire la « meilleure » approximation.

Une des notions clefs pour la suite est celle de subdivision d'intervalles. Cette notion sera encore utilisée par la suite à propos des intégrales.

Définition 20.3.
Si I est un intervalle d'extrêmes a et b avec $a < b$, nous appelons ***subdivision finie*** *de I un choix de nombres t_i tels que*

$$a = t_0 < t_1 < \ldots < t_n = b. \tag{20.3}$$

Nous disons qu'une subdivision σ' est ***plus fine*** *que la subdivision σ si l'ensemble des points de σ est inclus dans celui des points de σ'. Dans ce cas, la subdivision σ' est un* ***raffinement*** *de σ. Nous désignons par $\mathcal{S}(I)$ l'ensemble des subdivisions finies de l'intervalle I.*

Dans la suite, toutes les subdivisions que nous considérons seront des subdivisions finies. Aussi nous parlerons simplement de *subdivisions* sans préciser. Nous allons souvent noter $\sigma = (t_i)_{i=1}^n$ pour désigner la subdivision formée par les nombres t_i. Il faut garder en tête que dans une subdivision, les nombres *sont ordonnés*.

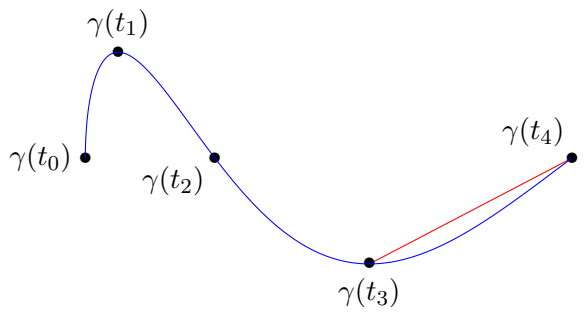


FIGURE 20.1 – La longueur d'un découpage. La somme des longueurs des segments droits est facile à calculer.

Définition 20.4.
Soit un arc paramétré compact (I, γ) et une subdivision $\sigma = (t_i)_{i=1}^n$ de $I = [a, b]$. À partir de γ et du découpage σ nous définissons le nombre (voir figure 20.1)

$$l_\sigma(\gamma) = \sum_{i=1}^n \big\|\gamma(t_i) - \gamma(t_{i-1})\big\|. \tag{20.4}$$

On appelle ***longueur*** *de l'arc γ le nombre*

$$l(\gamma) = \sup_\sigma l_\sigma(\gamma) \in [0, \infty]. \tag{20.5}$$

Nous disons que γ est ***rectifiable*** *lorsque $l(\gamma) < \infty$.*

Lorsque nous voulons spécifier sur quel intervalle nous considérons l'arc, nous noterons $l(I, \gamma)$ au lieu de $l(\gamma)$ pour être plus précis.

Par l'inégalité triangulaire, si σ_1 est plus fine que σ, nous avons

$$l_\sigma(\gamma) \leqslant l_{\sigma_1}(\gamma), \tag{20.6}$$

Comme cela peut être vu sur la figure 20.2.

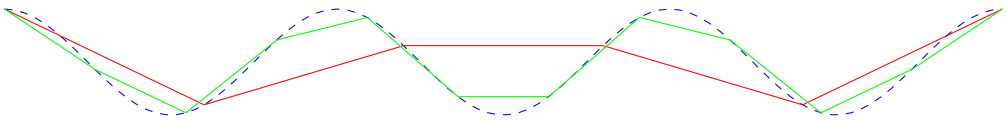


FIGURE 20.2 – Il est visible que la longueur donnée par l'approximation par des petits segments (verts) est plus longue et plus précise que celle donnée par les longs segments (rouge).

Proposition 20.5.
Si P et Q sont des points de $\mathbb{R}^2$, alors le segment de droite joignant P à Q est le plus court des arcs paramétrés passant par P et Q.

Démonstration. Si γ est un arc paramétré joignant P et Q, la longueur de γ est donné par un supremum dont un des éléments est la longueur du segment de droite. □

Dans la vie réelle, il est souvent difficile et peu pratique de calculer le supremum « à la main ». C'est pourquoi nous allons travailler à exprimer la longueur d'un arc à l'aide d'une intégrale (théorème 20.10).

Lemme 20.6.
Nous avons $l(\gamma) = 0$ si et seulement si $\gamma(t)$ est un vecteur constant.

Démonstration. Si l'application $\gamma(t)$ est constante, le résultat est évident. Supposons maintenant que γ ne soit pas constante. Cela signifie qu'il existe t_1 et t_2 dans I tels que $\gamma(t_1) \neq \gamma(t_2)$. Dans ce cas, si nous prenons le découpage $\sigma = \{a, t_1, t_2, b\}$, la somme (20.4) contient au moins le terme non nul $\|\gamma(t_2) - \gamma(t_1)\|$, et donc $l_\sigma(\gamma) > 0$. Par définition du supremum, nous avons alors $l(\gamma) \geqslant l_\sigma(\gamma) > 0$. □

Proposition 20.7.
Soit (I, γ) un arc paramétré compact.

(1) Si $\gamma' = (I', \gamma)$ avec $I' \subset I$, alors $l(\gamma') \leqslant l(\gamma)$.

(2) Soit $c \in [a, b]$, et considérons les arcs $\gamma_1 = \big([a, c], \gamma\big)$ et $\gamma_2 = \big([c, b], \gamma\big)$. Alors

$$l(\gamma) = l(\gamma_1) + l(\gamma_2). \tag{20.7}$$

En particulier, γ est rectifiable si et seulement si γ_1 et γ_2 le sont.

Démonstration. (1) Nous notons $I = [a, b]$ et $I' = [a', b']$. Étant donné que $I' \subset I$, nous avons

$$a \leqslant a' < b' \leqslant b. \tag{20.8}$$

Pour chaque subdivision $\sigma_0 : a' = t_0 < t_1 < \ldots < t_n = b'$ de I', nous pouvons construire une subdivision de I en « ajoutant » les points a et b, c'est-à-dire

$$\sigma : a \leqslant t_0 < \ldots < t_n \leqslant b. \tag{20.9}$$

Si nous calculons $l_\sigma(\gamma)$, nous avons tous les termes qui arrivent dans $l_{\sigma_0}(\gamma')$ plus le premier et dernier terme : $\|\gamma(t_0) - \gamma(a)\|$ et $\|\gamma(b) - \gamma(t_n)\|$. Nous avons donc

$$l_{\sigma_0}(\gamma') \leqslant l_\sigma(\gamma) \leqslant \sup_\sigma l_\sigma(\gamma) = l(\gamma). \tag{20.10}$$

Étant donné que pour toute subdivision σ_0 nous avons $l_{\sigma_0}(\gamma') \leqslant l(\gamma)$, en prenant le supremum sur les subdivisions σ_0 de I', nous avons comme annoncé

$$l(\gamma') \leqslant l(\gamma). \tag{20.11}$$

(2) Soit $\sigma = \{t_i\}$ une subdivision de $[a, b]$. Nous considérons les subdivisions σ_1 et σ_2 définies comme suit :

$$\begin{aligned} \sigma_1 &: \{t_i \text{ tel que } t_i < c\} \cup \{c\}, \\ \sigma_2 &: \{t_i \text{ tel que } t_i > c\} \cup \{c\}. \end{aligned} \tag{20.12}$$

L'inégalité triangulaire implique que

$$l_\sigma(\gamma) \leqslant l_{\sigma \cup \{c\}}(\gamma) = l_{\sigma_1}(\gamma_1) + l_{\sigma_2}(\gamma_2) \leqslant l(\gamma_1) + l(\gamma_2). \tag{20.13}$$

Nous avons donc

$$l(\gamma) \leqslant l(\gamma_1) + l(\gamma_2). \tag{20.14}$$

Nous prouvons maintenant l'inégalité inverse. Soit $\varepsilon > 0$. Étant donné que $l(\gamma_1)$ est le supremum des quantités $l_{\sigma_1}(\gamma_1)$ lorsque σ_1 parcours toutes les subdivisions possibles, il existe une partition σ_1^ε telle que (idem pour γ_2)

$$\begin{aligned} l_{\sigma_1^\varepsilon}(\gamma_1) + \frac{\varepsilon}{2} > l(\gamma_1), \\ l_{\sigma_2^\varepsilon}(\gamma_2) + \frac{\varepsilon}{2} > l(\gamma_2), \end{aligned} \tag{20.15}$$

où σ_1^ε est une subdivision de $[a, c]$ et σ_2^ε en est une de $[c, b]$. En faisant la somme des deux équations (20.15), nous trouvons

$$l(\gamma_1) + l(\gamma_2) < l_{\sigma_1^\varepsilon}(\gamma_1) + l_{\sigma_2^\varepsilon}(\gamma_2) + \varepsilon = l_{\sigma_1^\varepsilon \cup \sigma_2^\varepsilon}(\gamma) \leqslant l(\gamma) + \varepsilon. \tag{20.16}$$

L'inégalité $l(\gamma_1) + l(\gamma_2) < l(\gamma) + \varepsilon$ étant valable pour tout ε, nous avons

$$l(\gamma_1) + l(\gamma_2) \leqslant l(\gamma). \tag{20.17}$$

Cette inégalité, combinée avec l'inégalité (20.14), donne bien $l(\gamma) = l(\gamma_1) + l(\gamma_2)$.

□

20.3 Abscisse curviligne

Définition 20.8.
Soit (I, γ) un arc rectifiable compact avec $I = [a, b]$. L'application

$$\begin{aligned} \varphi\colon [a, b] &\to \mathbb{R}^+ \\ t &\mapsto l\big([a, t], \gamma\big) \end{aligned} \tag{20.18}$$

*est la **longueur d'arc** de γ.*

Cette fonction nous permet de calculer la distance (suivant la courbe) entre deux points arbitraires parce que si $a \leqslant t < u \leqslant b$, nous avons

$$l\big([t, u], \gamma\big) = \varphi(u) - \varphi(t). \tag{20.19}$$

En effet,

$$\varphi(u) - \varphi(t) = l\big([a, u], \gamma\big) - l\big([a, t], \gamma\big), \tag{20.20}$$

mais en utilisant la proposition 20.7, nous avons

$$l\big([a, u], \gamma\big) = l\big([a, t], \gamma\big) + l\big([t, u], \gamma\big). \tag{20.21}$$

Proposition 20.9.
La longueur d'arc d'un arc rectifiable compact est une fonction continue et croissante.

Démonstration. Soit (I, γ) un arc paramétré rectifiable compact avec $I = [a, b]$. Afin de montrer que φ est croissante, prenons $t \in I$ ainsi que $h > 0$ et montrons que $\varphi(t + h) \geqslant \varphi(t)$. La proposition 20.7 implique que

$$l\big([a, t + h], \gamma\big) = l\big([a, t], \gamma\big) + l\big([t, t + h], \gamma\big), \tag{20.22}$$

c'est-à-dire

$$\varphi(t + h) = \varphi(t) + l\big([t, t + h], \gamma\big) \geqslant \varphi(t). \tag{20.23}$$

Pour la continuité, soit t fixé dans $[a, b]$ et $\varepsilon > 0$. Il nous faut démontrer qu'il existe $\eta > 0$ tel que si s est dans $[0, \eta]$ alors

$$|\varphi(t + s) - \varphi(t)| \leqslant \varepsilon, \qquad \forall t \in [a, b].$$

Étant donné que $l\big([t,b],\gamma\big)$ est le supremum des $l_\sigma\big([t,b],\gamma\big)$, il existe une subdivision σ donnée par les points $t, t_1, \cdots, t_{n-1}, b$ telle que

$$l_\sigma\big([t,b],\gamma\big) > l\big([t,b],\gamma\big) - \frac{\varepsilon}{2} = \varphi(b) - \varphi(t) - \frac{\varepsilon}{2}. \tag{20.24}$$

La continuité de γ implique qu'il existe un η tel que

$$s \in [0,\eta] \Rightarrow \|\gamma(t+s) - \gamma(t)\| < \frac{\varepsilon}{2} \tag{20.25}$$

Quitte à prendre η encore plus petit, nous supposons que $t + \eta < t_1$. Soit $s \in [0,\eta]$ et considérons la subdivision de $[t,b]$ donnée par $\sigma' = \sigma \cup \{t+s\}$. Étant donné que σ' est plus fine que σ, le nombre $l_\sigma\big([t,b],\gamma\big)$ est inférieur ou égal à $l_{\sigma'}\big([t,b],\gamma\big)$. Nous avons donc les inégalités

$$\begin{aligned}
\varphi(b) - \varphi(t) - \frac{\varepsilon}{2} &\leqslant l_\sigma\big([t,b],\gamma\big) \\
&\leqslant l_{\sigma'}\big([t,b],\gamma\big) \\
&= \big\|\gamma(t+s) - \gamma(t)\big\| + l_{\sigma'\setminus\{t\}}\big([t+s,b]\gamma\big) \\
&\leqslant \|\gamma(t+s) - \gamma(t)\| + \varphi(b) - \varphi(t+s) \\
&\leqslant \frac{\varepsilon}{2} + \varphi(b) - \varphi(t+s).
\end{aligned} \tag{20.26}$$

Au final, nous avons trouvé que

$$\varphi(t+s) - \varphi(t) \leqslant \varepsilon, \tag{20.27}$$

ce qui prouve que φ est continue au point t. □

En guise de paramètre sur un arc, nous pouvons utiliser la longueur d'arc elle-même. En effet si (I,γ) est un arc de longueur l, nous pouvons donner le même arc avec le couple $\big([0,l],g\big)$ où g est la fonction qui au réel s fait correspondre l'élément $\gamma\big(\varphi^{-1}(s)\big)$ de $\mathbb{R}^n$. Dire

$$P = (\gamma \circ \varphi^{-1})(s) \tag{20.28}$$

revient à dire que le point P est le point sur la courbe sur lequel on tombe après avoir marché une distance s sur la courbe.

Nous allons revenir sur ce « changement de paramètre » plus tard, en particulier dans la section 20.7.

20.3.1 Formule intégrale de la longueur

Nous pouvons voir un chemin γ comme étant la trajectoire d'une particule en fonction du temps. Sa vitesse à l'instant t est le vecteur $\gamma'(t)$, tandis que sa vitesse *scalaire* est le nombre $\|\gamma'(t)\|$. Une question naturelle est de savoir quelle est la longueur de la trajectoire parcourue entre $t = a$ et $t = b$.

Si nous prenons un petit intervalle de temps dt, nous pouvons supposer que le mobile avance à la vitesse constante $\|\gamma'(t)\|$. Cela ferait un trajet parcouru de longueur $\|\gamma'(t)\|dt$. Nous nous attendons donc à une formule de la forme suivante pour la longueur de γ :

$$l(\gamma) = \int_a^b \|\gamma'(t)\| dt. \tag{20.29}$$

Plus explicitement, si $\gamma(t) = \big(x(t), y(t), z(t)\big)$, alors nous aurions la formule

$$l(\gamma) = \int_a^b \sqrt{x'(t)^2 + y'(t)^2 + z'(t)^2} dt. \tag{20.30}$$

Théorème 20.10.
Soit (I, γ) un arc paramétré compact de classe $\mathcal{C}^1$. Alors γ est rectifiable et

$$l(\gamma) = \int_a^b \|\gamma'(t)\| dt = \int_\gamma 1, \tag{20.31}$$

où $I = [a, b]$.

Démonstration. L'égalité avec l'intégrale le long de γ de la fonction 1 est simplement la définition 19.40 de l'intégrale curviligne.

Si $\sigma = \{t_i\}$ est une subdivision de l'intervalle $[a, b]$, alors

$$\begin{aligned} l_\sigma(\gamma) &= \sum_{i=1}^n \|\gamma(t_i) - \gamma(t_{i-1})\| \\ &= \sum_{i=1}^n \| \int_{t_{i-1}}^{t_i} \gamma'(t) dt\| \\ &\leqslant \sum_{i=1}^n \int_{t_{i-1}}^{t_i} \|\gamma'(t)\| dt \\ &= \int_a^b \|\gamma'(t)\| dt. \end{aligned} \tag{20.32}$$

Cela prouve déjà que

$$l(\gamma) = \sup_\sigma l_\sigma(\gamma) \leqslant \int_a^b \|\gamma'(t)\| dt. \tag{20.33}$$

Nous devons maintenant prouver l'inégalité inverse.

Notons φ l'abscisse curviligne $\varphi(t) = l\big([a, t], \gamma\big)$. Cette dernière vérifie

$$\varphi(t + h) - \varphi(t) = l\big([t, t + h], \gamma\big) \geqslant \|\gamma(t + h) - \gamma(t)\|, \tag{20.34}$$

et en particulier

$$\left\| \frac{\gamma(t + h) - \gamma(t)}{h} \right\| \leqslant \frac{\varphi(t + h) - \varphi(t)}{h}. \tag{20.35}$$

D'autre part, en utilisant (20.33) sur le segment $[t, t + h]$, nous avons

$$\varphi(t + h) - \varphi(t) = l\big([t, t + h], \gamma\big) \leqslant \int_t^{t+h} \|\gamma'(u)\| du. \tag{20.36}$$

Cela nous permet de continuer l'inéquation (20.35) en

$$\left\| \frac{\gamma(t + h) - \gamma(t)}{h} \right\| \leqslant \frac{\varphi(t + h) - \varphi(t)}{h} \leqslant \frac{1}{h} \int_t^{t+h} \|\gamma'(u)\| du. \tag{20.37}$$

Prenons la limite $h \to 0$. À gauche nous reconnaissons la formule de la dérivée, et nous obtenons $\|\gamma'(t)\|$; au centre nous avons $\varphi'(t)$ et à droite, si $n(u)$ représente une primitive de la fonction $u \mapsto \|\gamma'(u)\|$,

$$\lim_{h \to 0} \frac{n(t + h) - n(t)}{h} = n'(t) = \|\gamma'(t)\|. \tag{20.38}$$

Au final,

$$\|\gamma'(t)\| \leqslant \varphi'(t) \leqslant \|\gamma'(t)\|, \tag{20.39}$$

c'est-à-dire $\varphi'(t) = \|\gamma'(t)\|$ et donc par le théorème fondamental du calcul intégral 13.236,

$$\varphi(t) - \varphi(a) = \int_a^t \|\gamma'(u)\| du. \tag{20.40}$$

Par construction de la longueur d'arc, $\varphi(a) = 0$ et en posant $t = b$ nous obtenons la relation recherchée :

$$l(\gamma) = \varphi(b) = \int_a^b \|\gamma'(u)\| du. \tag{20.41}$$

□

Remarque 20.11.
Cela est cohérent avec 19.181, mais il faut garder en tête que $l(\gamma)$ n'est pas le mesure de Lebesgue de l'image de γ dans $\mathbb{R}^2$. Cette dernière est nulle.

Exemple 20.12.
Soient donc a et b deux points de $\mathbb{R}^m$, et γ la droite joignant a à b, c'est-à-dire

$$\gamma(t) = (1-t)a + tb \tag{20.42}$$

avec $t \in [0,1]$. Le théorème 20.10 nous enseigne que la longueur de ce chemin est

$$l\big([0,1],\gamma\big) = \int_0^1 \|\gamma'(t)\| dt = \int_0^1 \|-a+b\| = \|b-a\|, \tag{20.43}$$

qui est bien la distance entre a et b. △

Exemple 20.13 (Circonférence du cercle)**.**
Nous savons que l'image de

$$\begin{aligned} \gamma\colon [0,2\pi[&\to \mathbb{R}^2 \\ t &\mapsto \big(R\cos(t), R\sin(t)\big) \end{aligned} \tag{20.44}$$

est le cercle de centre $(0,0)$ et de rayon $R > 0$. Et de plus cet arc est de classe C^1 (et même C^∞) par la proposition 17.1. La longueur sera, d'après la formule (20.10)

$$l_\gamma = \int_0^{2\pi} \|\gamma'(t)\| dt = 2\pi R \tag{20.45}$$

grâce à la formule $\sin^2 + \cos^2 = 1$ du lemme 17.3.

Mais tout cela n'est pas satisfaisant parce que nous n'avons pas encore de valeur numérique de π.

Il y a une autre façon de faire en considérant le quart de cercle dont la longueur en fonction de π est vite calculée par

$$l_\gamma = \int_0^{pi/2} \|\gamma'(t)\| dt = \frac{\pi R}{2}. \tag{20.46}$$

Cette même longueur est calculée en termes de fonctions plus courantes avec le chemin

$$\begin{aligned} \sigma\colon]0,1[&\to \mathbb{R}^2 \\ t &\mapsto \left(\frac{t}{\sqrt{R^2-t^2}}\right) \end{aligned} \tag{20.47}$$

La longueur s'exprime avec

$$l_\sigma = \int_0^1 \sqrt{\frac{R^2}{R^2+t^2}} dt. \tag{20.48}$$

Notons que le changement de variables $t = R\sin(u)$ permet de retrouver l'expression $l_\sigma = \pi R/2$.

Pour avoir une approximation de π, il est loisible de calculer une approximation numérique de l'intégrale (20.48) (avec $R = 1$) et de l'égaler à $\pi/2$. △

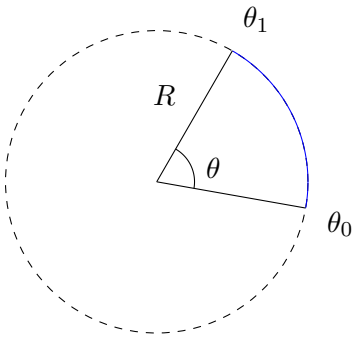


FIGURE 20.3 – Quelle est la longueur de la partie bleue de ce cercle de rayon R ?

Exemple 20.14.
Considérons l'arc de cercle de rayon R interceptée par l'angle θ présenté sur la figure 20.3.

Par définition, cette longueur sera

$$\int_{\theta_0}^{\theta_1} \sqrt{R^2 \sin^2(t) + R^2 \cos^2(t)} dt = R(\theta_1 - \theta_0). \tag{20.49}$$

Le radian comme unité de mesure d'angle est donc l'unité parfaite : elle est la longueur d'arc interceptée (si le rayon est $R = 1$). △

Une conséquence à peine indirecte de ce que nous venons de voir à propos de longueur d'arc de cercle est la proposition suivante[1].

Proposition 20.15.
Pour tout $x, y \in \mathbb{R}$, nous avons

$$|e^{ix} - e^{iy}| \leqslant |x - y|. \tag{20.50}$$

Démonstration. Évacuons tout de suite la différence entre $\mathbb{R}^2$ et $\mathbb{C}$: ils sont isométriques. Si vous n'êtes pas convaincu que tout se passe bien, vous pouvez récrire toute la démonstration en écrivant systématiquement $\big(\cos(x), \sin(x)\big)$ au lieu de e^{ix}. Cela serait au passage un bon exercice pour voir que les formules de dérivation fonctionnent bien.

Nous considérons les points e^{ix} et e^{iy} dans $\mathbb{C}$ et deux chemins différents les joignant. Le premier est le segment de droite

$$\begin{aligned} \sigma_1 \colon [0,1] &\to \mathbb{C} \\ t &\mapsto te^{ix} + (1-t)e^{iy}. \end{aligned} \tag{20.51}$$

Le second est l'arc de cercle

$$\begin{aligned} \sigma_1 \colon [0,1] &\to \mathbb{C} \\ t &\mapsto e^{i\left(tx+(1-t)y\right)}. \end{aligned} \tag{20.52}$$

Nous avons $\sigma_1'(t) = e^{ix} - e^{iy}$ qui ne dépend pas de t, et donc la longueur est facile à calculer à partir de la formule intégrale du théorème 20.10 :

$$l(\sigma_1) = \int_0^1 |\sigma_1'(t)| = |e^{ix} - e^{iy}|. \tag{20.53}$$

En ce qui concerne le second chemin,

$$\sigma_2'(t) = (x-y)e^{i\left(tx+(1-t)y\right)}. \tag{20.54}$$

Nous avons[2] $|\sigma_2'(t)| = |x - y|$ qui ne dépend pas non plus de t. Donc

$$l(\sigma_2) = |x - y|. \tag{20.55}$$

1. À mon avis il y a moyen de prouver ça avec un développement limité, mais je ne sais pas trop comment majorer l'erreur sans accepter que x soit arbitrairement proche de y. Si vous savez comment faire, écrivez-moi.
2. Si vous voulez citer des résultats, lemme 17.10 et proposition 1.167.

Étant donné la proposition 20.5 qui dit que le chemin le plus court est le segment de droite,

$$l(\sigma_1) < l(\sigma_2) \tag{20.56}$$

et donc le résultat annoncé. □

20.16.
Si on veut savoir la longueur d'une courbe donnée sous la forme d'une fonction $y = y(x)$, un chemin qui trace la courbe est évidemment donné par

$$\gamma(t) = (t, y(t)), \tag{20.57}$$

et le vecteur tangent au chemin est $\gamma'(t) = (1, y'(t))$. Donc

$$\|\gamma'(t)\| = \sqrt{1 + y'(t)^2}, \tag{20.58}$$

et

$$L = \int_a^b \sqrt{1 + y'(t)^2}. \tag{20.59}$$

Exemple 20.17.
La longueur de l'hélice

$$\sigma(t) = \begin{pmatrix} \cos(2t) \\ \sin(2t) \\ \sqrt{5}t \end{pmatrix} \tag{20.60}$$

pour $t \in [0, 2\pi]$ est donnée par

$$l(\sigma) = \int_0^{4\pi} \sqrt{4\sin^2(2t) + 4\cos^2(2t) + 5}dt = \int_0^{4\pi} \sqrt{9} = 12\pi. \tag{20.61}$$

△

Définition 20.18.
*Soit $\sigma_1 \colon [a, b] \to \mathbb{R}^3$, un chemin et $\sigma_2 \colon [c, d] \to \mathbb{R}^3$, un autre chemin. On dit que ces chemins sont **équivalents** s'il existe une fonction $\varphi \colon [a, b] \to [c, d]$ strictement croissante telle que $\sigma_1(t) = \sigma_2\big(\varphi(t)\big)$.*

Deux chemins équivalents parcourent la même courbe dans le même sens. Ils ne le parcourent toutefois pas à la même vitesse. On dit que les chemins sont **opposée** si la fonction φ de la définition est strictement décroissante. Dans ce cas, ils ont la même image, mais parcourue dans le sens opposés. Nous disons que deux chemins équivalents sont un **changement de paramétrage** pour la même courbe.

Dans le cas d'un paramétrage équivalente, nous avons $\varphi(a) = c$ et $\varphi(b) = d$. Les points de départ et d'arrivée des deux paramètres coïncident. Dans le cas d'un paramètre qui va dans le sens opposé par contre nous avons automatiquement $\varphi(a) = d$ et $\varphi(b) = c$.

Proposition 20.19.
La longueur d'une courbe ne dépend pas du paramètre (équivalent ou opposé) choisi.

Démonstration. Soient $\sigma_1 \colon [a, b] \to \mathbb{R}^3$ et $\sigma_2 \colon [c, d] \to \mathbb{R}^3$ tels que

$$\sigma_1(t) = \sigma_2\big(\varphi(t)\big) \tag{20.62}$$

où $\varphi \colon [a, b] \to [a, d]$ est une bijection strictement monotone. Par définition on a

$$l(\sigma_1) = \int_a^b \|\sigma_1'(t)\|dt. \tag{20.63}$$

Nous pouvons exprimer la dérivée de σ_1 en termes de celle de σ_2 en dérivant la relation (20.62) :

$$\sigma_1'(t) = \varphi'(t)\sigma_2'\big(\varphi(t)\big). \tag{20.64}$$

En ce qui concerne la norme,

$$\|\sigma_1'(t)\| = |\varphi'(t)|\|\sigma_2'(t)\|. \tag{20.65}$$

Notez dans cette relation que $\varphi'(t)$ est un nombre (et non un vecteur). Étant donné que nous avons supposé que φ était monotone, soit elle est monotone croissante et $\|\varphi'(t)\| = \varphi'(t)$ pour tout t, soit elle est monotone décroissante et $\|\varphi'(t)\| =' \varphi(t)$ pour tout t.

Considérons d'abord le premier cas, c'est-à-dire $\|\varphi'(t)\| = \varphi'(t)$. Nous posons $s = \varphi(t)$, $ds = \varphi'(t)dt$. En remplaçant cela dans la formule de la longueur est

$$\begin{aligned} l(\sigma_1) &= \int_a^b \varphi'(t)\|\sigma_2\big(\varphi(t)\big)\|dt \\ &= \int_{\varphi(a)}^{\varphi(b)} \|\sigma_2'(s)\|ds \\ &= \int_c^d \|\sigma_2'(s)\|ds \\ &= l(\sigma_2). \end{aligned} \tag{20.66}$$

Si nous considérons maintenant un paramétrage strictement décroissante. Dans ce cas, $\varphi'(t) \leqslant 0$ et $\|\varphi'(t)\| = -\varphi'(t)$. Nous posons encore une fois $s = \varphi(t)$, $ds = \varphi'(t)ds$. Ici il ne faut pas oublier que $\varphi(a) = d$ et $\varphi(b) = c$. Le calcul est à part cela le même en faisant attention au singe :

$$\begin{aligned} l(\sigma_1) &= \int_a^b \varphi'(t)\|\sigma_2\big(\varphi(t)\big)\|dt \\ &= -\int_{\varphi(a)}^{\varphi(b)} \|\sigma_2'(s)\|ds \\ &= -\int_d^c \|\sigma_2'(s)\|ds \\ &= \int_c^d \|\sigma_2'(s)\|ds \\ &= l(\sigma_2). \end{aligned} \tag{20.67}$$

Nous avons changé le signe en changeant l'ordre des bornes. □

20.4 Suite du chapitre

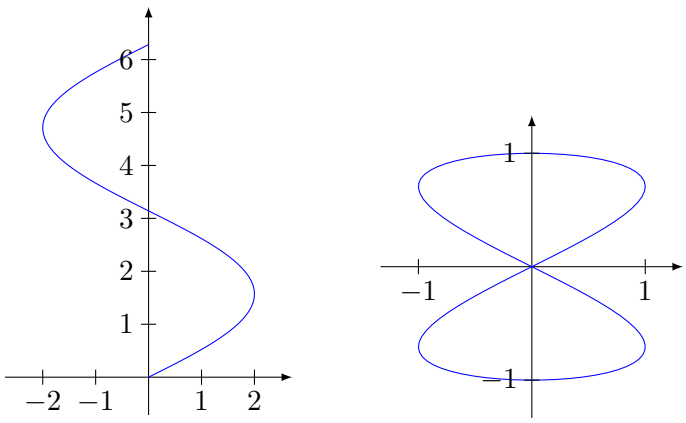
Le grand avantage des arcs paramétrés par rapports aux graphes de fonctions est le le graphe peut « faire des retours en arrière », ou bien des auto intersections. Outre les deux exemples typiques de la la figure 20.4, un exemple classique est la droite verticale. Les fonctions $y = ax + b$ permettent de décrire toutes les droites, sauf les droites verticales. Dans le cadre des courbes paramétrées, les droites verticales et horizontales sont sur pied d'égalité. Quelques exemples classiques :

Droite horizontale Une droite horizontale à la hauteur a est donnée par la courbe paramétrée $\gamma(t) = (t, a)$, avec $t \in I = \mathbb{R}$.

Droite verticale Une droite verticale à la distance b de l'origine est donnée par la courbe paramétrée $\gamma(t) = (b, t)$, avec $t \in I = \mathbb{R}$.

Graphe d'une fonction Le graphe d'une fonction $f\colon \mathbb{R} \to \mathbb{R}$ est donné par l'arc $\gamma(t) = \big(t, f(t)\big)$.

Un cercle Le cercle de rayon R est donné par l'arc $\gamma(t) = \big(R\cos(t), R\sin(t)\big)$.



(a) $\gamma(t) = (2\sin(t), t)$, avec $0 \leqslant t \leqslant 2\pi$ (b) $\gamma(t) = (sin(2t), sin(t))$ avec $-\pi \leqslant t \leqslant \pi$

FIGURE 20.4 – Des exemples d'arcs paramétrées. Ceux ne sont pas des graphes.

Remarque 20.20.
Afin d'alléger la notation, nous allons le plus souvent désigner l'arc (I, γ) simplement par la fonction γ. Il est cependant toujours *très* important de savoir sur quel intervalle nous considérons le chemin. Cela dépendra le plus souvent du contexte, et nous indiquerons l'intervalle I explicitement lorsqu'une ambigüité est à craindre.

Par exemple, lorsque nous considérons le cercle $\gamma(t) = \big(R\cos(t), R\sin(t)\big)$, le plus souvent l'intervalle de variation de t sera $I = [0, 2\pi]$. Par contre, si nous considérons la droite $\gamma(t) = (t, 2t)$, l'intervalle de variation de t sera naturellement $I = \mathbb{R}$.

20.5 Autres exemples

Exemple 20.21.
Soit $v \in \mathbb{R}^3$ et $x_0 \in \mathbb{R}^3$. Le chemin

$$\sigma(t) = x_0 + tv \tag{20.68}$$

est une droite. Sa vitesse est $\sigma'(t) = v$. △

Exemple 20.22.
La courbe

$$\sigma(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix} \in \mathbb{R}^2 \tag{20.69}$$

avec $t \in [0, 2\pi[$ est le cercle unité parcouru une fois dans le sens trigonométrique.

Notez que si on prend $t \in [0, 4\pi[$, nous avons un *autre* chemin ; c'est le même cercle unité, mais parcouru *deux* fois. Même si le « dessin » (le graphe) des deux est le même, le chemin n'est pas le même.

Le chemin

$$\gamma(t) = \begin{pmatrix} \cos(2\pi - t) \\ \sin(2\pi - t) \end{pmatrix} \tag{20.70}$$

est le cercle unité parcouru une fois dans le sens inverse. Encore une fois le « dessin » est le même, mais le chemin n'est pas le même. △

Exemple 20.23.
Le chemin

$$\sigma(t) = \begin{pmatrix} t \\ t^2 \end{pmatrix} \tag{20.71}$$

est un chemin dont l'image est la parabole d'équation $y = x^2$. △

L'importance de la dérivée du chemin réside en le fait qu'elle donne la tangente. En effet le vecteur $\sigma'(t)$ est tangent au graphe de σ au point $\sigma(t)$.

Corolaire 20.24.
La tangente à un cercle est perpendiculaire au rayon.

Démonstration. Nous savons que pour un cercle,

$$y'(x) = \frac{-x}{\sqrt{R^2 - x^2}}. \tag{20.72}$$

Un point général du cercle a pour abscisse $x = R\cos(\theta)$. En remplaçant nous trouvons le coefficient directeur suivant pour la tangente :

$$y'\big(R\cos(\theta)\big) = -\frac{1}{\tan(\theta)}. \tag{20.73}$$

Par conséquent une droite perpendiculaire à la tangente aurait comme coefficient directeur le nombre $\tan(\theta)$. Or cela est bien le coefficient directeur du rayon qui joint le point $(0,0)$ au point $\big(R\cos(\theta), R\sin(\theta)\big)$.

□

Exemple 20.25.
Pour le cercle,

$$\sigma(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}, \tag{20.74}$$

la dérivée est donnée par

$$\sigma'(t) = \begin{pmatrix} -\sin(t) \\ \cos(t). \end{pmatrix} \tag{20.75}$$

Le produit scalaire $\sigma(t) \cdot \sigma'(t)$ est nul. Le vecteur $\sigma'(t)$ est donc bien tangent (corolaire 20.24). △

Exemple 20.26.
Le courbe donnée par le chemin

$$\sigma(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \\ t \end{pmatrix} \tag{20.76}$$

est une hélice. Sa vitesse est

$$\sigma'(t) = \begin{pmatrix} -\sin(t) \\ \cos(t) \\ 1 \end{pmatrix}. \tag{20.77}$$

Notez que pour tout $t \in \mathbb{R}$, nous avons $\|\sigma'(t)\| = \sqrt{2}$. △

Remarque 20.27.
Lorsqu'on parle d'une courbe dans l'espace, l'intervalle sur lequel on considère la variation du paramètre est une donné fondamentale. Elle fait partie intégrante de la définition de la courbe.

20.6 Élément de longueur

20.6.1 Élément de longueur : cartésiennes

Étant donné que la longueur d'arc d'une courbe paramétrée (I, γ) est donnée par l'intégrale de $\|\gamma'(t)\|$, il est naturel d'appeler le nombre $\|\gamma'(t)\|\,dt$, **l'élément de longueur** de la courbe γ au point $\gamma(t)$.

En coordonnées cartésiennes dans le plan, une courbe paramétrée est donnée par

$$\gamma(t) = \big(x_1(t), x_2(t)\big), \tag{20.78}$$

et l'élément de longueur est

$$\|x'(t)\|\,dt = \sqrt{(x_1')^2 + (x_2')^2}\,dt. \tag{20.79}$$

20.6.2 Élément de longueur : polaires (1)

En coordonnées polaires, une courbe est donnée par

$$\gamma(t) = \big(\rho(t), \theta(t)\big), \tag{20.80}$$

et le passage aux cartésiennes se fait via les formules

$$\begin{cases} x(t) = \rho(t)\cos\big(\theta(t)\big) & (20.81a) \\ y(t) = \rho(t)\sin\big(\theta(t)\big). & (20.81b) \end{cases}$$

L'élément de longueur se trouve directement en remplaçant $x(t)$ et $y(t)$ dans la formule (20.79). Les dérivées sont données par

$$\begin{aligned} x'(t) &= \rho'(t)\cos\theta(t) - \rho(t)\theta'(t)\sin\theta(t) \\ y'(t) &= \rho'(t)\sin\theta(t) + \rho(t)\theta'(t)\cos\theta(t), \end{aligned} \tag{20.82}$$

et un calcul montre que

$$\big(x'(t)\big)^2 + \big(y'(t)\big)^2 = \big(\rho'(t)\big)^2 + \big(\rho(t)\big)^2\big(\theta'(t)\big)^2. \tag{20.83}$$

Nous reviendrons plus en détail sur le concept de changement de paramétrage (ici, les polaires) à la section 20.7.

20.6.3 Élément de longueur : polaires (2)

Parfois on utilise θ comme paramètre. L'équation de la courbe est alors donnée en coordonnées polaires sous la forme

$$\rho(\theta) = f(\theta), \tag{20.84}$$

où f est une fonction réelle et il faut comprendre que nous parlons de la courbe $\big(\rho(\theta), \theta\big)$ en coordonnées polaires. En coordonnées cartésiennes, cette courbe est donnée par

$$\begin{cases} x(t) = \rho(t)\cos(t) & (20.85a) \\ y(t) = \rho(t)\sin(t) & (20.85b) \end{cases}$$

avec t qui parcours le plus souvent l'intervalle $[0, 2\pi]$. Notez qu'il se peut que le domaine ne soit pas toujours $[0, 2\pi]$; cela peut dépendre des circonstances. Quoi qu'il en soit, la donnée du domaine fait partie de la donnée d'une courbe, et il ne peut donc pas y avoir d'équivoques à ce niveau.

Nous utilisons à nouveau la formule (20.79) en y mettant les valeurs (20.85) :

$$\begin{cases} x'(t) = \rho'(t)\cos(t) - \rho(t)\sin(t) & (20.86a) \\ y'(t) = \rho'(t)\sin(t) + \rho(t)\cos(t), & (20.86b) \end{cases}$$

et

$$\big(x'(t)\big)^2 + \big(y'(t)\big)^2 = \rho'(t)^2 + \rho(t)^2. \tag{20.87}$$

Remarque 20.28.
N'oubliez pas, en utilisant ces formules, que ce qui rentre dans l'intégrale est la *racine carré* de $(x')^2 + (y')^2$.

Exemple 20.29.
Calculons la circonférence du cercle. En coordonnées polaires, le graphe du cercle correspond à l'équation

$$\big(\rho(t), \theta(t)\big) = (R, t) \tag{20.88}$$

où R est constante (le rayon du cercle) et t va de 0 à 2π. En substituant dans l'équation (20.83), l'élément de longueur à intégrer est seulement

$$\sqrt{R^2} = R \tag{20.89}$$

parce que $\rho'(t) = 0$ et $\theta'(t) = 1$. La longueur du cercle est alors directement donnée par

$$l = \int_0^{2\pi} R dt = 2\pi R. \tag{20.90}$$

Nous pouvions aussi faire le calcul en coordonnées cartésiennes. Alors la courbe est donnée par les équations

$$\begin{aligned} x(t) &= R\cos(t) \\ y(t) &= R\sin(t) \end{aligned} \tag{20.91}$$

et $t \in [0, 2\pi]$. La circonférence du cercle est alors

$$l = \int_0^{2\pi} \sqrt{R^2\sin^2(t) + R^2\cos^2(t)}\, dt = \int_0^{2\pi} R\, dt = 2\pi R. \tag{20.92}$$

△

Remarque 20.30.
Il faut bien comprendre que quand on parle de courbes paramétrées en coordonnées cartésiennes on pense à une courbe dont le paramètre est, par exemple, t et les équations de la courbe sont $(x(t), y(t))$. Cela ne veut pas dire que x ou y soit le paramètre. La cas où x ou y est le paramètre est un cas particulier qui est possible seulement pour certaines courbes et notamment pour les graphes. Le cercle de rayon 1 n'est pas un graphe, donc si on veut utiliser x ou y comme paramètre il faut d'abord découper la courbe en deux morceaux, par exemple, la moitié inférieure ($y < 0$) et la moitié supérieure ($y > 0$).

Exemple 20.31.
Une **cycloïde** est une courbe paramétrée par

$$\begin{cases} x(t) = a(t - \sin(t)) & (20.93a) \\ y(t) = a(1 - \cos(t)) & (20.93b) \end{cases}$$

avec $a > 0$ et $t \in \mathbb{R}$. Comme montré sur la figure 20.5, la cycloïde donne lieu à un graphe périodique. Il est possible de montrer (le faire) que le premier arc correspond à $t \in [0, 2\pi]$. Nous voulons donc calculer la longueur de l'arc sur cet intervalle.

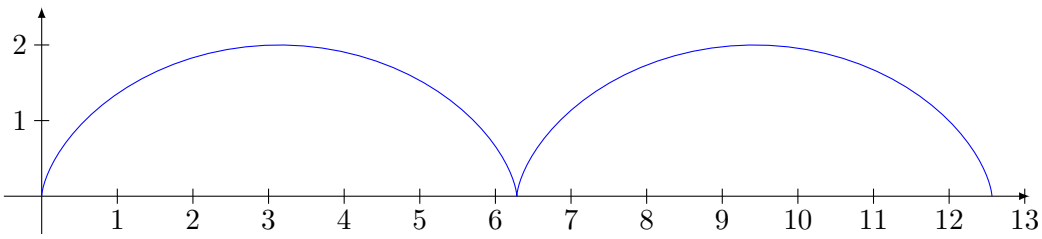


FIGURE 20.5 – La cycloïde de paramètre $a = 1$ entre 0 et 4π.

Nous avons $x'(t) = a(1 - \cos(t))$ et $y'(t) = a\sin(t)$, de telle façon que

$$\sqrt{(x')^2 + (y')^2} = a\sqrt{2 - 2\cos(t)} = a\sqrt{4\sin^2\left(\frac{t}{2}\right)} = 2a\left|\sin\frac{t}{2}\right|. \tag{20.94}$$

La longueur est donc donnée par

$$\int_0^{2\pi} 2a|\sin\frac{t}{2}|dt = 4a\int_0^{\pi} \sin(t)dt = 8a. \tag{20.95}$$

△

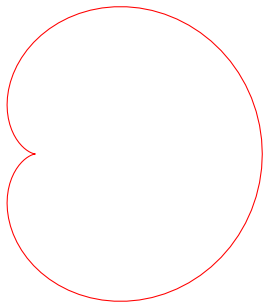


FIGURE 20.6 – Une cardioïde, $\rho = 1 + \cos(\theta)$.

Exemple 20.32.
La **cardioïde** est la courbe donnée par

$$\rho(\theta) = a(1 + \cos(\theta)). \tag{20.96}$$

avec $\theta \in [-\pi, \pi]$. Le nom de cette courbe provient de son graphe illustré à la figure 20.6.

L'équation (20.96) est donnée sous la forme (20.84), c'est-à-dire que $\theta(t) = t$ et $\theta'(t) = 1$, et par conséquent l'élément de longueur est donné par

$$\begin{aligned}(\rho')^2 + (\rho)^2 &= \big(-a\sin(\theta)\big)^2 + a^2\big(1 + \cos(\theta)\big)^2 \\ &= a^2\sin^2(\theta) + a^2\big(1 + 2\cos(\theta) + \cos^2(\theta)\big) \\ &= a^2\big(1 + 1 + 2\cos(\theta)\big) \\ &= 2a^2\big(1 + \cos(\theta)\big) \\ &= 4a^2\cos^2\frac{\theta}{2}.\end{aligned} \tag{20.97}$$

La longueur d'arc est donc donnée par

$$l = \int_{-\pi}^{\pi} 2a\cos\frac{\theta}{2}d\theta = 2a\int_{-\pi/2}^{\pi/2}\cos(t)2dt = 8a. \tag{20.98}$$

△

20.6.4 Approximation de la longueur par des cordes

Définition 20.33.
Soit un arc paramétré (I, γ). Un point $t \in I$ est dit ***régulier*** *si $\gamma'(t) \neq 0$, et il est dit* ***critique*** *si $\gamma'(t) = 0$. Le point $t \in I$ est dit* ***birégulier*** *si les vecteurs $\gamma'(t)$ et $\gamma''(t)$ sont linéairement indépendants et non nuls.*

Par extension, nous dirons également que le point $\gamma(t)$ lui-même est régulier, critique ou birégulier. Un arc est dit régulier *lorsque tous ses points sont réguliers.*

Note : dans le lemme 20.71 et ses dépendances, nous utilisons effectivement que l'arc γ est de classe C^2.

Nous savons que la longueur d'une courbe est donné par le supremum sur toutes les subdivisions de la longueur des cordes correspondantes. De plus l'inégalité triangulaire nous enseigne que plus la subdivision est fine, plus la longueur sera grande. Il est donc naturel de penser que sur un petit intervalle, la longueur de la courbe ne doit pas être très différente de la longueur de la corde correspondante.

La proposition suivante est un énoncé précis et quantitatif de ce fait.

Proposition 20.34.
Soit (I, γ) un arc de classe $\mathcal{C}^1$ et $t_0 \in I$ un point régulier (c'est-à-dire $\gamma'(t_0) \neq 0$). Alors pour tout

$\varepsilon > 0$, il y a un $\delta > 0$ tel que on trouve $t, t' \in I \cap (t_0, \delta)$ tels que

$$\left| \int_t^{t'} \|\gamma'(u)\| du - \|\gamma(t) - \gamma(t')\| \right| \leqslant 2\varepsilon |t - t'|. \tag{20.99}$$

Intuitivement, cette proposition signifie qu'au voisinage de t_0, la longueur d'arc est équivalente à celle de la corde.

Démonstration. Par la continuité de γ' (parce que γ est $\mathcal{C}^1$), pour tout ε, il existe un δ tel que

$$|t - t_0| < \delta \Rightarrow \left\|\gamma'(t) - \gamma'(t_0)\right\| \leqslant \varepsilon. \tag{20.100}$$

Nous considérons la fonction

$$u \mapsto \gamma(u) - \gamma(t_0) - (u - t_0)\gamma'(t_0), \tag{20.101}$$

dont la dérivée (par rapport à u) est

$$\gamma'(u) - \gamma'(t_0). \tag{20.102}$$

Nous y appliquons la formule des accroissements finis entre t et t' choisis dans $I \cap]t_0 - \delta, t_0 + \delta[$. Il existe un u entre t et t' tel que

$$\begin{aligned} \left\|\gamma(t) - \gamma(t_0) - (t - t_0)\gamma'(t_0) - \gamma(t') + \gamma(t_0) + (t' - t_0)\gamma'(t_0)\right\| \\ = |t - t'| \|\gamma'(u) - \gamma'(t_0)\| \\ \leqslant \varepsilon |t - t'|. \end{aligned} \tag{20.103}$$

En simplifiant ce qui peut être simplifié dans le membre de gauche, nous trouvons

$$\left\|\gamma(t) - \gamma(t') - (t - t')\gamma'(t_0)\right\| \leqslant \varepsilon |t - t'|. \tag{20.104}$$

Le membre de gauche peut être minoré en utilisant la proposition 7.114 :

$$\Big| \|\gamma(t) - \gamma(t')\| - \|(t - t')\gamma'(t_0)\| \Big| \leqslant \varepsilon |t - t'|. \tag{20.105}$$

D'autre part, les inégalités (7.84) montrent que

$$-\|\gamma'(u) - \gamma'(t_0)\| \leqslant \|\gamma'(u)\| - \|\gamma'(t_0)\| \leqslant \|\gamma'(u) - \gamma'(t_0)\|. \tag{20.106}$$

Si de plus u est compris entre t et t', ces inégalités sont encore coincées entre $-\varepsilon$ et ε. En intégrant (20.106) par rapport à u entre t et t', nous obtenons

$$\left| \int_t^{t'} \|\gamma'(u)\| - (t - t')\|\gamma'(t_0)\| \right| \leqslant \varepsilon |t - t'|. \tag{20.107}$$

Afin d'alléger les notations pour la ligne suivante, nous notons A le nombre positif $\int_t^{t'} \|\gamma'(u)\| du$. Nous avons

$$\begin{aligned} \Big| A - \|\gamma(t) - \gamma(t')\| \Big| &= \Big| A - |t - t'| \, \|\gamma'(t_0)\| + |t - t'| \, \|\gamma'(t_0)\| - \|\gamma(t) - \gamma(t')\| \Big| \\ &\leqslant \Big| A - |t - t'| \, \|\gamma'(t_0)\| \Big| + \Big| |t - t'| \, \|\gamma'(t_0)\| - \|\gamma(t) - \gamma(t')\| \Big|. \end{aligned} \tag{20.108}$$

L'équation (20.105) montre que le second terme est plus petit ou égal à $\varepsilon |t - t'|$. En ce qui concerne le premier terme, étant donné que A est positif,

$$\Big| A - |t - t'| \, \|\gamma'(t_0)\| \Big| \leqslant \Big| A - (t - t') \, \|\gamma'(t_0)\| \Big| \leqslant \varepsilon |t - t'|. \tag{20.109}$$

Au final, l'inéquation (20.108) donne

$$\Big| A - \|\gamma(t) - \gamma(t')\| \Big| \leqslant 2\varepsilon \, |t - t'|, \tag{20.110}$$

ce qu'il fallait démontrer. □

20.7 Arc géométrique

Définition 20.35.
Soient (I, γ) et (J, g) deux arcs paramétrés de classe $\mathcal{C}^k$. On dit qu'il sont ***équivalents*** *s'il existe une bijection $\theta\colon I \to J$ de classe $\mathcal{C}^k$, d'inverse de classe $\mathcal{C}^k$ telle que $g = \gamma \circ \theta$. Nous notons $\gamma \sim g$ lorsque γ et g sont équivalents (les ensembles I et J sont sous-entendus).*

Le passage d'un paramétrage (I, γ) à une autre (J, g) se fait selon le diagramme suivant :

$$\begin{tikzcd} I \arrow[r, "\gamma"] & \mathbb{R}^n \\ J \arrow[u, "\theta"] \arrow[ur, "g"'] & \end{tikzcd} \tag{20.111}$$

Proposition 20.36.
La relation donnée dans la définition 20.35 est une relation d'équivalence.

Démonstration. Les trois points d'une relation d'équivalence se vérifient en utilisant le fait que θ est inversible, et que l'inverse θ^{-1} jouit des mêmes propriétés de continuité ($\mathcal{C}^k$) que θ.

Réflexivité Nous avons $\gamma \sim \gamma$ avec $\theta = \mathrm{Id}$.

Symétrie Si $\gamma \sim g$, alors nous avons une application θ telle que $g = \gamma \circ \theta$, et donc $\gamma = g \circ \theta^{-1}$, ce qui montre que $g \sim \gamma$.

Transitivité Si $\gamma \sim g$ et $g \sim h$ avec $g = \gamma \circ \theta$ et $h = g \circ \omega$, alors $h = \gamma \circ (\theta \circ \omega)$, ce qui montre que $\gamma \sim h$.

□

Si les arcs (I, γ) et (J, g) sont équivalents, les images dans $\mathbb{R}^n$ sont identiques, et décrivent donc « le même dessin ». Nous allons préciser cette notion plus loin.

Définition 20.37.
Pour cette raison les classes d'équivalences sont appelées des ***arcs géométriques*** *(de classe $\mathcal{C}^k$).*

Si Γ est une arc géométrique, ses représentants sont dits des **paramétrages admissibles** ou, plus simplement *paramétrage*. On dit que l'application $\theta\colon J \to I$ est un **changement de variable**. Nous disons que un arc géométrique est *compact* quand ses représentants sont compacts.

Lemme 20.38.
Dans le cas d'un arc $\mathcal{C}^1$, les changements de variables sont strictement monotones (croissants ou décroissants).

Démonstration. Nous considérons (I, γ) et (J, g), deux paramétrages différents du même arc géométrique, et $\theta \in \mathcal{C}^1(J, I)$ le changement de variable. Nous allons noter t la variable sur I et s la variable sur J. Par définition, $\theta\big(\theta^{-1}(t)\big) = t$, et par conséquent,

$$\theta'\big(\theta^{-1}(t)\big)(\theta^{-1})'(t) = 1. \tag{20.112}$$

En particulier $\theta'\big(\theta^{-1}(t)\big)$ ne s'annule pas pour aucune valeur de t. Mais $\theta^{-1}(t)$ peut prendre n'importe quelle valeur dans J, donc nous avons $\theta'(s) \neq 0$ pour tout $s \in J$. Cela signifie bien que θ est strictement monotone. En effet, θ' étant continue, elle ne peut pas changer de signe sans passer par zéro (théorème 11.50 des valeurs intermédiaires). □

Théorème 20.39.
La longueur d'un arc est indépendante de son paramétrage, c'est-à-dire que les représentants d'un arc géométrique compact de classe $\mathcal{C}^1$ ont même longueur.

Démonstration. Nous utilisons les mêmes notations que celles du lemme 20.38. Nous savons déjà que le changement de variable $\theta\colon J \to I$ est strictement monotone. Supposons que θ soit croissante. En effectuant un changement de variable dans l'intégrale qui donne la longueur[3] nous avons

$$\begin{aligned} l(\gamma) &= \int_I \|\gamma'(t)\| dt \\ &= \int_J \|\gamma'\big(\theta(s)\big)\| \theta'(s) ds \\ &= \int_J \|\gamma'\big(\theta(s)\big)\theta'(s)\| ds \\ &= \int_J \|\frac{d}{ds}(\gamma \circ \theta)(s)\| ds \\ &= \int_J \|g'(s)\| ds \\ &= l(J, g). \end{aligned} \tag{20.113}$$

□

Définition 20.40.
*Nous nommons **longueur** d'un arc géométrique la longueur commune de tous ses représentants. On dit que l'arc géométrique est **rectifiable** si sa longueur est $< \infty$.*

20.7.1 Abscisse curviligne et paramétrage normal

Définition 20.41.
*Soit (I, γ) un arc paramétré continu rectifiable. Nous appelons **abscisse curviligne** de γ toute application $\phi\colon I \to \mathbb{R}$ telle que pour tout $t, t' \in I$ avec $t < t'$, nous ayons*

$$l\big([t, t'], \gamma\big) = \big|\phi(t') - \phi(t)\big|. \tag{20.114}$$

*Si il existe un $t_0 \in I$ tel que $\phi(t_0) = 0$, alors nous disons que t_0 est l'**origine** de l'abscisse ϕ.*

Définition 20.42.
*Un arc paramétré (I, γ_N) continu rectifiable est dit **normal** si l'identité est une abscisse curviligne.*

Lemme 20.43.
Si γ est de classe C^1 et est un paramétrage normal, alors

(1) pour tout choix de t et t' dans I avec $t < t'$, nous avons

$$l\big([t, t'], \gamma_N\big) = t' - t. \tag{20.115}$$

(2) $\|\gamma'(t)\| = 1$ pour tout t.

Démonstration. Pour tout x_1, x_2 dans le domaine nous avons

$$l\big([x_1, x_2], \gamma\big) = \int_{x_1}^{x_2} \|\gamma'(t)\| dt = x_2 - x_1. \tag{20.116}$$

Cela implique $\|\gamma'(t)\| = 1$ pour tout t. En effet, pour fixer les idées, supposons que $\|\gamma'(t)\| > 1$ en un point, par continuité, cela reste strictement supérieur à 1 sur un intervalle. L'intégrale sur cet intervalle ne peut alors pas être la taille de l'intervalle. □

Exemple 20.44.
Le cercle unitaire est donné par l'arc

$$\gamma(t) = \big(\cos(t), \sin(t)\big) \tag{20.117}$$

3. Théorème 20.10.

et $t \in [0, 2\pi]$. Pour tout choix de t et t' dans $[0, 2\pi]$, nous avons

$$l\big([t, t'], \gamma\big) = \int_t^{t'} \sqrt{\sin^2(u) + \cos^2(u)} du = t' - t. \tag{20.118}$$

Les angles exprimés en radians forment donc un paramétrage normal du cercle de rayon 1. △

Lemme 20.45.
Pour un arc paramétré compact, la longueur d'arc est une abscisse curviligne.

Démonstration. Par définition de la longueur d'arc φ, nous avons

$$\varphi(t') - \varphi(t) = l\big([a, t'], \gamma\big) - l\big([a, t], \gamma\big) = \diamondsuit. \tag{20.119}$$

Supposons pour fixer les idées que $t' > t$. En utilisant la proposition 20.7, nous avons

$$l\big([a, t'], \gamma\big) = l\big([a, t], \gamma\big) + l\big([t, t'], \gamma\big), \tag{20.120}$$

et donc après simplification de deux termes,

$$\diamondsuit = l\big([t, t'], \gamma\big), \tag{20.121}$$

ce qui est précisément la propriété demandée pour être une abscisse curviligne. □

Proposition 20.46.
Pour tout arc paramétré C^1 sans points critiques, il existe un changement de coordonnées qui rend l'arc normal.

Démonstration. Soit (I, γ) un arc de classe $\mathcal{C}^1$. Nous devons montrer qu'il existe un intervalle J et une application $\theta\colon J \to I$ de classe $\mathcal{C}^1$ et d'inverse $\mathcal{C}^1$ tel que l'arc (J, γ_N) soit $\mathcal{C}^1$ où $\gamma_N = \gamma \circ \theta$.

Si $I = [a, b]$, nous considérons la fonction

$$\begin{aligned} \phi\colon I &\to \mathbb{R}^+ \\ t &\mapsto \int_a^t \|\gamma'(u)\| du. \end{aligned} \tag{20.122}$$

Étant définie par l'intégrale d'une fonction $\mathcal{C}^0$, la fonction ϕ est $\mathcal{C}^1$, et nous avons $\phi'(t) = \|\gamma'(t)\| > 0$ pour tout $t \in I$. Vue comme application $\phi\colon [a, b] \to [0, l(\gamma)]$, l'application ϕ est bijective et d'inverse $\mathcal{C}^1$. Voyons cela point par point.

(1) La fonction ϕ est injective parce que strictement croissante.

(2) Elle est surjective parce que $\phi(a) = 0$ et $\phi(b) = l(\gamma)$.

(3) La continuité de l'inverse est plus délicate. Soit $l \in [0, l(\gamma)]$ et $\varepsilon > 0$. Pour prouver la continuité de ϕ^{-1} en s, nous devons trouver un δ tel que

$$|s - s'| < \delta \Rightarrow \big|\phi^{-1}(s) - \phi^{-1}(s')\big| < \varepsilon. \tag{20.123}$$

Étant donné que s et s' sont dans l'image de ϕ, nous considérons les uniques t et t' tels que $s = \phi(t)$ et $s' = \phi(t')$. La quantité $\phi(t) - \phi(t')$ devient

$$\int_a^t \|\gamma'(u)\| du - \int_a^{t'} \|\gamma'(u)\| du = \int_t^{t'} \|\gamma'(u)\| du. \tag{20.124}$$

D'autre part, $\phi^{-1}(s) = t$ et $\phi^{-1}(s') = t'$, donc la condition (20.124) devient

$$|\int_{t'}^t \|\gamma'(u)\| du| \leqslant \delta \Rightarrow |t - t'| < \varepsilon. \tag{20.125}$$

Cela revient à la continuité des fonctions définies par une intégrale.

(4) La dérivée de son inverse est donnée par[4]

$$(\phi^{-1})'(s) = \frac{1}{\phi'\big(\phi^{-1}(s)\big)}. \tag{20.126}$$

Nous avons vu que ϕ^{-1} et ϕ' étaient continues. La fonction $(\phi^{-1})'$ étant exprimée en termes de ces deux fonctions elle est également continue.

Nous considérons l'arc paramétré (J, γ_N) avec $J = [0, l(\gamma)]$ et

$$\gamma_N(s) = (\gamma \circ \phi^{-1})(s). \tag{20.127}$$

Nous montrons maintenant que ce nouveau paramétrage est normal. Soient $0 \leqslant s \leqslant s' \leqslant l(\gamma)$,

$$\begin{aligned}
l\big([s, s'], g\big) &= \int_s^{s'} \|\gamma_N'(u)\| du \\
&= \int_{\phi^{-1}(s)}^{\phi^{-1}(s')} \|(\gamma_N' \circ \phi)(t)\| \phi'(t) dt \\
&= \int_{\phi^{-1}(s)}^{\phi^{-1}(s')} \|(\gamma_N \circ \phi)'(t)\| dt \\
&= \int_{\phi^{-1}(s)}^{\phi^{-1}(s')} \|\gamma'(t)\| dt \\
&= \int_0^{\phi^{-1}(s')} \|\gamma'(t)\| \, dt - \int_0^{\phi^{-1}(s)} \|\gamma'(t)\| \, dt \\
&= \phi\big(\phi^{-1}(s')\big) - \phi\big(\phi^{-1}(s)\big) \\
&= s' - s,
\end{aligned} \tag{20.128}$$

ce qui prouve que le paramétrage (J, γ_N) est normale. □

Nous retenons que le paramétrage normal de γ est donnée par (J, γ_N) avec $J = [0, l(\gamma)]$ et

$$\gamma_N(s) = (\gamma \circ \phi^{-1})(s) \tag{20.129}$$

où

$$\begin{aligned}
\phi \colon I &\to \mathbb{R}^+ \\
t &\mapsto \int_a^t \|\gamma'(u)\| du.
\end{aligned} \tag{20.130}$$

Notons aussi que ϕ est une fonction croissante, étant l'intégrale d'une fonction positive.

Exemple 20.47.
Trouvons les coordonnées normales pour la cycloïde donnée par

$$\begin{cases} x(t) = a(t - \sin(t)), & \text{(20.131a)} \\ y(t) = a(1 - \cos(t)) & \text{(20.131b)} \end{cases}$$

et $t \in]0, 2\pi[$. Relire l'exemple 20.31.

D'abord nous trouvons ϕ avec la formule (20.130) avec $a = 0$. En utilisant le bout de calcul (20.94), nous avons

$$\phi(t) = 2a \int_0^t \sin \frac{u}{2} du = 4a \left(1 - \cos \frac{t}{2}\right). \tag{20.132}$$

Pour trouver $\phi^{-1}(s)$, nous résolvons l'équation

$$s = \phi\big(\phi^{-1}(s)\big) \tag{20.133}$$

4. Pour obtenir cette formule, dérivez les deux membres de l'équation $\phi\big(\phi^{-1}(s)\big) = s$.

par rapport à $\phi^{-1}(s)$. Dans un premier temps, nous trouvons

$$1 - \frac{s}{4a} = \cos\frac{\phi^{-1}(s)}{2}, \tag{20.134}$$

donc $\frac{\phi^{-1}(s)}{2} = \arccos(\frac{4a-s}{4a})$, et finalement

$$\phi^{-1}(s) = 2\arccos\left(\frac{4a-s}{4a}\right). \tag{20.135}$$

Il nous reste à injecter cela dans les expressions de $x(t)$ et $y(t)$ pour trouver $(\gamma_N)_x(s)$ et $(\gamma_N)_y(s)$. D'abord,

$$(\gamma_N)_x(s) = a\big[\phi^{-1}(s) - \sin\big(\phi^{-1}(s)\big)\big]. \tag{20.136}$$

Nous utilisons maintenant la formule trigonométrique $\sin(x) = 2\sin\frac{x}{2}\cos\frac{x}{2}$ afin de simplifier les expressions :

$$\begin{aligned}(\gamma_N)_x &= a\Big[2\arccos\left(\frac{4a-s}{4a}\right) - 2\sin\big(\arccos\left(\frac{4a-s}{4a}\right)\big)\cos\big(\arccos\left(\frac{4a-s}{4a}\right)\big)\Big]\\ &= a\Big[2\arccos\left(\frac{4a-s}{4a}\right) - \frac{4a-s}{2a}\sqrt{1-\left(\frac{4a-s}{4a}\right)^2}\Big]\\ &= 2a\arccos\left(\frac{4a-s}{4a}\right) - \sqrt{8as-s^2}\,\frac{4a-s}{8a}\end{aligned} \tag{20.137}$$

où nous avons utilisé la formule $\sin\big(\arccos(x)\big) = \sqrt{1-x^2}$. Ensuite, pour obtenir $(\gamma_N)_y$ nous devons calculer

$$(\gamma_N)_y(s) = a\big[1 - \cos\big(\phi^{-1}(s)\big)\big]. \tag{20.138}$$

Encore une fois, il est intéressant d'exprimer le cosinus en termes des angles divisés par deux : $\cos(x) = \cos^2\frac{x}{2} - \sin^2\frac{x}{2}$.

$$\begin{aligned}(\gamma_N)_y &= a\Big[1 - \cos^2\frac{\phi^{-1}(s)}{2} + \sin^2\frac{\phi^{-1}(s)}{2}\Big]\\ &= a\Big[2 - 2\cos^2\frac{\phi^{-1}(s)}{2}\Big]\\ &= 2a\Big[1 - \left(\frac{4a-s}{4a}\right)^2\Big].\end{aligned} \tag{20.139}$$

Dans ce paramétrage, $s \in]0, 8a[$. △

Exemple 20.48.
La cardioïde $\rho(\theta) = a\big(1+\cos(\theta)\big)$ avec θ entre $-\pi$ et π. Avant d'utiliser la formule (20.130), nous devons trouver l'élément de longueur de la cardioïde. Étant donné la façon dont l'équation de la cardioïde nous est donnée, l'élément de longueur est donné par [5] (20.87) :

$$\begin{aligned}\|\gamma'(u)\|^2 &= a^2\sin^2(u) + a^2(1+\cos(u))^2\\ &= 2a^2\big(1+\cos(u)\big),\end{aligned} \tag{20.140}$$

et par conséquent [6]

$$\begin{aligned}\phi(t) &= \int_0^t \sqrt{2a^2\big(1+\cos(u)\big)}\,du\\ &= \int_0^t \sqrt{2a^2\left(1+\cos^2\frac{u}{2} - \sin^2\frac{u}{2}\right)}\,du\\ &= 2a\int_0^t \cos\frac{u}{2}\,du\\ &= 4a\sin\frac{t}{2}.\end{aligned} \tag{20.141}$$

5. Nous vous déconseillons d'étudier cette formule par cœur. Sachez cependant la retrouver assez vite.
6. L'utilisation stricte de la formule (20.130) demanderait d'intégrer à partir de $-\pi$. Pour plus de simplicité, nous intégrons à partir de zéro, et nous verrons plus tard comment adapter l'intervalle du nouveau paramètre.

Pour trouver l'inverse, nous résolvons $\phi\big(\phi^{-1}(s)\big) = s$ par rapport à $\phi^{-1}(s)$:

$$\begin{aligned} 4a\sin\left(\frac{\phi^{-1}(s)}{2}\right) &= s, \\ \phi^{-1}(s) &= 2\arcsin\left(\frac{s}{4a}\right). \end{aligned} \tag{20.142}$$

Avant d'écrire trop brutalement $\gamma_N(s) = (\gamma\circ\phi^{-1})(s)$, il faut comprendre comment est γ. Nous avons reçu la courbe sous forme polaire, c'est-à-dire

$$\gamma(t) = \big(\gamma_r(t), \gamma_\theta(t)\big) = \Big(a\big(1+\cos(t)\big), t\Big). \tag{20.143}$$

C'est comme cela qu'il faut comprendre la donnée $\rho(\theta) = a\big(1+\cos(\theta)\big)$. Maintenant la formule $\gamma_N(s) = (\gamma\circ\phi^{-1})(s)$ devient

$$\begin{cases} (\gamma_N)_r(s) = \gamma_r\big(\phi^{-1}(s)\big) & \text{(20.144a)} \\ (\gamma_N)_\theta(s) = \gamma_\theta\big(\phi^{-1}(s)\big). & \text{(20.144b)} \end{cases}$$

Étant donné que $\gamma_\theta(t) = t$, la seconde est facile :

$$(\gamma_N)_\theta(s) = 2\arcsin\left(\frac{s}{4a}\right). \tag{20.145}$$

Pour la première,

$$(\gamma_N)_r(s) = a\big[1+\cos\big(2\arcsin\frac{s}{4a}\big)\big] = \frac{16a^2-s^2}{8a}. \tag{20.146}$$

Nous écrivons donc le nouveau paramétrage en coordonnées polaires sous la forme

$$\left(\frac{16a^2-s^2}{8a}, 2\arcsin\frac{s}{4a}\right). \tag{20.147}$$

La question qui arrive maintenant est de savoir quel intervalle parcours la nouvelle variable s. D'après le résultat de l'exemple 20.96, la longueur de la cardioïde est de $8a$ et nous avons donc $s\in[0,8a]$. Cependant, la condition d'existence de arcsin nous interdit d'avoir s plus grand que $4a$ en valeur absolue. Où est le problème ?

Le problème est que nous avons changé l'origine de notre paramètre en donnant $\phi(t)$ comme une intégrale à partir de 0 au lieu de $-\pi$. Cela se voit en regardant de quel point nous partons : en $s=0$ nous sommes sur le point $(2a,0)$ tandis qu'avec le paramètre original, c'est-à-dire $\theta\in[-\pi,\pi]$, nous avons pour $\theta=-\pi$ le point $(0,-\pi)$.

Il se passe donc que si nous commençons à parcourir la cardioïde avec $s=0$, nous partons du milieu, et nous ne parcourons donc pas tout. Étant donné que le « premier » point de la cardioïde est le point $(0,-\pi)$, le paramètre s commence en $s=-4a$, et nous avons comme intervalle :

$$s\in[-4a,4a], \tag{20.148}$$

ce qui est en accord avec la conditions d'existence. △

Quel enseignement tirer de cet exemple ? Lorsqu'on calcule $\phi(t)$ pour trouver les coordonnées normales, il y a deux solutions.

(1) Utiliser strictement la formule $\phi(t) = \int_a^t \|\gamma'(u)\|du$, en prenant bien comme borne de départ le point de départ de le paramétrage de γ. À ce moment la coordonnée normale construite aura $[0,l(\gamma)]$ comme intervalle de variation.

(2) Faire commencer l'intervalle d'intégration en zéro (ou ailleurs). Un bon choix peut simplifier quelques calculs, mais alors il faudra bien choisir la valeur de départ de la nouvelle coordonnées pour que le « premier » point de la courbe soit correct. Dans ce cas, la longueur de l'intervalle sera quand même $l(\gamma)$. Il n'y a donc pas de problèmes pour trouver la valeur du bout de l'intervalle de variation du paramètre normal.

Dans tous les cas, il faut bien préciser l'intervalle de variation du paramètre lorsqu'on donne une courbe paramétrée.

20.7.2 Tangente à une courbe paramétrée

Définition 20.49.
Soit (I, γ) un arc paramétré de classe $\mathcal{C}^k$ avec $k \geqslant 1$. Nous disons que la courbe admet une ***tangente*** *en $\gamma(t_0) \in \mathbb{R}^n$ lorsque les deux conditions suivantes sont remplies*

(1) $\gamma(t) \neq \gamma(t_0)$ pour tout t dans un voisinage de t_0 ;

(2) la direction de la droite qui passe par $\gamma(t)$ et $\gamma(t_0)$ admet une limite lorsque $t \to t_0$.

Dans ce cas, la tangente sera la droite passant par le point $\gamma(t_0)$ et dont la direction est donnée par la limite.

Dans cette définition, par **direction** d'une droite, nous entendons le vecteur de norme 1 parallèle à celle-ci sans tenir compte du signe. La tangente sera donc la droite passant par $\gamma(t_0)$ et parallèle au vecteur

$$\lim_{t \to t_0} \frac{\gamma(t) - \gamma(t_0)}{\|\gamma(t) - \gamma(t_0)\|}. \tag{20.149}$$

Évidemment si nous avions écrit $\gamma(t_0) - \gamma(t)$, ça n'aurait pas changé la droite. Par abus de langage, nous parlerons souvent de « la direction u » même lorsque u n'est pas de norme 1.

Formellement, une direction est une classe d'équivalence de vecteurs pour la relation $u \sim v$ s'il existe $\lambda \neq 0$ tel que $u = \lambda v$, mais nous n'aurons pas besoin de cette précision ici.

Sans surprises, la tangente est à peu près toujours donnée par la dérivée lorsqu'elle existe. Plus précisément nous avons le

Théorème 20.50.
Soit (I, γ), un arc paramétré de classe $\mathcal{C}^k$ $(k \geqslant 1)$ et $t_0 \in I$ tel que

$$\gamma'(t_0) = \gamma''(t_0) = \ldots = \gamma^{(q-1)}(t_0) = 0 \tag{20.150}$$

et

$$\gamma^{(q)}(t_0) \neq 0 \tag{20.151}$$

pour un entier $1 \leqslant q \leqslant k$. Alors γ admet une tangente en $\gamma(t_0)$ de direction $\gamma^{(q)}(t_0)$.

Démonstration. Le développement de $\gamma(t_0)$ en série de Taylor autour de t jusqu'à l'ordre q est

$$\begin{aligned}\gamma(t) = \gamma(t_0) + \gamma'(t_0)|t - t_0| + \frac{\gamma' t(t_0)}{2}|t - t_0|^2 + \cdots + \frac{\gamma^{(q)}(t_0)}{q!}|t - t_0|^q \\ + \varepsilon(t)|t - t_0|^q\end{aligned} \tag{20.152}$$

où ε est une application $\varepsilon \colon \mathbb{R} \to \mathbb{R}^n$ telle que $\lim_{t \to t_0} \varepsilon(t) = 0$. En utilisant les hypothèses, nous éliminons la majorité des termes dans le développement (20.152) :

$$\gamma(t) - \gamma(t_0) = \frac{1}{q!}\gamma^{(q)}(t_0)|t - t_0|^q + \varepsilon(t)|t - t_0|^q. \tag{20.153}$$

La direction de la droite qui joint $\gamma(t)$ à $\gamma(t_0)$ est donc donnée par

$$\frac{\gamma(t) - \gamma(t_0)}{\|\gamma(t) - \gamma(t_0)\|} = \frac{\frac{1}{q!}\gamma^{(q)}(t_0)|t - t_0|^q + \varepsilon(t)|t - t_0|^q}{\|\frac{1}{q!}\gamma^{(q)}(t_0)|t - t_0|^q + \varepsilon(t)|t - t_0|^q\|} \tag{20.154}$$

et la limite lorsque $t \to t_0$ donne $\gamma^{(q)}(t_0)$ comme direction de la tangente.

□

Lorsque le théorème s'applique, le vecteur

$$\tau = \frac{\gamma^{(q)}(t_0)}{\|\gamma^{(q)}(t_0)\|} \tag{20.155}$$

est appelé le **vecteur unitaire tangent** en $\gamma(t_0)$ à l'arc paramétré γ.

Corolaire 20.51.
Si (I, γ) est un arc paramétré de classe $\mathcal{C}^1$ régulier (c'est-à-dire $\gamma'(t) \neq 0$ pour tout t) alors l'arc admet une tangente en tout point et le vecteur unitaire de la tangente est donné par

$$\tau(t) = \frac{\gamma'(t)}{\|\gamma'(t)\|}, \tag{20.156}$$

pour tout t dans I.

Corolaire 20.52.
Si $\gamma = (J, \gamma_N)$ est un arc paramétré de classe $\mathcal{C}^1$, normal, alors le vecteur unitaire de la tangente au point $\gamma_N(s)$ est donné par $\tau(s) = \gamma_N'(s)$.

Démonstration. Nous devons démontrer que dans le cas d'un paramétrage normal nous avons $\|\gamma_N'(s)\| = 1$ pour tout s. Par définition,

$$l\big([s, s'], g\big) = \int_s^{s'} \|\gamma_N'(u)\| du = s' - s. \tag{20.157}$$

Par conséquent,

$$\lim_{h \to 0} \frac{1}{h} \int_s^{s+h} \|\gamma_N'(u)\| du = \lim_{y \to 0} \frac{s + h - s}{h} = 1. \tag{20.158}$$

Cela implique que $\|\gamma_N'(s)\| = 1$, et donc en particulier que (J, γ_N) est un arc régulier. Le corolaire précédent montre alors que $\tau(s) = \gamma_N'(s)/\|\gamma_N'(s)\| = \gamma_N'(s)$. □

Exemple 20.53.
Considérons la courbe $\gamma(t) = (t^2, t^3)$, et cherchons la tangente en $t_0 = 0$. En dérivant nous avons successivement

$$\begin{aligned} \gamma(t) &= (t^2, t^3) \\ \gamma'(t) &= (2t, 3t^2) \\ \gamma''(t) &= (2, 6t). \end{aligned} \tag{20.159}$$

En posant $t = 0$, nous trouvons que $\gamma'(0) = 0$ mais $\gamma''(0) = (2, 0) \neq 0$. Le théorème nous dit donc que la direction de la tangente est horizontale. Nous pouvons faire le calcul directement :

$$\frac{\gamma(t) - \gamma(t_0)}{\|\gamma(t) - \gamma(t_0)\|} = \frac{(t^2, t^3)}{\sqrt{t^4 + t^6}} = \frac{(t^2, t^3)}{t^2\sqrt{1 + t^2}} = \frac{(1, t)}{\sqrt{1 + t^2}}, \tag{20.160}$$

dont la limite $t \to 0$ est bien le vecteur horizontal $(1, 0)$.

La figure 20.7 montre quelques tangentes, c'est-à-dire quelques vecteurs dans la direction $\gamma'(t)$ (pour les $t \neq 0$, il ne faut pas aller à la dérivée seconde). Nous remarquons que de part et d'autres du sommet, les vecteurs ne sont pas dirigés dans le même sens. *En tant que vecteurs* de norme 1, ces vecteurs n'ont pas de limites quand $t \to 0$. Ce sont bien les *directions* qui ont une limite, parce que la direction ne tient pas compte du sens.

△

20.8 Un peu de topologie

La proposition 20.54 donne une sorte de théorème des valeurs intermédiaires pour le cas d'une application à valeurs dans un chemin.

Proposition 20.54 ([1]).
Soit une application continue et injective $\gamma\colon [0, 1] \to \mathbb{R}^n$. Nous posons $\Gamma = \gamma\big([0, 1]\big)$ sur lequel nous considérons la topologie induite[7] *de $\mathbb{R}^n$.*

Nous supposons que $\gamma^{-1}\colon \Gamma \to [0, 1]$ est continue[8]

Nous considérons un chemin $\alpha\colon [0, 1] \to \mathbb{R}^n$ tel que

7. Définition 7.15.
8. Je ne suis pas certain que cette hypothèse soit indispensable. Voiir la question (7) dans 0.7.5.2.

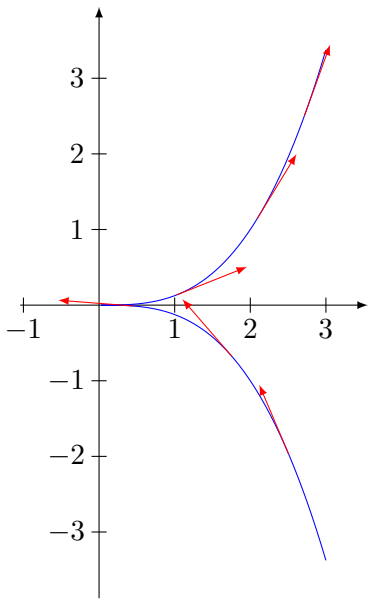


FIGURE 20.7 – Quelques tangentes de la courbe $\gamma(t) = (t^2, t^3)$.

(1) α *est continu,*

(2) $\alpha(0) = \gamma(0)$

(3) $\alpha\big([0,1]\big) \subset \Gamma$

(4) $\alpha(1) = \gamma(t_0)$ *pour un certain* $t_0 \in [0,1]$.

Alors pour tout $t \in [0, t_0]$, *il existe* $u \in [0,1]$ *tel que* $\alpha(u) = \gamma(1)$.

Démonstration. Nous commençons par montrer que l'application $\alpha\colon [0,1] \to \Gamma$ est encore continue lorsque nous voyons bien l'espace d'arrivée comme Γ muni de sa propre topologie et non comme $\mathbb{R}^2$ muni de sa topologie usuelle.

Soit un ouvert $\mathcal{O}$ de Γ ; il existe un ouvert $\mathcal{O}'$ de $\mathbb{R}^2$ tel que $\mathcal{O} = \mathcal{O}' \cap \Gamma$. Vu que α ne prend ses valeurs que dans Γ, nous avons $\alpha^{-1}(\Gamma \cap \mathcal{O}') = \alpha^{-1}(\mathcal{O}')$ et comme α est continue pour la topologie de $\mathbb{R}^2$, la partie $\alpha^{-1}(\mathcal{O}')$ est un ouvert de $[0,1]$.

Nous considérons donc l'application

$$\gamma^{-1} \circ \alpha\colon [0,1] \to [0,1] \tag{20.161}$$

qui est continue et vérifie donc le théorème des valeurs intermédiaires 11.50. Les hypothèses $\alpha(0) = \gamma(0)$ et $\alpha(1) = \gamma(t_0)$ donnent

$$(\gamma^{-1} \circ \alpha)(0) = 0 \tag{20.162a}$$
$$(\gamma^{-1} \circ \alpha)(1) = t_0. \tag{20.162b}$$

Donc pour tout $t \in [0, t_0]$, il existe $u \in [0,1]$ tel que $(\gamma^{-1} \circ \alpha)(u) = t$. En appliquant γ des deux côtés, nous voyons que ce u vérifie $\alpha(u) = \gamma(t)$ comme demandé. □

La proposition suivant dit essentiellement que la longueur d'un chemin est minoré par la longueur de son graphe.

Proposition 20.55 ([1]).
Soient $a, b \in \mathbb{R}^n$. *Soit une application continue et injective* $\gamma\colon [0,1] \to \mathbb{R}^n$ *telle que* $\gamma(0) = a$ *et* $\gamma(1) = b$.

Nous considérons maintenant un second chemin continu $\alpha\colon [0,1] \to \mathbb{R}^n$ *tel que* $\alpha(0) = a$, $\alpha(1) = b$ *et* $\sigma([0,1]) \subset \gamma([0,1])$.

Alors la longueur de γ *minore celle de* α : $l(\gamma) \leqslant l(\alpha)$.

Démonstration. Pour rappel, la longueur d'un chemin est donné par l'infimum des longueurs des lignes brisées reliant des points du chemin (définition 20.4). Soit une subdivision σ de $[0,1]$ et la « longueur » correspondante pour γ :

$$l_\sigma(\gamma) = \sum_{i=1}^{n} \|\gamma(t_i) - \gamma(t_{i-1})\|. \tag{20.163}$$

Nous allons construite une subdivision de $[0,1]$ pour laquelle la longueur de α sera la même. De cette façon, le supremum pour α sera un supremum sur un ensemble plus grand que celui du supremum pour γ.

Vu que le graphe de α est contenu dans celui de γ, et que les points de départ et d'arrivée sont les mêmes, la proposition 20.54 nous donne que pour tout $t \in [0,1]$, il existe $u \in [0,1]$ tel que $\alpha(u) = \gamma(t)$.

Cela ne suffit pas à considérer une subdivision $(u_i)_{i=1,\dots,n}$ de $[0,1]$ pour laquelle $\alpha(u_i) = \gamma(t_i)$ parce qu'il faut encore que les u_i soient ordonnés. Vu que α n'est pas injective, il n'y a pas de garanties de ce côté. Nous posons :

$$u_i = \min\{x \in [0,1] \text{ tel que } \alpha(x) = \gamma(t_i)\}. \tag{20.164}$$

Montrons que ce minimum existe. Ce minimum existe parce que α étant continue, l'ensemble des x sur lesquels $\alpha(x)$ a une valeur donnée est fermé (le complémentaire est ouvert par le théorème des valeurs intermédiaires). De plus cet ensemble est borné parce qu'inclus à $[0,1]$. Il est donc compact et possède un minimum.

Nous prouvons à présent que $u_{i+1} \geqslant u_i$. Nous avons $\alpha(u_{i+1}) = \gamma(t_{i+1})$ et il existe u dans $[0, u_{i+1}[$ tel que$\alpha(u) = \gamma(t_i)$ (proposition 20.54). Vu que u_i est le minimum de tels u, nous avons $u_i \in [0, u_{i+1}[$. Cela prouve bien que $u_i < u_{i+1}$.

Ces $(u_i)_{i=1,\dots,n}$ forment une subdivision de $[0,1]$ telle que $\alpha(u_i) = \gamma(t_i)$. La longueur associée à la subdivision (u_i) pour α est la même que celle associée à (t_i) pour γ.

Donc $l(\alpha) \geqslant l(\gamma)$. □

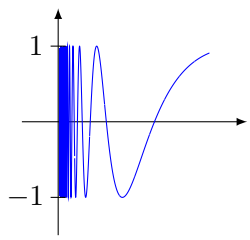


FIGURE 20.8 – La figure de la proposition 20.56.

Proposition 20.56 ([1])**.**
Nous considérons la partie suivante de $\mathbb{R}^2$:

$$A_2 = \big\{\big(t, \sin(1/t)\big)\big\}_{t\in]0,1]} \tag{20.165}$$

qui est dessinée est sur la figure 20.8. Ensuite nous posons

$$A = \big\{(0,0)\big\}. \tag{20.166}$$

La partie A est connexe, mais pas connexe par arcs.

Démonstration. En plusieurs points.

A_2 est connexe La partie A_2 est l'image de la fonction

$$\begin{aligned} f\colon]0,1] &\to \mathbb{R}^2 \\ t &\mapsto \big(t, \sin(1/t)\big). \end{aligned} \tag{20.167}$$

Vu que f est continue et que son ensemble de départ est connexe, A_2 est connexe (proposition 7.90).

A est connexe Soient deux ouverts disjoints $\mathcal{O}_1$ et $\mathcal{O}_2$ dont l'union contient A. Nous supposons que $(0,0) \in \mathcal{O}_1$. Si $\mathcal{O}_1$ contient $B\big((0,0),r\big)$, alors il contient tous les points de la forme $\big(\frac{1}{2k\pi},0\big)$ pour k assez grand. Ces points sont dans A_1.

Vu que $\mathcal{O}_1$ contient des points de A_1, il doit contenir tous les points de A_1 ; sinon les ouverts $\mathcal{O}_1$ et $\mathcal{O}_2$ contrediraient la connexité de A_1. Finalement, $A \subset \mathcal{O}_1$ et A est connexe.

Connexité par arcs : le chemin Pour faire aller le chemin dans le bon sens, nous considérons

$$\begin{aligned}\gamma\colon [0,1[&\to \mathbb{R}^2\\ t &\mapsto f(1-t).\end{aligned} \tag{20.168}$$

Si A était connexe par arcs, il existerait une application continue $\alpha\colon [0,1] \to A$ telle que $\alpha(0) = \big(1,\sin(1)\big) = \gamma(0)$ et $\alpha(1) = (0,0)$.

Minoration de la longueur de α Vu que α est continue, ses composantes le sont et vérifient le théorème des valeurs intermédiaires. Pour tout $t \in]0,1[$, il existe $u \in]0,1[$ tel que $\alpha(u) = \gamma(t)$. Par la proposition 20.55, la longueur de α est minorée par la longueur du chemin γ entre 0 et t_0 pour tout t.

La longueur de γ Vu que γ passe une infinité de fois par zéro et par 1, il est possible de construire une subdivision de $]0,1[$ par rapport à la quelle la longueur de γ est arbitrairement grande.

Conclusion Donc la longueur de α est minoré par tous les nombres, arbitrairement grand. Autrement dit, la longueur de α est infinie. Mais α étant un chemin continu depuis un compact, cela est impossible.

Nous déduisons qu'un chemin continu liant $\big(1,\sin(1)\big)$ à $(0,0)$ en restant sur A est impossible.

□

20.9 Repère de Frenet

Dans cette section, nous ne considérons que des courbes dans $\mathbb{R}^3$.

Proposition 20.57.
Soit $\gamma = (J,\gamma_N)$ un arc paramétré normal de classe $\mathcal{C}^2$. Alors pour toute valeur de s dans J, nous avons

$$\tau(s) \cdot \tau'(s) = 0 \tag{20.169}$$

où $\tau(s) = \gamma_N'(s)$. C'est-à-dire que la dérivée seconde est perpendiculaire à la dérivée première.

Démonstration. Le paramétrage étant normal, nous avons

$$\|\gamma_N'(s)\|^2 = \sum_{i=1}^{n} x_i'(s)^2 = 1; \tag{20.170}$$

ce qui implique, en dérivant les deux membres, que

$$0 = 2\sum_{i=1}^{n} x_i'(s)x_i''(s), \tag{20.171}$$

c'est-à-dire exactement $\gamma_N'(s) \cdot \gamma_N''(s) = 0$; d'où la thèse. □

Remarque 20.58.
Si nous n'utilisons pas des coordonnées normales, la proposition 20.57 n'est pas spécialement vraie. Prenons par exemple la courbe qui donne la parabole :

$$\gamma(t) = (t,t^2) \tag{20.172a}$$
$$\gamma'(t) = (1,2t) \tag{20.172b}$$
$$\gamma''(t) = (0,2) \tag{20.172c}$$

Nous avons $\gamma'(t) \cdot \gamma''(t) = 4t$. Par conséquent, la dérivée seconde n'est la normale à la courbe que en $t = 0$. Cela est une propriété très intéressante des coordonnées normales : la dérivée seconde d'une coordonnées normale donne un vecteur normal à la courbe, c'est-à-dire perpendiculaire à la tangente.

Définition 20.59.
Soit $\gamma = (J, \gamma_N)$ un arc paramétré normal de classe $\mathcal{C}^2$.

*(1) Le **vecteur unitaire tangent** est donné par le corolaire 20.51 : $\tau(t) = \gamma'(t)/\|\gamma'(t)\|$.*

*(2) La **normale principale** est le vecteur $\tau'(s)$. Le **vecteur unitaire normal** est le vecteur*

$$\nu(s) = \frac{\tau'(s)}{\|\tau'(s)\|} = \frac{\gamma''_N(s)}{\|\gamma''_N(s)\|}. \tag{20.173}$$

Nous déduirons une formule plus pratique en dehors des coordonnées normales en (20.201).

*(3) La **courbure** au point $\gamma_N(s)$ est le réel*

$$c(s) = \|\tau'(s)\| = \|\gamma''_N(s)\|. \tag{20.174}$$

Note : il y a une notion de courbure signée qui sera donnée dans la définition 20.73.

*(4) Le **rayon de courbure** est le réel*

$$R(s) = \frac{1}{c(s)} = \frac{1}{\|\gamma''_N(s)\|}. \tag{20.175}$$

Par la proposition 20.57, nous avons $\nu(s) \cdot \tau(s) = 0$. En combinant toutes les formules, nous avons les différentes expressions suivantes pour le vecteur normal unitaire :

$$\nu(s) = \frac{\gamma''_N(s)}{c(s)} = \frac{\tau'(s)}{\|\tau'(s)\|} = \frac{\tau'(s)}{c(s)} = R(s)\tau'(s) = R(s)\gamma''_N(s). \tag{20.176}$$

Proposition 20.60.
La fonction courbure s'écrit $c = \|\gamma'_N \times \gamma''_N\|$.

Démonstration. Par la proposition 7.212 nous avons :

$$\langle \gamma'_N, \gamma''_N \rangle^2 + \|\gamma'_N \times \gamma''_N\|^2 = \|\gamma'_N\|^2 \|\gamma''_N\|^2 = \|\gamma''_N\|^2 \tag{20.177}$$

parce que, le paramétrage étant normal, $\|\gamma'_N\| = 1$. Mais $\langle \gamma'_N, \gamma''_N \rangle = 0$, donc il reste $\|\gamma'_N \times \gamma''_N\|^2 = \|\gamma''_N\|^2$, d'où

$$c(s) = \|\gamma''_N(s)\| = \|\gamma'_N(s) \times \gamma''_N(s)\| \tag{20.178}$$

pour chaque s dans J. □

Définition 20.61.
*Soit s un point birégulier (c'est-à-dire $\gamma'_N(s) \neq 0$ et $\gamma''_N(s) \neq 0$) de l'arc normal $\gamma = (J, \gamma_N)$. Le **vecteur unitaire de la binormale** est le vecteur*

$$\beta(s) = \tau(s) \times \nu(s) \tag{20.179}$$

Par leurs définitions, τ et ν sont unitaires, tandis que la proposition 20.57 montre qu'ils sont également orthogonaux. Les propriétés du produit vectoriel font que β est également unitaire, et simultanément orthogonal à τ et à ν.

Définition 20.62.
*Le repère orthonormal $\{\gamma_N(s), \tau(s), \beta(s)\}$ est le **repère de Frenet** au point $\gamma_N(s)$.*

Lemme 20.63.
Le vecteur unitaire normal est donné par $\nu(s) = \beta(s) \times \tau(s)$.

Démonstration. Ceci est une application de la formule d'expulsion (7.226) et de l'orthonormalité de la base de Frenet :

$$\beta \times \tau = (\tau \times \nu) \times \tau = -(\nu \cdot \tau)\tau + (\tau \cdot \tau)\nu = \nu. \tag{20.180}$$

□

20.9.1 Torsion

Décomposons le vecteur $\beta'(s)$ dans la base de Frenet. Pour cela nous allons utiliser la proposition 9.270 et montrer que $\beta'(s) \cdot \tau(s) = \beta'(s) \cdot \beta(s) = 0$, ce qui voudra dire que, dans la base de Frenet, les composantes de β' le long de τ et β sont nulles. Le vecteur β' sera donc colinéaire à ν.

D'abord, étant donné que la norme de $\beta(s)$ est constante par rapport à s, nous avons

$$0 = \frac{d}{ds}\|\beta(s)\|^2 = 2\beta'(s) \cdot \beta(s). \tag{20.181}$$

Ensuite, nous dérivons la définition $\beta(s) = \tau(s) \times \nu(s)$ en utilisant la formule de Leibnitz (11.268) :

$$\beta'(s) = \tau'(s) \times \nu(s) + \tau(s) \times \nu'(s). \tag{20.182}$$

Mais $\tau'(s) = \gamma_N''(s)$ tandis que $\nu(s) = \frac{\gamma_N''(s)}{\|\gamma_N''(s)\|}$, de telle sorte que $\tau'(s) \times \nu(s) = 0$. Nous restons donc avec $\beta'(s) = \tau(s) \times \nu'(s)$, ce qui prouve que $\beta'(s)$ est perpendiculaire à $\tau(s)$ et donc que $\beta'(s) \cdot \tau(s) = 0$.

Le vecteur $\beta'(s)$ est donc un multiple de $\nu(s)$. Nous notons $t(s)$ le facteur de proportionnalité :

$$\beta'(s) = t(s)\nu(s). \tag{20.183}$$

Définition 20.64.
*Soit $\gamma = (J, \gamma_N)$ un arc paramétré normal de classe $\mathcal{C}^3$. La **torsion** de γ au point $\gamma_N(s)$ est le réel*

$$t(s) = \|\beta'(s)\| = \|\tau(s) \times \nu'(s)\|. \tag{20.184}$$

*Lorsque $t(s) \neq 0$, le réel $T(s) = \frac{1}{t(s)}$ est le **rayon de torsion** de γ en $\gamma_N(s)$.*

Étant donné que pour chaque s, l'ensemble $\{\tau(s), \nu(s), \beta(s)\}$ est une base, il est naturel de vouloir décomposer leurs dérivées dans cette base. D'abord, par définition de c et de t, nous avons

$$\begin{aligned} \tau'(s) &= c(s)\nu(s) \\ \beta'(s) &= t(s)\nu(s). \end{aligned} \tag{20.185}$$

Il reste à décomposer $\nu'(s)$. Définissons α_τ, α_ν et α_β (qui peuvent dépendre de s) par

$$\nu'(s) = \alpha_\tau \tau(s) + \alpha_\nu \nu(s) + \alpha_\beta \beta(s). \tag{20.186}$$

En vertu de la proposition 9.270, nous avons

$$\begin{aligned} \alpha_\tau &= \langle \nu'(s), \tau(s) \rangle = -\langle \nu(s), \tau'(s) \rangle = -\langle \nu(s), c(s)\nu(s) \rangle = -c(s), \\ \alpha_\nu &= \langle \nu'(s), \nu(s) \rangle = 0, \\ \alpha_\beta &= \langle \nu'(s), \beta(s) \rangle = -\langle \nu(s), \beta'(s) \rangle = -t(s), \end{aligned} \tag{20.187}$$

où nous avons utilisé le fait que $\langle \nu(s), \nu(s) \rangle = \|\nu(s)\|^2 = 1$. Si nous mettons ces résultats sous forme matricielle, nous avons les **formules de Frenet** :

$$\begin{pmatrix} \tau'(s) \\ \nu'(s) \\ \beta'(s) \end{pmatrix} = \begin{pmatrix} 0 & c(s) & 0 \\ -c(s) & 0 & -t(s) \\ 0 & t(s) & 0 \end{pmatrix} \begin{pmatrix} \tau(s) \\ \nu(s) \\ \beta(s) \end{pmatrix}. \tag{20.188}$$

Proposition 20.65.
Si s est un point birégulier, alors la torsion est donnée par

$$t(s) = -\frac{(\gamma_N' \times \gamma_N'') \times \gamma_N'''}{\|\gamma_N'(s) \times \gamma_N''(s)\|^2}. \tag{20.189}$$

Démonstration. Par l'équation (20.176), nous avons $\gamma_N''(s) = c'(s)\nu(s)$, et par conséquent

$$\gamma_N'''(s) = c'(s)\nu(s) + c(s)\nu'(s) = c'(s)\nu(s) + c(s)\big[-c(s)\tau(s) - t(s)\beta(s)\big], \tag{20.190}$$

où nous avons utilisé la formule de Frenet pour $\nu'(s)$. Par ailleurs, sachant le corolaire 20.52 et la formule de Frenet pour τ', nous avons

$$\gamma_N' \times \gamma_N'' = \tau(s) \times \tau'(s) = \tau(s) \times c(s)\nu(s) = c(s)\beta(s). \tag{20.191}$$

En combinant les deux dernières équations, et en se souvenant que la base de Frenet et orthonormale,

$$(\gamma_N' \times \gamma_N'') \cdot \gamma_N'''(s) = -c(s)^2 t(s), \tag{20.192}$$

et donc, en remplaçant $c(s)$ par la formule (20.178),

$$t(s) = -\frac{(\gamma_N' \times \gamma_N'') \cdot \gamma_N'''}{\|\gamma_N' \times \gamma_N''\|^2}. \tag{20.193}$$

□

20.10 Hors des coordonnées normales

Remarque 20.66.
Notons que la définition de τ est donnée pour tout arc $\mathcal{C}^1$ régulier (I, γ) par $\tau(t) = \gamma'(t)/\|\gamma'(t)\|$. La propriété $\tau = \gamma_N'$ n'est valable que lorsque le paramétrage est normal. Les autres définitions ont toutes été données dans le cas d'un paramétrage normal.

La remarque 20.66 nous incite à exprimer toute la base de Frenet en termes de γ lorsque le paramétrage n'est pas normal. Étant donné que nous pouvons toujours faire le changement de variable $\gamma(t) = \gamma_N\big(\phi(t)\big)$ (proposition 20.46), il est possible d'exprimer les vecteurs τ, ν et β ainsi que les réels c et t en fonction de γ et de ses dérivées.

Nous allons maintenant travailler à écrire les formules.

Pour plus de facilité, nous collectons les définitions. Afin d'alléger la notation, nous n'exprimons pas explicitement les dépendances en s :

Vecteur unitaire tangent Par le corolaire 20.52, τ est donné par $\tau = \gamma_N'$.

Vecteur unitaire normal Par la définition 20.59, ν est donné par $\nu = \frac{\tau'}{\|\tau'\|}$.

Vecteur unitaire de la binormale Par la définition 20.59, β est donné par $\beta = \tau \times \nu$.

Courbure Par la définition 20.59, c est donné par $c = \|\tau'\|$.

Torsion Par la définition 20.64, t est donné par $t = \|\beta'\|$.

Le schéma du changement de variable est

$$\begin{tikzcd} t \in I \arrow[r, "f"] \arrow[d, "\phi"'] & \mathbb{R}^3 \\ s \in J \arrow[ur, "g"'] & \end{tikzcd} \tag{20.194}$$

La difficulté ne sera pas d'éliminer γ_N de toutes les formules, mais bien de se débarrasser des fonctions ϕ qui arrivent quand nous exprimons γ_N en termes de γ, et en particulier lorsque nous voulons exprimer les dérivées de γ_N en termes de γ et de ses dérivées.

Regardons d'abord comment les dérivées de γ_N s'expriment en termes de γ. En utilisant le fait que $\gamma_N(s) = (\gamma \circ \phi^{-1})(s)$ et que $\|\gamma_N'(s)\| = 1$, nous avons

$$\gamma_N'(s) = \frac{\gamma_N'(s)}{\|\gamma_N'(s)\|} = \frac{(\gamma \circ \phi^{-1})'(s)}{\|(\gamma \circ \phi^{-1})'(s)\|} = \frac{\gamma'\big(\phi^{-1}(s)\big)(\phi^{-1})'(s)}{\|\gamma'\big(\phi^{-1}(s)\big)\|\,|(\phi^{-1})'(s)|} = \frac{\gamma'(t)}{\|\gamma'(t)\|} \tag{20.195}$$

où nous avons utilisé le fait que ϕ^{-1} étant croissante (parce que l'inverse d'une fonction croissante est croissante), $(\phi^{-1})'(s) = |(\phi^{-1})'(s)|$. Pourquoi écrivons nous $|\phi^{-1}(s)|$ et non $\|\phi^{-1}(s)\|$?

Pour la dérivée seconde, nous dérivons la relation (20.195) :

$$\gamma_N''(s) = \frac{\gamma''\big(\phi^{-1}(s)\big)(\phi^{-1})'(s)}{\|\gamma'\big(\phi^{-1}(s)\big)\|} + \gamma'\big(\phi^{-1}(s)\big)\frac{d}{ds}\Big[\|\gamma'\big(\phi^{-1}(s)\big)\|\Big]. \tag{20.196}$$

Le petit calcul suivant va nous permettre de simplifier cette expression :

$$(\phi^{-1})'(s) = (\phi^{-1})'\big(\phi(t)\big) = \frac{1}{\phi'(t)} = \frac{1}{\|\gamma'(t)\|}. \tag{20.197}$$

Donc

$$\gamma_N''(s) = \frac{\gamma''(t)}{\|\gamma'(t)\|^2} + \gamma'(t)\frac{d}{ds}\Big[\|\gamma'(t)\|\Big] \tag{20.198}$$

où il est entendu que $t = \phi^{-1}(s)$. Avec cette expression, nous ne nous sommes pas encore débarrassés de la fonction ϕ, mais nous allons voir que cela nous sera suffisant.

Pour le vecteur unitaire tangent $\tau(s)$, nous avons donc immédiatement

$$\tau(s) = \gamma_N'(s) = \frac{\gamma'(t)}{\|\gamma'(t)\|}. \tag{20.199}$$

Ici encore il est sous-entendu que le t dans le membre de droite est lié au s du membre de gauche par $t = \phi^{-1}(s)$. Il est donc naturel de nous demander si nous avons gagné quelque chose, étant donné que la formule (20.199) contient encore la fonction ϕ.

Géométriquement, le vecteur $\tau(s)$ est le vecteur normal unitaire de la courbe au point $\gamma_N(s)$. En utilisant les relations du diagramme (20.194), nous avons en réalité $\gamma_N(s) = \gamma_N\big(\phi(t)\big) = \gamma(t)$. Le vecteur $\frac{\gamma'(t)}{\|\gamma'(t)\|}$ représente donc le vecteur normal tangent au point $\gamma(t)$.

Pour calculer la courbure, nous devons d'abord calculer le produit vectoriel

$$\begin{aligned}\gamma_N'(s) \times \gamma_N''(s) &= \frac{\gamma'(t)}{\|\gamma'(t)\|} \times \left(\frac{\gamma''(t)}{\|\gamma'(t)\|^2} + \gamma'(t)\frac{d}{ds}\Big[\|\gamma'(t)\|\Big]\right)\\ &= \frac{\gamma'(t) \times \gamma''(t)}{\|\gamma'(t)\|^3}\end{aligned} \tag{20.200}$$

parce que le deuxième terme dans la parenthèse est un multiple de $\gamma'(t)$, de telle sorte à ce que son produit vectoriel avec $\gamma'(t)/\|\gamma'(t)\|$ soit nul. En prenant la norme,

$$c(s) = \frac{\|\gamma'(t) \times \gamma''(t)\|}{\|\gamma'(t)\|^3}. \tag{20.201}$$

Encore une fois, cette équation nous enseigne que la courbure au point $\gamma(t) \in \mathbb{R}^3$ est donnée par le membre de droite, qui ne dépend que de t.

Le vecteur unitaire binormal est donné par $\beta(s) = \tau(s) \times \nu(s)$. En utilisant (20.199) et (20.176),

$$\beta(s) = \tau(s) \times \nu(s) = \gamma_N'(s) \times \frac{\gamma_N''(s)}{c(s)}. \tag{20.202}$$

Les formules (20.200) pour le produit vectoriel et (20.201) pour la courbure donnent ensuite

$$\beta(s) = \frac{\gamma'(t) \times \gamma''(t)}{\|\gamma'(t)\|^3} \cdot \frac{1}{c(s)} = \frac{\gamma'(t) \times \gamma''(t)}{\|\gamma'(t) \times \gamma''(t)\|}. \tag{20.203}$$

Cela donne le vecteur unitaire binormal au point $\gamma(t)$ en termes de $\gamma'(t)$ et $\gamma''(t)$.

La torsion demande d'utiliser la dérivée troisième de γ_N. Nous avons

$$\begin{aligned}
\gamma_N'''(s) &= (\gamma\circ\phi^{-1})'''(s)\\
&= \Big(\gamma'\big(\phi^{-1}(s)\big)(\phi^{-1})'(s)\Big)''\\
&= \Big(\gamma''\big(\phi^{-1}(s)\big)(\phi^{-1})'(s)^2+\gamma'\big(\phi^{-1}(s)\big)(\phi^{-1})''(s)\Big)'\\
&= \gamma'''\big(\phi^{-1}(s)\big)(\phi^{-1})'(s)^3+v\\
&= \frac{\gamma'''\big(\phi^{-1}(s)\big)}{\|\gamma'(t)\|^3}+v \qquad \text{par (20.197)}
\end{aligned} \tag{20.204}$$

où v est un élément de $\langle\gamma''\big(\phi^{-1}(s)\big),\gamma'\big(\phi^{-1}(s)\big)\rangle$. Le vecteur v est donc perpendiculaire à $\gamma'\times\gamma''$ et donc à $\gamma_N'\times\gamma_N''$ à cause de la relation (20.200) qui montre que $\gamma'\times\gamma''$ est parallèle à $\gamma_N'\times\gamma_N''$. De ce fait, lorsque nous calculons $(\gamma_N'\times\gamma_N'')\cdot\gamma_N'''$, la partie v de γ_N''' n'entre pas en ligne de compte.

Nous avons donc le calcul suivant, en remplaçant les diverses occurrences de $\gamma_N'\times\gamma_N''$ par sa valeur (20.200) en termes de γ,

$$\begin{aligned}
t(s) &= -\frac{(\gamma_N'\times\gamma_N'')\cdot\gamma_N'''}{\|\gamma_N'\times\gamma_N''\|^2}\\
&= -\frac{(\gamma_N'\times\gamma_N'')\cdot\gamma'''(t)}{\|\gamma_N'\times\gamma_N''\|^2\,\|\gamma'(t)\|^2}\\
&= -\frac{(\gamma'\times\gamma'')\cdot\gamma'''}{\|\gamma'\times\gamma''\|^2}.
\end{aligned} \tag{20.205}$$

Dans cette expression, il est sous-entendu que tous les γ_N sont fonctions de s et tous les γ sont fonction de t où s et t sont liés par $s=\phi(t)$.

Ce que nous avons prouvé est le

Théorème 20.67.
Pour tout représentant (I,γ), les éléments métriques (τ,ν,β,c,t) au point $\gamma(t)$ s'expriment en fonction de $\gamma(t)$, $\gamma'(t)$, $\gamma''(t)$ et $\gamma'''(t)$.

Lemme 20.68.
Si γ est le graphe de la fonction y alors la courbure de γ est donnée par la formule

$$c\big(\gamma(t)\big)=\frac{|y''(t)|}{\big(1+y'(t)^2\big)^{3/2}} \tag{20.206}$$

Démonstration. Nous avons :

$$\gamma(t)=\big(t,y(t)\big) \tag{20.207a}$$
$$\gamma'(t)=\big(1,y'(t)\big) \tag{20.207b}$$
$$\gamma''(t)=\big(0,y''(t)\big). \tag{20.207c}$$

Il s'agit maintenant seulement d'utiliser la formule (20.201) en se souvenant comment on calcule un produit vectoriel[9] :

$$\gamma'\times\gamma''=\begin{vmatrix} e_1 & e_2 & e_3\\ 1 & y' & 0\\ 0 & y'' & 0\end{vmatrix}=y''. \tag{20.208}$$

□

9. Définition 7.204.

20.11 Tracer des courbes paramétriques dans $\mathbb{R}^2$

Nous allons maintenant voir comment les concepts introduits nous aident à effectivement tracer des courbes dans le plan. Les courbes que nous regardons sont de la forme $\gamma(t) = \big(x(t), y(t)\big)$, et nous supposons que ces fonctions soient suffisamment régulières (disons trois fois continument dérivables). Nous ne supposons pas que la courbe soit donnée en coordonnées normales, en particulier, $\gamma''(t)$ n'est pas le vecteur normal en $\gamma(t)$.

La notion clef qui va jouer est le **cercle osculateur** de la courbe γ au point $\gamma(t)$. Sans rentrer dans les détails, disons que c'est le cercle qui « colle » le mieux possible la courbe. Le rayon de ce cercle est le rayon de courbure :

$$R(t) = \frac{\|\gamma(t)\|^3}{\|\gamma'(t) \times \gamma''(t)\|}. \tag{20.209}$$

En pratique, le produit vectoriel se calcule comme ceci :

$$\gamma'(t) \times \gamma''(t) = \begin{vmatrix} e_x & e_y & e_z \\ x'(t) & y'(t) & 0 \\ x''(t) & y''(t) & 0 \end{vmatrix} = (x'y'' - x''y')e_z. \tag{20.210}$$

Le centre du cercle osculateur va se trouver quelque part sur la normale. Le vecteur normal est donné par

$$n(t) = J\frac{\gamma'(t)}{\|\gamma'(t)\|} \tag{20.211}$$

où J est la rotation d'angle $\frac{\pi}{2}$:

$$J\begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix} = \begin{pmatrix} y'(t) \\ -x'(t) \end{pmatrix}. \tag{20.212}$$

Cela nous laisse deux possibilités pour le centre du cercle osculateur : $\gamma(t) + R(t)n(t)$ ou bien $\gamma(t) - R(t)n(t)$. Il faut savoir de quel côté de la courbe est situé le centre du cercle osculateur. Il faut choisir le côté de la concavité, c'est-à-dire le côté de la dérivée seconde.

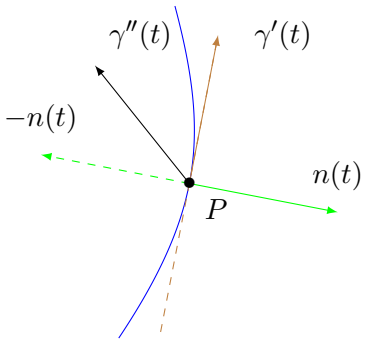


FIGURE 20.9 – De quel côté de $\gamma'(t)$ se trouvent $n(t)$ et $-n(t)$?

La difficulté maintenant est de savoir qui de $n(t)$ ou $-n(t)$ est du côté de $\gamma''(t)$. Il faut savoir si $n(t)$ est du même côté de la droite tangente que $\gamma''(t)$ ou non. Par construction, si nous regardons la figure 20.9, le vecteur $n(t)$ sera toujours à gauche de $\gamma'(t)$. Le fait que $\gamma''(t)$ soit à gauche ou à droite de $\gamma'(t)$ est donné par le signe du produit vectoriel $\gamma'(t) \times \gamma''(t)$. Si ce produit vectoriel est positif, il faut choisir $-n(t)$ et s'il est négatif, il faut choisir $n'(t)$.

Le truc pour obtenir le signe de $x'y'' - x''y'$ est de faire

$$\frac{(\gamma' \times \gamma'') \cdot e_z}{\|\gamma' \times \gamma''\|}. \tag{20.213}$$

Le centre de courbure sera donc situé à la position

$$\Omega(t) = \gamma(t) - n(t)\frac{\|\gamma(t)\|^3}{\|\gamma'(t) \times \gamma''(t)\|^2}(\gamma' \times \gamma'') \cdot e_z \tag{20.214}$$

Nous pouvons écrire cela plus explicitement en nous souvenant que $\gamma' \times \gamma'' = (x'y'' - x''y')e_z$, par conséquent $\frac{(\gamma' \times \gamma'') \cdot e_z}{\|\gamma' \times \gamma''\|^2} = \frac{1}{x'y'' - x''y'}$. Nous avons

$$\Omega_x(t) = x(t) - y'(t)\frac{x'^2 + y'^2}{x'y'' - x''y'} \tag{20.215a}$$

$$\Omega_y(t) = y(t) + x'(t)\frac{x'^2 + y'^2}{x'y'' - x''y'}. \tag{20.215b}$$

Quelques exemples de cercles osculateurs sont sur la figure 20.10.

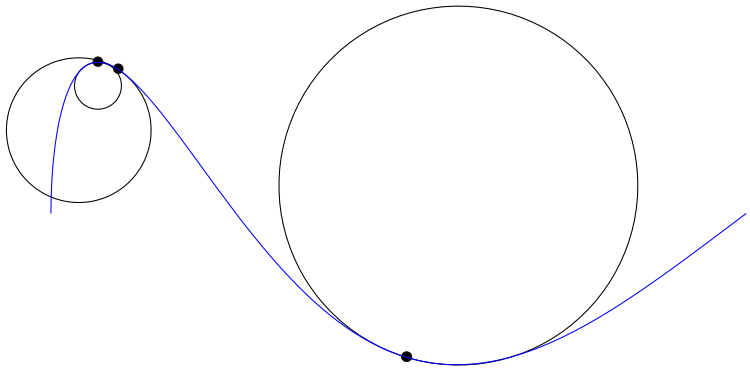


FIGURE 20.10 – Exemple de cercles osculateurs.

20.12 Courbes planes

Définition 20.69.
*Une courbe $\gamma\colon [a,b] \to \mathbb{R}^n$ est **fermée** si $\gamma(a) = \gamma(b)$. Elle est **simple** si $\gamma(t) \neq \gamma(t')$ dès que $t, t' \in]a,b[$ et $t \neq t'$.*

Définition 20.70.
Nous disons qu'une courbe fermée est continue, de classe C^1, de classe C^2 ou autre condition de régularité si son extension périodique comme application $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ a cette régularité.

20.12.1 Angle

Lemme 20.71 ([352])**.**
Soient des courbes régulières γ et σ de classe C^2 de l'intervalle ouvert I vers $\mathbb{R}^2$. Soit $\theta_0 \in \mathbb{R}$ tel que

$$\frac{\gamma'(t_0) \cdot \sigma'(t_0)}{\|\gamma'(t_0)\|\|\sigma'(t_0)\|} = \cos(\theta_0) \tag{20.216a}$$

$$\frac{\gamma'(t_0) \cdot J\sigma'(t_0)}{\|\gamma'(t_0)\|\|\sigma'(t_0)\|} = \cos(\theta_0). \tag{20.216b}$$

Alors il existe une unique fonction différentiable $\theta\colon I \to \mathbb{R}$

$$\frac{\gamma'(t) \cdot \sigma'(t)}{\|\gamma'(t)\|\|\sigma'(t)\|} = \cos\big(\theta(t)\big) \tag{20.217a}$$

$$\frac{\gamma'(t) \cdot J\sigma'(t)}{\|\gamma'(t)\|\|\sigma'(t)\|} = \sin\big(\theta(t)\big). \tag{20.217b}$$

Démonstration. Il suffit de prendre

$$f(t) = \frac{\gamma'(t) \cdot \sigma'(t)}{\|\gamma'(t)\|\|\sigma'(t)\|} \tag{20.218}$$

et

$$g(t) = \frac{\gamma'(t) \cdot J\sigma'(t)}{\|\gamma'(t)\|\|\sigma'(t)\|} \tag{20.219}$$

dans la proposition 17.27. Ces courbes sont de classe C^1 parce que γ et σ sont de classe C^2. □

20.12.2 Courbure signée

Nous avons déjà défini la courbure d'une courbe en la définition 20.59. Nous introduisons maintenant la courbure signée qui est propre à la dimension deux.

Définition 20.72.
*La **structure complexe** sur $\mathbb{R}^2$ est l'application*

$$\begin{aligned} J\colon \mathbb{R}^2 &\to \mathbb{R}^2 \\ (x, y) &\mapsto (-y, x). \end{aligned} \tag{20.220}$$

Définition 20.73.
*La **courbure signée** de la courbe $\gamma\colon I \to \mathbb{R}^2$ (I est un intervalle dans $\mathbb{R}$) est la fonction*

$$\kappa(t) = \frac{\gamma''(t) \cdot J\gamma'(t)}{\|\gamma'(t)\|^3} \tag{20.221}$$

où J est la structure complexe de la définition 20.72.

Cette définition est motivée par le fait qu'en identifiant $\mathbb{R}^2$ à $\mathbb{C}$, l'application J revient à l'application $z \mapsto iz$.

Si $v, w \in \mathbb{R}^2$ nous avons formellement

$$v \times w = -(v \cdot Jw)e_3. \tag{20.222}$$

En particulier pour tout $v \in \mathbb{R}^2$ nous avons

$$v \cdot Jv = 0. \tag{20.223}$$

Lemme 20.74.
Soir une courbe régulière $\gamma\colon [a, b] \to \mathbb{R}^2$ et un difféomorphisme $h\colon [c, d] \to [a, b]$. Si nous posons $\sigma = \gamma \circ h$ alors

$$\kappa_\sigma(u) = \operatorname{sgn}\big(h'(u)\big)\kappa_\gamma\big(h(u)\big). \tag{20.224}$$

Démonstration. Nous utilisons la définition (20.221) de la courbure signée. La règle de dérivation en chaine donne :

$$\sigma'(u) = \gamma'\big(h(u)\big)h'(u) \tag{20.225a}$$

$$\sigma''(u) = \gamma''\big(h(u)\big)h'(u)^2 + \gamma'\big(h(u)\big)h''(u). \tag{20.225b}$$

La numérateur de $\kappa_\sigma(u)$ est :

$$(\gamma'' \circ h)h'^2 \cdot J(\gamma' \circ h)h' + h''(\gamma' \circ h) \cdot J(\gamma' \cdot h)h' \tag{20.226}$$

dont le second terme est nul parce que $v \cdot Jv = 0$. Il nous reste donc

$$\kappa_\sigma(u) = \frac{(h')^3}{|h'|^3}\frac{(\gamma'' \circ h) \cdot J(\gamma' \circ h)}{\|\gamma' \circ h\|^3} = \operatorname{sgn}(h')\kappa_\gamma\big(h(u)\big). \tag{20.227}$$

□

Lemme 20.75 ([352]).
Si γ_N est un arc paramétré normal, alors

$$\gamma_N''(s) = \kappa(s)J\gamma_N'(s). \tag{20.228}$$

Démonstration. Vu que le paramétrage est normal, $\gamma'_N \cdot \gamma'_N = 1$, et en dérivant, $\gamma''_N \cdot \gamma'_N = 0$. Donc γ''_N est un multiple de $J\gamma'$. En tenant compte du fait que le paramétrage est normal, la courbure est

$$\kappa(s) = \gamma''_N(s) \cdot J\gamma'(s). \tag{20.229}$$

En y injectant $\gamma''(s) = \lambda(s)J\gamma'(s)$ nous trouvons

$$\kappa(s) = \lambda(s)J\gamma'(s) \cdot J\gamma'(s) = \lambda(s). \tag{20.230}$$

Donc le facteur de proportionnalité est $\kappa(s)$. □

Théorème 20.76.
Soit une courbe $\gamma\colon I \to \mathbb{R}^2$ de classe C^2.

(1) γ est une partie de droite si et seulement si $\kappa(t) = 0$ pour tout t.

(2) γ est une partie d'un cercle de rayon $r > 0$ si et seulement si $|\kappa(s)| = \frac{1}{r}$.

Démonstration. Si γ est une droite, la dérivée seconde est nulle et la courbure est nulle. Supposons pour la réciproque que $\kappa(t) = 0$ pour tout t. Nous utilisant un paramétrage normal de γ, ce qui ne change pas que la courbure reste nulle. Nous avons par le lemme 20.75 que $\gamma''(s) = 0$ et donc l'existence de $a, b \in \mathbb{R}^2$ tels que $\gamma(t) = at + b$.

Passons au cas du cercle. Si γ est un cercle, le paramétrage normal est

$$\gamma(t) = R\begin{pmatrix}\cos(\frac{t}{R})\\ \sin(\frac{t}{R})\end{pmatrix} \tag{20.231a}$$

$$\gamma'(t) = \begin{pmatrix}-\sin(\frac{t}{R})\\ \cos(\frac{t}{R})\end{pmatrix} \tag{20.231b}$$

$$\gamma''(t) = -\frac{1}{R}\begin{pmatrix}\cos(\frac{t}{R})\\ \sin(\frac{t}{R})\end{pmatrix}. \tag{20.231c}$$

Avec tout cela nous avons $\kappa(s) = \gamma''(s) \cdot J\gamma'(s) = \frac{1}{R}$.

Nous supposons enfin que $\kappa(t) = 1/R$ et que le paramétrage soit normal (encore une fois, un reparamétrage ne change pas la courbure lorsqu'elle est constante). Nous définissons la courbe

$$\begin{aligned}\beta\colon [b, c] &\to \mathbb{R}^2\\ t &\mapsto \gamma(t) + rJ\gamma'(t).\end{aligned} \tag{20.232}$$

Nous avons $\beta'(t) = \gamma'(t) + rJ\gamma''(t)$. Mais par le lemme 20.75 nous avons $\gamma'' = kJ\gamma' = \frac{1}{r}J\gamma'$. Donc

$$\beta'(t) = \gamma'(t) - \gamma'(t) = 0. \tag{20.233}$$

Du coup β est constante : $\beta(t) = a$. Alors $a = \gamma(t) + rJ\gamma'(t)$ et en particulier

$$\|\gamma(t) - a\| = \|rJ\gamma'(t)\| = r. \tag{20.234}$$

Donc effectivement γ reste sur un cercle de rayon r et de centre a. □

Définition 20.77 ([352]).
*La **courbure totale** de la courbe $\gamma\colon [a, b] \to \mathbb{R}^2$ est le nombre*

$$K = \int_a^b \kappa(t)\|\gamma'(t)\|dt. \tag{20.235}$$

Lemme 20.78.
La courbure signée ne change pas sous reparamétrage positif, et change de signe sous reparamétrage négatif.

Démonstration. Soit la courbe $\gamma\colon [a,b] \to \mathbb{R}^2$ et un difféomorphisme $h\colon [a,b] \to [c,d]$. Il s'agit d'intégrer la relation (20.224) en effectuant le changement de variables $t = h(u)$:

$$K_\gamma = \int_a^b \kappa_\gamma(t)\|\gamma'(t)\|dt \tag{20.236a}$$

$$= \int_c^d \underbrace{\kappa_\gamma\big(h(u)\big)}_{\kappa_\sigma(u)} \|\gamma'\big(h(u)\big)\|h'(u)du. \tag{20.236b}$$

En utilisant le fait que $\sigma'(u) = \gamma'\big(h(u)\big)h'(u)$ nous avons alors

$$K_\gamma = \int_c^d \kappa_\sigma(u) \left\| \frac{\sigma'(u)}{h'(u)} \right\| h'(u)du \tag{20.237a}$$

$$= \int_c^d \kappa_\sigma(u)\|\sigma'(u)\| \frac{h'(u)}{|h'(u)|} du \tag{20.237b}$$

$$= \operatorname{sgn}(h')K_\sigma. \tag{20.237c}$$

□

Lemme-définition 20.79 ([352]).
Soit une courbe régulière $\gamma\colon I \to \mathbb{R}^2$ de classe C^2 et t_0 dans l'intérieur de I. Soit $\theta_0 \in \mathbb{R}$ tel que

$$\frac{\gamma'(t_0)}{\|\gamma'(t_0)\|} = \big(\cos(\theta_0), \sin(\theta_0)\big). \tag{20.238}$$

Alors il existe une unique fonction différentiable $\theta\colon I \to \mathbb{R}$ telle que $\theta(t_0) = \theta_0$ et

$$\frac{\gamma'(t)}{\|\gamma'(t)\|} = \begin{pmatrix} \cos\big(\theta(t)\big) \\ \sin\big(\theta(t)\big) \end{pmatrix} \tag{20.239}$$

pour tout $t \in I$.

*Cette fonction est l'**angle** de γ déterminé par θ_0.*

Démonstration. Soit $\beta(t) = (t, 0)$; alors $\beta'(t) = (1, 0)$ et nous avons

$$\gamma' \cdot \beta' = \gamma'_x \tag{20.240a}$$

$$\gamma' \cdot J\beta' = \gamma'_y. \tag{20.240b}$$

Par la proposition 20.71 il existe une unique fonction $\theta\colon I \to \mathbb{R}$ telle que $\theta(t_0) = \theta_0$ et

$$\begin{cases} \cos\big(\theta(t)\big) = \dfrac{\gamma' \cdot \beta'}{\|\gamma'\|\|\beta'\|} = \dfrac{\gamma'_x}{\|\gamma'\|} & (20.241a) \\ \sin\big(\theta(t)\big) = \dfrac{\gamma' \cdot J\beta'}{\|\gamma'\|\|\beta'\|} = \dfrac{\gamma'_y}{\|\gamma'\|}. & (20.241b) \end{cases}$$

Une telle fonction est bien celle que l'on demande ici. □

Lemme 20.80.
Si γ est une courbe régulière de classe C^2, alors sa courbure et son angle vérifient la relation

$$\theta'(t) = \|\gamma'(t)\|\kappa(t). \tag{20.242}$$

Démonstration. Par définition de l'angle (lemme 20.79) nous avons

$$\frac{\gamma'(t)}{\|\gamma'(t)\|} = \big(\cos\theta(t), \sin\theta(t)\big). \tag{20.243}$$

Dérivant cela,

$$\frac{\gamma''(t)}{\|\gamma'(t)\|} + \gamma'(t)\frac{d}{dt}\left(\frac{1}{\|\gamma'(t)\|}\right) = \theta'(t)\begin{pmatrix}-\sin\big(\theta(t)\big)\\ \cos\big(\theta(t)\big)\end{pmatrix} \tag{20.244a}$$

$$= \theta'(t)J\begin{pmatrix}\cos\big(\theta(t)\big)\\ \sin\big(\theta(t)\big)\end{pmatrix} \tag{20.244b}$$

$$= \theta'(t)\frac{J\gamma'(t)}{\|\gamma'(t)\|}. \tag{20.244c}$$

Nous prenons le produit scalaire de cette égalité avec $J\gamma'$ en tenant compte du fait que $\gamma' \cdot J\gamma' = 0$:

$$\frac{\gamma'' \cdot J\gamma'}{\|\gamma'\|} = \frac{\theta'}{\|\gamma'\|}J\gamma' \cdot J\gamma'. \tag{20.245}$$

En remarquant que $Jv \cdot Jv = \|v\|^2$ nous trouvons $\theta'\|\gamma'\| = \|\gamma'\|^2\kappa_\gamma$ et donc

$$\theta'(t) = \|\gamma'(t)\|\kappa_\gamma(t), \tag{20.246}$$

ce qu'il fallait prouver. □

Ce lemme nous fournit la formule attendue pour la courbure totale.

Lemme 20.81 ([352]).
Soit une courbe régulière de classe C^2 $\gamma\colon [a,b] \to \mathbb{R}^2$. Sa courbure totale est donnée par

$$K = \theta(b) - \theta(a). \tag{20.247}$$

Démonstration. Il suffit de remplacer dans la définition (20.235) de la courbure totale l'intégrante par son expression du lemme 20.80 :

$$K = \int_a^b \kappa(t)\|\gamma'(t)\|dt = \int_a^b \theta'(t)dt = \theta(b) - \theta(a) \tag{20.248}$$

par le théorème 13.236. □

20.12.3 Degré, indice et homotopie

Définition 20.82 ([352]).
*Le **nombre de tours** d'une courbe fermée de classe C^1 $\gamma\colon [a,b] \to \mathbb{R}^2$ est le nombre*

$$\mathrm{Turn}(\gamma) = \frac{1}{2\pi}\int_a^b \kappa(t)\|\gamma'(t)\|dt \tag{20.249}$$

où κ est la courbure signée définie en 20.73.

Lemme-définition 20.83.
*Soit une application continue $\phi\colon S^1 \to S^1$. Un **relèvement** de σ est une application $\varphi\colon \mathbb{R} \to \mathbb{R}$ une application continue telle que*

$$\phi\big(\cos(t), \sin(t)\big) = \begin{pmatrix}\cos\big(\varphi(t)\big)\\ \sin\big(\varphi(t)\big)\end{pmatrix}. \tag{20.250}$$

*Le **degré** est l'entier $\deg(\phi)$ tel que*

$$\varphi(2\pi) - \varphi(0) = 2\deg(\phi)\pi. \tag{20.251}$$

Ce nombre ne dépend pas du choix du relèvement φ.

Démonstration. Soient φ_1 et φ_2, deux applications qui satisfont les contraintes. Nous avons une fonction continue $n\colon S^1 \to \mathbb{R}$ telle que

$$\varphi_1(t) - \varphi_2(t) = 2\pi n(t). \tag{20.252}$$

La fonction n ne pouvant prendre que des valeurs entières et étant continue, elle est constante. Par conséquent

$$\varphi_1(2\pi) - \varphi_1(0) = \varphi_2(2\pi) - \varphi_2(0), \tag{20.253}$$

ce que nous voulions. □

Définition 20.84.
Soit une courbe fermée $\gamma\colon [0, L] \to \mathbb{R}^2$ de classe C^1. Nous posons

$$\Phi_\gamma(t) = \frac{\tilde{\gamma}'(t)}{\|\tilde{\gamma}'(t)\|} \tag{20.254}$$

*où $\tilde{\gamma}(u) = \gamma\big(\frac{L}{2\pi}u\big)$. L'**indice de rotation** de γ est le degré de Φ_γ, c'est-à-dire*

$$\mathrm{Ind}(\gamma) = \deg(\Phi_\gamma). \tag{20.255}$$

Notons que dans cette définition, Φ_γ n'est rien d'autre que le vecteur unitaire tangent à γ, ramené à $[0, 2\pi]$.

Proposition 20.85.
Pour une courbe fermée de classe[10] *C^2, l'indice de rotation est égal au nombre de tours.*

Démonstration. Soit $\gamma\colon [0, L] \to \mathbb{R}^2$ une courbe vérifiant les hypothèses. Par définition, $\mathrm{Ind}(\gamma) = \deg(\Phi_\gamma)$ où

$$\Phi_\gamma(t) = \frac{\tilde{\gamma}'(t)}{\|\tilde{\gamma}'(t)\|} = \begin{pmatrix} \cos\tilde{\theta}(t) \\ \sin\tilde{\theta}(t) \end{pmatrix} \tag{20.256}$$

où nous avons noté $\tilde{\theta}$ l'angle tournant[11] de $\tilde{\gamma}$ et θ celui de γ. Vu que $\tilde{\theta}$ vérifie les hypothèses de la définition 20.83 nous pouvons calculer le degré de Φ_γ par

$$\deg(\Phi_\gamma) = \frac{1}{2\pi}\big(\tilde{\theta}(2\pi) - \tilde{\theta}(0)\big) = \frac{1}{2\pi}\int_0^{2\pi} \tilde{\theta}'(s)ds. \tag{20.257}$$

Il faut trouver le lien entre θ et $\tilde{\theta}$. Pour cela nous notons que

$$\frac{\tilde{\gamma}'(t)}{\|\tilde{\gamma}'(t)\|} = \frac{\gamma'\left(\frac{L}{2\pi}t\right)}{\|\gamma'\left(\frac{L}{2\pi}\right)\|}. \tag{20.258}$$

En comparant avec (20.256) il vient

$$\tilde{\theta}(t) = \theta\big(\frac{L}{2\pi}t\big) \tag{20.259}$$

et

$$\tilde{\theta}'(s) = \frac{L}{2\pi}\theta'\left(\frac{L}{2\pi}s\right). \tag{20.260}$$

Le changement de variables $t = \frac{L}{2\pi}s$ est donc tout vu dans l'intégrale (20.257) :

$$\deg(\Phi_\gamma) = \frac{1}{2\pi}\int_0^{2\pi} \frac{L}{2\pi}\theta'\big(\frac{L}{2\pi}s\big)ds = \frac{1}{2\pi}\int_0^L \theta'(t)dt = \frac{1}{2\pi}\int_0^L \kappa_\gamma(t)\|\gamma'(t)\|dt = \mathrm{Turn}(\gamma) \tag{20.261}$$

où nous avons aussi utilisé le lemme 20.80 qui donne le lien entre θ' et κ. □

10. La dérivée seconde arrive dans la définition de la courbure ; il faudrait donc supposer au moins C^2 pour avoir la continuité de la courbure.
11. Définition 20.79.

Définition 20.86 (homotopie de chemins fermés)**.**
Les courbes fermées $\gamma_0, \gamma_1 \colon [0, L] \to Y$ (Y est un espace topologique) sont ***homotopes*** *s'il existe une application continue*

$$F \colon [0,1] \times [0, L] \to Y \tag{20.262}$$

telle que

(1) $F(0,t) = \gamma_0(t)$ *pour tout* t*,*
(2) $F(1,t) = \gamma_1(t)$ *pour tout* t*,*
(3) $F(u, L) = F(u, 0)$ *pour tout* u*.*

L'application F est l'homotopie entre γ_0 et γ_1.

Proposition 20.87 (Homotopie, degré et indice[352])**.**
Il y a deux résultats à ne pas confondre.

(1) Si $\phi_i \colon S^1 \to S^1$ sont homotopes, alors $\deg(\phi_1) = \deg(\phi_2)$*.*
(2) Si $\gamma_i \colon [0, L] \to \mathbb{R}^2$ sont homotopes, alors $\mathrm{Ind}(\phi_1) = \mathrm{Ind}(\phi_2)$*.*

Démonstration. Nous allons décomposer la preuve en deux parties.

Le degré pour les applications $S^1 \to S^1$ Soient deux applications homotopes $\gamma_i \colon S^1 \to S^1$. Si F est l'homotopie[12] entre γ_0 et γ_1, nous posons $\gamma_u(t) = F(u,t)$ qui est encore une courbe fermée $\gamma_u \colon S^1 \to S^2$. Nous pouvons donc considérer le degré de γ_u. C'est un entier n_u qui vérifie

$$\tilde{\gamma}_u(2\pi) - \tilde{\gamma}_u(0) = 2\pi n_u. \tag{20.263}$$

Le membre de gauche est une fonction continue de u ; dans le membre de droite n_u ne pouvant prendre que des valeurs entières, elle est alors constante.

Indice de rotation pour des courbes dans $\mathbb{R}^2$ Soient maintenant γ_0 et γ_1 deux courbes homotopes de $[0, L]$ dans $\mathbb{R}^2$. Par définition, $\mathrm{Ind}(\gamma_i) = \deg(\Phi_{\gamma_i})$. Prouvons alors que Φ_{γ_0} et homotope à Φ_{γ_1}. De cette façon, la première partie de la preuve conclura à

$$\mathrm{Ind}(\gamma_0) = \deg(\Phi_{\gamma_0}) = \deg(\Phi_{\gamma_1}) = \mathrm{Ind}(\gamma_1). \tag{20.264}$$

Nous savons que

$$\Phi_\gamma = \frac{\gamma'\left(\frac{L}{2\pi}t\right)}{\left\|\gamma'\left(\frac{L}{2\pi}t\right)\right\|}, \tag{20.265}$$

donc en posant

$$\Phi_F(u,t) = \frac{\gamma_u'\left(\frac{L}{2\pi}t\right)}{\left\|\gamma_u'\left(\frac{L}{2\pi}t\right)\right\|} \tag{20.266}$$

nous avons une homotopie entre Φ_{γ_0} et Φ_{γ_1}.

□

Théorème 20.88 ([352])**.**
Le nombre de tours d'une courbe simple fermée de classe C^2 dans $\mathbb{R}^2$ est ± 1.

Démonstration. Soit une telle courbe γ et son image Γ.

Choix d'un point et d'une tangente Si ℓ est une droite dans le plan, soit p le point de Γ le plus proche de ℓ. Il n'est peut-être pas unique, mais la parallèle à ℓ passant par p est une tangente à Γ telle que tout Γ se trouve d'un seul côté de ℓ_p. Pour voir cela, il suffit de choisir un système d'axes pour lequel ℓ est l'axe $y = 0$. La distance entre ℓ et les points de Γ est donnée par γ_y, et donc les extrémums sont atteints là où $\gamma_y' = 0$, c'est-à-dire pour les points sur lesquels la tangente est parallèle à ℓ.

Étant donné qu'il y a (au moins) un maximum et un minimum distincts, pour pouvons choisir le point p de telle sorte que pour tout point $q \in \Gamma$, l'angle du vecteur $q - p$ soit entre α_0 et

12. Définition 20.86.

$\alpha_0 + \pi$ et non entre $\alpha_0 - pi$ et α_0. Ce choix revient à choisir p de telle sorte que Γ soit d'un côté ou de l'autre de ℓ_p.

Pour simplifier les notations plus tard nous choisissons ℓ_p horizontale, de telle sorte que $\alpha_0 = 0$, et que Γ est au dessus de ℓ_p. Les angles des vecteurs $q - p$ pour $q \in \Gamma$ sont donc tous entre 0 et π.

Définition de Σ Soit L la longueur[13] de γ et β, un paramétrage de γ telle que $\beta(0) = p$. Nous considérons le triangle

$$\mathcal{T} = \{(t_1, t_2) \in \mathbb{R}^2 \text{ tel que } 0 \leqslant t_1 \leqslant t_2 \leqslant L\} \tag{20.267}$$

et l'application sécante $\Sigma\colon \mathcal{T} \to S^1$ définie par

$$\Sigma(t_1, t_2) = \begin{cases} \dfrac{\beta'(t)}{\|\beta'(t)\|} & \text{si } t_1 = t_2 = t \\ -\dfrac{\beta'(0)}{\|\beta'(0)\|} & \text{si } t_1 = 0 \text{ et } t_2 = L \\ \dfrac{\beta(t_2) - \beta(t_1)}{\|\beta(t_2) - \beta(t_1)\|} & \text{sinon.} \end{cases} \tag{20.268}$$

Continuité de Σ Nous devons prouver les limites[14] suivantes :

$$\lim_{\substack{(t_1,t_2)\to(t,t)\\ 0\leqslant t_1\leqslant t\leqslant t_2\leqslant L}} \frac{\beta(t_1) - \beta(t_2)}{\|\beta(t_1) - \beta(t_2)\|} = \frac{\beta'(t)}{\|\beta'(t)\|} \tag{20.269a}$$

$$\lim_{\substack{(t_1,t_2)\to(0,L)\\ 0\leqslant t_1\leqslant t\leqslant t_2\leqslant L}} \frac{\beta(t_1) - \beta(t_2)}{\|\beta(t_1) - \beta(t_2)\|} = -\frac{\beta'(0)}{\|\beta'(0)\|} \tag{20.269b}$$

Nous commençons par (20.269a) en multipliant et divisant par $t_2 - t_1$:

$$\frac{\beta(t_2) - \beta(t_1)}{\|\beta(t_2) - \beta(t_1)\|} = \frac{\beta(t_2) - \beta(t_1)}{t_2 - t_1} \frac{t_2 - t_1}{\|\beta(t_2) - \beta(t_1)\|}. \tag{20.270}$$

Si chacune des limites des deux facteurs existent dans $\mathbb{R}$, la limite du produit sera le produit des limites. Pour le premier facteur nous développons $\beta(t_2)$ autour de $t = t_1$ via la formulation (11.1028) :

$$\beta(t_2) = \beta(t_1) + (t_2 - t_1)\beta'(t_1) + \alpha(t_2 - t_1) \tag{20.271}$$

où α est une fonction ayant la propriété $\lim_{t\to 0} \frac{\alpha(t)}{t} = 0$. Nous avons à calculer la limite de

$$\frac{(t_1 - t_1)\beta'(t_1) + \alpha(t_2 - t_1)}{t_2 - t_1} = \beta'(t_1) + \frac{\alpha(t_2 - t_1)}{t_2 - t_1}. \tag{20.272}$$

Prendre la limite $(t_1, t_2) \to (t, t)$ donne bien $\beta'(t)$ parce que β est de classe C^1. En ce qui concerne la limite de la deuxième partie, nous allons la faire plus en détail pour la limite (20.269b).

La limite (20.269b). Nous prenons la prolongation périodique de β. Alors si $t_2 = L - \epsilon$ nous pouvons écrire $\beta(-\epsilon)$ au lieu de $\beta(t_2)$. Nous développons $\beta(t_1)$ autour de $t = -\epsilon$ (parce que t_1 est petit) :

$$\beta(t_1) = \beta(-\epsilon) + (t_1 + \epsilon)\beta'(-\epsilon) + \alpha(t_1 + \epsilon). \tag{20.273}$$

Après multiplication et division par $t_1 + \epsilon$, la première limite à calculer est celle de

$$\frac{\beta(-\epsilon) - \beta(t_1)}{t_1 + \epsilon} = -\beta'(-\epsilon) + \frac{\alpha(t_1 + \epsilon)}{t_1 + \epsilon}, \tag{20.274}$$

13. Définition 20.4.
14. Limite au sens de la définition 7.67, en sachant qu'elle est unique par la proposition 7.70.

pour $(\epsilon, t_1) \to (0,0)$. Cela donne bien $-\beta'(0)$. La seconde limite à calculer est celle de

$$\frac{t_1+\epsilon}{\|\beta(-\epsilon)-\beta(t_1)\|} = \frac{t_1+\epsilon}{\|-(t_1+\epsilon)\beta'(-\epsilon)-\alpha(t_1+\epsilon)\|} = \frac{t_1+\epsilon}{\|(t_1+\epsilon)\beta'(-\epsilon)+\alpha(t_1+\epsilon)\|}. \tag{20.275}$$

Nous calculons la limite de l'inverse (qui, si elle est non nulle donnera la réponse en inversant à nouveau) en nous souvenant de la formule

$$\big|a-|b|\big| \leqslant |a+b| \leqslant \big|a+|b|\big|. \tag{20.276}$$

Nous avons l'encadrement

$$\frac{\big|(t_1+\epsilon)\beta'(-\epsilon)-|\alpha(t_1+\epsilon)|\big|}{t_1+\epsilon} \leqslant \frac{\|(t_1+\epsilon)\beta'(-\epsilon)+\alpha(t_1+\epsilon)\|}{t_1+\epsilon} \tag{20.277a}$$

$$\leqslant \frac{\big|(t_1+\epsilon)\beta'(-\epsilon)+|\alpha(t_1+\epsilon)|\big|}{t_1+\epsilon} v \tag{20.277b}$$

Les limites des deux extrêmes existent et valent $\beta'(0)$; la règle de l'étau 11.158 conclu.

Deux chemins homotopes Nous considérons dans $\mathcal{T}$ les points $A=(0,0)$, $B=(0,L)$ et $C=(L,L)$. Nous allons considérer les chemins direct de A à C et celui passant via B. Et comme ces chemins doivent être paramétrés de 0 à 2π, il faut faire un peu attention. Nous définissons les chemins

$$\sigma_i \colon S^1 \to S^1 \tag{20.278}$$

par

$$\sigma_1(t) = \Sigma\left(\frac{L}{2\pi}t, \frac{L}{2\pi}t\right) = \frac{\beta'(tL/2\pi)}{\|\beta'(tL/2\pi)\|} = \frac{\tilde{\beta}'(t)}{\|\tilde{\beta}'(t)\|} = \Phi_\beta \tag{20.279a}$$

$$\sigma_2(t) = \begin{cases} \Sigma(0, \frac{L}{\pi}t) & \text{si } 0 \leqslant t \leqslant \pi \\ \Sigma\big(\frac{L}{\pi}(t-\pi), L\big) & \text{si } \pi \leqslant t \leqslant 2\pi \end{cases} \tag{20.279b}$$

où nous avons repris les notations de la définition 20.84. Notons que pour σ_2, en $t=\pi$ les deux expressions donnent

$$\sigma_2(\pi) = \Sigma(0,L) = -\frac{\beta'(0)}{\|\beta'(0)\|}. \tag{20.280}$$

Ce sont des chemins fermés parce que

$$\sigma_1(2\pi) = \Sigma(L,L) = \frac{\beta'(L)}{\|\beta'(L)\|} = \frac{\beta'(0)}{\|\beta'(0)\|} = \Sigma(0,0) = \sigma_1(0). \tag{20.281}$$

Notons que dans toutes ces définitions et calculs, nous avons utilisé de façon assez cruciale la définition 20.70 pour définir la dérivée de β en $t=0$.

Les chemins σ_1 et σ_2 sont homotopes par construction.

Indices et degrés La proposition 20.87(1) nous donne $\deg(\sigma_1) = \deg(\sigma_2)$. Nous avons alors la chaine d'égalités

$$\deg(\sigma_2) = \deg(\sigma_1) = \deg(\Phi_\beta) = \mathrm{Ind}(\beta) = \mathrm{Ind}(\gamma) = \mathrm{Turn}(\gamma) \tag{20.282}$$

où les justifications sont :

(1) $\deg(\sigma_2) = \deg(\sigma_1)$ par homotopie : proposition 20.87.

(2) $\deg(\sigma_1) = \deg(\Phi_\beta)$ parce que $\sigma_1 = \Phi_\beta$.

(3) $\deg(\Phi_\beta = \mathrm{Ind}(\beta)$ par définition 20.84 de l'indice.

(4) $\mathrm{Ind}(\beta) = \mathrm{Ind}(\gamma)$ par invariance de l'indice sous reparamétrage.

(5) $\mathrm{Ind}(\gamma) = \mathrm{Turn}(\gamma)$ par la proposition 20.85.

Il nous reste à montrer que $\deg(\sigma_2) = \pm 1$.

La géométrie de σ_2 Notons que par définition les valeurs de $\sigma_2(t)$ sont les vecteurs (unitaires) joignant p aux points de Γ lorsque $0 \leqslant t \leqslant \pi$, et les vecteurs inverses pour $\pi \leqslant t \leqslant 2\pi$. Plus précisément nous avons, si $0 < a < \pi$:

$$\sigma_2(a) = \Sigma(0, \frac{L}{\pi}a) = \frac{\beta(\frac{L}{\pi}a) - \beta(L)}{\|\beta(\frac{L}{\pi}a) - \beta(L)\|} \tag{20.283}$$

et

$$\sigma_2(\pi + a) = \Sigma(\frac{L}{\pi}a, L) = \frac{\beta(L) - \beta(\frac{L}{\pi}a)}{\|\beta(L) - \beta(\frac{L}{\pi}a)\|}. \tag{20.284}$$

En sachant que $\beta(0) = \beta(L)$ nous avons alors

$$\sigma_2(\pi + a) = -\sigma_2(a). \tag{20.285}$$

Notons aussi que le fait que γ soit une courbe simple assure que le numérateur et le dénominateur de σ_2 ne s'annulent pas autrement que pour $t = L$ ou $t = 0$.

Un relèvement pour σ_2 Ces propriétés motivent cette idée pour le relèvement de σ_2 :

$$\begin{aligned} \theta\colon [0, 2\pi[&\to [0, 2\pi[\\ t &\mapsto \text{l'angle du vecteur } \sigma_2(t), \end{aligned} \tag{20.286}$$

avec $\theta(2\pi)$ définit par continuité. Nous allons cependant voir, en étant plus prudent, que cette définition n'assure pas la continuité (surtout en $t = \pi$).

Soyons donc plus prudent et construisons θ petit à petit.

Étant donné que ℓ_p est tangente à Γ au point p, la droite ℓ_p est parallèle à $\beta'(0)$, et l'angle entre ℓ_p et $\beta'(0)$ est soit 0 soit π. Donc $\theta(0)$ devrait valoir soit 0 soit π.

Nous commençons par définir ceci :

$$\begin{aligned} \theta\colon [0, \pi[&\to [0, \pi] \\ t &\mapsto \text{angle de } \sigma_2(t) = \arccos\big(\sigma_2(t)_x\big). \end{aligned} \tag{20.287}$$

C'est le choix d'avoir ℓ_p horizontale et Γ au dessus de ℓ_p qui nous assure que pour tout $t \in [0, \pi]$, l'angle de $\sigma_2(t)$ peut être choisi entre 0 et π. De plus cette fonction est continue en tant que partie de la fonction $\arccos\colon [-1, 1] \to [0, \pi]$ qui est elle-même continue par la proposition 7.83.

Le fait que θ soit continue est assuré par le fait que σ_2 est C^∞.

Si $\theta(0) = 0$ Cela correspond à la situation des vecteurs rouges sur la figure 20.11.

Vu que $\sigma_2(\pi) = -\sigma_2(0)$, l'angle de σ_2 en $t = \pi$ est le supplémentaire de celui en $t = 0$. Mais pour $0 \leqslant t < \pi$, $\theta(t)$ prend ses valeurs entre 0 et π, le seul supplémentaire de 0 à être disponible est $\theta(\pi) = \pi$ (et non $\theta(\pi) = -\pi$ par exemple). Nous définissons donc $\theta(\pi) = \pi$ pour la continuité.

En ce qui concerne les $\pi < t < 2\pi$ nous savons que $\sigma_2(\pi + a) = -\sigma_2(a)$. Les angles sont donc les supplémentaires de ceux pour $0 \leqslant t \leqslant \pi$. Pour assurer la continuité en $t = pi$ nous sélectionnons la place $]\pi, 2\pi]$; et nous définissons

$$\begin{aligned} \theta\colon]\pi, 2\pi[&\to]\pi, 2\pi] \\ t &\mapsto \text{angle de } \sigma_2(t). \end{aligned} \tag{20.288}$$

Le nombre $\theta(2\pi)$ est défini par continuité. Il doit valoir $\theta(2\pi) = 2\pi$.

La fonction θ ainsi définie est un relèvement pour σ_2, et le degré peut être calculé :

$$\deg(\sigma_2) = \frac{1}{2\pi}\big(\theta(2\pi) - \theta(0)\big) = 1. \tag{20.289}$$

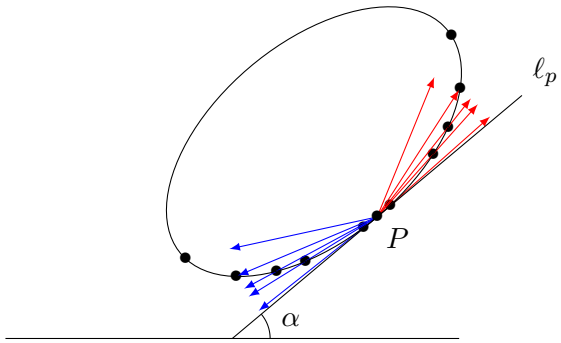


FIGURE 20.11 – Les vecteurs représenant σ_2 dans le cas où $\beta'(0)$ est dans le sens de ℓ_p ou dans le sens inverse. Pour le sport nous avons dessiné la situation avec une droite ℓ quelconque plutôt que horizontale.

Si $\theta(0) = \pi$ Cela correspond à la situation des vecteurs bleus sur la figure 20.11.

Alors $\theta(\pi) = 0$ est obligatoire parce qu'il doit être supplémentaire à $\theta(0)$. Les angles atteints par $\sigma_2(t)$ pour $t \in]\pi, 2\pi[$ sont encore les complémentaires, mais cette fois la continuité en $t = \pi$ nous impose de les chercher dans $]-\pi, 0]$ et nous définissons

$$\begin{aligned} \theta\colon]\pi, 2\pi[&\to]-\pi, 0] \\ t &\mapsto \text{angle de } \sigma_2(t). \end{aligned} \tag{20.290}$$

Par continuité nous devons avoir $\theta(2\pi) = \theta(\pi) - \pi$ et donc $\theta(2\pi) = -\pi$.

Le degré de σ_2 est alors

$$\deg(\sigma_2) = \frac{1}{2\pi}\big(\theta(2\pi) - \theta(0)\big) = -1. \tag{20.291}$$

Conclusion Nous avons montré que le degré de σ_2 est 1 ou -1, et en remontant les égalités (20.282) nous déduisons que $\mathrm{Turn}(\gamma) = \pm 1$.

□

20.13 Courbes fermées planes

20.13.1 Cercle circonscrit

La proposition suivante est dans le même esprit que l'ellipse de John-Loewer [15].

Proposition-définition 20.89 (Cercle circonscrit[406, 1]).
*Soit une courbe fermée simple et continue $\gamma\colon [0,1] \to \mathbb{R}^2$. Soit Γ son image. Il existe un unique cercle de rayon minimum contenant Γ. Ce cercle est le **cercle circonscrit** à γ.*

Il a les propriétés suivantes :

(1) Le cercle circonscrit à γ coupe Γ en au moins deux points distincts.

(2) Tout arc du cercle circonscrit plus grand que le demi-cercle intersection Γ.

Démonstration. Division de la preuve.

Existence L'application γ étant continue, l'ensemble Γ est compact (théorème 7.92). Nous considérons l'ensemble Q des formes quadratiques de la forme

$$q_{a,r}(x) = \|a - x\|^2 - r^2 \tag{20.292}$$

avec $a \in \mathbb{R}^2$ et $r \in \mathbb{R}$. Nous mettons sur cet ensemble la topologie de $\mathbb{R}^2 \times \mathbb{R} = \mathbb{R}^3$. Le nombre $q_{a,r}(x)$ est continu en a, r et x. Soit A l'ensemble des formes quadratiques de cette forme et vérifiant

$$q\big(\gamma(t)\big) \leqslant 0 \tag{20.293}$$

15. Proposition 16.117.

pour tout $t \in [0,1]$.

Cet ensemble A est non vide parce que Γ est compact et donc borné ; il existe donc une boule qui contient Γ en son intérieur.

Montrons que A est fermé dans Q. Si $q_{a,r} \notin A$ alors il existe t_0 tel que $q_{a,r}\big(\gamma(t_0)\big) > 0$. Par continuité, il existe un voisinage de (a,r) dans $\mathbb{R}^3$ et donc de $q_{a,r}$ dans Q tel que $q_{a',r'}\big(\gamma(t_0)\big)$ reste strictement positif pour tout (a',r') dans ce voisinage.

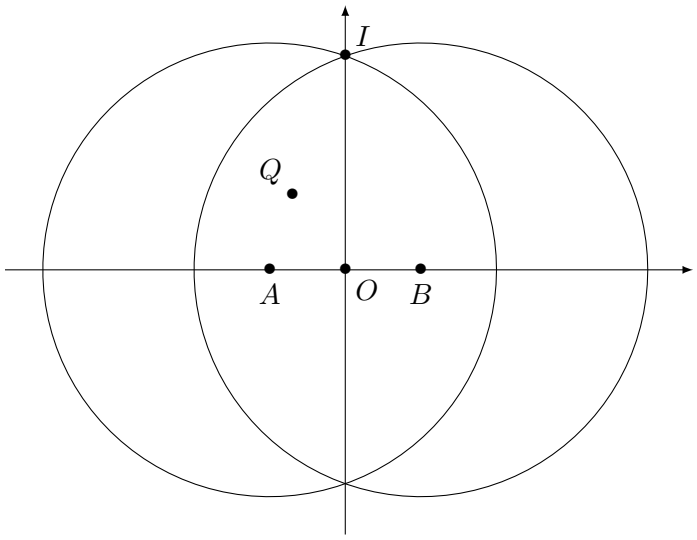
En particulier l'ensemble

$$\{r \in \mathbb{R}^+ \text{ tel que } \exists\,(a,r) \text{ tel que } q_{a,r} \in A\} \tag{20.294}$$

est fermé et borné vers le base. De plus $r = 0$ n'est pas dans cet ensemble. Il possède donc un minimum strictement positif. Le cercle correspondant donne l'existence.

Unicité En ce qui concerne l'unicité, si Γ est contenu dans les boules $B(a,R)$ et $B(b,R)$ alors

$$\Gamma \subset B(a,R) \cap B(b,R). \tag{20.295}$$



Nous choisissons les axes comme indiqué sur le dessin et nous montrons que l'intersection est dans le cercle de centre O et de rayon $\|OI\| < R$. Soit Q un point de l'intersection ; par symétrie il est suffisant de supposer $Q_x < 0$ et $Q_y > 0$. Vu que Q est dans le cercle de centre B et de rayon R, il doit satisfaire

$$(B_x - Q_x)^2 + Q_y^2 < R. \tag{20.296}$$

D'autre part le point I est d'abscisse $I_x = 0$ et d'ordonnée donnée par $I_y^2 = R^2 - B_x^2$.

Nous devons prouver que $Q_x^2 + Q_y^2 \leqslant I_y^2$. Il s'agit simplement de calculer

$$Q_x^2 + Q_y^2 \leqslant Q_x^2 + R^2 - B_x^2 + 2B_xQ_x - Q_x^2 = R^2 - B_x^2 + 2B_xQ_x \leqslant R^2 - B_x^2 = I_y^2 \tag{20.297}$$

parce que $B_x > 0$ et $Q_x < 0$.

Nous concluons que Γ est inclus dans un cercle de rayon plus petit que R et donc que R n'est pas minimum. D'où l'unicité.

Au moins deux intersections Nous nommons C le cercle circonscrit à γ, et nous écrivons, pour $p \in \Gamma$

$$r(p) = d(p,C) \tag{20.298}$$

la distance entre p et C. Cela est une fonction continue sur le compact Γ. Elle atteint donc ses bornes sur Γ.

Si $C \cap \Gamma = \varnothing$ alors $r(p) > 0$ pour tout p et le minimum est également strictement positif. Soit r_0 ce minimum. Alors le cercle C_2 même centre que C mais de rayon $r_0/2$ n'intersecte pas non plus Γ parce que

$$d(p,C) \leqslant d(p,C_2) + d(C_2,C) \tag{20.299}$$

où $d(p,C) = r(p)$ et $d(C_2,C) = r_0/2$.

$$d(p,C_2) \geqslant r(p) - \frac{r_0}{2} \geqslant \frac{r_0}{2} > 0. \tag{20.300}$$

Si $C \cap \Gamma = \{q\}$ alors un choix d'axe place le centre de C en $(0,0)$, le point q en $(1,0)$ et fixe le rayon de C à 1. Pour $p \in \Gamma$ nous notons $r(p)$ la distance entre p et le point de C situé sur sa gauche, c'est-à-dire, si $p = (p_x, p_y)$,

$$r(p) = \sqrt{1 - p_y^2} + p_x. \tag{20.301}$$

Cela est encore une fonction continue sur Γ qui atteint son minimum valant r_0. Alors le cercle de centre $(\frac{r_0}{2}, 0)$ et de même rayon contient encore Γ mais n'a plus de points d'intersection avec Γ.

Enfin, tout nombre de points d'intersection entre C et Γ est possible à partir de 2. Pour en avoir deux, prendre une ellipse, et pour en avoir plus, prendre des polynômes dont les angles sont un peu modifiés de façon à rester C^1.

Intersection avec les demi-arcs Supposons, en fixant encore les axes, que le cercle circonscrit soir encore centré en $(0,0)$ et que Γ n'intersecte pas le demi-cercle $x < 0$. Alors pour tout $p \in \Gamma$ la distance entre p et ce demi-cercle est strictement positive. Il y a un minimum r_0. En décalant le centre du cercle de $r_0/2$ vers la droite, nous obtenons un nouveau cercle contenant Γ mais ne l'intersectant pas.

□

20.13.2 Description locale

Définition 20.90.
Une courbe plane différentiable est ***convexe*** *si son graphe est en tout point d'un seul côté de sa tangente.*

Lemme 20.91 ([1])**.**
Soit une courbe simple, convexe $\gamma\colon [0,1] \to \mathbb{R}^2$ que nous supposons être de classe C^1. Nous notons Γ l'image de γ. Alors Γ est localement le graphe d'une fonction convexe (définition 16.76).

Démonstration. Soit $p \in \Gamma$ et l_p la tangente à Γ en p. Nous considérons un système d'axe centré en p de telle sorte que l_p soit l'axe des abscisses et l'axe des ordonnées soit dirigé de tel manière que Γ se trouve dans la partie $y > 0$. De plus nous paramétrons γ de telle sorte à avoir $\gamma(0) = p = (0,0)$.

Vu que $l_p \equiv y = 0$ est la tangente à Γ nous avons $\gamma_y' = 0$ et $\gamma_x'(0) > 0$. Nous en déduisons que $\gamma_x'(t) > 0$ pour tout $t \in B(0,\delta)$ pour δ suffisamment petit. Nous posons alors

$$g(x) = \gamma_y\big(\gamma_x^{-1}(x)\big) \tag{20.302}$$

qui est bien définie parce que γ_x est une bijection entre $B(0,\delta)$ et son image. La fonction g est continue par la proposition 7.84 et même dérivable par la proposition 11.121. De plus, vu la formule (11.277), la fonction g^{-1} est de classe C^1 parce que $(g^{-1})'$ est une composée d'applications continues.

Si g est C^2 Dans ce cas, g'' ne peut pas changer de signe, sinon la tangente coupe le graphe. Par positivité de g (et le fait que $g(0) = g'(0) = 0$), il n'est pas possible d'avoir $g'' < 0$ partout. Donc $g'' \geqslant 0$ partout. Cela prouve que g est convexe par la caractérisation 16.82.

Si g est seulement de classe C^1 Le graphe de g correspond au graphe de Γ. Nous montrons que g est convexe en utilisant la caractérisation de la proposition 16.87.

La tangente au graphe de g en $x = x_0$, que nous notons l_0, est la tangente à Γ en $t = \gamma_x^{-1}(x_0)$. Le graphe de g, qui est une partie de Γ se trouve donc d'un seule côté de l_0.

Nous nous restreignons g à un compact I et nous considérons la fonction

$$d_a(x) = g(x) - l_a(x) \tag{20.303}$$

qui donne la distance entre le graphe de g et la tangente à g en $x = a$. Cela est une fonction continue en x et en a. Le graphe de g est au dessus de la tangente en $x = 0$ (par construction des axes). Supposons que le graphe de g soit en dessous de la tangente en $x = x_2$. Alors nous avons, pour tout x :

$$\begin{cases} d_0(x) \geqslant 0 & (20.304a) \\ d_{x_1}(x) \leqslant 0. & (20.304b) \end{cases}$$

Nous posons

$$s(a) = \sup_{x \in I} d_a(x), \tag{20.305}$$

qui est une fonction continue par la proposition 11.31. Vu le définitions, $s(0) \geqslant 0$ et $s(x_2) \leqslant 0$. Il existe donc $m \in [0, m_2]$ tel que $s(m) = 0$. À ce moment nous avons $g(x) = l_m(x)$ pour tout $x \in I$ et donc g est une droite, et en réalité toutes les inégalités sont des égalités. La fonction g est alors bien convexe (mais pas strictement).

□

20.13.3 Enveloppe convexe

Proposition 20.92 ([407, 1]).
Soit une courbe simple, fermée et convexe $\gamma\colon [0,1] \to \mathbb{R}^2$ que nous supposons être de classe C^2. Nous notons Γ l'image de γ. Alors il existe un convexe D tel que $\partial D = \Gamma$.

Démonstration. Pour $p \in \Gamma$ nous notons l_p la tangente à Γ en p et H_p le demi-plan (fermé) contenant Γ. Nous posons

$$D = \bigcap_{p \in \Gamma} H_p. \tag{20.306}$$

Cet ensemble est convexe comme intersection de convexes et fermé comme intersection de fermés. Nous prouvons que $\Gamma = \partial D$.

L'inclusion $\Gamma \subset \partial D$ est la plus facile. Si $p \in \Gamma$ alors p est dans chacun des H_q et donc dans D. De plus tout voisinage de p contient des points en dehors de H_p, donc p n'est pas dans l'intérieur de D. Ce dernier étant fermé, un point hors de l'intérieur est sur le bord. Ergo $p \in \partial D$.

Pour l'inclusion inverse, soit $p \in \partial D$.

Il existe q tel que $p \in l_q$ Vu que D est fermé, le point p est dans D, et donc dans tous les H_q. Supposons qu'il soit dans l'intérieur de tous les H_q. Alors nous considérons la fonction

$$r(q) = \frac{d(p, H_q)}{2} \tag{20.307}$$

définie sur Γ. C'est une fonction continue [16] strictement positive définie sur le compact Γ qui possède donc un minimum strictement positif. Si r_0 est ce minimum, alors $B(p, r_0)$ est inclue à tous les H_q, ce qui ferait que p est à l'intérieur de D. Nous concluons à l'existence de $q \in \Gamma$ tel que $p \in l_q$.

Le point où ça décolle Nous supposons que $p \notin \Gamma$, sinon ce serait trop facile. Nous paramétrons γ de telle sorte à avoir $q = \gamma(0)$ et nous posons

$$T = \{t \in \mathbb{R}^+ \text{ tel que } l_{\gamma(t)} = l_q\}. \tag{20.308}$$

Cela est un fermé dans $\mathbb{R}$ parce que γ est de classe C^1. Nous posons $t_0 = \inf(T^c)$ et pour tout ϵ suffisamment petit,

$$\begin{cases} l_{\gamma(t_0+\epsilon)} \neq l_q, & (20.309a) \\ l_{\gamma(t_0-\epsilon)} = l_q. & (20.309b) \end{cases}$$

La seconde est parce que si $l_{\gamma(t_0-\epsilon)} \neq l_q$ nous aurions $t_0 - \epsilon \in T^c$. Soit $r = \gamma(t_0)$; nous avons $l_r = l_q$ parce que si $l_r \neq l_q$ alors par continuité de γ' nous aurions $l_{r'} \neq l_q$ pour tout $r' \in \gamma\big(B(t_0, \delta)\big)$.

16. L'équation de la droite l_q a des coefficients continus parce que γ est de classe C^1.

Graphe d'une fonction strictement convexe En suivant le lemme 20.91, l'ensemble Γ est localement (autour de r) le graphe d'une fonction convexe au-dessus de l_r. Soit $g\colon B(0,\delta) \to \mathbb{R}^2$ cette fonction convexe.

Soit $\epsilon > 0$. Si $g''(x) = 0$ sur $]0,\epsilon]$ alors g' y est constante. Mais $g'(0) =$, ce qui signifierait que sur $[0,\epsilon]$ nous ayons $g'(x) = 0$ et donc $g(x) = 0$. Cela ferait que $l_{r'} \equiv y = 0$ pour tout r' de la forme $(x,0)$ avec $x \in [0,\epsilon]$ (qui sont des points de Γ). Cela est en contradiction avec la définition de r. Donc il existe un point $x \in [0,\epsilon[$ tel que $g''(x) > 0$.

Rappelons que $p \in l_q$, ce qui fait que p a pour coordonnées $(p_x, 0)$. Nous restreignons δ et ϵ de telle sorte que p_x soit plus grand à la fois que ϵ et δ.

Il existe donc un intervalle $[a,b]$ avec $a,b \geqslant 0$ et $a,b < p_x$ sur lequel g est strictement convexe.

La tangente qui tue En particulier $g(b) > 0$ et le théorème des accroissements finis 11.130(1) nous donne l'existence de $m \in [0,b]$ tel que la tangente à g en $x = m$ est parallèle au segment joignant $(0,0)$ à $(b, f(b))$. Cette tangente, que nous nommons l_m, est en dessous de la corde, par strict convexité. En particulier, son point d'intersection avec $y = 0$ est strictement entre 0 et m.

L'ensemble D est d'un seul côté de l_m. Ce côté est forcément celui de $q = (0,0)$ (parce que $q \in \Gamma \subset D$), et donc le points de coordonnées $(x,0)$ avec $x > m$ ne sont pas dans D. Pas de chance, p est un point de ce type.

La contradiction Nous avons prouvé que si $p \in \partial D \backslash \Gamma$ alors p n'est pas dans D, ce qui est impossible parce que, l'ensemble D étant fermé, nous avons $\partial D \subset D$.

□

Remarque 20.93.
Bien que cela puisse paraitre évident dès le début, nous ne démontrerons que dans la proposition 20.95 que D est l'enveloppe convexe de Γ.

Corolaire 20.94.
Si $p \in \mathrm{Int}(D)$ alors toute droite passant par p intersecte Γ en exactement 2 points.

Démonstration. Vu que la partie D est bornée, toute droite passant par son intérieur coupe ∂D en au moins deux points (un dans chaque sens, et en utilisant le lemme de passage de douane 7.243).

Soit ℓ une droite passant par p et supposons qu'elle coupe Γ en trois points distincts. Alors au moins deux d'entre eux sont du même côté de p. Soient q_1 et q_2 ces points. Nous avons donc dans l'ordre $p \in \mathrm{Int}(D)$, $q_1 \in \partial D$ et $q_2 \in \partial D$.

Vu que tout Γ est d'un seul côté de ses tangentes, lesdites tangentes ne passent pas par l'intérieur de D. Ni p ni q_2 ne sont sur ℓ_{q_1}, parce que si $q_2 \in \ell_{q_1}$ alors $\ell_{q_1} = \ell$, ce qui est impossible parce que ℓ passe par $p \in \mathrm{Int}(D)$.

Or p et q_2 sont de part et d'autres de ℓ_{q_1}, ce qui est impossible parce que Γ est d'un seul côté de cette droite. □

Proposition 20.95.
Soit D l'ensemble définit en (20.306).

(1) $D = \mathcal{C}\mathrm{onv}(\Gamma)$ ($\mathcal{C}\mathrm{onv}(\Gamma)$ désigne l'enveloppe convexe de Γ)

(2) $\partial\, \mathcal{C}\mathrm{onv}(\Gamma) = \Gamma$.

Pour la définition d'enveloppe convexe, voir la définition 8.36.

Démonstration. L'ensemble D est un convexe contenant Γ. Donc $\mathcal{C}\mathrm{onv}(\Gamma) \subset D$. L'inclusion inverse est à prouver.

Soit $x \in D$. Si $x \in \partial D$ alors $x \in \Gamma$ (proposition 20.92) et donc $x \in \mathcal{C}\mathrm{onv}(\Gamma)$. Nous ne devons donc traiter que le cas $x \in \mathrm{Int}(D)$.

Le corolaire 20.94 nous dit que toute droite passant par x coupe Γ en exactement deux points. Soient p et q ces points. Alors $p, q \in \Gamma \subset \mathcal{C}\mathrm{onv}(\Gamma)$. Vu que $\mathcal{C}\mathrm{onv}(\Gamma)$ est convexe, tout le segment $[p,q]$ est dans $\mathcal{C}\mathrm{onv}(\Gamma)$, et en particulier $p \in \mathcal{C}\mathrm{onv}(\Gamma)$.

Nous passons à la seconde affirmation. Nous savons que $D = \mathcal{C}\mathrm{onv}(\Gamma)$. En prenant le bord des deux côtés, $\partial D = \partial\,\mathcal{C}\mathrm{onv}(\Gamma)$, donc $\Gamma = \partial\,\mathcal{C}\mathrm{onv}(\Gamma)$. □

Lemme 20.96 (Des tangentes parallèles[408])**.**
Une courbe fermée γ de classe C^1 est convexe si et seulement si elle ne possède pas 3 tangentes parallèles distinctes.

Démonstration. Si le graphe Γ possédait trois tangentes parallèles distinctes, une serait entre les deux autres et le graphe Γ serait de part et d'autres de cette tangente. Dans ce cas, γ n'est pas convexe.

Nous montrons maintenant que si γ n'est pas convexe, alors elle possède trois tangentes parallèles distinctes. Pour $p \in \Gamma$ tel que Γ soit des deux côtés de l_p.

Soient Γ_1 et Γ_2 les parties de Γ délimitées par l_p. Nous notons q_i le point de Γ_i le plus éloigné de la droite l_p, il existe parce que G est compact et que la fonction distance à l_p est continue sur Γ. Les points p, q_1 et q_2 sont distincts (sinon l_p ne couperait pas Γ en deux parties).

Montrons que $l_{q_i} \parallel l_p$. Pour cela nous choisissons un système d'axe dans lequel $l_p \equiv y = 0$. Dans ce système, la distance entre $\gamma(t)$ et l_p est $\gamma_y(t)$ et les extrémums de cette fonction ont lieu aux points t tels que $\gamma_y'(t) = 0$, c'est-à-dire aux points sur lesquels la tangente est parallèle à l_p.

À quel moment avons nous utilisé le fait que la courbe soit fermée ? Au moment de dire que le point le plus éloigné devait vérifier $\gamma_y'(t) = 0$. En effet un extrémum peut ne pas vérifier cette condition s'il n'est pas à l'intérieur du domaine. Dans notre cas, nous avons $\gamma\colon [0,1] \to \mathbb{R}^2$ qui est fermée : $\gamma(0) = \gamma(1)$. Donc en réalité nous pouvons considérer $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ et tous les points du domaine sont intérieurs au domaine. L'extrémum doit donc vérifier la condition d'annulation de la dérivée. □

Exemple 20.97.
Si la courbe n'est pas fermée, alors le lemme 20.96 ne tient pas comme le montre le contre-exemple du graphe de $f(x) = x^3 - x$. Il est non convexe et pourtant ne présente que 2 tangentes parallèles (dans chaque directions). △

Lemme 20.98.
Soit γ une courbe convexe et ℓ une droite qui intersecte Γ mais qui n'est pas tangente. Alors l'intersection entre ℓ et Γ comprend au maximum 2 points.

Démonstration. Supposons que ℓ coupe Γ en trois points p, q, r (dans cet ordre). Vu que ℓ n'est pas tangente à Γ, la tangente ℓ_q est distincte de ℓ (et intersecte ℓ en l'unique point q). Les points p et r sont dans Γ et sont pourtant de deux côtés différents de ℓ_q. Contradiction avec la convexité de γ. □

La proposition suivante nous dit que si deux points de Γ ont la même tangente, alors entre ces deux points, Γ est le segment de droite les joignant.

Proposition 20.99 ([1])**.**
Soit $\gamma\colon [0,L] \to \mathbb{R}^2$ une courbe fermée simple et convexe de classe C^1. Si la tangente en $p = \gamma(s_p)$ et la tangente en $q = \gamma(s_q)$ sont identiques (pas seulement parallèles), alors soit

$$\gamma\big([s_p, s_q]\big) = [p, q] \tag{20.310}$$

soit

$$\gamma\big([s_q, L]\big) = [p, q] \tag{20.311}$$

Démonstration. Nous considérons un système d'axe dans lequel $p = (0,0)$, $\ell_p = \ell_q \equiv y = 0$ et tel que $\gamma_y(t) \geqslant 0$ pour tout t. Nous choisissons enfin un paramétrage de γ telle que $p = \gamma(0)$.

Si $\gamma_y\big([0, s_q]\big) = \{0\}$ alors par le théorème des valeurs intermédiaires 11.50, tous les points du type $(t, 0)$ avec $0 \leqslant t \leqslant q_x$ sont atteints par $\gamma(t)$ avec $0 \leqslant t \leqslant s_q$. De plus aucun autre point ne peut être atteint parce que γ étant simple, elle ne peut pas faire de retour en arrière.

Nous supposons donc que γ_y n'est pas identiquement nulle sur $[0, s_q]$; il existe donc un s_M avec $0 < s_M < q_q$ qui maximise γ_y sur $[0, s_q]$. La tangente à Γ en $\gamma(s_M)$ est horizontale. Les droites $y = 0$ et $y = \gamma_y(s_M)$ sont donc deux tangentes parallèles à Γ. Par le lemme des tangentes parallèles 20.96, il n'y a pas d'autres tangentes horizontales. Donc pour tout $n \in \mathbb{N}$ la droite $y = \frac{1}{n}$ n'est tangente nulle part [17] à Γ.

Par le lemme 20.98, la droite $y = \frac{1}{n}$ ne peut intersecter Γ qu'en seulement deux points. Or le théorème des valeurs intermédiaires appliqué à γ_y sachant que $\gamma_y(0) = \gamma_y(s_q) = 0$ et $\gamma_y(s_M) > 0$ nous donne a_n, b_n tels que

$$0 < a_n < s_M \tag{20.312a}$$
$$s_M < b_n < s_q \tag{20.312b}$$

et $\gamma_y(a_n) = \gamma_y(b_n) = \frac{1}{n}$. Donc la droite $y = \frac{1}{n}$ intersecte γ deux fois dans $[0, s_q]$. En conséquence de quoi $\gamma_y(t) < \frac{1}{n}$ pour tout $t \in [q_q, L]$. Cela étant valable pour tout n nous avons $\gamma_y\big([s_q, L]\big) = \{0\}$ et nous sommes ramenés essentiellement au premier cas. □

Proposition 20.100.
Soit une courbe fermée simple γ de classe C^1 et une droite ℓ. Alors il existe au moins deux points distincts q_1, q_2 tels que $\ell_{q_1} \parallel \ell_{q_2} \parallel \ell$ avec $\ell_p \neq \ell_q$.

Démonstration. Pour $p \in \Gamma$ nous considérons le nombre $r(p) = d(p, \ell)$. En tant que fonction continue sur un compact, elle possède un minimum et un maximum. Dans un système d'axe pour lequel $\ell \equiv y = 0$, la fonction r s'écrit $r(p) = p_y$ et les extrémums arrivent en $\gamma(s)$ avec $\gamma_y'(s) = 0$, ce qui signifie que les tangentes aux extrémums sont parallèles à ℓ. Vu que le maximum et le minimum ne peuvent pas être égaux (sinon la courbe serait horizontale et pas simple), les tangentes en ces points sont distinctes. □

Corolaire 20.101.
Soit une courbe convexe fermée simple γ de classe C^2. L'ensemble $\mathcal{C}\mathrm{onv}(\Gamma)$ est compact.

Démonstration. Nous savons que $\mathcal{C}\mathrm{onv}(\Gamma)$ n'est autre que D par la proposition 20.95. Nous savons déjà que D est fermé. Il nous suffit donc de prouver qu'il est borné (théorème de Borel-Lebesgue 7.127). Nous considérons deux droites perpendiculaires et les 4 tangentes correspondantes par la proposition 20.100. Vu que Γ est d'un seul côté de chacune de ces tangentes, elle est contenue dans le rectangle délimité par ces 4 droites. □

20.13.4 Courbure et convexité

Lemme 20.102.
Soit $\gamma\colon [0, L] \to \mathbb{R}^2$ une courbe simple, fermée en paramétrage normal. Alors l'application

$$\begin{aligned} \sigma\colon [0, L] &\to S^1 \\ s &\mapsto \gamma(s) \end{aligned} \tag{20.313}$$

est surjective.

Démonstration. Le théorème 20.88 nous dit que si $\theta(0) = a$ alors $\theta(2\pi) = a + 2\pi$ ou $a - 2\pi$. Le théorème des valeurs intermédiaires nous dit alors que θ prend toutes les valeurs entre a et $a + 2\pi$ ou $a - 2\pi$. □

Proposition 20.103 ([408]).
Une courbe fermée simple de classe C^2 est convexe si et seulement si sa courbure est soit toujours positive, soit toujours négative.

17. À part $n = 0$ et si par manque de chance, $\gamma_y(s_M)$ est un nombre de la forme $1/n$.

Démonstration. Nous considérons la courbe γ munie d'un paramétrage de vitesse 1, c'est-à-dire avec $\|\gamma'(t)\| = 1$ pour tout t. Si θ est sa fonction d'angle, alors nous avons $\theta' = \kappa$ par le lemme 20.80. Donc la fonction θ est monotone si et seulement si la courbure ne change pas de signe. Nous allons montrer que θ est monotone si et seulement si γ est convexe.

$\Rightarrow$ Nous supposons que θ est monotone et γ non convexe. Soient $p \in \Gamma$ tel que Γ soit des deux côtés de ℓ_p, et soient les points q_1, q_2 donnés par le lemme des tangentes parallèles 20.96 tels que $\ell_p \parallel \ell_{q_1} \parallel \ell_{q_2}$. Parmi les vecteurs tangents en p, q_1 et q_2, deux au moins ont la même direction ; supposons que ce sont q_1 et q_2. C'est-à-dire que si $p = \gamma(s_0)$, $q_1 = \gamma(s_1)$ et $q_2 = \gamma(s_2)$ alors nous avons $\gamma'(s_1) = \gamma'(s_2)$ et donc aussi

$$\theta(s_1) = \theta(s_2) + 2n\pi \tag{20.314}$$

pour un certain n. Mais θ est monotone et la différence entre sa première et sa dernière valeur doit valoir 2π ou -2π par le théorème 20.88. Donc n ne peut valoir que -1, 0 et 1.

Si $n = 0$ alors θ est constante sur $[s_1, s_2]$. Si $n = 1$ alors $\theta(s_1) = \sigma(s_2) + 2\pi$ alors que sur toute la courbe, θ ne peut faire que 2π. Donc θ est constant sur $[0, s_1]$ et sur $[s_2, L]$ (où L est le bord de le paramétrage). Si $n = -1$, même conclusion.

Dans tous les cas, Γ contient une ligne droite, soit de q_1 à q_2, soit de q_2 à q_1. Et dans ces cas nous avons $\ell_{q_1} = \ell_{q_2}$, ce qui est contraire à la construction de q_i.

Nous concluons que γ est convexe.

$\Leftarrow$ Nous supposons que γ est convexe, mais que θ n'est pas monotone. Il existe donc $s_1 < s_0 < s_2$ tels que

$$\theta(s_1) = \theta(s_2) \neq \theta(s_0). \tag{20.315}$$

Et vu le lemme 20.102, il existe s_3 tel que $\gamma'(s_3) = -\gamma'(s_1)$.

Donc en s_1, s_2 et s_3 nous avons trois tangentes parallèles. La proposition 20.96 est alors formelle, γ étant convexe, deux de ces tangentes doivent être identiques.

La proposition 20.99 dit qu'entre deux points dont les tangentes sont identiques, la courbe doit être un segment de droite. Or sur un segment de droite, $\kappa = 0$ et θ est constante.

- La partie $\gamma\big([s_1, s_2]\big)$ ne peut pas être droite parce que nous avions supposé l'existence d'un $s_0 \in]s_1, s_2[$ tel que $\theta(s_1) = \theta(s_2) \neq \theta(s_0)$.
- La partie $\gamma\big([s_1, s_3]\big)$ ne peut pas être droite parce que $\theta(s_3) \neq \theta(s_1)$.
- La partie $\gamma\big([s_2, s_3]\big)$ ne peut pas être droite parce que $\theta(s_2) \neq \theta(s_1)$.

Nous sommes donc devant une contradiction.

Nous en concluons que θ doit être monotone.

□

20.13.5 Théorème des quatre sommets

Lemme 20.104.
Soit une droite ℓ du plan. Il existe $a, c \in \mathbb{R}^2$ avec $c \neq 0$ tels que $z \in \ell$ si et seulement si $(z-a) \cdot c = 0$.

Démonstration. Une droite est paramétrée par $\gamma(t) = p + tq$. En posant $a = p$ et $c = Jq$ nous avons la réponse. En effet nous allons montrer qu'avec ces valeurs de a et c, nous avons $z \in \Gamma$ si et seulement si $(z - a) \cdot c = 0$.

D'abord un point de γ est de la forme $z = \gamma(t) = p + tq$. Nous avons :

$$\big(\gamma(t) - a\big) \cdot c = \big(\gamma(t) - p\big) \cdot Jq = tq \cdot Jq = 0. \tag{20.316}$$

Et dans l'autre sens, si $(z-a) \cdot c = 0$ nous devons prouver que $z \in \Gamma$. Nous avons : $(z-p) \cdot Jq = 0$, ce qui fait que $z - p$ est un multiple de q. Autrement dit : $z - p = \lambda q$ ou encore $z = \alpha q + p$, qui est sur la droite Γ. □

Définition 20.105.
*Un **sommet** d'une courbe est un point d'extrémum local de la courbure.*

Théorème 20.106 (Théorème des quatre sommets[54, 352])**.**
Soit un arc paramétrique $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ fermé, simple et convexe[18] *de classe C^3 et T-périodique.*

Alors γ possède au moins 4 points critiques sur chaque période.

Démonstration. Nous supposons que le paramétrage de γ soit normale.

Si la courbure κ est constante sur une partie ouverte de la (du paramétrage de la) courbe, alors tous les points de cette partie sont des sommets et le théorème est fait. Nous supposons que κ n'est pas constante et en particulier que Γ ne contient ni bouts de droites ni bouts de cercles (théorème 20.76).

La fonction κ étant de classe C^1 sur le compact Γ, elle admet au moins un maximum et un minimum distincts. Vu que ces points sont intérieurs, ils correspondent au changement de signe de κ'. Soient p et q ces points. Pour la simplicité nous supposons que γ est paramétré de telle sorte que $\gamma(0) = p$, et $q = \gamma(s_q)$ avec $0 < s_q < T$.

Nous supposons que p et q sont les seuls points de changement de signe de κ'.

Soit ℓ la droite passant par p et q. Tous les points du segment $[p, q]$ (qui sont dans $\mathcal{C}\mathrm{onv}(\Gamma)$) ne peuvent pas être sur Γ (sinon nous aurions un morceau de droite). Donc certains points sont dans l'intérieur de $\mathcal{C}\mathrm{onv}(\Gamma)$. Donc la droite ℓ passe par l'intérieur de $\mathcal{C}\mathrm{onv}(\Gamma)$ et le corolaire 20.94 nous dit que la droite ℓ ne coupe Γ en seulement deux points.

Par conséquent, les ensembles $\gamma\big([0, s_q]\big)$ et $\gamma\big([s_q, T]\big)$ sont de part et d'autre de Γ. Vu qu'en ces points, κ' change de signe et qu'il ne change de signe en aucun autre points, la fonction κ' est positive d'un côté de ℓ et négative de l'autre.

D'autre part par le lemme 20.104, il existe $a \in \mathbb{R}^2$ et $c \neq 0$ tels que $z \in \ell$ si et seulement si $(z - a) \cdot c = 0$. La fonction $z \mapsto (z - a) \cdot c$ est donc positive d'un côté de ℓ et négative de l'autre.

En résumé les fonctions

$$s \mapsto \kappa'(s) \tag{20.317a}$$
$$s \mapsto \big(\gamma(s) - a\big) \cdot c \tag{20.317b}$$

changent de signe en même temps et le produit a donc un signe constant. Ce produit n'est de plus pas nul parce que κ' n'est nul sur aucun intervalle (sinon κ y serait constant et Γ un segment de droite) et $\big(\gamma(s) - a\big) \cdot c$ ne s'annule pour aucun s sauf ceux qui correspondent à p et q.

Nous avons donc

$$0 \neq \int_0^T \kappa'(s)\big(\gamma(s) - a\big) \cdot c\, ds \tag{20.318a}$$
$$= \underbrace{\Big[\big(\gamma(s) - a\big) \cdot c\kappa(s)\Big]_0^T}_{A=0} - \int_0^T \kappa(s)\gamma'(s) \cdot c\, ds \tag{20.318b}$$
$$= -\int_0^T \kappa(s)\big(\gamma'(s) \cdot c\big) ds \tag{20.318c}$$
$$= \int_0^T J\gamma''(s) \cdot c\, ds \tag{20.318d}$$
$$= J\int_0^T \gamma''(s) \cdot c\, ds \tag{20.318e}$$
$$= J\big[\gamma'(s) \cdot c\big]_0^T \tag{20.318f}$$
$$= 0. \tag{20.318g}$$

Justifications :

18. Par la proposition 20.103 nous pouvons aussi bien demander à la courbure d'être toujours strictement positive, comme le fait [54].

— L'expression A est nulle parce que les valeurs en 0 et en T sont identiques.

— Nous utilisons le lemme 20.75 pour faire $-\kappa(s)\gamma'(s) = J\gamma''(s)$.

Le tout est une contradiction de la forme $0 \neq a = 0$.

Nous avons donc au moins un troisième point de changement de signe de κ'. Vu que la courbe est périodique, il en faut un nombre pair et donc un quatrième. □

L'exemple de l'ellipse montre qu'il n'y a pas lieu de chercher d'autres extrémums de κ à part les 4 déjà trouvés.

Exemple 20.107.
Nous trouvons les sommets de l'ellipse.

$$\gamma(t) = \big(a\cos(t), b\sin(t)\big) \tag{20.319a}$$
$$\gamma'(t) = \big(-a\sin(t), b\cos(t)\big) \tag{20.319b}$$
$$\gamma''(t) = -\big(a\cos(t), b\sin(t)\big) \tag{20.319c}$$
$$\tag{20.319d}$$

La courbure est

$$\kappa(t) = \frac{\gamma'' \cdot J\gamma'}{\|\gamma'\|^3} \tag{20.320a}$$
$$= \frac{-1}{[a^2\sin^2(t) + b^2\cos^2(t)]^{3/2}} \begin{pmatrix} a\cos(t) \\ b\sin(t) \end{pmatrix} \cdot \begin{pmatrix} -b\cos(t) \\ -a\sin(t) \end{pmatrix} \tag{20.320b}$$
$$= \frac{ab}{[a^2\sin^2(t) + b^2\cos^2(t)]^{3/2}}. \tag{20.320c}$$

Vu que $ab > 0$, les extrémums de cela sont ceux du dénominateur et il suffit donc d'étudier les extrémums de

$$f(t) = a^2\sin^2(t) + b^2\cos^2(t). \tag{20.321}$$

Nous avons

$$f'(t) = 2(a^2 - b^2)\cos(t)\sin(t), \tag{20.322}$$

fonction qui s'annule effectivement 4 fois sur une période. Deux maximums et deux minima. △

20.13.6 Le théorème de Jordan

Définition 20.108 ([409]).
*Une **courbe de Jordan** est une courbe simple fermée dans le plan.*

Définition 20.109.
*Une **courbe de Jordan** dans le plan est une application $\gamma\colon S^1 \to \mathbb{R}^2$ qui est continue et injective.*

Une telle courbe peut évidemment être vue comme une application $\gamma\colon [0, 2\pi] \to \mathbb{R}^2$ telle que $\gamma(0) = \gamma(2\pi)$. En particulier il n'est jamais mauvais de se rappeler qu'on peut choisir un paramétrage normal par la proposition 20.46.

Le théorème suivant a un énoncé relativement simple, mais la démonstration est en réalité très longue.

Théorème 20.110 (Théorème de Jordan[409, 410]).
Le complémentaire d'une courbe de Jordan Γ dans un plan affine réel est formé de exactement deux composantes connexes distinctes, dont l'une est bornée et l'autre non. Toutes deux ont pour frontière la courbe Γ.

Chapitre 21

Géométrie hyperbolique

21.1 Inversion

Lemme-définition 21.1.
Soit un cercle C et un point A du plan $\mathbb{R}^2$. Soit une droite passant par A et coupant C en deux points P et P' (pas spécialement distincts). Alors le nombre

$$\|AP\|\|AP'\| \tag{21.1}$$

ne dépend pas du choix de la droite et est nommé la ***puissance*** *du point A par rapport au cercle C.*

21.1.1 Cercles perpendiculaires

Définition 21.2.
Deux cercles sont ***perpendiculaires*** *lorsque leurs tangentes aux points d'intersection sont perpendiculaires.*

Lemme 21.3 ([1])**.**
Soient deux cercles perpendiculaires C_1 et C_2. Alors

(1) le centre de C_1 est hors de C_2.

(2) Si ℓ est une droite passant par ce le centre de C_1 (nommé O) et si ℓ coupe C_2 en les deux points P et P', alors P et P' sont situés du même côté de O.

Démonstration. Nous nommons O_1 le centre de C_1 ainsi que Q et Q' les points d'intersection de C_1 avec C_2. Si ℓ_Q et $\ell_{Q'}$ sont les tangentes à C_1 en Q et Q', alors ce sont des rayons de C_2 (parce que les cercles sont perpendiculaires). Par conséquent le centre O_2 de C_2 est le point d'intersection $Q' = \ell_Q \cap \ell_{Q'}$.

Le triangle O_1O_2Q est rectangle en Q et donc $\|O_1O_2\| > \|QO_2\|$. Or le nombre $\|QO_2\|$ est le rayon de C_2, donc O_1 est en dehors de C_2.

Ceci achève de prouver le point (1) ; nous démontrons le point (2). Les points P et P' sont sur le cercle C_2, donc tous les points du segment $[PP']$ sont dans le cercle. Or le centre de C_1 doit être en dehors de C_2 ; il ne peut donc pas être dans le segment $[PP']$, ce qui prouve que P et P' ne sont pas de part et d'autre de O sur la droite (OPP'). □

Proposition 21.4 ([1])**.**
Soit un cercle C_1 de centre O et de rayon R.

(1) Un cercle C_2 est perpendiculaire à C_1 si et seulement s'il existe une droite ℓ passant par O telle que les points d'intersection $\{P, P'\} = \ell \cap C_2$ soient situés du même côté de O et vérifient

$$\|OP\|\|OP'\| = R^2. \tag{21.2}$$

(2) Dans ce cas, toutes les droites passant par O et coupant C_2 en deux points P, P' vérifient le fait que P et P' soient du même côté de O et $\|OP\|\|OP'\| = R^2$.

Démonstration. Nous commençons par prouver le point (2). Le fait que les points P et P' soient du même côté de O est le lemme 21.3(2). Pour la relation sur les distances, soit $Q \in C_1 \cap C_2$. Vu que C_1 et C_2 sont perpendiculaires, la droite (OQ) ne coupe C_2 qu'en ce point, et la puissance de O par rapport au cercle C_2 est $\|OQ\|^2 = R^2$.

La même puissance peut être calculée via la droite ℓ :

$$\|OP\|\|OP'\|. \tag{21.3}$$

Donc $\|OP\|\|OP'\| = R^2$.

Soit un cercle C_2 passant par P et P'. Notons que P et P' ne sont pas sur C_1 parce qu'ils ne pourraient pas être alignés avec O. De plus l'un est à l'intérieur de C_1 et l'autre à l'extérieur de C_1. Les cercles C_1 et C_2 possèdent donc deux points distincts d'intersections. La puissance de O par rapport à C_2 est :

$$\|OP\|\|OP'\| = R^2 \tag{21.4}$$

parce que (OP) est une droite coupant C_2 en les points P et P'.

Soit Q un point d'intersection de C_1 et C_2, et Q' l'autre point d'intersection de C_2 avec la droite (OQ). La puissance de O par rapport à C_2 peut également être calculée à partir de cette droite (lemme 21.1) et nous avons

$$\|OQ\|\|OQ'\| = R^2, \tag{21.5}$$

mais $Q \in C_1$, donc $\|OQ\| = R$ et partant $\|OQ'\| = R$. Nous en déduisons que $Q' \in C_1$ également. Or Q' ne peut pas être l'autre point d'intersection de C_1 avec C_2 (sinon O, Q, Q' ne seraient pas alignés). Donc $Q = Q'$ et nous déduisons que la droite (OQ) est tangente à C_2.

Le rayon de C_1 est tangent à C_2. Cela signifie que C_1 est perpendiculaire à C_2. □

21.1.2 Inversion

Proposition-définition 21.5 ([354]).
Soit un cercle C de centre O dans $\mathbb{R}^2$. Il existe une unique application

$$i_C \colon \mathbb{R}^2 \backslash O \to \mathbb{R}^2 \backslash O \tag{21.6}$$

telle que

(1) $i_C(x) = x$ pour tout $x \in C$

(2) i_C échange l'intérieur et l'extérieur de C.

(3) i_C laisse invariants les droites et les cercles orthogonales à C.

*Cette application est l'**inversion** de cercle C.*

Démonstration. Soit P à l'intérieur de C, mais différent de O. Nous notons ℓ la droite (OP) et nous considérons un cercle C_2 passant par P et perpendiculaire à C (existence par la proposition 21.4). Nous avons $\ell \perp C$ (parce que ℓ est un rayon) et $C_2 \perp C$. Donc $i_C(\ell) = \ell$ et $i_C(C_2) = C_2$ par l'exigence (3). Mais comme $P \in \ell \cap C_2$ nous avons aussi

$$i_C(P) \in \ell \cap C_2. \tag{21.7}$$

Mais ℓ et C_2 se coupent en exactement deux points. Vu que $i_C(P)$ doit être hors de C (exigence (2)), avoir $i_C(P) = P$ est impossible. Nous en concluons que $i_C(P)$ doit être l'autre intersection.

Nous avons prouvé que les conditions (2) et (3) fixent l'image d'un point situé dans l'intérieur de C.

Si P est extérieur au cercle C, la même procédure fonctionne : nous considérons la droite $\ell = (OP)$ et un cercle C_2 perpendiculaire à C et passant par P. Encore une fois, ces deux objets sont fixés par i_C, et vu que $i_C(P)$ doit être à l'intérieur de C, il est fixé.

L'unicité est montrée.

En ce qui concerne l'existence, si $P \neq O$, la procédure suivante donne P' :

— Soit la droite $\ell = (OP)$.

— Soit un cercle C_2 perpendiculaire à C et passant par P.

— Le point P' est le point de l'intersection $C_2 \cap \ell$ qui n'est pas P.

Il est aisé de vérifier que poser $i_C(P) = P'$ donne une application qui vérifie toutes les propriétés. □

21.6.
Nous recopions la construction de l'inversion d'un point par rapport à un cercle. Si C_1 est un cercle de centre O et P est un point différent de O, alors la procédure suivante construit $P' = i_{C_1}(P)$:

— Soit la droite $\ell = (OP)$.

— Soit un cercle C_2 perpendiculaire à C_1 et passant par P.

— Le point P' est le point de l'intersection $C_2 \cap \ell$ qui n'est pas P.

Que se passe-t-il si ℓ et C_2 n'ont qu'une seule intersection ? Alors la droite $\ell = (OP)$ est tangente à C_2. Or de O il n'existent que deux droites tangentes à C_2, et ce sont les rayons passant par les intersections parce que les cercles sont perpendiculaires. En d'autres mots, cette situation se présente lorsque P est sur le cercle C_1. Dans ce cas, $i_{C_1}(P) = P$.

Remarque 21.7.
Lorsque nous disons qu'une inversion « conserve les droites passant par O », il y a pour sous-entendu que nous considérons la droite privée du point O, parce que de toutes façons, l'inversion n'est pas définie sur O.

Nous allons résoudre cet intéressant problème en 22.6.1, en ajoutant le point ∞ à toutes les droites.

Corolaire 21.8.
L'inversion i_C est une involution ($i_C^2 = \mathrm{Id}$).

Démonstration. Soit un cercle C de centre O et un point P. Si C_2 est un cercle perpendiculaire à C passant par P, alors nous avons vu en 21.6 que $P' = i_C(P)$ est l'autre intersection entre C_2 et la droite (OP).

Pour construire l'image de P', il faut un cercle perpendiculaire à C passant par P'. Le cercle C_2 déjà utilisé fait l'affaire. Ensuite, la droite (OP') est la même que la droite (OP). Donc l'image de P' est P. □

Soit C_1 le cercle dans $\mathbb{C}$ de centre 0 et de rayon 1. Nous notons $\alpha\colon \mathbb{C} \to \mathbb{C}$ la dilatation de rapport R.

Proposition 21.9 ([1])**.**
Soit C le cercle de rayon R centré en 0 ($C = \alpha(C_1)$). Alors

$$i_{\alpha(C_1)} \circ \alpha = \alpha \circ i_{C_1}. \tag{21.8}$$

Démonstration. Soit $z = re^{i\theta}$; nous devons prouver que

$$i_C\big(\alpha(z)\big) = \alpha\big(i_{C_1}(z)\big). \tag{21.9}$$

Nous avons :

$$i_C\big(\alpha(z)\big) = i_C(Rre^{i\theta}) = R^2\frac{1}{Rr}e^{i\theta} = R\frac{1}{r}e^{i\theta} = \alpha\big(\frac{1}{r}e^{i\theta}\big) = \alpha\big(i_{C_1}(z)\big). \tag{21.10}$$

□

La définition 22.67 pour l'inversion sur $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ sera basée sur la proposition suivante.

Proposition 21.10 ([1]).
Soit le cercle de rayon R et de centre $a \in \mathbb{C}$. Alors

$$i_C(z) = \frac{R^2}{\bar{z} - \bar{a}} + a. \tag{21.11}$$

De plus si t_a est la translation de vecteur a, nous avons la décomposition

$$i_{t_a(C_R)} = t_a \circ i_{C_R} \circ t_{-a} \tag{21.12}$$

où C_R est le cercle de rayon R centré en 0.

Démonstration. Si $z \in \mathbb{C}$ et z' est son image par i_C, alors non seulement

$$\|a - z'\|\|a - z\| = R^2, \tag{21.13}$$

mais en plus $a - z' = \lambda(a - z)$ pour un certain $\lambda > 0$. Cela est l'expression du fait que z' est sur la demi-droite qui joint a à z. Nous avons donc

$$\lambda\|a - z\|^2 = R^2 \tag{21.14}$$

et alors

$$\lambda = \frac{R^2}{(a - z)(\bar{a} - \bar{z})}. \tag{21.15}$$

En récrivant $a - z' = \alpha(a - z)$ avec cette valeur de λ nous trouvons

$$a - z' = \frac{R^2}{\overline{a - z}}, \tag{21.16}$$

ce qu'il fallait démontrer.

La décomposition demandée est une simple vérification en utilisant $i_{C_R}(z) = \frac{R^2}{\bar{z}}$ qui découle de la proposition 21.9. □

Avant d'aller plus loin, donnons l'équation d'un cercle dans $\mathbb{C}$. Si C est un cercle de entre ω et de rayon r, alors $z \in C$ si et seulement si $d(z, \omega) = r$. En développant, et en passant au carré sans perte d'information (les deux membres sont positifs), $z \in C$ si et seulement si

$$(z - \omega)(\bar{z} - \bar{\omega}) = r^2. \tag{21.17}$$

Proposition 21.11 ([354, 411]).
Soit un cercle C de centre O. L'inversion

$$i_C \colon \mathbb{R}^2 \backslash \{0\} \to \mathbb{R}^2 \backslash \{0\} \tag{21.18}$$

transforme

(1) les droites passant par O sur elles-mêmes ;

(2) les cercles passant par O en des droites ne passant pas par O ;

(3) les droites ne passant pas par O en des cercles passant par O ;

(4) les cercles ne passant pas par O en des cercles ne passant pas par O.

Démonstration. Point par point.

(1) Une droite passant par O est une droite perpendiculaire à C. Par le point (3) de la définition 21.5, elle est invariante.

(2) Nous construisons successivement :

— Un cercle C_1 de centre O_1 et passant par O. Le but est de déterminer l'image de ce cercle.

— Le point P de C_1 tel que $[OP]$ en soit un diamètre.

— Le point $P' = i_C(P)$.

— La droite ℓ perpendiculaire à (OP) et passant par P'.

Nous montrons maintenant que $i_C(C_1\backslash\{O\}) = \ell$. Soit $Q \in C_1\backslash\{P, O\}$. Nous posons

$$Q' = (QO) \cap \ell. \tag{21.19}$$

Vu que $[OP]$ est un diamètre de C_1 et que $Q \in C_1$, le triangle OPQ est rectangle en Q. Et étant donné que Q' est sur ℓ nous savons que $OP'Q'$ est rectangle en P'.

De plus les angles en O de ces deux triangles sont identiques (parce que c'est l'angle formé par les droites (OQ) et (OP)) ; les triangles OPQ et $QP'Q'$ sont donc semblables et nous pouvons utiliser le théorème de Thalès [1] :

$$\frac{OP}{OQ} = \frac{OQ'}{OP'}. \tag{21.20}$$

Donc

$$\|OP\|\|OP'\| = \|OQ\|\|OQ'\|, \tag{21.21}$$

mais P' est l'image de P par l'inversion du cercle C, c'est-à-dire $\|OP\|\|OP'\| = R^2$. Nous en déduisons que

$$\|OQ\|\|OQ'\| = R^2, \tag{21.22}$$

c'est-à-dire que Q' est l'image de Q par i_C, et donc que

$$i_C\big(C_1\backslash\{O\}\big) \subset \ell. \tag{21.23}$$

Pour avoir l'inclusion inverse, il faut remarquer que ℓ est parallèle à la tangente à C_1 en O. Donc si $Q \in \ell$, la droite (OQ) intersecte le cercle C_1 en un point Q'. En refaisant le cheminement du résultat (21.23) à l'envers, il est loisible de prouver que $i_C(Q') = Q$ et donc que ℓ est bien inclue à l'image de C_1 par i_C.

(3) Nous commençons par prouver que toutes les droites ne passant pas par O sont des images de cercles passant par O.

Nous considérons :

— une droite ℓ ne passant pas par O.

— la droite d, perpendiculaire à ℓ passant par O

— le point $P' = \ell \cap d$,

— le point $P = i_C(P')$,

— le cercle C_1 dont $[OP]$ est un diamètre.

Par tout ce que nous avons fait jusqu'à présent, la droite ℓ est l'image du cercle C_1. Or si $\|OP\| = r$ alors

$$\|OP'\| = \frac{R^2}{r}. \tag{21.24}$$

Donc quelle que soit la valeur de $\|OP'\|$ dans $]0, \infty[$, il existera un point P tel que le cercle passant par O et P ait pour image la droite perpendiculaire à (OP) passant par $i_C(P)$.

Étant donné que i_C est une involution surjective des cercles passant par O vers les droites ne passant pas par O, elle transforme également toutes les droites ne passant pas par O en un cercle passant par O.

(4) Pour cette partie, nous allons utiliser un peu de géométrie analytique dans $\mathbb{C}$ [2].

1. Faites bien le dessin : ce n'est pas une situation de Thalès ultra-standard de collège.
2. Principalement parce que je ne comprends pas le raisonnement fait dans [354].

Le cas centré Nous supposons que C est centré en 0 et de rayon, 1, ce telle sorte que $i_C(z) = \frac{1}{\bar{z}}$. Soit C_1 un cercle de centre ω et de rayon r, ne passant pas par 0, en particulier tel que $|\omega| \neq r$.

Si $z \in i(C_1)$ alors $i(z) \in C_1$ et nous avons

$$\big(i(z) - \omega\big)\overline{\big(i(z) - \omega\big)} = r^2. \tag{21.25}$$

En développant et en multipliant par $z\bar{z}$ nous trouvons

$$\bar{\omega}z + \omega\bar{z} + z\bar{z}(r^2 - \omega\bar{\omega}) = 1. \tag{21.26}$$

Nous pouvons diviser par $(r^2 - |\omega|^2)$ parce que C_1 ne passe pas par 0. En remettant en ordre, et en notant $s = r^2 - \omega\bar{\omega}$ pour plus de clarté,

$$\big(z - \frac{\omega}{s}\big)(\bar{z} - \frac{\bar{\omega}}{s}) - \frac{\omega\bar{\omega}}{s^2} = \frac{1}{s^2}, \tag{21.27}$$

ou encore

$$\big(z - \frac{\omega}{s}\big)(\bar{z} - \frac{\bar{\omega}}{s}) = \frac{1}{s} + \frac{\omega\bar{\omega}}{s}. \tag{21.28}$$

Pour que cela soit l'équation du cercle de centre $\frac{\omega}{s}$ et de rayon $\sqrt{\frac{1}{s} + \frac{\omega\bar{\omega}}{s^2}}$, il faut vérifier que

$$\frac{1}{s} + \frac{\omega\bar{\omega}}{s^2} \geqslant 0. \tag{21.29}$$

En multipliant par s^2, il s'agit de vérifier que $s + \omega\bar{\omega} \geqslant 0$, ce qui est correct parce que $s + \omega\bar{\omega} = r$.

En résumé, si $z \in i(C_1)$ alors z est dans le cercle C_2 de centre $\frac{\omega}{s}$ et de rayon $\frac{r}{s}$. Étant donné que $r \neq \omega$ nous savons que ce dernier cercle ne passe pas par 0.

Nous avons prouvé que $i(C_1) \subset C_2$. Pour prouver l'inclusion inverse, vu que i est une involution, il faut prouver $i(C_2) \subset C_1$. Pour cela nous écrivons l'équation qui donne $i(z) \in C_2$ et en développant nous devons conclure que $z \in C_1$. Nous ne le faisons pas ici.

Le cas de rayon non unité Si C est un cercle quelconque, nous écrivons l'inversion du cercle C via la formule (21.11). Si $z \in i_C(C_1)$ alors $i(z) \in C_1$ et nous pouvons écrire

$$\big(i(z) - a\big)\big(\overline{i(z)} - \bar{a}\big) = R^2. \tag{21.30}$$

De là il faut déduire que z est sur un cercle ne passant pas par 0. Bons calculs...

□

Chapitre 22

Espaces projectifs

Sur les espaces projectifs : [412].

Définition 22.1.
*Soit E un espace vectoriel de dimension finie sur le corps commutatif $\mathbb{K}$. Nous définissons sur $E\backslash\{0\}$ la relation d'équivalence $u \sim v$ si et seulement si $u = \lambda v$ pour un certain $\lambda \in \mathbb{K}$. Cette relation est la relation de **colinéarité**. L'ensemble des classes d'équivalence de $\sim$ est l'**espace projectif** de E et sera noté $P(E)$.*

Définition 22.2.
*Si $\dim E = 2$, l'ensemble $P(E)$ est la **droite projective**, et si $\dim E = 3$ nous parlons du **plan projectif**.*

Étant donné que tous les $\mathbb{K}$-espaces vectoriels de dimensions $n+1$ sont isomorphes à $\mathbb{K}^{n+1}$, nous noterons $P_n(\mathbb{K})$ ou P_n l'espace projectif $P(\mathbb{K}^{n+1})$.

Exemple 22.3.
Si $n = 1$ et $\mathbb{K} = \mathbb{R}$, l'espace projectif est l'ensemble des droites vectorielles dans le plan usuel. Il y en a une pour chaque point du type $(x, 1)$ avec $x \in \mathbb{R}$ et ensuite une horizontale, passant par le point $(1, 0)$. Nous avons donc

$$P_1(\mathbb{R}) = \{(1,0)\} \cup \{(x,1) \text{ tel que } x \in \mathbb{R}\}. \tag{22.1}$$

Le point $(1, 0)$ est dit « point à l'infini ». △

22.1 Sous espaces projectifs

Un **sous-espace projectif** de $P(E)$ est une partie de la forme $P(F)$ où F est un sous-espace vectoriel de E.

Proposition 22.4.
Si F et G sont des sous-espaces vectoriels de E, alors

$$P(F) \cap P(G) = P(F \cap G) \tag{22.2}$$

et nous avons

$$\dim P(F) + \dim P(G) = \dim P(F+G) + \dim P(F \cap G). \tag{22.3}$$

Démonstration. Nous avons

$$P(F) = \{[v] \text{ tel que } v \in F\} \tag{22.4}$$

où les crochets signifient la classe par rapport à la relation de colinéarité. Nous avons alors

$$P(F) \cap P(G) = \{[v] \text{ tel que } v \in F \cap G\} = P(F \cap G). \tag{22.5}$$

Cela prouve le premier point.

En ce qui concerne l'équation (22.3), en considérant $\dim P(E) = \dim E - 1$ nous devons prouver l'égalité

$$\dim F + \dim G = \dim(F + G) + \dim(F \cap G) \tag{22.6}$$

concernant les dimensions des espaces vectoriels usuelles. Si nous considérons une base de E telle que $B_1 = \{e_1, \dots, e_{k_1}\}$ est une base de $F \cap G$, $B_2 = \{e_{k_1+1}, \dots, e_{k_2}\}$ complète B_1 en une base de F et $B_3 = \{e_{k_2+1}, \dots, e_n\}$ complète $B_1 \cup B_2$ en une base de G.

Nous avons alors

$$\dim F + \dim G = 2\,\mathrm{Card}(B_1) + \mathrm{Card}(B_2) + \mathrm{Card}(b_3) \tag{22.7a}$$
$$\dim(F + G) = \mathrm{Card}(B_1) + \mathrm{Card}(b_2) + \mathrm{Card}(B_3) \tag{22.7b}$$
$$\dim(F \cap G) = \mathrm{Card}(B_1). \tag{22.7c}$$

De là la relation (22.3) se déduit immédiatement. □

Théorème 22.5 (incidence)**.**
Soient F et F deux sous-espaces vectoriels de E tels que

$$\dim P(F) + \dim P(G) \geqslant \dim P(E). \tag{22.8}$$

Alors $P(F) \cap P(G) \neq \varnothing$.

Démonstration. En utilisant les hypothèses et la proposition 22.4 nous avons

$$\dim P(E) + \dim P(G) = \dim P(F + G) + \dim P(F \cap G) \geqslant \dim P(E). \tag{22.9}$$

En passant aux espaces vectoriels correspondants,

$$\dim(F + G) + \dim(F \cap G) \geqslant \dim(E) + 1. \tag{22.10}$$

Mais nous avons aussi $\dim(F + G) \leqslant \dim(E)$ et par conséquent $\dim(F \cap G) \geqslant 1$. Au final, $\dim P(F \cap G) \geqslant 0$. Cela prouve que $P(F \cap G)$ contient au moins un élément (nous rappelons que lorsqu'un espace projectif contient un seul élément, sa dimension est zéro). □

Exemple 22.6.
Soient les plans $\Pi_1 \equiv x = 0$ et $\Pi_2 \equiv y = 0$. Nous avons

$$P(\Pi_1) = \{[0, y, 1]\} \cup \{[0, 1, 0]\} \tag{22.11a}$$
$$P(\Pi_2) = \{[x, 0, 1]\} \cup \{[1, 0, 0]\} \tag{22.11b}$$

où le crochet signifie la classe pour la colinéarité. Ces deux droites projectives ont comme point d'intersection le point $[0, 0, 1]$. △

Définition 22.7.
*Un **hyperplan projectif** est un sous-espace projectif de $P(E)$ de la forme $P(V)$ où V est un hyperplan de E.*

Définition 22.8.
*Soit E un espace vectoriel de dimension au moins 3. Nous disons que $d \subset P(E)$ est une **droite projective** de $P(E)$ si $d = P(D)$ pour une plan vectoriel $D \subset E$.*

*Nous disons que trois points de $P(E)$ sont **alignés** lorsqu'il existe une droite projective les contenant.*

22.9.
Dans la définition 22.8 nous voyons $P(D)$ comme inclus dans $P(E)$ dès que D est un sous-espace vectoriel de E. Cela est possible parce que si la direction de $v \in D$, c'est-à-dire la classe $[v]$ est également une direction dans E.

Le lemme suivant peut paraitre idiot, mais ce qui serait surement idiot est de l'utiliser sans s'en rendre compte.

Lemme 22.10.
Deux points dans $P(E)$ sont toujours alignés.

Démonstration. Soient deux points $A, B \in P(E)$. Si $A = \pi(a)$ et $B = \pi(b)$ alors le plan D passant par a, b et 0 est vectoriel et $P(D)$ contient A et B.

Note : si a, b et 0 sont trois points alignés, alors $A = B$. Il suffit de prendre les points a, c et 0 où $c \in E$ est un point quelconque non aligné avec 0 et a. Nous avons de toutes façons $A = B = \pi(a)$. □

Lemme 22.11.
Trois points distincts A, B, C dans $P(E)$ sont alignés si et seulement si il existe trois points non alignés $a, b, c \in E$ tels que

(1) le plan passant par a, b et c est vectoriel (c'est-à-dire passe par 0),
(2) $A = \pi(a)$, $B = \pi(b)$, et $C = \pi(c)$.

Démonstration. Deux implications à montrer.

Sens direct Soient A, B, C distincts et alignés dans $P(E)$. Alors il existe un plan vectoriel D tel que $A, B, D \in P(D)$.
La condition $A \in P(D)$ implique qu'il existe $a \in D$ tel que $A = \pi(A)$. Idem pour B et C. Les points a, b et c ainsi construits sont distincts parce que A, B et C sont distincts. Si par malheur ces trois points étaient alignés, ce n'est pas grave : il suffit de remplacer a par λa avec $\lambda \neq 0$ pour qu'ils ne le soient plus (cette manipulation ne change pas le fait que le nouveau choix de point a reste dans D parce que D est vectoriel). Nous avons donc trois points non alignés a, b et c tous contenus dans D. Le plan D répond à la question.

Sens réciproque Soient a, b et c non alignés dans E tels que $A = \pi(a)$, $B = \pi(b)$ et $C = \pi(c)$. Le plan D les contenant tous trois est vectoriel par hypothèse. Nous avons $A, B, C \in P(D)$ et donc A, B et C sont alignés dans $P(E)$.

□

Proposition 22.12.
Soit $H = P(V)$ un hyperplan projectif de $P(E)$ et soit m hors de H. Alors toute droite projective passant par m coupe H en un et un seul point.

Démonstration. Si $\dim E = n$ nous avons $\dim V = n - 1$. Soit $d = P(D)$ une droite projective passant par m, c'est-à-dire que D est de dimension 2 dans E. Si $D \subset V$ alors $m \in P(D) \subset P(V)$; or nous avons demandé que m soit hors de $P(V)$. Par conséquent D n'est pas inclus dans V et en particulier $\dim(D + V) = \dim(E)$.

Nous recopions la formule (22.3) pour notre cas :

$$\underbrace{\dim d}_{=1} + \underbrace{\dim H}_{=n-2} = \underbrace{\dim P(D + V)}_{=n-1} + \dim P(D \cap V). \tag{22.12}$$

Nous avons donc $\dim P(D \cap V) = 0$, ce qui signifie que l'ensemble $P(D \cap V) = P(D) \cap P(V) = d \cap H$ contient un et un seul point. □

22.2 Espace projectifs comme « complétés » d'espaces affines

Soit E un espace vectoriel de dimension 2 et $P(E)$ la droite projective correspondante, et soit $\{e_1, e_2\}$ une base de E. Nous considérons la droite affine $d \equiv y = 1$. Nous avons la bijection

$$\begin{aligned} \phi\colon d \cup \{\infty\} &\to P(E) \\ (x, 1) &\mapsto \text{la droite vectorielle passant par } (x, 1) \\ \infty &\mapsto \text{la droite vectorielle passant par } (1, 0). \end{aligned} \tag{22.13}$$

Lemme 22.13.
Si nous munissons l'ensemble $d \cup \{\infty\}$ de la topologie compactifiée d'Alexandrov[1]*, la bijection* (22.13) *est un homéomorphisme.*

Soient maintenant les plans affines dans l'espace vectoriel E de dimension 3

$$\Pi_1 \equiv z = 0 \tag{22.14a}$$
$$\Pi_2 \equiv z = 1. \tag{22.14b}$$

Une droite (vectorielle) de E coupe Π_2 en un et un seul point, sauf si elle est contenue dans Π_1. Nous avons donc une bijection

$$\begin{aligned} \phi\colon P(E) &\to \Pi_2 \cup P(\Pi_1) \\ d &\mapsto \begin{cases} \Pi_2 \cap d & \text{si cette intersection est non vide} \\ d & \text{sinon.} \end{cases} \end{aligned} \tag{22.15}$$

La droite projective $P(\Pi_1)$ est la droite à l'infini du plan projectif $P(E)$. Nous voyons que le plan projectif $P(E)$ peut être vu comme un plan affine (Π_2) « complété » par une droite affine $P(\Pi_1)$. Cette dernière droite est elle-même une droite affine complétée par un point à l'infini.

Nous pouvons généraliser cette démarche en considérant un espace affine $\mathcal{E}$ de direction E sur le corps $\mathbb{K}$. Nous construisons $F = E \times \mathbb{K}$ et nous considérons un repère affine sur F tel que $E \equiv x_{n+1} = 0$. Nous pouvons donc identifier $\mathcal{E}$ à l'hyperplan affine d'équation $x_{n+1} = 1$ dans F.

Une droite vectorielle de F non contenue dans E coupe $\mathcal{E}$ en un unique point ; nous avons donc une bijection

$$\mathcal{E} \cup P(E) \to P(F). \tag{22.16}$$

Dans ce cadre, $P(E)$ est l'hyperplan à l'infini et nous disons que $P(E)$ est la **complétion projective** de $\mathcal{E}$.

Exemple 22.14.
Nous considérons les plans affines

$$\Pi_1 \equiv z = 0 \tag{22.17a}$$
$$\Pi_2 \equiv z = 1 \tag{22.17b}$$

et nous avons la bijection

$$P(E) = \Pi_2 \cup P(\Pi_1). \tag{22.18}$$

Un plan affine D a deux possibilités : soit il coupe Π_2 en une droite, soit il est égal à Π_1. Si $D \cap \Pi_2 = d$ (d est une droite affine), alors nous avons

$$P(D) = d \cup \{\infty_D\}, \tag{22.19}$$

ce qui justifie la terminologie comme quoi $P(D)$ est une droite dans $P(E)$. △

Soit E un espace vectoriel de dimension 3 et le plan projectif $P(E)$. Nous avons deux types de droites projectives :

(1) D'abord nous avons la droite à l'infini, donnée[2] par $P(z = 0)$.
(2) Ensuite nous avons toutes les droites affines du plan $z = 1$. Chacune de ces droites est complétée par un point à l'infini.

Exemple 22.15.
Étudions un peu le second type de droites. D'abord si deux droites sont parallèles, leurs points à

1. Définition 7.66.
2. Dans notre représentation usuelle du plan projectif $z = 1$.

l'infini sont identiques. Prenons par exemple les droites $d = \{z = 1, x = 1\}$ et $d' = \{z = 1, x = 2\}$. Elles décrivent les directions des vecteurs

$$\begin{pmatrix} 1 \\ y \\ 1 \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} 2 \\ y \\ 1 \end{pmatrix}. \tag{22.20}$$

En normalisant, ce sont les vecteurs

$$\frac{1}{\sqrt{2+y^2}} \begin{pmatrix} 1 \\ y \\ 1 \end{pmatrix} \quad \text{et} \quad \frac{1}{\sqrt{5+y^2}} \begin{pmatrix} 2 \\ y \\ 1 \end{pmatrix}, \tag{22.21}$$

et toutes deux tendent vers le vecteur $(0, 1, 0)$ pour $y \to \infty$. △

Lemme 22.16.
Deux droites d'un plan projectif ont toujours une intersection.

Démonstration. Si les deux droites sont des droites affines non parallèles, le résultat est évident. Si elles sont parallèles, alors l'intersection est donnée par le point à l'infini comme indiqué dans l'exemple 22.15.

Supposons que d est la droite à l'infini tandis que d' est une droite affine. Dans notre représentation usuelle du plan affine, la droite à l'infini d a contient les vecteurs $(1, y, 0)$ et le point à l'infini $(0, 1, 0)$. La droite affine d' a pour équation paramétriques

$$\begin{cases} x = at + c & \text{(22.22a)} \\ y = bt + d & \text{(22.22b)} \\ z = 1. & \text{(22.22c)} \end{cases}$$

Les directions données par la droite d' sont donc

$$\frac{1}{a^2t^2 + b^2t^2 + c^2 + d^2} \begin{pmatrix} at + c \\ bt + d \\ 1 \end{pmatrix} \tag{22.23}$$

Son point à l'infini est la direction du vecteur $(a, b, 0)$, qui est bien un point de la droite à l'infini (éventuellement son point à l'infini [3]). □

La plupart du temps nous considérons le plan projectif comme étant le plan affine $z = 1$ de l'espace affine de dimension 3 complété par la droite affine $x = 1, z = 0$, elle-même complétée par le point $(0, 1, 0)$. Ce n'est évidemment pas la seule manière. Tout plan peut être considéré comme le plan à l'infini et pour une droite projective, tout point peut être considéré comme point à l'infini.

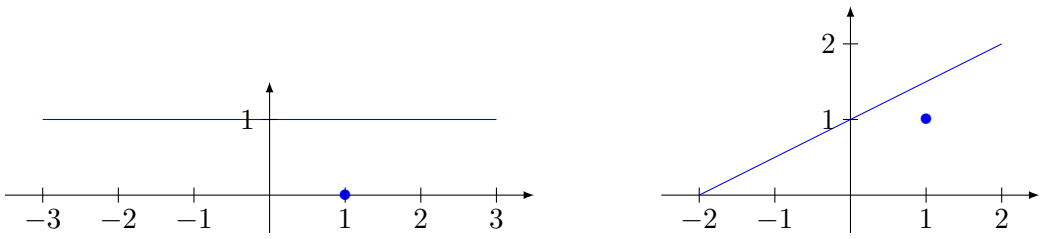
Sur la figure 22.1(a), le point à l'infini est la direction $(1, 0)$ tandis que la direction $(1, 1)$ n'a rien de spécial. À l'inverse sur la figure 22.1(b), la direction à l'infini est $(1, 1)$ tandis que la direction $(1, 0)$ est une direction usuelle.

Remarque 22.17.
Du point de vue de la topologie, si nous mettons celle de la compactification d'Alexandrov, tous les points de la droite projective sont équivalents.

Du point de vue de la géométrie différentielle, c'est la même chose. En effet nous pouvons mettre sur la droite projective un système de deux cartes en pensant aux angles. La première sur $]-a, a[$ avec par exemple $a < \pi/4$. La seconde carte serait $]a/2, \pi[$. Dans ce cas la direction $\theta = 0$ semble jouer un rôle spécial, mais il n'en est rien.

Nous pouvons également considérer les cartes $]\pi/4 - a, \pi/4 + a[$ et $]\pi/4 + a/2, 5\pi/4[$. Dans ces cartes, c'est plutôt le point $\theta = \pi/4$ qui semble différent (encore qu'il soit bien centré dans une carte).

3. D'accord, aller chercher le point à l'infini de la droite à l'infini, c'est chercher loin, mais n'empêche que ça existe.



(a) Ici le point à l'infini est la direction $(1, 0)$.

(b) Ici le point à l'infini est la direction $(1, 1)$.

FIGURE 22.1 – Deux façons de voir la droite projective. Étant donné que les points de la droite projective doivent être interprétés comme des directions (des classes d'équivallence), en réalité les deux dessins représentent les mêmes ensembles.

22.3 Théorème de Pappus

Théorème 22.18.
Soient deux droites d et d' dans un plan affine. Soient $A, B, C \in d$ et $A', B', C' \in d'$ tels que $AB' \parallel BA'$ et $BC' \parallel B'C$. Alors $AC' \parallel A'C$.

Démonstration. Si d et d' ne sont pas parallèles nous considérons o, le point d'intersection. Les relations de parallélisme des hypothèses impliquent qu'il existe λ_1 et λ_2 tels que

$$\begin{cases} A = \lambda_1 B & (22.24a) \\ B' = \lambda_1 A' & (22.24b) \end{cases}$$

et

$$\begin{cases} B' = \lambda_2 C' & (22.25a) \\ C = \lambda_2 B. & (22.25b) \end{cases}$$

En substituant nous trouvons

$$\begin{cases} C = \dfrac{\lambda_2}{\lambda_1} A & (22.26a) \\ A' = \dfrac{\lambda_2}{\lambda_1} C', & (22.26b) \end{cases}$$

ce qui implique que $A'C \parallel AC'$.

Si les droites d et d' sont parallèles, alors nous avons les translations

$$\begin{cases} B = A + x & (22.27a) \\ A' = B' + x & (22.27b) \end{cases}$$

et

$$\begin{cases} B = C + y & (22.28a) \\ C' = B' + y, & (22.28b) \end{cases}$$

ce qui montre que

$$\begin{cases} C = A + x - y & (22.29a) \\ A' = C' + x - y, & (22.29b) \end{cases}$$

et donc que $A'C \parallel AC'$. □

Le théorème suivant est une version projective.

Théorème 22.19.
Soient d et d' deux droites projectives d'un plan projectif. Soient $A, B, C \in d$ et $A', B', C' \in d'$. Alors les points $B'C \cap C'B$, $C'A \cap A'C$ et $A'B \cap B'A$ sont alignés.

Démonstration. Soient $E = BC' \cap C'B$ et $E' = C'A \cap A'C$. Ces deux points existent parce que deux droites projectives distinctes ont toujours un unique point d'intersection. Nous allons prendre EE' comme droite à l'infini et prouver que le point $A'B \cap B'A$ est dessus. Étant donné que le point d'intersection de $B'C$ et $C'B$ est à l'infini nous avons $B'C \parallel C'B$ (cela est un exemple de la flexibilité de la notion de parallélisme en géométrie projective). De la même façon nous avons $C'A \parallel A'C$.

Par le théorème de Pappus affine nous avons alors $A'B \parallel B'A$ et par conséquent le point d'intersection est sur la droite à l'infini, c'est-à-dire sur la droite EE'. □

22.4 Homographies

22.4.1 Homographies

Définition 22.20.
Soient E et F deux espaces vectoriels avec leurs projections naturelles

$$\pi_E \colon E\backslash\{0\} \to P(E) \tag{22.30a}$$
$$\pi_F \colon F\backslash\{0\} \to P(F). \tag{22.30b}$$

Une application $g\colon P(E) \to P(F)$ est une ***homographie*** *s'il existe un isomorphisme d'espaces vectoriels $\bar{g}\colon E \to F$ tel que le diagramme*

$$\begin{tikzcd} E\backslash\{0\} \arrow[r, "\bar{g}"] \arrow[d, "\pi_E"'] & F\backslash\{0\} \arrow[d, "\pi_F"] \\ P(E) \arrow[r, "g"'] & P(F) \end{tikzcd} \tag{22.31}$$

commute, c'est-à-dire s'il existe $\bar{g}\colon E \to F$ telle que

$$\pi_F\big(\bar{g}(v)\big) = g\big(\pi_E(v)\big) \tag{22.32}$$

pour tout $v \in E$.

Lemme 22.21.
Si $\bar{g}\colon E \to F$ est linéaire et si $\ker \bar{g} = \{0\}$ alors l'application g définie par

$$g\big(\pi_E(v)\big) = \pi_F\big(\bar{g}(v)\big) \tag{22.33}$$

est une homographie.

Démonstration. Nous devons simplement vérifier que l'équation (22.33) définit bien une application. Soient $v, w \in E$ tels que $\pi_E v = \pi_E w$; nous devons montrer que

$$\pi_F \bar{g} v = \pi_F \bar{g} w. \tag{22.34}$$

L'équation (22.34) sera vérifiée si et seulement s'il existe $\lambda \in \mathbb{R}$ tel que $\bar{g}v = \lambda \bar{g}w$, c'est-à-dire si et seulement si $\bar{g}(v - \lambda w) = 0$. Étant donné que nous supposons que le noyau de $\bar{g}$ est réduit à $\{0\}$, l'équation (22.34) sera vérifiée si et seulement si $v = \lambda w$, ce qui signifie exactement $\pi_E(v) = \pi_E(w)$. □

La proposition suivante donne les premières propriétés des homographies.

Proposition 22.22.
Quelques propriétés des homographies.

(1) Une homographie est bijective.

(2) Si deux espaces projectifs sont homographes, alors ils ont même dimension.

(3) L'ensemble des homographies $P(E) \to P(F)$ est un groupe (pour la composition).

(4) Une homographie conserve l'alignement des points.

Démonstration. Nous considérons une homographie $g\colon P(E) \to P(F)$, et $\bar{g}$ l'isomorphisme d'espaces vectoriels correspondant.

(1) Pour l'injectivité, si $g\big([v]\big) = g\big([w]\big)$ alors en utilisant la définition d'une homographie, $\pi_F \bar{g} v = \pi_F \bar{g} w$, ce qui implique que $\bar{g} v = \lambda \bar{g} w$, et donc $v = \lambda w$, ce qui signifie $[v] = [w]$.
Pour la surjectivité, un élément général de $P(F)$ prend la forme $\pi_F \bar{g} v$ pour un certain $v \in E$. Nous avons $g\big(\pi_E v\big) = \pi_F \bar{g} v$. Par conséquent l'élément $\pi_F \bar{g} v$ est bien dans l'image de g.

(2) Une homographie $P(E) \to P(F)$ n'existe que s'il existe un isomorphisme $E \to F$. Les dimensions sont donc automatiquement égales.

(3) Il suffit de vérifier que l'application

$$\begin{aligned} \varphi\colon P(E) &\to P(E) \\ \pi_F \bar{g} v &\mapsto \pi_E v \end{aligned} \tag{22.35}$$

est bien définie et donne l'inverse de g.

(4) Soient les points A, B, C alignés dans $P(E)$; ils correspondent à des directions de E qui sont données par des vecteurs situés sur la même droite affine. Autrement dit, il existe trois points $a, b, c \in E$ situés sur la même droite affine tels que $A, B, C = \pi_E(a, b, c)$. Les images par g sont données par $\pi_F \bar{g} a$, $\pi_F \bar{g} b$, et $\pi_F \bar{g} c$.
Étant donné qu'un isomorphisme d'espaces vectoriels conserve l'alignement affin, les points $\bar{g} a$, $\bar{g} b$ et $\bar{g} c$ sont alignés dans F. Cela implique que les projections par π_F sont alignés dans $P(F)$.

□

22.4.2 Le groupe projectif

Définition 22.23.
*Le groupe des homographies de l'espace $P(E)$ est le **groupe projectif**, noté* $\mathrm{PGL}(E)$.

Nous avons une surjection naturelle

$$\begin{aligned} \mathrm{GL}(E) &\to \mathrm{PGL}(E) \\ \bar{g} &\mapsto g \end{aligned} \tag{22.36}$$

qui s'avère être un morphisme de groupes.

Proposition 22.24.
Nous avons l'isomorphisme de groupes

$$\frac{\mathrm{GL}(E)}{\{\textit{homothéties}\}} \simeq \mathrm{PGL}(E). \tag{22.37}$$

Démonstration. Nous devons prouver que le noyau de l'application (22.36) est constitué des homothéties. Considérons un automorphisme d'espace vectoriel $f\colon E \to E$ dont l'homographie associée est l'identité, et prouvons que f est une homothétie. Nous avons le diagramme commutatif suivant :

$$\begin{tikzcd} E\backslash\{0\} \arrow[r, "f"] \arrow[d, "\pi_E"'] & E\backslash\{0\} \arrow[d, "\pi_E"] \\ P(E) \arrow[r, "\mathrm{Id}"'] & P(E). \end{tikzcd} \tag{22.38}$$

Pour tout vecteur $v \in E$ nous avons $\pi_E(v) = \pi_E\big(f(v)\big)$. Cela implique qu'il existe $\lambda \in \mathbb{R}$ tel que $f(v) = \lambda v$. Tous les vecteurs de E sont donc des vecteurs propres de f. Cela n'est possible que si toutes les valeurs propres sont identiques, c'est-à-dire que f est une homothétie. □

22.4.3 Repères projectifs

Définition 22.25 ([413]).
Des éléments $\{P_i\}_{i\in I}$ sont ***projectivement independents*** *si en choisissant $v_i \in \pi^{-1}(P_i)$ nous obtenons des vecteurs $\{v_i\}_{i\in I}$ linéairement indépendants.*

Définition 22.26.
Soit E un espace vectoriel de dimension $n+1$. Un ***repère projectif*** *de $P(E)$ est la donnée de $n+2$ points $m_0, \dots, m_{n+1}$ tels que*

(1) les vecteurs m_i, $i \neq 0$, sont les images d'une base $\{e_i\}$ de E

(2) $m_0 = \pi_E(e_1 + e_2 + \dots + e_{n+1})$.

Note que si $m_k = \pi_E(v_k)$ $(k = 0, \dots, n+1)$, alors tout choix de $n+1$ vecteurs parmi les v_k est une base de E.

Exemple 22.27.
Un repère projectif de l'espace $P(\mathbb{R}^3)$ est par exemple les éléments $\{m_i\}_{i=1,\dots,3}$ donnés par

$$m_1 = \pi(e_1) \tag{22.39a}$$
$$m_2 = \pi(e_2) \tag{22.39b}$$
$$m_3 = \pi(e_3) \tag{22.39c}$$
$$m_0 = \pi(e_1 + e_2 + e_3). \tag{22.39d}$$

△

Exemple 22.28.
Pour $P(\mathbb{C}^2)$, un repère projectif possible est $m_1 = [1, 0]$, $m_2 = [0, 1]$, $m_0 = [1, 1]$. △

22.29.
Pourquoi voulons nous des repères projectifs ? Pourquoi demander un quatrième élément alors que trois devraient suffire ? Le fait est que si E est de dimension 3, nous voudrions pouvoir identifier E et $P(\mathbb{R}^3)$.

Plus précisément, si E est de dimension $n+1$ et possède une base $\{f_i\}_{i=1,\dots,n+1}$, il existe un unique isomorphisme d'espaces vectoriels $E \to \mathbb{R}^{n+1}$ qui envoie cette base sur la base canonique de $\mathbb{R}^{n+1}$. La base de E étant fixée, nous pouvons donner à un point de E les coordonnées de son image dans $\mathbb{R}^{n+1}$ par cet isomorphisme *qui est unique*.

Dans le cas des espaces projectifs, nous voudrions avoir une unique homographie $\phi\colon P(E) \to P(\mathbb{R}^{n+1})$ qui permet de donner à un point $A \in E$ les coordonnées de $\pi^{-1}\big(\phi(A)\big)$. Bien entendu ce dernier n'est pas un élément bien défini de $\mathbb{R}^{n+1}$ parce qu'il y a toute une droite d'éléments de $\mathbb{R}^n$ qui se projettent sur $\phi(A)$.

L'idée d'imposer un point de plus est la bonne. Si nous imposons un point de plus, nous pouvons dire que les coordonnées de $A \in P(E)$ sont celles dans $\mathbb{R}^{n+1}$ de l'élément de $\pi^{-1}\big(\phi(A)\big)$ dont la dernière coordonnée est par exemple 1.

Nous allons maintenant mettre ça en musique.

D'abord nous donnons un exemple de non unicité.

Exemple 22.30.
Soit un espace vectoriel E de dimension 2 et une base $\{b_1, b_2\}$ de E. Nous considérons également l'espace $\mathbb{R}^2$ muni de sa base canonique $\{e_1, e_2\}$.

Soit une homographie $\phi\colon P(E) \to P(\mathbb{R}^2)$ telle que

$$\phi\big(\pi(b_i)\big) = \pi(e_i) \tag{22.40}$$

pour $i = 1, 2$. Nous allons facilement construire une autre homographie qui vérifie les mêmes conditions.

L'idée est la suivante. L'espace $P(E)$ peut être vu comme la droite complétée $\{(x,1)\}_{x\in\mathbb{R}} \cup \{(1,0)\}$ et l'espace $P(\mathbb{R}^2)$ également. Une homographie respectant (22.40) doit envoyer le $(1,0)$ de E vers le $(1,0)$ de $\mathbb{R}^2$ et le $(1,1)$ de E vers le $(1,1)$ de $\mathbb{R}^2$. Mais en ce qui concerne le reste de la droite, l'homographie peut la parcourir à la vitesse qu'elle veut.

Il faut envoyer

Soit donc une homographie $\phi\colon P(E) \to P(\mathbb{R}^2)$, et nous définissons

$$\begin{aligned} \phi'\colon P(E) &\to P(\mathbb{R}^2) \\ \pi(xb_1 + yb2)\phi\big(\pi(xb_1 + \lambda y b_2)\big) \end{aligned} \tag{22.41}$$

pour un certain $\lambda \neq 1$. En ce qui concerne le relèvement, l'application $\bar{\phi}'\colon E \to \mathbb{R}^2$ donnée par

$$\bar{\phi}'(xb_1 + yb_2) = \bar{\phi}(xb_1 + \lambda y b_2) \tag{22.42}$$

est bien définie et vérifie

$$\pi_{\mathbb{R}^2} \circ \bar{\phi} = \phi \circ \pi_E. \tag{22.43}$$

Donc ϕ' est une homographie. De plus

$$\phi'\big(\pi(b_1)\big) = \phi\big(\pi(b_1)\big) \tag{22.44a}$$

$$\phi'\big(\pi(b_2)\big) = \phi\big(\pi(\lambda b_2)\big) = \phi\big(\pi(b_2)\big) \tag{22.44b}$$

parce que $\pi(\lambda b_2) = \pi(b_2)$.

Nous n'avons donc pas l'unicité. △

C'est pour rétablir cette unicité que nous demandons d'avoir un point de plus pour avoir un repère projectif. De cette façon nous aurons une unique homographie $\phi\colon P(E) \to P(\mathbb{R}^{n+1})$ vérifiant $\phi\big(\pi_E(b_i)\big) = \pi_{\mathbb{R}^{n+1}}(e_i)$ pour tout $i = 0, \ldots, n+1$.

Lemme 22.31.
Soit un espace vectoriel E de dimension $n+1$ muni de deux bases $\{e_i\}_{i=1,\ldots,n+1}$ et $\{f_i\}_{i=1,\ldots,n+1}$. Soit un repère projectif $\{m_0, m_i\}_{i=1,\ldots,n+1}$ de $P(E)$.

Si $\pi(e_i) = \pi(f_i) = m_i$ pour tout $i = 1, \ldots, n+1$ et si

$$\pi(e_1 + \ldots + e_{n+1}) = \pi(f_1 + \ldots + f_{n+1}) \tag{22.45}$$

alors les deux bases sont proportionnelles : il existe λ tel que $f_i = \lambda e_i$ pour $i = 1, \ldots, n+1$.

Démonstration. Nous avons $\pi(e_i) = \pi(f_i)$ pour tout $i = 1, \ldots, n+1$. Donc pour chaque $i = 1, \ldots, n+1$ il existe $\lambda_i \in \mathbb{K}$ tel que $e_i = \lambda f_i$. Nous devons voir que les λ_i sont en réalité tous égaux.

Pour cela nous avons aussi l'égalité pour $i = 0$:

$$\pi(e_1 + \ldots + e_{n+1}) = \pi(f_1 + \ldots + f_{n+1}), \tag{22.46}$$

ce qui donne un $\mu \in \mathbb{K}$ tel que $e_1 + \ldots + e_{n+1} = \mu(f_1 + \ldots + f_{n+1})$, c'est-à-dire

$$\lambda_1 f_1 + \ldots + \lambda_{n+1} f_{n+1} = \mu f_1 + \ldots + \mu f_{n+1}. \tag{22.47}$$

Du fait que les f_i forment une base, cette égalité impose à tous les λ_i d'être égal à μ. □

Théorème 22.32 ([414]).
Soient $P(E)$ et $P(F)$ deux espaces projectifs de dimensions n.

(1) Une homographie $P(E) \to P(F)$ envoie un repère projectif sur un repère projectif.

(2) Si $(m_0, \dots, m_{n+1})$ est un repère projectif de $P(E)$, si $(m'_0, \dots, m'_{n+1})$ est un repère projectif de $P(F)$ alors il existe une unique homographie $g\colon P(E) \to P(F)$ telle que $g(m_i) = m'_i$ pour tout $i = 0, 1, \dots, n+1$

Démonstration. Un point à la fois.

(1) Soit une homographie $\phi\colon P(E) \to P(F)$ et un repère projectif $\{m_0, m_1, \dots, m_{n+1}\}$ de $P(E)$. Nous posons $m'_i = \phi(m_i)$ pour tout $i = 0, \dots, n+1$. Nous devons prouver que ces m'_i forment un repère projectif de $P(F)$.

D'abord pour $i = 1, \dots, n+1$ nous avons $m'_i = \phi\big(\pi_E(e_i)\big) = \pi_F\big(\bar{\phi}(e_i)\big)$, mais $\{\bar{\phi}(e_i)\}_{i=1,\dots,n+1}$ est une base de F parce que $\bar{\phi}$ est un isomorphisme d'espaces vectoriels. Donc oui : les m'_i $(i = 1, \dots, n+1)$ sont les projetés d'une base de F.

Nous posons au passage $f_i = \bar{\phi}(e_i)$. En ce qui concerne m_0 nous savons que $m_0 = \pi_E(e_1 + \dots + e_{n+1})$ et

$$m'_0 = \phi\big(\pi_E(e_1 + \dots + e_{n+1})\big) = \pi_F\big(\bar{\phi}(e_1 + \dots + e_{n+1})\big) = \pi_F(f_1 + \dots + f_{n+1}), \tag{22.48}$$

ce qui termine de montrer que $\{m'_i\}_{i=0,\dots,n+1}$ est un repère projectif de $P(F)$.

(2) Soient un repère projectif $(m_0, \dots, m_{n+1})$ de $P(E)$ et un repère projectif $(m'_0, \dots, m'_{n+1})$ de $P(F)$. Nous choisissons des bases $\{e_i\}$ de E et $\{f_i\}$ de F telles que

$$m_i = \pi_E(e_i) \tag{22.49a}$$
$$m'_i = \pi_F(f_i) \tag{22.49b}$$

pour $i = 1, \dots, n+1$ et

$$m_0 = \pi_E(e_1 + \dots + e_{n+1}) \tag{22.50a}$$
$$m'_0 = \pi_F(f_1 + \dots + f_{n+1}). \tag{22.50b}$$

Nous considérons un isomorphisme d'espace vectoriel $\bar{\phi}\colon E \to F$ tel que $\bar{\phi}(e_i) = f_i$ pour tout i, et nous voulons définir $\phi\colon P(E) \to P(F)$ par

$$\phi\big(\pi_E(v)\big) = \pi_F\big(\bar{\phi}(v)\big). \tag{22.51}$$

Cela est bien définit parce que si $\pi_E(v) = \pi_E(w)$ alors $w = \lambda v$ et

$$\pi_F\big(\bar{\phi}(\lambda v)\big) = \pi_F\big(\lambda\bar{\phi}(v)\big) = \pi_F\big(\bar{\phi}(v)\big). \tag{22.52}$$

L'application définie par (22.51) est une homographie qui envoie m_i sur m'_i pour tout $i = 0, \dots, n+1$. Ceci prouve la partie « existence » du point (2).

Pour l'unicité, soient des homographies

$$\phi_1\colon P(E) \to P(F) \tag{22.53a}$$
$$\phi_2\colon P(E) \to P(F) \tag{22.53b}$$

telles que $\phi_1(m_i) = \phi_2(m_i)$ pour tout $i = 0, \dots, n+1$. Soit aussi une base $\{e_i\}_{i=1,\dots,n+1}$ de E adaptée au repère projectif, c'est-à-dire $m_i = \pi_E(e_i)$ pour $i = 1, \dots, n+1$ et $\pi_E(e_1 + \dots + e_{n+1}) = m_0$. Nous considérons aussi les isomorphismes d'espaces vectoriels $\bar{\phi}_1$ et $\bar{\phi}_2$. Avec tout ce beau monde nous avons

$$\phi_1(m_i) = \pi_E\big(\bar{\phi}_1(e_i)\big) \tag{22.54a}$$
$$\phi_2(m_i) = \pi_E\big(\bar{\phi}_2(e_i)\big). \tag{22.54b}$$

Mais nous savons que $\phi_1(m_i) = \phi_2(m_i)$, donc nous savons que $\pi_E\big(\bar{\phi}_1(e_i)\big) = \pi_E\big(\bar{\phi}_2(e_i)\big)$, ce qui nous fait conclure que

$$\bar{\phi}_1(e_i) = \lambda_i\bar{\phi}_2(e_i) \tag{22.55}$$

pour certaines constantes $\lambda_i \in \mathbb{K}$. Le même raisonnement appliqué à m_0 nous donne un $\mu \in \mathbb{K}$ tel que

$$\bar{\phi}_1(e_1) + \ldots + \bar{\phi}_1(e_{n+1}) = \mu\big(\bar{\phi}_2(e_1) + \ldots + \bar{\phi}_2(e_{n+1})\big). \tag{22.56}$$

En mettant l'un dans l'autre :

$$\lambda_1\bar{\phi}_2(e_1) + \ldots + \lambda_{n+1}\bar{\phi}_2(e_{n+1}) = \mu\big(\bar{\phi}_2(e_1) + \ldots + \bar{\phi}_2(e_{n+1})\big). \tag{22.57}$$

Sachant que $\{\bar{\phi}_2(e_i)\}_{i=1,\ldots,n+1}$ est une base de F et nous souvenant de l'unicité de la décomposition d'un élément dans une base[4], nous en déduisons que tous les λ_i doivent être égaux à μ. Donc pour tout $v \in E$ nous avons $\bar{\phi}_1(v) = \lambda\bar{\phi}_2(v)$.

Cela a pour conséquence que $\phi_1 = \phi_2$.

□

22.33.

Si nous avons une droite projective, trois points sont nécessaires pour créer un repère et donc pour construire une homographie de la droite sur elle-même. Soit E un espace vectoriel de dimension 2 et $P(E)$ la droite projective qui lui est associée. Soit une homographie $f\colon P(E) \to P(E)$ et $\bar{f}\colon E \to E$, l'isomorphisme d'espaces vectoriels associé (par $f \circ \pi_E = \pi_E \circ \bar{f}$). Si $\{e_1, e_2\}$ est une base de E alors l'application $\bar{f}$ a une matrice

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbb{M}(2, \mathbb{K}) \tag{22.58}$$

avec $\det A \neq 0$ parce que $\bar{f}$ est un isomorphisme.

La plupart des points de $P(E)$ sont représentés par des points de la forme $(z, 1)$. Nous voudrions savoir quelle est la direction représentée par le point $\bar{f}(z, 1)$; c'est-à-dire que nous voudrions savoir $f([z, 1])$ sous la forme $[z', 1]$ (si possible). Nous avons

$$\bar{f}(z, 1) = (a_{11}z + a_{12}, a_{21}z + a_{22}). \tag{22.59}$$

Nous posons $\lambda = a_{21}z + a_{22}$ et nous avons

$$\bar{f}(z, 1) = \lambda\left(\frac{a_{11}z + a_{12}}{\lambda}, 1\right). \tag{22.60}$$

Il y a plusieurs possibilités suivant les valeurs de λ et de z.

(1) Si $\lambda = 0$ c'est que nous avons $\bar{f}(z, 1) = (a_{11}z + a_{12}, 0)$. L'application f envoie donc le point $(z : 1)$ sur le point à l'infini.

(2) Si $\lambda \neq$, alors f envoie le point $(z : 1)$ vers un autre point « normal ».

(3) Si le point de départ est le point à l'infini alors $\bar{f}(1, 0) = (a_{11}, a_{21})$. Cela peut être le point à l'infini ou non selon les valeurs des a_{ij}.

Dans tous les cas si nous posons

$$\begin{cases} \varphi_f(z) = \dfrac{a_{11}z + a_{12}}{a_{21}z + a_{22}} & \text{(22.61a)} \\ \varphi_f(\infty) = \dfrac{a_{11}}{a_{21}} & \text{(22.61b)} \end{cases}$$

alors nous avons

$$\bar{f}(z, 1) = \big(\varphi_f(z), 1\big). \tag{22.62}$$

Si nous prenons la convention que $\frac{1}{0} = \infty$ et que $(\infty, 0)$ est le point à l'infini, alors cette application φ_f donne bien toutes les valeurs de f, y compris les cas à l'infini.

4. Proposition 4.6.

Lemme 22.34 ([415]).
Trois points distincts d'une droite projective forment un repère projectif.

Démonstration. Soit une droite projective $d = P(E)$ où E est un espace vectoriel de dimension 2 sur le corps $\mathbb{K}$. Soient trois points distincts A, B et C de d. Nous avons $a, b, c \in E$ tels que $A = \pi(a)$, $B = \pi(b)$ et $C = \pi(c)$. Vu que $A \neq B$, les vecteurs a et b ne sont pas proportionnels et la partie $\{a, b\}$ est libre dans E. Autrement dit, c'est une base [5].

Il existe donc $\alpha, \beta \in \mathbb{K}$ tels que $c = \alpha a + \beta b$. De plus α et β ne sont pas nuls parce que $C \neq A$ et $C \neq B$. En prenant $a' = \alpha a$, $b' = \beta b$ et $c' = c$ nous avons : $A = \pi(a')$, $B = \pi(b')$, $C = \pi(c')$ en même temps que $\{a', b'\}$ est une base de E et $c' = a' + b'$. Donc A, B, C est un repère projectif de $d = P(E)$. □

Corolaire 22.35 ([416]).
Soient E et F des espaces vectoriels de dimension 2. Soient A_i ($i = 1, 2, 3$) des points distincts sur $P(E)$ et B_i distincts sur $P(F)$. Alors il existe une unique homographie $P(E) \to P(F)$ portant A_i sur B_i pour tout $i = 1, 2, 3$.

Démonstration. Il s'agit de mettre en conjonction le lemme 22.34 qui dit que les A_i forment un repère projectif de $P(E)$ (idem : les B_i forment un repère projectif de $P(F)$) et le théorème 22.32(2) qui dit que l'une va sur l'autre par une unique homographie. □

22.4.4 Identifications $P(\mathbb{K}^2)$ vers $\mathbb{K} \cup \{\infty\}$

Pour rappel, une droite projective est l'espace projectif modelé sur un espace vectoriel de dimension deux (définition 22.8).

22.36.
Nous allons faire un usage assez intense de bijections entre $P(\mathbb{K}^2)$ et $\hat{\mathbb{K}} = \mathbb{K} \cup \{\infty\}$. Une possible est

$$\begin{aligned} \varphi_0 \colon P(\mathbb{K}^2) &\to \mathbb{K} \cup \{\infty\} \\ [k_1, k_2] &\mapsto \begin{cases} \frac{k_1}{k_2} & \text{si } k_2 \neq 0 \\ \infty & \text{si } k_2 = 0. \end{cases} \end{aligned} \tag{22.63}$$

Notons que nous utilisons ici le fait que $\mathbb{K}$ soit commutatif, sinon il aurait fallu choisir $k_1 k_2^{-1}$ ou $k_2^{-1} k_1$ au lieu d'écrire gentiment k_1/k_2.

Corolaire 22.37.
Soit une bijection $\varphi \colon P(\mathbb{K}^2) \to \hat{\mathbb{K}}$. Les points $\{\varphi^{-1}(\infty), \varphi^{-1}(0), \varphi^{-1}(1)\}$ forment un repère projectif de $P(\mathbb{K}^2)$.

Démonstration. Il s'agit seulement d'une application du lemme 22.34.

Juste pour l'amusement, nous allons le prouver explicitement pour la bijection $\varphi = \varphi_0$ donnée en (22.63). Un repère projectif est la définition 22.26. Nous avons

$$\varphi_0^{-1}(\infty) = [1, 0] = \pi_{K^2}\big((1, 0)\big) \tag{22.64a}$$
$$\varphi_0^{-1}(0) = [0, 1] = \pi_{K^2}\big((0, 1)\big) \tag{22.64b}$$
$$\varphi_0^{-1}(1) = [1, 1] = \pi_{K^2}\big((1, 1)\big) \tag{22.64c}$$

Les points $(1, 0)$ et $(0, 1)$ forment un base de $\mathbb{K}^2$ et nous avons bien $(1, 1) = (1, 0) + (0, 1)$. Donc le tout vérifie bien la définition d'un repère projectif. □

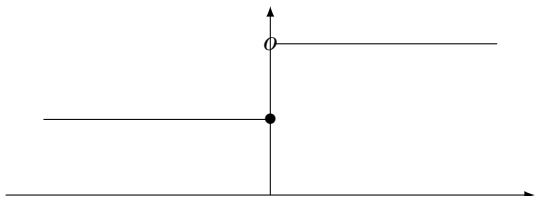
D'autre part, comme il est plus agréable de travailler avec $\hat{\mathbb{K}}$ qu'avec $P(\mathbb{K}^2)$ nous avons envie de voir $\hat{\mathbb{K}}$ comme un espace projectif (qu'il n'est pas). Il y a cependant nombre d'autres identifications

5. Il convient de citer ici la proposition 4.16.

possibles. En voici un autre :

$$\begin{aligned}\varphi_1\colon P(\mathbb{K}^2) &\to \hat{\mathbb{K}}\\ [k_1,k_2] &\mapsto \begin{cases}\frac{k_2}{k_1} & \text{si } k_1\neq 0\\ \infty & \text{si } k_1=0.\end{cases}\end{aligned}\tag{22.65}$$

Vous en voulez une plus compliquée ? En voici une pour $\mathbb{K}=\mathbb{R}$, basée sur le dessin suivant :



La bijection associée est :

$$\begin{aligned}\varphi_d\colon P(\mathbb{R}^2) &\to \hat{\mathbb{R}}\\ [k_1,k_2] &\mapsto \begin{cases}\frac{k_1}{k_2} & \text{si } k_1k_2\leqslant 0\\ 2\frac{k_1}{k_2} & \text{si } k_1k_2<0\\ \infty & \text{si } k_2=0.\end{cases}\end{aligned}\tag{22.66}$$

Pour mettre un peu d'ordre dans toutes ces identifications possibles, nous introduisons une classe.

Définition 22.38.
Pour une bijection $\varphi\colon P(\mathbb{K}^2)\to\hat{\mathbb{K}}$ nous définissons

$$A(\varphi)=\left\{\varphi_a\colon P(\mathbb{K}^2)\to\hat{\mathbb{K}}\text{ tel que }\varphi_a^{-1}\circ\varphi\text{ soit une homographie }\right\}\tag{22.67}$$

Les classes sont assez larges parce que pour toute homographie $\phi\colon P(\mathbb{K}^2)\to P(\mathbb{K}^2)$, nous avons $\varphi\circ\phi^{-1}\in A(\varphi)$. Mieux, nous avons le lemme suivant.

Lemme 22.39.
L'application

$$\begin{aligned}\psi\colon A(\varphi) &\to \mathrm{PGL}(\mathbb{K}^2)\\ \varphi_a &\mapsto \varphi_a^{-1}\circ\varphi\end{aligned}\tag{22.68}$$

est une bijection.

Démonstration. Pour rappel, $\mathrm{PGL}(E)$ est le groupe des homographies de E, voir la définition 22.23.

Surjectif Si $\phi\in\mathrm{PGL}(\mathbb{K}^2)$ nous avons $\phi=\psi(\varphi\circ\phi^{-1})$.

Injectif Si $\psi(\varphi_a)=\psi(\varphi_b)$ alors

$$\varphi_a^{-1}\circ\varphi=\varphi_b^{-1}\circ\varphi,\tag{22.69}$$

d'où nous déduisons $\varphi_a^{-1}=\varphi_b^{-1}$ parce que φ est une bijection.

□

22.4.5 Birapport

Proposition-définition 22.40.
Soit une droite projective $d=P(E)$ et trois points distincts A, B et C sur cette droite. Soit une bijection $\varphi\colon P(\mathbb{K}^2)\to\hat{\mathbb{K}}$. Si X est un point de d alors nous nommons le ***birapport*** *de X par rapport à A, B et C l'élément de $\hat{\mathbb{K}}$ donné par*

$$[A,B,C,X]_\varphi=(\varphi\circ\phi)(X)\tag{22.70}$$

où $\phi\colon d \to P(\mathbb{K}^2)$ est l'unique homographie telle que

$$\phi(A) = \varphi^{-1}(\infty) \tag{22.71a}$$
$$\phi(B) = \varphi^{-1}(0) \tag{22.71b}$$
$$\phi(C) = \varphi^{-1}(1). \tag{22.71c}$$

Démonstration. Nous devons prouver qu'il existe effectivement une unique homographie vérifiant les conditions (22.71).

A, B, C est un repère projectif de $P(E)$ Voir le lemme 22.34.

$\varphi^{-1}(\infty), \varphi^{-1}(0), \varphi^{-1}(1)$ est un repère projectif de $P(\mathbb{K}^2)$ Lemme 22.34 ou corolaire 22.37 au choix.

Conclusion Le théorème 22.32(2) nous donne existence et unicité d'une homographie $P(E) \to P(\mathbb{K}^2)$ envoyant le premier repère sur le second.

□

Remarque 22.41.
La majorité des sources ne parlent pas de la dependence du birapport en le choix de φ parce que tout le monde ne semble ne considérer que $\varphi = \varphi_0$ définie en (22.63). Il est cependant naturel de se demander si la définition dépend effectivement du choix de φ. La réponse est oui : ça dépend du choix.

Exemple 22.42 (Une autre identification qui ne va pas bien)**.**
Nous montrons que l'identification $\varphi_d\colon P(\mathbb{K}^2) \to \hat{\mathbb{K}}$ donnée en (22.66) ne donne pas lieu au même birapport que celui de φ_0.

Nous travaillons le birapport sur la droite projective la plus simple : $P(\mathbb{K}^2)$ avec $\mathbb{K} = \mathbb{R}$. Prenons pour la simplicité $A = [1, 0]$, $B = [0, 1]$ et $C = [1, 1]$. Alors l'homographie demandée dans la définition de $[.,.,.,.]_{\varphi_0}$ est $\varphi_0 = \mathrm{Id}$. Par conséquent,

$$\big[A, B, C, [k_1, k_2]\big]_{\varphi_0} = (\varphi_0 \circ \phi_0)[k_1, k_2] = \varphi_0[k_1, k_2] = \frac{k_1}{k_2}. \tag{22.72}$$

En ce qui concerne le birapport définit par φ_d nous avons

$$\varphi_d^{-1}(\infty) = [1, 0] \tag{22.73a}$$
$$\varphi_d^{-1}(0) = [0, 1] \tag{22.73b}$$
$$\varphi_d^{-1}(1) = [1, 2], \tag{22.73c}$$

de telle sorte que nous cherchons une homographie $\phi_d\colon P(\mathbb{K}^2) \to P(\mathbb{K}^2)$ telle que

$$\phi_d[1, 0] = [1, 0] \tag{22.74a}$$
$$\phi_d[0, 1] = [0, 1] \tag{22.74b}$$
$$\phi_d[1, 1] = [1, 2] \tag{22.74c}$$

L'homographie $\phi_d[k_1, k_2] = [k_1, 2k_2]$ convient et nous avons

$$\big[A, B, C, D, [k_1, k_2]\big]_{\varphi_d} = (\varphi_d \circ \phi_d)[k_1, k_2] = \varphi_d[k_1, 2k_2] = \frac{k_1}{2k_2} \tag{22.75}$$

dès que $k_1 k_2 < 0$. Nous avons donc bien trouvé

$$[A, B, C, X]_{\varphi_0} \neq [A, B, C, X]_{\varphi_d}. \tag{22.76}$$

△

Le birapport n'est pas un objet tout à fait canonique parce qu'il dépend effectivement du choix de l'identification entre $P(\mathbb{K}^2)$ et $\hat{\mathbb{K}}$.

Proposition 22.43.
Soit une bijection $\varphi\colon P(\mathbb{K}^2) \to \hat{\mathbb{K}}$. Si $\varphi_a \in A(\varphi)$[6] alors les birapports construits sur φ et φ_a coïncident.

Démonstration. Soient trois points distincts $A, B, C \in P(E)$, et $X \in P(E)$. Nous avons

$$[A, B, C, X]_\varphi = (\varphi \circ \phi)(X) \tag{22.77}$$

où $\phi\colon P(E) \to P(\mathbb{K}^2)$ est l'unique homographie telle que

$$\phi(A) = \varphi^{-1}(\infty) \tag{22.78a}$$
$$\phi(B) = \varphi^{-1}(0) \tag{22.78b}$$
$$\phi(C) = \varphi^{-1}(1). \tag{22.78c}$$

Et

$$[A, B, C, X]_{\varphi_a} = (\varphi_a \circ \phi_a)(X) \tag{22.79}$$

où $\phi_a\colon P(E) \to P(\mathbb{K}^2)$ est l'unique homographie telle que

$$\phi_a(A) = \varphi_a^{-1}(\infty) \tag{22.80a}$$
$$\phi_a(B) = \varphi_a^{-1}(0) \tag{22.80b}$$
$$\phi_a(C) = \varphi_a^{-1}(1). \tag{22.80c}$$

Il est facile de voir que $\phi_a = \varphi_a^{-1} \circ \varphi\phi$. En effet, cela est une homographie parce que $\varphi_a \in A(\varphi)$, et parce que la composée d'homographies est une homographie. De plus,

$$(\varphi_a^{-1} \circ \varphi \circ \phi)(A) = (\varphi_a^{-1} \circ \varphi)\phi^{-1}(\infty) = \varphi_a(\infty) \tag{22.81a}$$
$$(\varphi_a^{-1} \circ \varphi \circ \phi)(B) = (\varphi_a^{-1} \circ \varphi)\phi^{-1}(0) = \varphi_a(0) \tag{22.81b}$$
$$(\varphi_a^{-1} \circ \varphi \circ \phi)(C) = (\varphi_a^{-1} \circ \varphi)\phi^{-1}(1) = \varphi_a(1). \tag{22.81c}$$

Au final nous avons :

$$[A, B, C, X]_{\varphi_a} = (\varphi_a \circ \varphi_a^{-1} \circ \varphi \circ \phi)(X) = (\varphi \circ \phi)(X) = [A, B, C, X]_\varphi. \tag{22.82}$$

□

Remarque 22.44.
Tout le monde semble ne considérer que l'identification usuelle $\varphi_0\colon P(\mathbb{K}^2) \to \hat{\mathbb{K}}$ donnée par $\varphi_0[k_1, k_2] = k_1/k_2$. Toute la discussion concernant la dépendance du birapport en le choix de l'identification (y compris la définition des classes $A(\varphi)$) peut être sautée en disant qu'on ne considéra que φ_0.

Et c'est ce que nous allons faire : sauf avis contraire, nous utiliserons le birapport associé à l'identification φ_0.

Lemme 22.45 ([414])**.**
Nous avons

$$[A, B, C, X]_\varphi = \begin{cases} \infty & \text{si et seulement si } X = A \\ 0 & \text{si et seulement si } X = B \\ 1 & \text{si et seulement si } X = C. \end{cases} \tag{22.83}$$

Démonstration. Par définition $[A, B, C, X]_\varphi = \varphi \circ \phi(X)$. Nous avons donc équivalence entre les affirmations suivantes :

— $[A, B, C, X]_\varphi = \infty$
— $(\varphi \circ \phi)(X) = \infty$

6. $A(\varphi)$ définie en 22.38.

— $\phi(X) = \varphi^{-1}(\infty)$
— $\phi(X) = \phi(A)$
— $A = X$

parce que ϕ et φ sont des bijections.

Le même raisonnement tient pour les deux autres. □

Proposition 22.46.
Autres petites propriétés faciles ... Soit une droite projective $d = P(E)$ et trois points distincts $A, B, C \in d$.

(1) Les points A, B, C et X sont distincts si et seulement si $[A, B, C, X] \in \mathbb{K}\backslash\{0, 1\}$.

(2) Pour tout $k \in \hat{\mathbb{K}}$, il existe un unique $X \in d$ tel que $[A, B, C, X] = k$.

Démonstration. Notons pour le point (1) que l'énoncé demande déjà que A, B et C soient distincts. Sinon le birapport n'est pas définit.

(2) Les points A, B et C sont distincts par hypothèse. Vu le lemme 22.45, pour que X soit distincts de A, B et C il faut et il suffit que le birapport ne soit ni ∞ ni 1 ni 0. Donc $\mathbb{K}\backslash\{0, 1\}$.

(2) Nous avons $[A, B, C, X] = \phi(X)$ où $\phi\colon P(E) \to P(\mathbb{K}^2)$ est une homographie et donc une bijection par la proposition 22.22(1). Donc oui, pour tout éléments de $P(\mathbb{K}^2)$ il existe un unique élément de $P(E)$ dont le birapport par rapport à A, B et C soit cet élément.

Notons encore une fois que nous avons identifié $P(\mathbb{K}^2)$ à $\hat{\mathbb{K}}$ par la bijection (22.63). □

Proposition 22.47 ([414])**.**
Nous considérons deux droites projectives d et d' ainsi que 4 points sur chacune. Nous les nommons $A_1, A_2, A_3, A_4 \in d$ et $A'_1, A'_2, A'_3, A'_4 \in d'$. Nous supposons que A_1, A_2, A_3 sont distincts et que $A'_1, A'_2, A'3$ également. Alors il y a équivalence entre

(1) Il existe une homographie $\phi\colon d \to d'$ telle que $\phi(A_i) = A'_i$ pour $i = 1, 2, 3, 4$,

(2) égalité des birapports :

$$[A_1, A_2, A_3, A_4] = [A'_1, A'_2, A'_3, A'_4]. \tag{22.84}$$

Dans ce cas, l'homographie est unique.

Démonstration. Nous divisons la preuve en trois parties évidentes.

(1) implique (2) Nous avons une homographie $\mu'\colon d' \to \hat{\mathbb{K}}$ telle que $\mu'(A'_1) = \infty$, $\mu'(A'_2) = 0$ et $\mu'(A'_3) = 1$. En composant[7] avec l'homographie $\phi\colon d \to d'$ de l'hypothèse nous avons une homographie $\mu' \circ \phi\colon d \to \hat{\mathbb{K}}$ qui vérifie

$$(\mu' \circ \phi)(A_1) = \mu'(A'_1) = \infty \tag{22.85a}$$
$$(\mu' \circ \phi)(A_2) = \mu'(A'_2) = 0 \tag{22.85b}$$
$$(\mu' \circ \phi)(A_3) = \mu'(A'_3) = 1, \tag{22.85c}$$

ce qui signifie que $\mu' \circ \phi$ est l'homographie qui définie le birapport sur d. Par conséquent

$$[A_1, A_2, A_3, A_4] = (\mu' \circ \phi)(A_4) = \mu'(A'_4) = [A'_1, A'_2, A'_3, A'_4]. \tag{22.86}$$

(2) implique (1) La partie $\{A_1, A_2, A_3\}$ est un repère projectif de d par le lemme 22.34. Idem pour $\{A'_1, A'_2, A'_3\}$. Nous considérons les homographies $\mu\colon d \to \hat{\mathbb{K}}$ et $\mu'\colon d' \to \hat{\mathbb{K}}$ définissant les birapports par rapport à ces repères. Ces homographies vérifient, en utilisant l'hypothèse :

$$\mu(A_4) = [A_1, A_2, A_3, A_4] = [A'_1, A'_2, A'_3, A'_4] = \mu'(A'_4). \tag{22.87}$$

7. Proposition 22.22(3), la composition est encore une homographie.

Et de plus

$$\mu(A_1) = [A_1, A_2, A_3, A_1] = \infty = \mu'(A'_1) \tag{22.88a}$$
$$\mu(A_2) = [A_1, A_2, A_3, A_2] = 0 = \mu'(A'_2) \tag{22.88b}$$
$$\mu(A_3) = [A_1, A_2, A_3, A_3] = 1 = \mu'(A'_3). \tag{22.88c}$$

Autrement dit : $\mu(A_i) = \mu'(A'_i)$ pour tout $i = 1, 2, 3, 4$. Nous nous inspirons de ce diagramme :

$$\begin{tikzcd} d \arrow[rr, dotted, "\phi"] \arrow[dr, "\mu"'] & & d' \arrow[dl, "\mu'"] \\ & \mathbb{K} \cup \{\infty\} & \end{tikzcd} \tag{22.89}$$

La composée $\phi = \mu'^{-1} \circ \mu$ vérifie

$$(\mu'^{-1} \circ \mu)(A_i) = A'_i \tag{22.90}$$

pour tout i, et est une homographie.

Unicité Le fait qu'une homographie vérifiant $\phi(A_i) = A'_i$ pour $i = 1, 2, 3$ soit unique découle du fait qu'il existe une unique homographie portant un repère projectif sur un autre. A fortiori la condition $\phi(A_4) = A'_4$ ne retire rien à l'unicité.

□

Théorème 22.48 ([414]).
Une bijection entre deux droites projectives est une homographie si et seulement si elle conserve le birapport.

Démonstration. Chacun des deux sens séparément.

$\Rightarrow$ Soit une homographie $\phi\colon d \to d'$ entre deux droites projectives. Nous devons prouver que pour tout choix 4 points A, B, C, X dans d (dont A, B et C sont distincts) nous avons

$$[A, B, C, X] = [\phi(A), \phi(B), \phi(C), \phi(X)]. \tag{22.91}$$

Nous nommons $\mu\colon d \to \hat{\mathbb{K}}$ l'homographie qui donne le birapport sur d par rapport à A, B et C, et $\mu'\colon d' \to \hat{\mathbb{K}}$ celle qui donne le birapport sur d' par rapport à $\phi(A)$, $\phi(B)$, $\phi(C)$. Voici un diagramme de la situation :

$$\begin{tikzcd} d \arrow[rr, dotted, "\phi"] \arrow[dr, "\mu"'] & & d' \arrow[dl, "\mu'"] \\ & \mathbb{K} \cup \{\infty\} & \end{tikzcd} \tag{22.92}$$

Nous prouvons maintenant que $\mu' = \mu \circ \phi^{-1}$. En effet :

$$(\mu \circ \phi^{-1})\big(\phi(A)\big) = \mu(A) = \infty \tag{22.93a}$$
$$(\mu \circ \phi^{-1})\big(\phi(B)\big) = \mu(B) = 0 \tag{22.93b}$$
$$(\mu \circ \phi^{-1})\big(\phi(C)\big) = \mu(C) = 1. \tag{22.93c}$$

Par conséquent le birapport à droite dans (22.91) peut se calculer à l'aide de $\mu \circ \phi^{-1}$:

$$[\phi(A), \phi(B), \phi(C), \phi(X)] = \mu'\big(\phi(X)\big) = (\mu \circ \phi^{-1})\big(\phi(X)\big) = \mu(X) = [A, B, C, X]. \tag{22.94}$$

La première implication est prouvée.

$\Leftarrow$ Soit une bijection $f\colon d \to d'$ conservant le birapport, ainsi que trois points distincts A, B et C dans d. Vu que f est une bijection les points $f(A)$, $f(B)$ et $f(C)$ sont distincts dans d'. Par le lemme 22.34 et le théorème 22.32(2), il existe une unique homographie $\phi\colon d \to d'$ telle que $\phi(A) = f(a)$, $\phi(B) = f(B)$ et $\phi(C) = f(C)$. Pour tout $X \in d$ nous avons

$$\begin{aligned}[f(A), f(B), f(C), f(X)] &= [A, B, C, X] && \text{(22.95a)}\\ &= [\phi(A), \phi(B), \phi(C), \phi(X)] && \text{(22.95b)}\\ &= [f(A), f(B), f(C), \phi(X)]. && \text{(22.95c)}\end{aligned}$$

Justifications :

— (22.95a) parce que f conserve le birapport par hypothèse.

— (22.95b) parce que ϕ conserve le birapport étant une homographie (c'est le premier sens du présent théorème)

Nous nommons $\mu'\colon d' \to \hat{\mathbb{K}}$ l'homographie donnant le birapport par rapport aux points $f(A)$, $f(B)$, $f(C)$. Alors le résultat (22.95) se lit

$$\mu'\big(f(X)\big) = \mu'\big(\phi(X)\big). \tag{22.96}$$

Mais comme μ' est une bijection (proposition 22.22(1)) cela implique $f(X) = \phi(X)$. Vu que nous avons fait ce raisonnement pour un X quelconque dans d nous avons $f = \phi$, ce qui prouve que f est une homographie.

□

Lemme 22.49.
Soient a, b, c distincts sur la droite projective $D = P(E)$. Soient $x, y \in E$ tels que $\pi_E(x) = a$, $\pi_E(y) = b$, $\pi_E(x + y) = c$. Alors

$$d = \pi_E(\lambda x + \mu y) \tag{22.97}$$

si et seulement si

$$[a, b, c, d] = \pi_{\mathbb{K}^2}(\lambda, \mu). \tag{22.98}$$

Démonstration. Étant donné que a et b sont distincts, les vecteurs x et y forment une base de E. Soit $f\colon E \to \mathbb{K}^2$ un isomorphisme qui envoie (x, y) sur e_1, e_2 où e_i sont les vecteurs de base de $\mathbb{K}^2$. Ensuite nous considérons $g\colon P(E) \to P(\mathbb{K}^2)$, l'homographie associée à f. Par définition $f\big(\pi_E z\big) = \pi_{\mathbb{K}^2}\big(f(z)\big)$. Par f nous avons

$$a \mapsto \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad b \mapsto \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{22.99}$$

Donc par g nous avons

$$a \mapsto \infty \quad b \mapsto 0. \tag{22.100}$$

Nous avons aussi $f(\lambda x + \mu y) = (\lambda, \mu)$ et

$$\begin{aligned} g(c) &= g\big(\pi_E(x + y)\big) && \text{(22.101a)}\\ &= \pi_F f(x + y) && \text{(22.101b)}\\ &= \pi_F(f(x) + f(y)) && \text{(22.101c)}\\ &= \pi_F \begin{pmatrix} 1 \\ 1 \end{pmatrix} && \text{(22.101d)}\\ &= 1. && \text{(22.101e)}\end{aligned}$$

La dernière inégalité est le fait que la direction $(1, 1)$ dans $\mathbb{R}^2$ est représentée par le point $x = 1$ sur la droite $y = 1$ qui est notre « représentation » de la droite affine. L'application g a donc toutes les propriétés qu'il faut pour être l'application qui définit le birapport. Nous avons donc bien $g(d) = [a, b, c, d]$.

D'une part si $d = \pi_E(\lambda x + \mu y)$ alors

$$g(d) = \pi_{\mathbb{K}^2} f(\lambda x + \mu y) = \pi_{\mathbb{K}^2}(\lambda, \mu). \tag{22.102}$$

Dans l'autre sens si $[a, b, c, d] = \pi_{\mathbb{K}^2}(\lambda, \mu)$ alors supposons que $g(d) = \pi_{\mathbb{K}^2}(\lambda, \mu)$ avec $d = \pi_E(v)$ alors

$$g\pi_E v = \pi_{\mathbb{K}^2} f(v), \tag{22.103}$$

ce qui implique $f(v) = \alpha(\lambda, \mu)$ pour un certain $\alpha \in \mathbb{K}$. Par conséquent $v = \alpha(\lambda x + \mu y)$ et $d = \pi_E(\lambda x + \mu y)$. □

22.5 Coordonnées homogènes

Soit E un espace vectoriel de dimension $n+1$ et une base $\{e_0, \dots, e_n\}$ de E. Soit $M \in P(E)$ et $u \in E$ un élément engendrant M. Au point M nous voudrions associer les coordonnées $(x_0, \dots, x_n)$ de u dans E. Notons que toutes les coordonnées de u ne sont jamais nulles en même temps parce que u doit indiquer une direction. Nous savons par ailleurs que les coordonnées $(x_0, \dots, x_n)$ indiquent le même point de $P(E)$ que les coordonnées $(x'_0, \dots, x'_n)$ si et seulement si $x_i = \lambda x_i$.

Définition 22.50.
La classe d'équivalence de $(x_0, \dots, x_n)$ est la ***coordonnées homogène*** *de M. Nous la notons $(x_0 : \dots : x_n)$.*

Si nous avons une base $\{e_i\}$ de $\mathbb{R}^n$ nous associons à $M \in P(E)$ les coordonnées $(X : Y : T)$. Mais si on prend la base $\{2e_1, e_2, \dots, e_n\}$, les coordonnées du même point deviennent $(X/2 : Y : T)$ alors que du point de vue de l'espace projectif, rien n'a été changé : la classe de e_1 est la même que celle de $2e_1$. Les coordonnées homogènes[8] ne sont donc pas intrinsèques.

22.5.1 Curiosité : matrice de translation

Si E est un espace vectoriel, l'espace projectif $P(E)$ est l'ensemble des classes d'équivalence dans E pour la relation $v \sim \lambda v$ pour tout $\lambda \neq 0$.

Il se fait que l'étude de $\mathbb{R}^3$ peut être fait à partir de $\mathbb{R}^4$ en considérant les coordonnées homogènes sur $P(\mathbb{R}^4)$. Plus précisément, nous considérons

$$\begin{aligned} \varphi\colon \mathbb{R}^3 &\to P(\mathbb{R}^4) \\ (x, y, z) &\mapsto \big[(x, y, z, 1)\big]. \end{aligned} \tag{22.104}$$

Cela est injectif mais pas surjectif parce que les éléments de la forme $[(x, y, z, 0)]$ ne sont pas atteints. Ces éléments sont alors dits « à l'infini » .

Nous aurions pu placer $\mathbb{R}^3$ dans $\mathbb{R}^4$ de nombreuses autres manières ; chacune aurait donné une notion différente de « point à l'infini ».

Nous allons maintenant montrer une petite curiosité qui a une grande importance en informatique, lors de la manipulation d'objets 3D. Nous considérons la bijection

$$\begin{aligned} \varphi\colon \mathbb{R}^3 &\to \mathbb{R}^4 \\ (x, y, z) &\mapsto (x, y, z, 1). \end{aligned} \tag{22.105}$$

Soit l'opérateur de translation $T_a\colon \mathbb{R}^3 \to \mathbb{R}^3$. En considérant la matrice

$$T_a^h = \begin{pmatrix} 1 & 0 & 0 & a_x \\ 0 & 1 & 0 & a_y \\ 0 & 0 & 1 & a_z \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{22.106}$$

8. Définition 22.50.

nous avons

$$T_a = \varphi^{-1} \circ T_a^h \circ \varphi. \tag{22.107}$$

Autrement dit, ce passage de $\mathbb{R}^3$ à $\mathbb{R}^4$ permet de voir les translations comme des matrices, et c'est bien pratique.

Si $R\colon \mathbb{R}^3 \to \mathbb{R}^3$ est une rotation, la matrice correspondante sur $\mathbb{R}^4$ est

$$R^h = \begin{pmatrix} R & 0 \\ 0 & 1 \end{pmatrix}. \tag{22.108}$$

Elle se combine assez bien avec une translation parce que le produit donne

$$T_a^h R^h = \begin{pmatrix} & R & & \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix} \\ \begin{pmatrix} 0 & 0 & 0 \end{pmatrix} & & & 1 \end{pmatrix}. \tag{22.109}$$

C'est-à-dire que la translation et la rotation restent assez visible dans la matrice composée.

Note : pour la composition $R^h T_a^h$, c'est beaucoup moins vrai.

22.5.2 Dualité

Soit E un espace vectoriel de dimension $n+1$. Une forme linéaire non nulle est un élément de E^*, mais aussi un représentant d'un élément de $P(E^*)$.

Le noyau d'une forme linéaire ω est un hyperplan. Le noyau de la forme linéaire $\lambda\omega$ étant le même hyperplan, l'hyperplan est donné par toute la classe de ω dans $P(E^*)$. Nous avons donc une bijection

$$P(E^*) \leftrightarrow \{\text{hyperplans vectoriels de } E\}. \tag{22.110}$$

Soit un espace vectoriel E de dimension 3 muni d'une base $\{e_1, e_2, e_3\}$ à partir de laquelle nous construisons la base duale $\{e_1^*, e_2^*, e_3^*\}$ de l'espace dual E^*. À un élément $m \in P(E^*)$ nous associons la droite

$$H_m\{(X : Y : T) \text{ tel que } m(X, Y, T) = 0\} \tag{22.111}$$

dans $P(E)$. Si les coordonnées homogènes de m étaient $(u : v : w)$ alors l'équation de la droite H_m est

$$uX + vY + wT = 0. \tag{22.112}$$

En effet si $\omega \in E^*$ est un représentant de m alors $\omega = \lambda(ue_1^* + ve_2^* + we_3^*)$ et l'équation (22.112) est indépendante de λ ainsi que du choix du représentant dans E du point $(X : Y : T)$ dans $P(E)$.

Si les points m_1 et m_2 sont distincts dans $P(E^*)$, ils donnent deux droites $m_1(X, Y, T) = 0$ et $m_2(X, Y, T) = 0$. Les points de la droite qui joint m_1 à m_2 dans $P(E^*)$ sont de la forme $\lambda m_1 + \mu m_2$ et ils sont associés à l'équation

$$\lambda m_1(X, Y, T) + \mu m_2(X, Y, T) = 0 \tag{22.113}$$

qui sont encore des droites dans $P(E)$. Toutes ces droites passent par le point d'intersection des droits associées à m_1 et m_2. Nous avons donc

$$\bigcap_{\lambda,\mu} H_{\lambda m_1 + \mu m_2} = H_{m_1} \cap H_{m_2}. \tag{22.114}$$

Lemme 22.51.
L'application

$$\begin{aligned} P(E^*) &\to \{\textit{droites dans } P(E)\} \\ m &\mapsto H_m \end{aligned} \tag{22.115}$$

est une bijection.

Démonstration. Une droite dans $P(E)$ est donnée en coordonnées homogènes par une équation $aX + bY + cT = 0$. Cette droite est décrite par le point $(a : b : c)$ dans $P(E^*)$. Ce dernier correspond à la direction de la forme $ae_1^* + be_2^* + ce_3^*$. Cela prouve que l'application est surjective.

Pour l'injectivité, si $m_1 \neq m_2$ dans $P(E^*)$, les formes ω_1 et ω_2 associées dans E^* ne sont pas multiples l'une de l'autre. Donc les équations

$$a_1X + b_1Y + z_1T = 0 \tag{22.116}$$

et

$$a_2X + b_2Y + z_2T = 0 \tag{22.117}$$

n'ont pas de solutions communes et décrivent donc des droites distinctes. □

Lemme 22.52.
Trois points distincts m_1, m_2 et m_3 dans $P(E^*)$ sont alignés si et seulement si les droites H_{m_1}, H_{m_2} et H_{m_3} sont distinctes et concourantes.

Démonstration. Supposons avoir trois points alignés, c'est-à-dire

$$m_3 = m_1 + \mu(m_2 - m_1). \tag{22.118}$$

Soit $X : Y : T$ le point d'intersection de H_{m_1} avec H_{m_2}. Alors $m_1(X, Y, T) = m_2(X, Y, T) = 0$. En tenant compte de (22.118) nous avons alors évidemment $m_3(X, Y, T) = 0$.

Supposons maintenant que les trois droites H_{m_i} soient concourantes. Nous avons donc un point $(X : Y : T)$ dans $P(E)$ tel que $m_i(X, Y, T) = 0$. Si m_i est la classe de $a_ie_1^* + b_ie_2^* + c_ie_3^*$ alors nous avons le système

$$\begin{cases} a_1X + b_1Y + c_1T = 0 & \text{(22.119a)} \\ a_2X + b_2Y + c_2T = 0 & \text{(22.119b)} \\ a_3X + b_3Y + c_3T = 0. & \text{(22.119c)} \end{cases}$$

Afin que cela ait une solution non triviale nous devons avoir

$$\det \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} \neq 0, \tag{22.120}$$

c'est-à-dire que les points (a_i, b_i, c_i) soient alignés. □

En tenant compte de ce qui a été dit, une droite dans $P(E^*)$ est constituée de points qui fournissent des droites concourantes dans $P(E)$. Donc une droite de $P(E^*)$ se caractérise par un point de $P(E)$ (l'intersection) de la façon suivante. Un point $M_d \in P(E)$ donne lieu à un **faisceau de droites** passant par M_d. Chacune de ces droites donne lieu à un point de $P(E^*)$ et tous ces points sont alignés. Nous avons ainsi construit la droite d dans $P(E^*)$ correspondante au point M_d de $P(E)$.

22.5.3 Polynômes

Soit l'espace projectif de dimension n avec ses coordonnées homogènes $(X_0 : \ldots : X_n)$. Nous considérons l'espace affine $H \equiv X_n = 1$ dans l'espace vectoriel E de dimension $n + 1$. Nous considérons pour H un repère affine ayant pour origine le point $(0, \ldots, 0, 1)$. Considérons un polynôme homogène P sur le corps $\mathbb{K}$. L'équation

$$P(X_0, \ldots, X_n) = 0 \tag{22.121}$$

sur l'espace vectoriel E descend immédiatement à l'espace projectif : étant donné que P est homogène nous avons $P(u) = 0$ si et seulement si $P(\lambda u) = 0$.

Nous essayons de décrire l'ensemble A des points de $P(E)$ satisfaisant $P(X_0, \dots, X_n) = 0$. Nous savons que les éléments de $P(E)$ ont chacun un représentant soit dans H soit sur la droite à l'infini. Ceux de A ayant un représentant dans H sont d'équation

$$Q(x_0, \dots, x_{n-1}) = 0 \tag{22.122}$$

où Q est le polynôme donné par $Q(X_0, \dots, x_{n-1}) = P(x_0, \dots, x_{n-1}, 1)$. Les points de A ayant un représentant sur la droite à l'infini s'obtiennent par l'équation

$$R(x_0, \dots, x_{n-1}) = 0 \tag{22.123}$$

où R est le polynôme donné par $R(x_0, \dots, x_{n-1}) = P(x_0, \dots, x_{n-1}, 0)$.

Exemple 22.53.
Nous considérons la conique projective

$$X^2 - XT - Y^2 - T^2 = 0. \tag{22.124}$$

Elle est décomposée en deux parties : une dans l'espace affine « normale » et une à l'infini. La première s'obtient en posant $T = 1$ dans (22.124) :

$$x^2 - x - y^2 - 1 = 0. \tag{22.125}$$

L'autre est obtenue en posant $T = 0$:

$$x^2 - y^2 = 0. \tag{22.126}$$

La partie à l'infini est donc composée de deux points : $(1 : 1 : 0)$ et $(1 : -1 : 0)$.

Le graphique de l'équation (22.125) est donné à la figure 22.2. Nous y voyons que les asymptotes sont effectivement données par les directions $(1, 1)$ et $(1, -1)$ dans le plan.

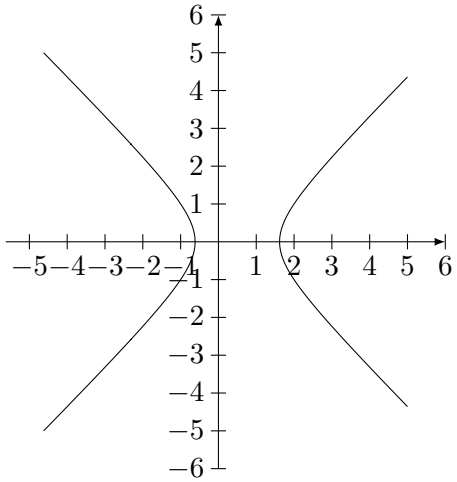


FIGURE 22.2 – Le graphique de $x^2 - x - y^2 - 1 = 0$.

△

Nous pouvons tenter de faire l'exercice inverse : considérer une conique dans $\mathbb{R}^2$, la voir comme une partie d'une conique dans l'espace projectif et trouver les points à l'infini qui la complètent.

Exemple 22.54.
La droite projective usuelle est donnée par la droite affine $y - 1 = 0$. L'homogénéisation donne $y - z = 0$ et par conséquent la partie à l'infini est donnée par $y = 0$, c'est-à-dire la direction $(1, 0)$ comme il se doit. △

Exemple 22.55.
Prenons la conique

$$x^2 + xy + y^3 - 2 = 0. \tag{22.127}$$

D'abord nous homogénéisons cette équation pour la voir dans $\mathbb{R}^3$:

$$x^2z + xyz + y^3 - 2z^3 = 0. \tag{22.128}$$

Les points à l'infini sont ceux qui correspondent à $z = 0$, c'est-à-dire la droite donnée en coordonnées homogènes par $(1 : 0 : 0)$. △

22.6 La sphère de Riemann $P_1(\mathbb{C})$

Définition 22.56.
*La **sphère de Riemann** est l'espace projectif modelé sur $\mathbb{C}^2$: en vertu des notations données à la page 1579, c'est*

$$P_1(\mathbb{C}) = P(\mathbb{C}^2). \tag{22.129}$$

L'ensemble $P_1(\mathbb{C})$ est le quotient $\mathbb{C}^2\backslash\{(0,0)\}/\sim$ où $\sim$ est la relation d'équivalence de $\mathbb{C}$-colinéarité dans $\mathbb{C}^2$.

Lemme 22.57.
L'application

$$\begin{aligned} \varphi_0 \colon P_1(\mathbb{C}) &\to \mathbb{C} \cup \{\infty\} \\ [z_1, z_2] &\mapsto \begin{cases} \frac{z_1}{z_2} & si\ z_2 \neq 0 \\ \infty & si\ z_2 = 0 \end{cases} \end{aligned} \tag{22.130}$$

est une bijection qui respecte la conjugaisons complexe : $\varphi_0\big([z_1, z_2]^*\big) = \varphi_0\big([z_1, z_2]\big)^*$.

Démonstration. Notons d'abord que la définition a un sens parce que si un représentant que $[z_1, z_2]$ est de la forme $(z, 0)$ alors ils sont tous de cette forme. L'affirmation « $z_1 \neq 0$ dans $[z_1, z_2]$ » a donc un sens.

Injectif Supposons $\varphi_0\big([z_1, z_2]\big) = \varphi_0\big([t_1, t_2]\big)$.

Si les deux membres sont égaux à ∞ alors nous avons $z_2 = t_2 = 0$, et alors avec $\lambda = z_1/t_1$ nous avons $(z_1, z_2) = \lambda(t_1, t_2)$, ce qui prouve que $[z_1, z_2] = [t_1, t_2]$.

Si les deux membres sont égaux à zéro alors $z_1 = t_1 = 0$ et le même raisonnement tient.

Sinon nous avons $z_1/z_2 = t_1/t_2$ où tous les nombres sont non nuls. Cela donne

$$z_2 = \frac{t_2}{t_1} z_1, \tag{22.131}$$

et donc

$$\frac{t_1}{z_1}(z_1, z_2) = (t_1, t_2), \tag{22.132}$$

qui montre qu'au niveau des classes, $[z_1, z_2] = [t_1, t_2]$.

Surjectif Nous avons

$$\infty = \varphi_0\big([1, 0]\big) \tag{22.133}$$

et si $z \neq \infty$ nous avons $z = \varphi_0\big([z, 1]\big)$.

□

22.6.1 Éléments de géométrie dans $P_1(\mathbb{C})$

Étant donné que nous sommes partis pour faire de la géométrie dans $\mathbb{C}$ et même dans $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, autant nous armer des équations de cercles et de droites dans $\mathbb{C}$, ainsi que de quelques notions adjacentes.

Remarque 22.58.
La définition 22.2 parle de plan et de droites projectives. Ici nous ne sommes pas dans ce cadre parce que nous travaillons sur $P_1(\mathbb{C})$ où $\mathbb{C}$ n'est certainement pas un espace de dimension 3. Les droites dont nous allons parler ne sont pas des droites projectives avec leur point à l'infini.

22.6.1.1 Équation complexe d'une droite

L'équation d'une droite dans $\mathbb{R}^2$ est $d \equiv ax + by = c$ avec $a, b, c \in \mathbb{R}$ et a, b non nuls en même temps. En posant $z = x + iy$ nous voulons exprimer l'équation en termes de z au lieu de x et y. Nous avons[417]

$$x = \frac{z + \bar{z}}{2}, \quad y = \frac{z - \bar{z}}{2i}, \tag{22.134}$$

et nous pouvons écrire $d \equiv a(z + \bar{z}) - ib(z - \bar{z}) = 2c$, ou encore $d \equiv (a - bi)z + (a + ib)\bar{z} = 2c$. En posant $\omega = a + bi \in \mathbb{C}^*$ et $k = 2c \in \mathbb{R}$ nous avons l'équation

$$\bar{\omega}z + \omega z = k. \tag{22.135}$$

Définition 22.59.
*Une **droite** est une partie de $\hat{\mathbb{C}}$ de la forme*

$$d(\omega, k) = \{z \in \mathbb{C} \text{ tel que } \bar{\omega}z + \omega\bar{z} = k\} \cup \{\infty\} \tag{22.136}$$

avec $\omega \in \mathbb{C}^*$ et $k \in \mathbb{R}$.

Dans $\hat{\mathbb{C}}$, toutes les droites contiennent le point ∞.

Problèmes et choses à faire

La proposition 21.11 montre que toute inversion transforme un cercle-droite en un cercle-droite, nonobstant d'accepter de prolonger toute droite par ∞.

Est-ce que l'on peut dire que toutes les droites contiennent le point ∞ ?

En donnant ∞ à toutes les droites et à aucun cercle, la proposition 21.11 fonctionne partout en posant $i_C(O) = \infty$ et $i_C(\infty) = O$.

De plus en pensant à la projection stéréographique, ce serait logique : quelle que soit la direction dans laquelle un point s'éloigne de $z = 0$, son image par l'inverse de la projection stéréographique s'approche du pôle nord.

22.6.1.2 Équation complexe d'un cercle

Un cercle de centre $\omega \in \mathbb{C}$ et de rayon r a pour équation $|z - \omega| = r$, et nous avons les équivalences suivantes :

$$|z - \omega| = r \Leftrightarrow |z - \omega|^2 = r^2 \Leftrightarrow (z - \omega)(\bar{z} - \bar{\omega}) = r^2 \Leftrightarrow z\bar{r} - \bar{\omega}z - \omega\bar{z} = r^2 - |\omega|^2. \tag{22.137}$$

Donc un cercle de centre $\omega \in \mathbb{C}$ et de rayon $r \in \mathbb{R}$ a pour équation

$$z\bar{z} - \bar{\omega}z - \omega\bar{z} = r^2 - |\omega|^2. \tag{22.138}$$

Définition 22.60.
*Un **cercle** dans $\hat{\mathbb{C}}$ est une partie de la forme*

$$C(\omega, r) = \{z \in \mathbb{C} \text{ tel que } z\bar{z} - \bar{\omega}z - \omega\bar{z} = r^2 - |\omega|^2\} \tag{22.139}$$

avec $\omega \in \mathbb{C}$ et $r \in \mathbb{R}$.

Dans la sphère de Riemann, aucun cercle ne contient le point ∞.

Exemple 22.61.
Trouvons le centre et le rayon du cercle d'équation

$$\bar{\omega}z + \omega\bar{z} = kz\bar{z} \tag{22.140}$$

avec $k \neq 0$. En divisant par k et en posant $\sigma = \omega/k$ nous avons :

$$z\bar{z} - \bar{\sigma}z - \sigma\bar{z} = 0. \tag{22.141}$$

Cela est un cercle de centre σ et de rayon $|\sigma|$. En effet si $z \in \mathbb{C}$ vérifie cette équation,

$$|z-\sigma|^2 = (z-\sigma)(\bar{z}-\bar{\sigma}) = \underbrace{z\bar{z} - \bar{\sigma}z - \sigma\bar{z}}_{=0} + |\sigma|^2 = |\sigma|^2, \tag{22.142}$$

c'est-à-dire que tous les points de $\mathbb{C}$ qui vérifient l'équation donnée sont à la distance $|\sigma|$ de σ. En particulier $z = 0$ est sur le cercle. △

22.6.1.3 Cercle-droite

Une chose de bien avec les équations complexes, c'est que nous pouvons écrire les droites et les cercles avec le même type d'équations.

Lemme-définition 22.62 ([417]).
*Un **cercle-droite** est l'ensemble des points $z \in \mathbb{C}$ tels que*

$$az\bar{z} - \bar{\omega}z - \omega\bar{z} = k \tag{22.143}$$

avec $a, k \in \mathbb{R}$ et $\omega \in \mathbb{C}$.

(1) Si $a = 0$, cela est une droite ;

(2) si $a \neq 0$, cela est un cercle.

(3) Un cercle-droite peut être l'ensemble vide.

Démonstration. Si $a = 0$ alors nous tombons tout de suite sur l'équation (22.135). Si $a \neq 0$ alors nous pouvons diviser par a, poser $\sigma = \omega/a$ et $l = k/a$ pour obtenir

$$z\bar{z} - \bar{\sigma} - \sigma\bar{z} = l, \tag{22.144}$$

qui est l'équation (22.138) d'un cercle ... ou pas tout à fait. En effet, (22.144) serait l'équation du cercle de centre σ et de rayon r donné par $l = r^2 - |\sigma|^2$, c'est) dire

$$r^2 = l + |\sigma|^2, \tag{22.145}$$

alors que rien n'assure que le nombre $l + |\sigma|^2$ soit positif. Dans le cas où c'est positif, nous avons bien un cercle. Sinon c'est l'ensemble vide. □

Remarque 22.63.
Lorsque nous parlons de cercle-droite, nous parlons de partie de $\mathbb{C}$ et non de $\hat{\mathbb{C}}$ parce que l'équation (22.143) a du mal à traiter le cas $z = \infty$. À cause du fait que nous avons décider de donner le point ∞ à toutes les droites, la fusion des notions de droites et de cercles n'est pas totale ; en tout cas pas en une seule équation.

Exemple 22.64 ([417]).
Soit le cercle de centre $\omega = ir$ et de rayon r. Quelle que soit la valeur de $r > 0$, ce cercle passe par le point 0 et l'axe réel lui est tangent. L'équation de ce cercle est :

$$z\bar{r} + irz - ir\bar{z} = 0. \tag{22.146}$$

Vu que $ir \neq 0$ nous pouvons diviser et obtenir

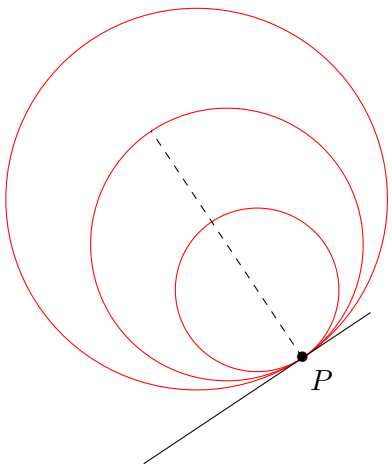
$$\frac{z\bar{z}}{ir} + z - \bar{z} = 0. \tag{22.147}$$

En faisant tendre r vers ∞ nous obtenons $z - \bar{z} = 0$, c'est-à-dire l'équation de la droite réelle.

Cela explique pourquoi il est souvent dit qu'une droite est un cercle dont le rayon est à l'infini. △

22.65.
Notons que l'exemple 22.64 est générique : prenez une droite ℓ, un point P sur ℓ, et considérez un cercle dont le centre est situé sur la perpendiculaire à ℓ passant par P, et dont le rayon est tel que le cercle passe par P. En prenant $|\omega - P| \to \infty$, l'équation du cercle devient celle de la droite ℓ.

Cela est particulièrement pratique lorsque nous travaillons dans $\hat{\mathbb{C}}$ parce nous y avons une notion précise du point à l'infini. Notons que (peut-être contre-intuitivement), il existe un seul point à l'infini dans $\hat{\mathbb{C}}$. Et ce point est le centre de tous les cercles que l'on veut transformer en droites. Cela pose évidemment la question de savoir comment on définit précisément un cercle dont le centre est réellement ∞.



22.6.1.4 Rotation-homothétie

Définition 22.66.
*Une **rotation-homothétie** est une application $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$ de la forme $z \mapsto \lambda z$ avec $\lambda \in \mathbb{C}$.*

Le nom provient du fait que si λ est réel, alors $z \mapsto \lambda z$ est une vraie homothétie, et si $\lambda = e^{i\theta}$ alors $z \mapsto e^{i\theta} z$ est une vraie rotation. Pour une valeur $\lambda \in \mathbb{C}$ générique, l'application $z \mapsto \lambda z$ est une composée des deux.

22.6.1.5 Application linéaire

Nous nous en voudrions de ne pas parler d'applications linéaires lorsque nous parlons de géométrie sur $\hat{\mathbb{C}}$. Soit $\alpha \in \mathbb{C}^*$ et $\beta \in \mathbb{C}$. Lorsque nous parlons de l'application linéaire

$$\begin{aligned} f\colon \hat{\mathbb{C}} &\to \hat{\mathbb{C}} \\ z &\mapsto \alpha z + \beta, \end{aligned} \tag{22.148}$$

nous entendons implicitement que $f(\infty) = \infty$.

22.6.1.6 Inversion

L'inversion d'un cercle de $\mathbb{R}^2$ est définie par la proposition 21.5. De nombreuses propriétés y sont décrites, y compris son écriture complexe dans la proposition 21.10. Tout cela était du temps de $\mathbb{R}^2$ ou de $\mathbb{C}$, mais maintenant nous sommes dans $\hat{\mathbb{C}}$ et nous voulons plus.

Définition 22.67 ([417]).
*Soient $\omega \in \mathbb{C}$ et $R \in \mathbb{R}^*$. L'**inversion** de centre ω et de **puissance** R^2 est l'application*

$$\begin{aligned} i\colon \hat{\mathbb{C}} &\to \hat{\mathbb{C}} \\ z &\mapsto \begin{cases} \dfrac{R^2}{\bar{z}-\bar{\omega}} + \omega & si\ z \in \mathbb{C}\backslash\{\omega\} \\ \infty & si\ z = \omega \\ \omega & si\ z = \infty. \end{cases} \end{aligned} \tag{22.149}$$

Notons que grâce aux conventions type $1/0 = \infty$ et $1/\infty = 0$, nous pouvons nous contenter de la première formule pour tout $z \in \hat{\mathbb{C}}$, et nous n'avons en réalité pas besoin de décrire $i(\infty)$ et $i(\omega)$ séparément.

Exemple 22.68.
L'inversion de cercle de centre 0 et de rayon 1 est l'application $z \mapsto \frac{1}{\bar{z}}$, que l'on prolonge avec $0 \mapsto \infty$ et $\infty \mapsto 0$. △

22.6.2 Homographies

La notion d'homographie est la définition 22.20. Pour une homographie $\phi\colon P_1(\mathbb{C}) \to P_1(\mathbb{C})$ nous avons un isomorphisme d'espace vectoriel $\bar{\phi}\colon \mathbb{C}^2 \to \mathbb{C}^2$. Vue la bijection (22.130), nous voulons plutôt travailler avec $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ qui est un ensemble avec lequel nous sommes plus familier. Nous allons donc travailler avec

$$\begin{aligned} &\tilde{\phi}\colon \hat{\mathbb{C}} \to \hat{\mathbb{C}} \\ &\tilde{\phi} = \varphi_0 \circ \phi \circ \varphi_0^{-1}. \end{aligned} \tag{22.150}$$

Proposition 22.69.
L'application $\phi = \varphi_0^{-1} \circ \tilde{\phi} \circ \varphi_0$ est une homographie de $P(\mathbb{C}^2)$ si et seulement si l'application $\tilde{\phi}\colon \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ est de la forme

$$\tilde{\phi}(z) = \frac{az+b}{cz+d} \tag{22.151}$$

avec $a, b, c, d \in \mathbb{C}$ tels que $ad - bc \neq 0$.

Par convention nous posons $z/0 = \infty$ dès que $z \neq 0$; en particulier $\tilde{\phi}(\infty) = a/c$ et $\tilde{\phi}(-d/c) = \infty$.

Démonstration. Nous séparons la condition suffisante de la condition nécessaire.

$\Rightarrow$ La condition $\pi \circ \bar{\phi} = \phi \circ \pi$ (de (22.32)) nous dit que

$$\big[\bar{\phi}(z_1, z_2)\big] = \phi\big([z_1, z_2]\big), \tag{22.152}$$

et comme $\bar{\phi}$ est un isomorphisme d'espace vectoriel nous avons $a, b, c, d \in \mathbb{C}$ vérifiant $ad - cb \neq 0$ pour lesquels

$$\bar{\phi}(z_1, z_2) = \begin{pmatrix} az_1 + bz_2 \\ cz_1 + dz_2 \end{pmatrix}. \tag{22.153}$$

Soit $z \in \mathbb{C}$. Alors nous avons

$$\tilde{\phi}(z) = (\varphi_0 \circ \phi)[z, 1] = \varphi_0\Big(\big[\bar{\phi}(z, 1)\big]\Big) = \varphi_0\big([az+b, cz+d]\big) = \frac{az+b}{cz+d}. \tag{22.154}$$

Il est important de comprendre que cette formule fonctionne pour tout $z \in \mathbb{C}$. En effet nous pourrions avoir un doute sur $z = -d/c$. D'abord si $c = 0$ alors $d \neq 0$ et ce problème n'existe pas : le dénominateur est toujours non nul. Nous avons donc seulement un doute lorsque $c \neq 0$. Dans ce cas,

$$\tilde{\phi}(-d/c) = \frac{-\frac{ad}{c} + b}{0}. \tag{22.155}$$

Mais $c \neq 0$, donc le numérateur est non nul. Or lorsque $z \neq 0$ nous avons posé $z/0 = \infty$, donc dans notre cas,

$$\tilde{\phi}(-d/c) = \infty \tag{22.156}$$

automatiquement, et cela est encodé dans la formule (22.154).
Il nous reste à déterminer $\tilde{\phi}(\infty)$. Nous avons :

$$\tilde{\phi}(\infty) = (\varphi_0 \circ \phi)\big([1,0]\big) = \varphi_0\big[\bar{\phi}(1,1)\big] = \varphi_0\big([a,c]\big) = \begin{cases} a/c & \text{si } c \neq 0 \\ \infty & \text{si } c = 0 \end{cases} \tag{22.157}$$

où la distinction entre les deux cas n'est pas fondamentale parce que si $c = 0$ alors $a \neq 0$ et $a/c = \infty$.

$\Leftarrow$ En notant $Z = [z_1, z_2]$ (pour $z_1, z_2 \in \mathbb{C}$) nous avons

$$\phi\big([z_1, z_2]\big) = \varphi^{-1}\left(\frac{a\varphi_0(Z) + b}{c\varphi_0(Z) + d}\right) = \big[a\varphi_0(Z) + b, c\varphi_0(Z) + d\big]. \tag{22.158}$$

Nous définissons $\bar{\phi}$ par son action sur les vecteurs de base : $\bar{\phi}(1,0) = (a,c)$ et $\bar{\phi}(0,1) = (b,d)$. Nous avons bien l'isomorphisme d'espace vectoriel $\bar{\phi}\colon \mathbb{C}^2 \to \mathbb{C}^2$ dont nous avons besoin pour la définition 22.20. D'abord le fait que $ad - cb$ soit non nul assure que le $\bar{\phi}$ ainsi défini est bien bijectif[9]. Et de plus ce $\bar{\phi}$ vérifie la condition

$$\begin{aligned} \pi\big(\bar{\phi}(z_1, z_2)\big) &= \pi\big((az_1, cz_1) + (bz_2, dz_2)\big) && \text{(22.159a)} \\ &= [az_1 + bz_2, cz_1 + dz_2] && \text{(22.159b)} \\ &= \big[a\varphi_0(Z) + b, c\varphi_0(Z) + d\big] && \text{(22.159c)} \\ &= \phi\big(\pi(z_1, z_2)\big). && \text{(22.159d)} \end{aligned}$$

Nous avons utilisé la notion de classe pour diviser par z_2 et faire apparaitre $\varphi_0(Z)$.

□

Remarque 22.70.
En prenant les conventions relativement claires $\infty \times a = \infty$ (pour $a \neq 0$) et $\infty \pm a = \infty$ (avec $a \neq \infty$), alors tout est dans la formule

$$\bar{\phi}(z) = \frac{az + b}{cz + d} \tag{22.160}$$

avec $ad - cb \neq 0$. Il n'y a pas besoin de traiter séparément le cas $z = \infty$ ou $z = -d/c$.

Définition 22.71.
*Nous aimons tellement l'identification $\varphi_0\colon P(\mathbb{C}^2) \to \hat{\mathbb{C}}$ que nous allons parler d'**homographie** sur $\hat{\mathbb{C}}$ pour les applications $\tilde{\phi}$ de la forme*

$$\tilde{\phi}(z) = \frac{az + b}{cz + d} \tag{22.161}$$

avec $ad - cb \neq 0$.

22.72.
La proposition 22.69 nous indique que les homographies de $\hat{\mathbb{C}}$ sont de la forme $\phi = \varphi_0^{-1} \circ \tilde{\phi} \circ \varphi_0$ pour une homographie $\phi\colon P(\mathbb{C}^2) \to P(\mathbb{C}^2)$.

Proposition 22.73 ([417]).
L'application $h\colon \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ associée à une homographie est soit linéaire, soit de la forme $h = l_1 \circ \iota \circ l_2$ où l_i sont linéaires et ι est l'application $z \mapsto 1/z$.

9. Si vous n'en êtes pas convaincu, écrivez la matrice de l'application qui envoie $(1,0)$ sur (a,c) et $(0,1)$ sur (b,d), et demandez-vous sous quelle condition elle est inversible.

Démonstration. Commençons par une remarque : lorsque nous parlons d'une application linéaire, c'est au sens de la note 22.6.1.5 qui explique qu'une application linéaire sur $\mathbb{C}$ est automatiquement prolongée à $\hat{\mathbb{C}}$ par $f(\infty) = \infty$.

Soit donc l'application

$$h(z) = \frac{az + b}{cz + d}. \tag{22.162}$$

Si $c = 0$, alors c'est une application linéaire et la preuve est terminée. Nous supposons que $c \neq 0$. Nous posons

$$l_2(z) = cz + d, \tag{22.163}$$

et ensuite $l_1(z) = \alpha z + \beta$ avec α et β à déterminer. Un peu de calcul :

$$(l_1 \circ \iota \circ l_2)(z) = l_1\left(\frac{1}{cz + d}\right) = \frac{\alpha + \beta cz + \beta d}{cz + d}, \tag{22.164}$$

et en imposant que cela soit égal à $\frac{az+b}{cz+d}$ nous trouvons $\beta = a/c$ et $\alpha = b - ad/c$. Il est vite vérifier que ces choix donnent le bon résultat. □

22.74.
Vu que les applications linéaires sont des composées d'une translation et d'une rotation-homothétie, et que l'application i est une composée d'une inversion $z \mapsto 1/\bar{z}$ et d'une réflexion $z \mapsto \bar{z}$, toutes les homographies sont des composées des éléments suivants :

— inversion[10] $z \mapsto 1/\bar{z}$, prolongée par $i(\infty) = 0$ et $i(0) = \infty$;

— réflexion $z \mapsto \bar{z}$;

— translation $z \mapsto z + \alpha$ avec $\alpha \in \mathbb{C}$;

— rotation-homothétie $z \mapsto \lambda z$ avec $\lambda \in \mathbb{C}$.

Toutes ces opérations sont prolongées à $\hat{\mathbb{C}}$ par $1/\infty = 0$, $\lambda \cdot \infty = \omega$ (si $\lambda \neq 0$) et $\infty + \lambda = \infty$. Nous ne définissons pas $0 \cdot \infty$ et $\infty - \infty$.

Certes nous pouvons construire des homographies à partir d'ingrédients dont la conjugaison complexe. Il ne faudrait cependant pas déduire que cette conjugaison est une homographie.

Lemme 22.75.
La conjugaison complexe n'est pas une homographie.

Démonstration. Si elle l'était nous aurions des nombres $a, b, c, d \in \mathbb{C}$ tels que $ad - bc \neq 0$ et

$$\frac{az + b}{cz + d} = \bar{z} \tag{22.165}$$

pour tout $z \in \mathbb{C}$.

En posant $z = 0$ nous avons déjà $b/d = 0$, c'est-à-dire $b = 0$. Avec $z = 1$ nous trouvons alors $a/(c + d) = 1$, c'est-à-dire

$$a = c + d. \tag{22.166}$$

Si $ci + d \neq 0$ Dans ce cas nous pouvons évaluer (22.165) en $z = i$ et avoir $a = -ci + d$. Mais comme nous avions déjà $a = c + d$ nous déduisons $c = 0$. Nous restons donc avec

$$\frac{a}{d} z = \bar{z} \tag{22.167}$$

pour tout z. En prenant $z = 1$ puis $z = i$, il est vite remarqué que cela n'est pas possible.

10. Oui, c'est l'inversion de la géométrie hyperbolique, voir 22.6.1.6.

Si $ci + d = 0$ Nous rappelons que $ad \neq 0$. Nous écrivons l'équation avec $z = -i$ pour trouver

$$\frac{-ai}{-ci + d} = i, \tag{22.168}$$

qui donne immédiatement $a = ci - d$. Nous avons donc les trois équations

$$\begin{cases} c = id & (22.169a) \\ a = ci - d & (22.169b) \\ a = c + d. & (22.169c) \end{cases}$$

Une tentative de résolution tombe rapidement sur une impossibilité (en substituant la première dans les deux autres et en comparant les deux valeurs de a par exemple).

□

La proposition suivante ressemble à s'y méprendre à la proposition 21.11, mais elle diffère en deux points. D'abord elle ne traite que de l'inversion par rapport à l'origine, mais surtout, elle traite le point $z = \infty$. C'est un avantage de travailler sur $\hat{\mathbb{C}}$ plutôt que sur $\mathbb{R}^2$.

Proposition 22.76 (Inversion de cercles et de droites)**.**
L'inversion dans $\hat{\mathbb{C}}$ envoie

(1) une droite passant par 0 sur elle-même

(2) une droite ne passant pas par 0 sur un cercle passant par 0.

(3) un cercle ne passant pas par 0 en un cercle ne passant pas par 0.

(4) un cercle passant par 0 en une droite ne passant pas par 0.

Démonstration. Décomposition en tous les cas possibles.

Droite passant par 0 La façon la plus simple de traiter la droite passant par 0 est de l'écrire sous forme paramétrique :

$$z(t) = te^{i\theta} \tag{22.170}$$

pour θ fixé et $t \in \mathbb{R} \cup \{\infty\}$. En appliquant l'inversion :

$$\iota\big(z(t)\big) = \frac{1}{(te^{i\theta})^*} = \frac{1}{t}e^{i\theta}. \tag{22.171}$$

Notons que les cas particuliers fonctionnent : pour $t = 0$ nous avons le point ∞ et pour $t = \infty$ nous avons 0.

Droite ne passant pas par 0 Une droite ne passant par par $z = 0$ est un ensemble de la forme

$$d(\omega, k) = \{z \in \mathbb{C}^* \text{ tel que } \bar{\omega}z + \omega\bar{z} = k\} \cup \{\infty\} \tag{22.172}$$

avec $k \neq 0$. Étant donné que ι est une bijection et même une involution nous avons $z \in \iota\big(d(\omega, k)\big)$ si et seulement si $\tau(z) \in d(\omega, k)$. L'équation est donc, pour $z \neq 0$:

$$\frac{\bar{\omega}}{z} + \frac{\omega}{\bar{z}} = k. \tag{22.173}$$

Et comme $z \neq 0$ nous pouvons multiplier par $z\bar{z}$ pour trouver $\bar{z}\bar{\omega} + z\omega = kz\bar{z}$. Donc

$$\iota\big(d(\omega, k)\big) = \{z \in \mathbb{C}^* \text{ tel que } \bar{z}\bar{\omega} + z\omega = kz\bar{z}\} \cup \{0\}. \tag{22.174}$$

Dans l'ensemble, nous pouvons renommer z et $\bar{z}$ pour avoir une forme plus symétrique. De plus il se fait que $z = 0$ vérifie l'équation donnée ; nous pouvons donc lever la condition $z \in \mathbb{C}^*$ et ne plus ajouter $\{0\}$ à côté :

$$\iota\big(d(\omega, k)\big) = \{z \in \mathbb{C} \text{ tel que } \bar{\omega}z + \omega\bar{z} = kz\bar{z}\}. \tag{22.175}$$

Cela est l'équation d'un cercle passant par l'origine (définition 22.60).

Cercle ne passant pas par 0 Nous considérons le cercle $C(\omega, r)$ avec $r^2 \neq |\omega|^2$. Il ne contient ni ∞ ni 0 et nous avons alors

$$\iota\big(C(\omega, r)\big) = \{z \in \mathbb{C}^* \text{ tel que } \frac{1}{z\bar{z}} - \bar{\omega}\frac{1}{\bar{z}} - \omega\frac{1}{z} = r^2 - |\omega|^2\}. \tag{22.176}$$

Vu que z n'est jamais nul nous pouvons multiplier l'équation par $z\bar{z}$:

$$\iota\big(C(\omega, r)\big) = \{z \in \mathbb{C} \text{ tel que } \big(|\omega|^2 - r^2\big)z\bar{z} - \bar{\omega}z - \omega\bar{z} = -1\}. \tag{22.177}$$

Le coefficient $|\omega|^2 - r^2$ est non nul par hypothèse et cet ensemble est un cercle par le lemme 22.62. Il ne passe manifestement pas par $z = 0$.

Cercle passant par 0 Le cercle passe par 0, et donc son image par ∞. Nous écrivons alors

$$\iota\big(C(\omega, |\omega|)\big) = \iota\big(C(\omega, |\omega|)\backslash\{0\}\big) \cup \{\infty\}. \tag{22.178}$$

Nous avons

$$C(\omega, |\omega|)\backslash\{0\} = \{z \in \mathbb{C}^* \text{ tel que } z\bar{z} - \bar{\omega}z - \omega\bar{z} = 0\}, \tag{22.179}$$

et un calcul usuel donne

$$\iota\big(C(\omega, |\omega|)\backslash\{0\}\big) = \{z \in \mathbb{C}^* \text{ tel que } 1 - \bar{\omega}z - \omega\bar{z} = 0\}, \tag{22.180}$$

et donc

$$\iota\big(C(\omega, |\omega|)\big) = d(\omega, 1), \tag{22.181}$$

en nous souvenant que le point ∞ est contenu dans $d(\omega, 1)$.

□

Proposition 22.77.
Une homographie conserve l'ensemble des droites et cercles de $\hat{\mathbb{C}}$.

Attention : cela ne veut pas dire qu'une homographie transforme une droite en une droite et un cercle en un cercle. Ça veut dire qu'une homographie transforme une droite en une droite ou un cercle et un cercle en une droite ou un cercle.

Démonstration. Nous savons par la proposition 21.11 et 22.74 que les homographies se décomposent en inversion, réflexion, translation et rotation-homothétie.

À part pour l'inversion, tout est clair comment ça fonctionne hein. En ce qui concerne l'inversion, nous avons la proposition 22.76 qui donne déjà toutes les réponses.

□

22.78.
Les homographies préservent les angles, c'est l'objet du théorème suivant. Il ne faudrait cependant pas croire que si A, B et C sont trois points distincts, l'angle entre $\overline{AC}$ et $\overline{BC}$ est le même que celui entre $\overline{f(A)f(C)}$ et $\overline{f(B)f(C)}$ dès que f est une homographie. Cela serait préserver les angles globalement, c'est-à-dire préserver les angles lorsque les points sont déplacés par f.

Nous allons regarder les angles locaux, c'est-à-dire lorsque les points sont déplacés par df.

Définition 22.79.
Nous disons qu'une application $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ préserve localement les angles non orientés lorsque

$$\cos\big(df_a(u), df_a(v)\big) = \cos(u, v) \tag{22.182}$$

pour tout $a \in \mathbb{R}^2$ et $u, v \in \mathbb{R}^2$. Ici il est mieux de penser à $u, v \in T_a\mathbb{R}^2$ pour qui sait les espaces tangents en géométrie différentielle.

Voir la définition de l'angle 17.41.

Théorème 22.80.
Les homographies de $P(\mathbb{C}^2)$ préservent localement les angles non orientés.

Démonstration. En ce qui concerne les translations, dilatations et rotations, les choses sont claires. Vérifions pour l'inversion, qu'il faut interpréter dans $\mathbb{R}^2$. Pour $z = x + iy$ nous avons

$$\iota(z) = \frac{1}{z} = \frac{x}{x^2+y^2} - i\frac{y}{x^2+y^2}. \tag{22.183}$$

Nous devons donc étudier la fonction

$$\begin{aligned} F\colon \mathbb{R}^2 &\to \mathbb{R}^2 \\ (x,y) &\mapsto \left(\frac{x}{r^2}, -\frac{y}{r^2}\right) \end{aligned} \tag{22.184}$$

où nous avons posé $r^2 = x^2 + y^2$ pour simplifier les notations.

Soient deux vecteurs $u, v \in \mathbb{R}^2$ et un point $a \in \mathbb{R}^2$. Nous devons prouver que

$$\frac{u \cdot v}{\|u\|\|v\|} = \frac{df_a(u) \cdot df_a(v)}{\|df_a(u)\|\|df_a(v)\|}. \tag{22.185}$$

Pour cela, nous pourrions calculer dF_a et passer en coordonnées polaires[414] mais nous préférons faire les calculs à la dure parce que nous avons Sage avec nous.

Nous notons A la matrice de dF en $a = (x, y)$ et nous avons

$$Au \cdot Av = A^t Au \cdot v \tag{22.186}$$

ainsi que $\|u\| = \sqrt{u \cdot u} = \sqrt{A^t Au \cdot u}$, de telle sorte qu'il devienne urgent de calculer $A^t A$. Voici le calcul :

```
var('y')

# les fonctions coordonnées
F1(x,y)=x/(x**2+y**2)
F2(x,y)=-y/(x**2+y**2)

# La matrice différentielle :
A=matrix( [  [F1.diff(x).simplify_full(),F1.diff(y).simplify_full↩
    ()],[F2.diff(x).simplify_full(),F2.diff(y).simplify_full()]   ↩
    ] )

# Quelqu'un peut expliquer pourquoi ceci ne fonctionne pas ?
# A=matrix( [  [F1.diff(x),F1.diff(y)],[F2.diff(x),F2.diff(y)]   ]↩
     ).simplify_full()

# A^tA
S=A.transpose()*A
S=S.simplify_full()      # Mais ça, ça marche !!
print(S)
```

tex/sage/sageSnip009.sage

Le résultat est que

$$A^t A = \begin{pmatrix} \frac{1}{r^4} & 0 \\ 0 & \frac{1}{r^4} \end{pmatrix} = \frac{1}{r^4}\,\mathrm{Id}\,. \tag{22.187}$$

La vérification de (22.185) est alors immédiate. □

22.6.3 Birapport

Nous introduisons maintenant quelque chose qui s'appelle le « birapport » et qui n'est à priori pas du tout lié au birapport définit en 22.40.

Définition 22.81 (Birapport dans $\hat{\mathbb{C}}$[354])**.**
*Soient $a, b, c, x \in \hat{\mathbb{C}}$ où a, b et c sont distincts. Le **birapport** de ces quatre nombres est l'élément de $\hat{\mathbb{C}}$ donné par, si $a, b, c \neq \infty$:*

$$[a, b, c, x] = \frac{(a-c)(b-x)}{(b-c)(a-x)}, \tag{22.188}$$

et

$$[\infty, b, c, x] = \frac{b-x}{a-x} \tag{22.189a}$$

$$[a, \infty, c, x] = \frac{a-c}{a-x} \tag{22.189b}$$

$$[a, b, \infty, x] = \frac{b-x}{a-x} \tag{22.189c}$$

22.82.
Notons la « logique » des cas particuliers. Pour le premier, si $a \to \infty$ tandis que les autres restent dans $\mathbb{C}$ alors $a - c$ et $a - x$ deviennent du même ordre de grandeur et se simplifient. Il reste les deux autres parties de la fraction.

C'est cette même logique qui, partant de $[a, b, \infty, x] = \frac{b-x}{a-x}$ donne

$$[a, b, \infty, \infty] = 1 \tag{22.190}$$

comme il se doit si nous avons l'intention de ressembler au lemme 22.45.

L'objet « birapport » introduit ici est évidemment lié au birapport sur $P(\mathbb{C}^2)$ définit plus haut. Le lien est la proposition suivante.

Proposition 22.83.
Soit l'application $\varphi_0 \colon P(\mathbb{C}^2) \to \hat{\mathbb{C}}$ définie en 22.36. Si $A, B, C, X \in P(\mathbb{C}^2)$ alors

$$[A, B, C, X]_{\varphi_0} = [\varphi_0(A), \varphi_0(B), \varphi_0(C), \varphi_0(X)]. \tag{22.191}$$

Cela est une égalité dans $\hat{\mathbb{C}}$.

Démonstration. Nous écrivons $A = [a_1, a_2]$, $B = [b_1, b_2]$, $C = [c_1, c_2]$ avec $a_1, a_2, b_2, b_2, c_1, c_2 \in \mathbb{C}$. Par définition $[A, B, C, X]_{\varphi_0} = (\varphi_0 \circ \phi)(X)$ où $\phi \colon P(\mathbb{C}^2) \to P(\mathbb{C}^2)$ est l'unique homographie telle que

$$\phi[a_1, a_2] = [1, 0] \tag{22.192a}$$

$$\phi[b_1, b_2] = [0, 1] \tag{22.192b}$$

$$\phi[c_1, c_2] = [1, 1] \tag{22.192c}$$

Une des difficultés de cette preuve va être de calculer ce ϕ. D'abord nous pouvons introduire $\tilde{\phi} = \varphi_0 \circ \phi \circ \varphi^{-1}$ qui est obligatoirement (proposition 22.69) de la forme

$$\tilde{\phi}(z) = \frac{\alpha z + \beta}{\gamma a + \delta}. \tag{22.193}$$

Nous allons imposer les relations (22.192) pour déterminer les coefficients α, β, γ et δ.

D'abord

$$(\varphi_0^{-1} \circ \tilde{\phi} \circ \varphi_0)\big([a_1, a_2]\big) = (\varphi_0^{-1} \circ \tilde{\phi})(a_1/a_2) \tag{22.194a}$$

$$= \varphi_0^{-1}\left(\frac{\alpha \frac{a_1}{a_2} + \beta}{\gamma \frac{a_1}{a_2} + \delta}\right) \tag{22.194b}$$

$$= \big[\alpha \frac{a_2}{a_2} + \beta, \gamma \frac{a_1}{a_2} + \delta\big]. \tag{22.194c}$$

Égaler cela à $[1,0]$ donne

$$\begin{cases} \alpha\dfrac{a_1}{a_2}+\beta\neq 0 & (22.195a)\\ \gamma\dfrac{a_1}{a_2}+\delta=0. & (22.195b)\end{cases}$$

Donc nous avons déjà

$$\tilde{\phi}(z)=\frac{\alpha z+\beta}{\gamma(z-\frac{a_1}{a_2})}=\frac{\alpha z+\beta}{\gamma\big(z-\varphi_0(A)\big)}. \tag{22.196}$$

En y imposant la contrainte $(\varphi^{-1}\circ\tilde{\phi}\circ\varphi_0)([b_1,b_2])=[0,1]$ nous trouvons les contraintes

$$\begin{cases} \gamma\big(\varphi_0(B)-\varphi_0(A)\big)\neq 0 & (22.197a)\\ \beta=-\alpha\varphi_0(B). & (22.197b)\end{cases}$$

Nous avons décidé d'écrire $\varphi_0(A)$ au lieu de a_1/a_2 à la fois pour un soucis de simplification d'écriture et dans le but de ressembler à (22.191). En substituant :

$$\tilde{\phi}(z)=\frac{\alpha\big(z-\varphi_0(B)\big)}{\gamma\big(z-\varphi_0(A)\big)}. \tag{22.198}$$

La condition pour $[c_1,c_2]$ donne

$$\big[\alpha\big(\varphi_0(C)-\varphi_0(B)\big),\gamma\big(\varphi_0(C)-\varphi_0(A)\big)\big]=[1,1], \tag{22.199}$$

ce qui donne

$$\alpha\big(\varphi_0(C)-\varphi_0(B)\big)=\gamma\big(\varphi_0(C)-\varphi_0(A)\big). \tag{22.200}$$

Nous avons alors

$$\gamma=\alpha\frac{\varphi_0(C)-\varphi_0(B)}{\varphi_0(C)-\varphi_0(A)}, \tag{22.201}$$

et les α se simplifient dans la formule pour $\tilde{\phi}$:

$$\tilde{\phi}(z)=\frac{\big(z-\varphi_0(B)\big)\big(\varphi_0(C)-\varphi_0(A)\big)}{\big(z-\varphi_0(A)\big)\big(\varphi_0(C)-\varphi_0(B)\big)}. \tag{22.202}$$

Par la proposition 22.69, l'application $\varphi_0^{-1}\circ\tilde{\phi}\circ\varphi_0$ est une homographie. Nous pouvons donc calmement calculer le birapport $[A,B,C,X]_{\varphi_0}$ de la façon suivante :

$$\begin{aligned}
[A,B,C,X]_{\varphi_0}&=(\varphi_0\circ\phi)(X) && (22.203a)\\
&=(\varphi_0\circ\varphi_0^{-1}\circ\tilde{\phi}\circ\varphi_0)(X) && (22.203b)\\
&=\tilde{\phi}\big(\varphi_0(X)\big) && (22.203c)\\
&=\frac{\big(\varphi_0(X)-\varphi_0(B)\big)\big(\varphi_0(C)-\varphi_0(A)\big)}{\big(\varphi_0(X)-\varphi_0(A)\big)\big(\varphi_0(C)-\varphi_0(B)\big)} && (22.203d)\\
&=[\varphi_0(A),\varphi_0(B),\varphi_0(C),\varphi_X(A)]. && (22.203e)
\end{aligned}$$

□

Proposition 22.84.
Les homographies de $\hat{\mathbb{C}}$ conservent le birapport.

Démonstration. Ici le mot « homographie » réfère à la définition 22.71 et le birapport à 22.81. Soient $a,b,c,x\in\hat{\mathbb{C}}$ et une homographie $\tilde{\phi}\colon\hat{\mathbb{C}}\to\hat{\mathbb{C}}$. Il existe une homographie $\phi\colon P(\mathbb{C}^2)\to P(\mathbb{C}^2)$ telle que $\tilde{\phi}=\varphi_0\circ\phi\varphi_0^{-1}$. Alors

$$\begin{aligned}
\big[\tilde{\phi}(a),\tilde{\phi}(b),\tilde{\phi}(c),\tilde{\phi}(x)\big]&=\big[(\phi\circ\varphi_0^{-1})(a),(\phi\circ\varphi_0^{-1})(b),(\phi\circ\varphi_0^{-1})(c),(\phi\circ\varphi_0^{-1})(x),\big]_{\varphi_0} && (22.204a)\\
&=\big[\varphi_0^{-1}(a),\varphi_0^{-1}(b),\varphi_0^{-1}(c),\varphi_0^{-1}(x),\big]_{\varphi_0} && (22.204b)\\
&=[a,b,c,x]. && (22.204c)
\end{aligned}$$

Justifications :

— Identification des birapports sur $P(\mathbb{C}^2)$ et sur $\hat{\mathbb{C}}$, proposition 22.83.
— Invariance du birapport sour les homographies (dans $P(\mathbb{C}^2)$), proposition 22.47.

□

Proposition 22.85.
Soient des points a, b, c, x dans $\hat{\mathbb{C}}$ avec a, b, c distincts. Ils sont alignés ou cocycliques si et seulement si $[a, b, c, x] \in \hat{\mathbb{R}}$.

Démonstration. Nous allons faire plusieurs cas. Mais dans tous les cas vous pouvez relire la définition des angles orientés 17.142 et la partie sur les angles dans les nombres complexes 17.9.12.

Tous les points sont distincts et dans $\mathbb{C}$ D'une part, nous savons que le nombre complexe $re^{i\theta}$ est réel si et seulement si $\theta \in [0]_\pi$, et d'autre part l'argument du birapport (22.188) est

$$[\overrightarrow{ca}, \overrightarrow{cb}] + [\overrightarrow{xb}, \overrightarrow{xa}]. \tag{22.205}$$

Le birapport est réel si et seulement si

$$[\overrightarrow{ca}, \overrightarrow{cb}] - [\overrightarrow{xa}, \overrightarrow{xb}] \in [0]_\pi. \tag{22.206}$$

À gauche nous avons une classe modulo 2π et à droite une classe modulo π. L'égalité signifie qu'il y a un représentant du membre de gauche qui appartient au membre de droite. Si vous aimez faire très attention à ce que signifient les notations, voici trois manières d'écrire la condition, par ordre croissant de précision :

$$[\overrightarrow{ca}, \overrightarrow{cb}] - [\overrightarrow{xa}, \overrightarrow{xb}] = [0]_\pi, \tag{22.207a}$$
$$[\overrightarrow{ca}, \overrightarrow{cb}] - [\overrightarrow{xa}, \overrightarrow{xb}] \in [0]_\pi, \tag{22.207b}$$
$$[\overrightarrow{ca}, \overrightarrow{cb}] - [\overrightarrow{xa}, \overrightarrow{xb}] \subset [0]_\pi. \tag{22.207c}$$

Nous avons donc que le birapport est réelle si et seulement si la condition (22.207c) est vérifiée. D'après le théorème 17.150, cette dernière condition est équivalente à dire que les points a, b, c et x sont alignés.

Pas quatre points distincts, dans $\mathbb{C}$ Nous supposons encore que a, b, c et x sont dans $\mathbb{C}$. Mais nous supposons que x est un de a, b ou c. Vu que par hypothèse a, b et c sont distincts, c'est le seul cas à considérer dans la catégorie des 4 points non distincts.

Trois points sont toujours alignés ou cocycliques[11]. Donc nous devons seulement montrer que dans ce cas le birapport est toujours dans $\hat{\mathbb{R}}$. Par définition,

— Si $x = a$ alors $[a, b, c, x] = \infty$,
— Si $x = b$ alors $[a, b, c, x] = 0$,
— Si $a = c$ alors $[a, b, c, x] = 1$.

Dans tous les cas de figure le birapport est dans $\hat{\mathbb{R}}$.

À ce niveau de la preuve nous devons encore vérifier les cas où a, b, c ou x valent ∞. Si l'un de a, b ou c est ∞ et si x l'est aussi, alors, comme ∞ est aligné avec tout, nous avons seulement une droite passant par deux points. Il nous faut donc seulement regarder les cas où un seul des 4 points est ∞.

Si $a = \infty$ Le birapport est alors

$$[\infty, b, c, x] = \frac{b - x}{a - x}, \tag{22.208}$$

qui est un nombre à priori complexe dont le dénominateur est supposé non nul parce que le cas $a = x$ est déjà traité. L'argument de ce nombre est dans la classe de l'angle orienté

$$\arg\left(\frac{b - x}{a - x}\right) \in [\overrightarrow{xa}, \overrightarrow{xb}]. \tag{22.209}$$

11. Si ils ne sont pas alignés, prendre la médiatrice du segment $[a, b]$ et celle de $[b, c]$, et l'intersection vous donnera le centre d'un cercle passant par a, b et c.

Le birapport est réel si et seulement si le membre de gauche est dans $[0]_\pi$. Et cela est justement le cas où le membre de droite donne des points alignés.

Les cas $b = \infty$, $c = \infty$ et $x = \infty$ se traitent de la même manière. □

22.6.4 Division harmonique

Définition 22.86.
Nous disons que les éléments a, b, c et x de $\hat{\mathbb{C}}$ sont en ***division harmonique*** *lorsque $[a, b, c, x] = -1$.*

22.87.
Une chose qui sera utile par la suite est de remarquer que $[a, b, c, \infty] = -1$ lorsque $c = \frac{a+b}{2}$.

Nous allons maintenant voir comment, pour $a, b, c \in \mathbb{C}$ donnés nous pouvons construire x tels que a, b, c, x soient en division harmonique. Vu que trois points sont soit cocycliques soit alignés nous divisons la construction en deux parties.

Notons que si nous trouvons une construction qui donne une point x vérifiant $[a, b, c, x] = -1$ alors nous prouvons au passage que la construction ne dépend pas des choix intermédiaires parce que il n'existe qu'un unique x tel que $[a, b, c, x] = -1$ lorsque a, b, c sont donnés.

Lemme 22.88 ([418, 419])**.**
Soient a, b, c cocycliques dans $\mathbb{C}$. Nous nommons $\mathcal{C}$ le cercle contenant a, b et c ainsi que T_a et T_b les tangentes à $\mathcal{C}$ en a et b. Soit $m = T_a \cap T_b$ et la droite $L = (mc)$. Alors le point

$$x = (mc) \cap \mathcal{C} \tag{22.210}$$

vérifie $[a, b, c, x] = -1$.

Si $m = \infty$ (arrive lorsque $T_a \parallel T_b$) alors en guise de L nous prenons la parallèle à T_a passant par c.

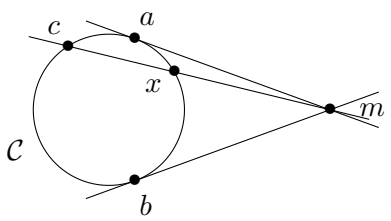
Démonstration. Nous séparons les cas suivant que $m = \infty$ ou non.

<u>$m = \infty$</u> Les tangentes à $\mathcal{C}$ en a et en b sont parallèles, c'est-à-dire que ces points sont diamétralement opposés sur $\mathcal{C}$. Les homographies préservent le birapport (proposition 22.84), et les rotations, dilatations et translations sont des homographies (voir 22.74).

Nous pouvons donc nous ramener au cas où $\mathcal{C}$ est centré en 0 et de rayon 1 avec $a = i$ et $b = -i$. Dans ce cas, $c = e^{i\theta}$. Vu que x est donné par l'intersection entre le cercle et la droite horizontale passant par c nous avons $x = e^{i(\pi-\theta)}$. Le birapport se calcule explicitement :

$$[a, b, c, x] = \frac{(i - e^{i\theta})(-i - e^{i(\pi-\theta)})}{(-i - e^{i\theta})(i - e^{i(\pi-\theta)})} = -1. \tag{22.211}$$

<u>$m \neq \infty$</u> Nous sommes dans la situation suivante où à une translation près nous supposons $x = 0$:



Nous allons prouver que dans ce cas, $[a, b, c, x] = -1$. Pour cela nous considérons l'inversion de centre x (qui est $x = 0$ par translation). Soit $\tilde{\phi}$ cette homographie. Elle conserve le birapport, il nous allons voir que calculer $[\tilde{\phi}(a), \tilde{\phi}(b), \tilde{\phi}(c), \tilde{\phi}(x)]$ se révèle être plus facile [12].

12. Si vous n'avez peur d'aucun calculs, il suffit de poser $a = e^{i\theta}$, $b = e^{-i\theta}$ et $c = e^{i\sigma}$ et vous êtes théoriquement capable de calculer les coordonnées de tous les points, y compris de x en termes de θ et σ. Ensuite le calcul du birapport est explicite.

Nous nommons $A = (am)$, $B = (bm)$, $C = (cm)$ et $\mathcal{C}$, le cercle. Nous allons maintenant faire intensément usage de la proposition 22.76. Nous avons :

— $\tilde{\phi}(A)$ est un cercle passant par 0.
— $\tilde{\phi}(B)$ est un cercle passant par 0.
— $\tilde{\phi}(C)$ est la droite C.
— $\tilde{\phi}(\mathcal{C})$ est une droite ne passant pas par 0.

Les droites A et B se coupent en m et en ∞ (qui sont des points distincts). Donc les cercles $\tilde{\phi}(A)$ et $\tilde{\phi}(B)$ se coupent en 0 et $\tilde{\phi}(m)$, aucun de ces deux points n'est sur la droite $\tilde{\phi}(\mathcal{C})$.

Par tangence, la droite A et le cercle $\mathcal{C}$ se coupent en un seul point (a). Donc $\tilde{\phi}(A)$ coupe $\tilde{\phi}(\mathcal{C})$ en un seul point, $\tilde{\phi}(a)$. Idem pour le cercle $\tilde{\phi}(B)$.

Nous avons donc que les cercles $\tilde{\phi}(A)$ et $\tilde{\phi}(B)$ sont tangents à la droite $\tilde{\phi}(\mathcal{C})$ et se coupent en exactement deux points distincts (qui sont donc du même côté de la droite).

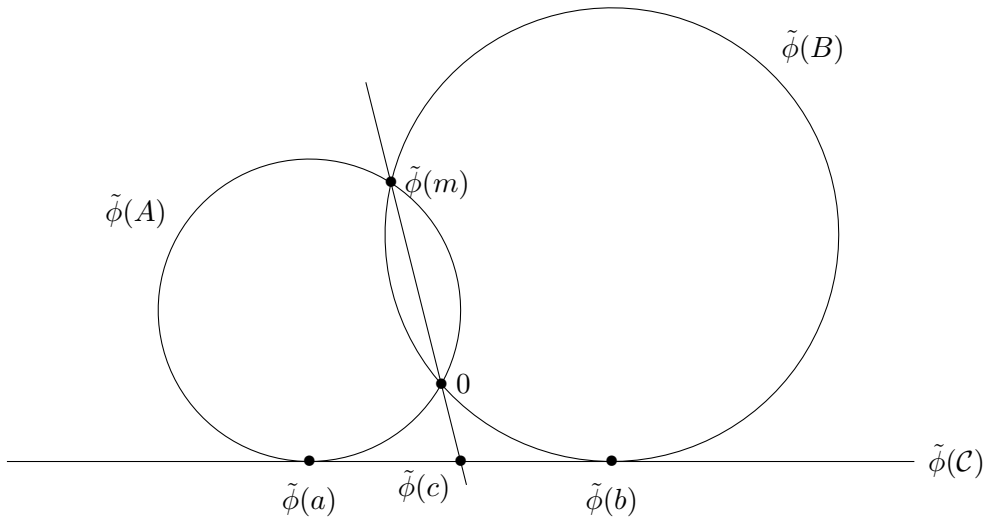
Nous nous intéressons à la droite $\tilde{\phi}(C)$. C'est une droite parce que c'est l'image d'une droite passant par 0. Elle passe par 0, par $\tilde{\phi}(c)$ et $\tilde{\phi}(m)$. Le fait qu'elle passe par 0 et $\tilde{\phi}(m)$ fait que c'est la droite passant par les deux intersections des cercles. Vu que $c \in \mathcal{C} \cap C$, le point d'intersection $\tilde{\phi}(C) \cap \tilde{\phi}(\mathcal{C})$ est $\tilde{\phi}(c)$.

Quelle est la puissance du point $\tilde{\phi}(c)$ par rapport au cercle $\tilde{\phi}(A)$? En la calculant avec la droite $\tilde{\phi}(C)$, qui intersecte les deux cercles aux points déjà étudiés, la puissance est :

$$k = d\big(\tilde{\phi}(c), 0\big) \times d\big(\tilde{\phi}(c), \tilde{\phi}(m)\big). \tag{22.212}$$

Vu que ces points sont également sur le cercle $\tilde{\phi}(B)$, la puissance de $\tilde{\phi}(c)$ par rapport à ce second point est la même.

Tout cela justifie le dessin suivant [13] :



Mais la droite passant par $\tilde{\phi}(a)$ et $\tilde{\phi}(c)$ (qui est tangente au cercle) permet également de calculer cette puissance :

$$k = d\big(\tilde{\phi}(a), \tilde{\phi}(c)\big). \tag{22.213}$$

Idem pour la puissance par rapport à l'autre cercle :

$$k = d\big(\tilde{\phi}(a), \tilde{\phi}(b)\big). \tag{22.214}$$

Nous en déduisons que $\tilde{\phi}(c)$ est le milieu entre $\tilde{\phi}(a)$ et $\tilde{\phi}(b)$.

Du coup

$$\big[\tilde{\phi}(a), \tilde{\phi}(b), \tilde{\phi}(c), \tilde{\phi}(x)\big] = \big[\tilde{\phi}(a), \tilde{\phi}(b), \tilde{\phi}(c), \infty\big] = -1 \tag{22.215}$$

en vertu de ce que nous avons raconté en 22.87.

13. Bien que ce ne soit pas strictement nécessaire à la preuve, est-ce que vous savez si les deux cercles ont le même rayon ? Et si par hasard la droite $\big(\tilde{\phi}(m)\tilde{\phi}(c)\big)$ n'arrive pas perpendiculairement à $\tilde{\phi}(\mathcal{C})$?

□

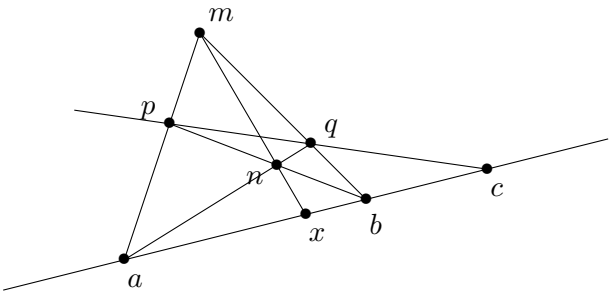
Lemme 22.89 ([420]).
Soit $a, b, c \in \mathbb{C}$ colinéaires. Soit m un point hors de cette droite. Nous considérons une droite issue de c coupant $[ma]$ en p et $[mb]$ en q.

Nous construisons $n = (aq) \cap (pb)$ et finalement $x = (mn) \cap (ab)$.

À la fin nous avons

$$[a, b, c, x] = -1. \tag{22.216}$$

Démonstration. Commençons par un dessin de la situation :



Les points a, b et m ne sont pas alignés, et nous pouvons les utiliser comme repère barycentrique (voir 8.48 pour savoir en deux mots ce que c'est). Nous nommons (α, β, γ) les coordonnées de n dans ce système, c'est-à-dire que

$$\alpha\overrightarrow{na} + \beta\overrightarrow{nb} + \gamma\overrightarrow{nm} = 0. \tag{22.217}$$

Dans notre contexte, nous pouvons voir le vecteur $\overrightarrow{st}$ comme une façon d'écrire le nombre $t - s$. Par la proposition 8.53 nous savons les coordonnées barycentriques de p, x et q en regardant le triangle acb. Voici les coordonnées et les relations qu'elles signifient :

$$n = (\alpha, \beta, \gamma), \qquad \alpha\overrightarrow{na} + \beta\overrightarrow{nb} + \gamma\overrightarrow{nm} = 0 \tag{22.218a}$$
$$p = (\alpha, 0, \gamma), \qquad \alpha\overrightarrow{pa} + \gamma\overrightarrow{pm} = 0 \tag{22.218b}$$
$$q = (0, \beta, \gamma), \qquad \beta\overrightarrow{qb} + \gamma\overrightarrow{qm} = 0 \tag{22.218c}$$
$$x = (\alpha, \beta, 0). \qquad \alpha\overrightarrow{xa} + \beta\overrightarrow{xb} = 0. \tag{22.218d}$$

Nous voudrions maintenant voir les coordonnées de c. Nous posons $c = (\lambda, \mu, \sigma)$:

$$\lambda\overrightarrow{ca} + \mu\overrightarrow{cb} + \sigma\overrightarrow{cm} = 0. \tag{22.219}$$

Mais a, b et c sont alignés, donc $\overrightarrow{ca}$ et $\overrightarrow{cb}$ sont colinéaires, alors que $\overrightarrow{cm}$ n'est pas aligné avec les deux autres. L'annulation (22.219) demande donc l'annulation séparément

$$\begin{cases} \sigma\overrightarrow{cm} = 0 & (22.220a) \\ \lambda\overrightarrow{ca} + \mu\overrightarrow{cb} = 0. & (22.220b) \end{cases}$$

Nous en déduisons que $\sigma = 0$ et aussi que λ et μ ne sont pas nuls. Nous posons arbitrairement $\lambda = 1$ parce que les coordonnées barycentriques sont définies à coefficient multiplicatif près.

Nous imposons à présent le fait que p, q et c sont alignés. Pour cela nous devons faire apparaitre les vecteurs $\overrightarrow{pq}$, $\overrightarrow{pc}$, $\overrightarrow{pc}$. Vu le dessin et les relations disponibles (22.218) le mieux est d'utiliser les relations de Chasles (proposition 8.4) pour faire $\overrightarrow{ca} = \overrightarrow{cp} + \overrightarrow{pq}$ et $\overrightarrow{cb} = \overrightarrow{cq} + \overrightarrow{qb}$. La relation (22.220b) devient :

$$\overrightarrow{cp} + \overrightarrow{pa} + \mu(\overrightarrow{cq} + \overrightarrow{qb}) = 0. \tag{22.221}$$

Les vecteurs $\overrightarrow{cp}$ et $\overrightarrow{cq}$ sont alignés, donc nous les écrivons ensemble. Les vecteurs $\overrightarrow{pa}$ et $\overrightarrow{qb}$ se transforment en utilisant les relations (22.218) :

$$\overrightarrow{cp} + \mu\overrightarrow{cq} - \frac{\gamma}{\alpha}\overrightarrow{pm} - \mu\frac{\gamma}{\beta}\overrightarrow{qm}. \tag{22.222}$$

Enfin nous voulons faire la somme du terme $\overrightarrow{pm}$ avec le terme $\overrightarrow{qm}$. D'abord on change le signe :

$$\overrightarrow{cp} + \mu\overrightarrow{cq} - \frac{\gamma}{\alpha}\overrightarrow{pm} + \frac{\mu\gamma}{\beta}\overrightarrow{mq} \tag{22.223}$$

ensuite nous écrivons

$$\frac{\mu\gamma}{\beta} = \frac{\mu\gamma}{\beta} + \frac{\gamma}{\alpha} - \frac{\gamma}{\alpha}, \tag{22.224}$$

et

$$\overrightarrow{cp} + \mu\overrightarrow{cq} - \frac{\gamma}{\alpha}(\overrightarrow{pm} + \overrightarrow{mq}) + \left(\frac{\mu\gamma}{\beta} + \frac{\gamma}{\alpha}\right)\overrightarrow{mq} = 0. \tag{22.225}$$

Tout cela pour

$$\overrightarrow{cp} + \mu\overrightarrow{cq} - \frac{\gamma}{\alpha}\overrightarrow{pq} + \left(\frac{\mu\gamma}{\beta} + \frac{\gamma}{\alpha}\right)\overrightarrow{mq} = 0. \tag{22.226}$$

Le dernier terme n'est pas colinéaire aux deux premiers et s'annule donc séparément :

$$\frac{\mu\gamma}{\beta} + \frac{\gamma}{\alpha} = 0. \tag{22.227}$$

Cela donne $\mu = -\beta/\alpha$.

Au finale nous avons

$$\overrightarrow{ca} - \frac{\beta}{\alpha}\overrightarrow{cb} = 0 \tag{22.228}$$

et donc

$$\alpha\overrightarrow{ca} - \beta\overrightarrow{cb} = 0, \tag{22.229}$$

ce qui donne les coordonnées $(\alpha, -\beta, 0)$ pour le point c.

Vu que nous somme dans l'espace vectoriel $\mathbb{C}$, ce que nous notons $\overrightarrow{AB}$ n'est rien d'autre que la différence $B - A$ dans $\mathbb{C}$ [14]. La relation (22.229) signifie donc

$$\alpha(a - c) = \beta(b - c). \tag{22.230}$$

Nous avons alors :

$$\frac{a - c}{b - c} = \frac{\beta}{\alpha}. \tag{22.231}$$

Par ailleurs, la relation (22.218d) à propos des coordonnées de x donne

$$\frac{a - x}{b - x} = -\frac{\beta}{\alpha}. \tag{22.232}$$

En égalisant les deux valeurs de β/α nous trouvons :

$$\frac{a - c}{b - c} = \frac{x - a}{b - x}, \tag{22.233}$$

ce qui donne (via un petit jeu de signes)

$$\frac{(c - a)(x - b)}{(c - b)(x - a)} = -1. \tag{22.234}$$

C'est cela que nous voulions. □

14. Les mauvaise langues diront que tout le chapitre sur les espaces affines, et surtout la partie sur les barycentres ne sont rien d'autres que le snobisme d'écrire $\overrightarrow{xy}$ au lieu de $y - x$. C'est aussi une facilité d'écriture.

22.6.5 Groupe circulaire

Nous avons vu que les homographies présent l'ensemble des cercles et droites. Nous pouvons nous demander quel est le groupe maximum préservant l'ensemble des cercles et droites.

Définition 22.90.
*Le **groupe circulaire** de $\mathbb{C}$ est le groupe de transformations de $\hat{\mathbb{C}}$ engendré par les homographies*[15] *et la conjugaison complexe. Le groupe circulaire de l'espace projectif est l'ensemble des applications de la forme $\varphi_0^{-1} \circ f \circ \varphi_0$ où f est un élément du groupe circulaire de eC.*

Plusieurs remarques à propos de cette définition.

(1) Vu le lemme 22.75, la conjugaison complexe n'est pas une homographie. Donc cette définition n'est pas stupide : le groupe circulaire est strictement plus grand que le groupe des homographies.

(2) Vous vous souvenez de la définition d'un sous-groupe engendré ? C'est la définition 2.6.

Lemme 22.91.
Soit une application $\alpha\colon \mathbb{C} \to \mathbb{C}$ fixant 1 et 0 et préservant les divisions harmoniques (c'est-à-dire tel que son prolongement à $\hat{\mathbb{C}}$ donné par $\alpha(\infty) = \infty$ préserve les divisions harmoniques). Alors α est un automorphisme de corps[16].

Démonstration. Nous savons que si $a, b, c \in \mathbb{C}$ nous avons $c = (a + b)/2$ si et seulement si $[a, b, c, \infty] = -1$. Vu que α préserve les divisions harmoniques nous avons équivalence entre les affirmations suivantes :

$$c = \frac{a+b}{2} \tag{22.235a}$$

$$[a, b, c, \omega] = -1 \tag{22.235b}$$

$$[\alpha(a), \alpha(b), \alpha(c), \alpha(\infty)] = -1 \tag{22.235c}$$

$$\frac{\alpha(a) + \alpha(b)}{2} = \alpha(c). \tag{22.235d}$$

Donc α préserve les milieux : pour tout $a, b \in \mathbb{C}$ nous avons

$$\alpha\left(\frac{a+b}{2}\right) = \frac{\alpha(a) + \alpha(b)}{2}. \tag{22.236}$$

En particulier, cette relation avec $b = 0$ donne (parce que $\alpha(0) = 0$) : $\alpha(a/2) = \alpha(a)/2$. Nous avons au final, en utilisant cela en conjonction avec (22.236) :

$$\frac{\alpha(a) + \alpha(b)}{2} = \alpha\left(\frac{a+b}{2}\right) = \frac{\alpha(a+b)}{2}. \tag{22.237}$$

Cela démontre déjà que

$$\alpha(a + b) = \alpha(a) + \alpha(b). \tag{22.238}$$

En particulier $\alpha(-a) = \alpha(0 - a) = \alpha(0) - \alpha(a) = -\alpha(a)$.

Nous passons maintenant à la démonstration du fait que $\alpha(ab) = \alpha(a)\alpha(b)$. Pour tout a différent de 0 et ± 1 nous avons

$$[a, -a, a^2, 1] = \frac{(a - a^2)(-a - 1)}{(-a - a^2)(a - 1)} = -1. \tag{22.239}$$

Et en prenant $\alpha(a)$ en guise de a nous avons aussi

$$[\alpha(a), -\alpha(a), \alpha(a)^2, \alpha(1)] = -1. \tag{22.240}$$

15. Homographie de $\hat{\mathbb{C}}$: définition 22.71.
16. Définition 1.74.

Vu que α préserve les divisions harmoniques, l'équation (22.239) donne aussi

$$\big[\alpha(a), \alpha(-a), \alpha(a^2), \alpha(1)\big] = -1, \tag{22.241}$$

c'est-à-dire

$$\big[\alpha(a), -\alpha(a), \alpha(a^2), 1\big] = -1. \tag{22.242}$$

Comparant (22.240) avec (22.242) et en tenant compte de l'unicité du birapport [17] nous avons

$$\alpha(a^2) = \alpha(a)^2. \tag{22.243}$$

Avec cela nous pouvons y aller en remarquant que

$$ab = \left(\frac{a+b}{2}\right)^2 - \left(\frac{a-b}{2}\right)^2. \tag{22.244}$$

Nous appliquons α à cette dernière équations en tenant compte de ce que nous savons déjà

$$\alpha(ab) = \left(\frac{\alpha(a)+\alpha(b)}{2}\right)^2 - \left(\frac{\alpha(a)-\alpha(b)}{2}\right)^2 = \alpha(a)\alpha(b). \tag{22.245}$$

□

Théorème 22.92 ([420, 421])**.**
Le groupe circulaire de $\mathbb{C}$ est le groupe des bijections $\hat{\mathbb{C}} \to \hat{\mathbb{C}}$ préservant l'ensemble des cercles-droites.

Démonstration. L'inclusion dans un sens est facile : les homographies conservent l'ensemble des cercles et droites par la proposition 22.77. Et la conjugaison complexe aussi.

Soit une bijection $f\colon \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ préservant les cercles-droites. Nous supposons dans un premier temps que $f(0) = 0$, $f(1) = 1$ et $f(\infty) = \infty$.

Pour f vérifiant $f(0,1,\infty) = 0,1,\infty$ Si $\mathcal{C}$ est un cercle alors $f(\mathcal{C})$ est un cercle ou une droite, mais vu que $\mathcal{C}$ ne contient pas ∞, l'ensemble $f(\mathcal{C})$ ne le contient pas non plus. Donc f transforme un cercle en un cercle et une droite en une droite.

f préserve les divisions harmoniques Soient a, b, c, x dans $\hat{\mathbb{C}}$ tels que $[a,b,c,x] = -1$. Nous allons prouver que $[f(a), f(b), f(c), f(x)] = -1$.

Si a, b et c sont colinéaires, nous suivons la construction du lemme 22.89. Soit m hors de la droite (ab) et une droite D passant par c et coupant $[ma]$ en p et $[mb]$ en q. Nous posons $n = (pb) \cap (qa)$. Alors $x = (mn) \cap (ac)$.

Vues les propriétés de f (en particulier c'est une bijection qui respecte les intersectons, tangences, cercles et droites). Le point $f(m)$ est hors de la droite $\big(f(a)f(b)\big)$. La droite $f(D)$ passe par $f(c)$ et coupe les segments $[f(m)f(a)]$ en $f(p)$ et $[f(m)f(b)]$ en $f(q)$. Alors nous avons

$$f(n) = \big(p(p)f(b)\big) \cap \big(f(q)f(a)\big) \tag{22.246}$$

et aussi

$$f(x) = \big(p(m)f(n)\big) \cap \big(f(a)f(c)\big) \tag{22.247}$$

Donc $f(x)$ se construit à partir de $f(a)$, $f(b)$ et $f(c)$ en suivant la même construction que x à partir de a, b et c. Nous en concluons que $[f(a), f(b), f(c), f(x)] = -1$.

Si a, b et c sont cocycliques, le même raisonnement, en suivant le lemme 22.88 nous donne le même résultat.

f est un automorphisme du corps $\mathbb{C}$ C'est le lemme 22.91.

17. C'est-à-dire que si trois éléments du birapport sont donnés, le quatrième est fixé. C'est une variation sur la thème de la proposition 22.46(2).

Et enfin ... Notre application f est un automorphisme du corps $\mathbb{C}$ qui fixe $\mathbb{R}$ parce qu'elle laisse invariante les droites dans $\mathbb{C}$. Donc la proposition 6.6 nous dit que f est soit l'identité soit la conjugaison complexe. Dans les deux cas, f est dans le groupe circulaire.

Pour f plus générale Nous ne supposons plus que f fixe 0, 1 et ∞. En tout cas les nombres $f^{-1}(1)$, $f^{-1}(0)$ et $f^{-1}(\infty)$ sont distincts parce que f est une bijection. Nous pouvons considérer une homographie[18] $\phi\colon \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ telle que $\phi(1) = f^{-1}(1)$, $\phi(0) = f^{-1}(0)$ et $\phi(\infty) = f^{-1}(\infty)$. Dans ce cas l'application

$$g = f \circ \phi \tag{22.248}$$

vérifie $g(1) = 1$, $g(0) = 1$ et $g(\infty) = \infty$ tout en continuant à transformer un cercle-droite en un cercle-droite. Donc $f \circ \phi$ est soit l'identité soit la conjugaison complexe. Avec ça, l'application

$$f = g \circ \phi^{-1} \tag{22.249}$$

est la composée d'une homographie avec soit l'identité soit la conjugaison complexe. Elle est donc dans le groupe circulaire.

□

22.6.6 Action du groupe modulaire

Le **demi-plan de Poincaré** est l'ensemble

$$P = \{z \in \mathbb{C} \text{ tel que } \Im(z) > 0\}. \tag{22.250}$$

Le **groupe modulaire** est le quotient de groupes

$$\mathrm{PSL}(2, \mathbb{Z}) = \frac{\mathrm{SL}(2, \mathbb{Z})}{\mathbb{Z}_2}. \tag{22.251}$$

Ce sont donc les matrices au signe près de la forme

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{22.252}$$

où a, b, c et d sont entiers tels que $ad - cb = 1$.

Théorème 22.93 ([422]).
Le groupe modulaire agit fidèlement (définition 2.49) sur le demi-plan de Poincaré par

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} * z = \frac{az + b}{cz + d}. \tag{22.253}$$

L'ensemble $D = D_1 \cup D_2$ avec

$$D_1 = \{z \in P \text{ tel que } |z| > 1, -\frac{1}{2} \leqslant \Re(z) < \frac{1}{2}\} \tag{22.254a}$$

$$D_2 = \{z \in P \text{ tel que } |z| = 1, -\frac{1}{2} \leqslant \Re(z) \leqslant 0\} \tag{22.254b}$$

est un domaine fondamental (définition 2.56) de cette action.

De plus si nous notons

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{22.255}$$

alors pour tout $z \in P$, il existe $A \in \mathrm{gr}(S, T)$ telle que $A * z \in D$.

18. Attention : ici nous parlons d'homographies de $\hat{\mathbb{C}}$, pas de $P(\mathbb{C}^2)$. L'existence d'une telle application demande de composer le corolaire 22.35 avec l'application φ_0 et la définition 22.71 et 22.72.

Démonstration. Nous divisions la preuve en plusieurs étapes.

Bien définie D'abord il faut remarquer que l'action (22.253) est bien définie par rapport au quotient : $A * z = (-A) * z$. La vérification est immédiate.

Interne Montrons que si $A \in \mathrm{PSL}(2, \mathbb{Z})$ et $z \in P$ alors $A * z \in P$. Nous avons

$$A * z = \frac{az + b}{cz + d} = \frac{(az + b)(c\bar{z} + d)}{|cz + d|^2} = \frac{a|z|c + azd + bc\bar{z} + bd}{|cz + d|^2}, \tag{22.256}$$

et donc en décomposant $z = \Re(z) + i\Im(z)$,

$$\Im(A * z) = \Im\left(\frac{azd + bc\bar{z}}{|cz + d|^2}\right) = \frac{ad - bc}{|cz + d|^2}\Im(z) = \frac{\Im(z)}{|cz + d|^2} \tag{22.257}$$

où nous avons tenu compte de $ad - bc = 1$. Donc l'action respecte la (stricte) positivité de la partie imaginaire.

Action Nous vérifions maintenant que la formule donne bien une action : $A * (B * z) = (AB) * z$. Cela est un bon calcul :

$$\begin{aligned}
A * (B * z) &= A * \left(\frac{a'z + b}{c'z + d'}\right) && (22.258a)\\
&= \frac{a\left(\frac{a'z+b}{c'z+d'}\right) + b}{c\left(\frac{a'z+b}{c'z+d'}\right) + d} && (22.258b)\\
&= \frac{a(a'z + b') + b(c'z + d')}{c(a'z + b') + d(c'z + d')} && (22.258c)\\
&= \frac{(aa' + bc')z + (ab' + bd')}{(ca' + dc')z + (cb' + dd')} && (22.258d)\\
&= \begin{pmatrix} aa' + bc' & ab' + bd' \\ a'c + dc' & cb' + dd' \end{pmatrix} * z && (22.258e)\\
&= (AB) * z. && (22.258f)
\end{aligned}$$

Fidèle Soit $A \in \mathrm{PSL}(2, \mathbb{Z})$ tel que pour tout $z \in P$ nous ayons

$$\frac{az + b}{cz + d} = z. \tag{22.259}$$

Alors nous avons

$$cz^2 + (d - a)z + b = 0. \tag{22.260}$$

Cela est donc un polynôme en z qui s'annule sur un ouvert [19] (le demi-plan de Poincaré). Il doit donc être identiquement nul, donc $c = b = a - d = 0$. Si vous n'y croyez pas, écrivez pour $z = \epsilon i$ (avec $\epsilon > 0$) :

$$-c\epsilon^2 + \epsilon(d - a)i + b = 0 \tag{22.261}$$

pour tout ϵ. Le fait d'avoir $c\epsilon^2 = b$ pour tout ϵ implique que $c = b = 0$. Donc A est de la forme

$$A = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}, \tag{22.262}$$

avec la contrainte supplémentaire que $ad = 1$, les nombres a et b étant entiers. Nous avons donc soit $a = d = 1$ soit $a = d = -1$. Étant donné le quotient par $\mathbb{Z}_2$, ces deux possibilités donnent le même élément de $\mathrm{PSL}(2, \mathbb{Z})$.

19. On ne peut pas dire que $b = 0$ simplement en justifiant qu'on l'obtient en posant $z = 0$ parce que $z = 0$ n'est pas dans le demi-plan de Poincaré.

Les orbites intersectent D Soit $z \in P$. Nous devons trouver $A \in \mathrm{PSL}(2,\mathbb{Z})$ tel que $A * z \in D$. Nous savons déjà que

$$\Im(A * z) = \frac{\Im(z)}{|cz+d|^2}. \tag{22.263}$$

Nous notons $\mathcal{O}_z$ l'orbite de z sous le groupe modulaire et nous posons

$$I_z = \{\Im(u) \text{ tel que } u \in \mathcal{O}_z\} = \{\Im(A * z) \text{ tel que } A \in \mathrm{PSL}(2,\mathbb{Z})\}, \tag{22.264}$$

l'ensemble des parties imaginaires des éléments de l'orbite de z. Nous allons montrer que cet ensemble est borné vers le haut en montrant que la quantité $|cz+d|$ ne peut,) z donné, prendre qu'un nombre fini de valeurs plus grandes que $\Im(z)$ [20]. Nous cherchons donc les couples $(c,d) \in \mathbb{Z}^2$ tels que $|cz+d| < 1$.

Nous avons $\Im(cz+d) = c\Im(z)$, donc $|cz+d| \geqslant |c\Im(z)|$, mais il n'y a qu'un nombre fini de $c \in \mathbb{Z}$ tels que $|c\Im(z)| < 1$. De la même façon, pour la partie réelle nous avons

$$\Re(cz+d) = c\Re(z) + d, \tag{22.265}$$

et pour chaque c, il n'y a qu'un nombre fini de $d \in \mathbb{Z}$ qui laissent cette quantité plus petite que 1 (en valeur absolue).

Donc I_z possède un maximum. Soit $A_1 \in \mathrm{PSL}(2,\mathbb{Z})$ tel que $\Im(A_1 * z) = \max I_z$. Nous notons $z_1 = A_1 * z$, et que nous n'avons à priori pas l'unicité. Nous allons maintenant agir sur z_1 avec l'élément

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{22.266}$$

pour ramener z_1 dans le domaine D. Si $u \in P$ nous avons $T * u = u + 1$ et donc

$$T^n * u = u + n. \tag{22.267}$$

Vu que D est de largeur 1, il existe un n (éventuellement négatif) tel que

$$\Re(T^n * z_1) \in [-\frac{1}{2}, \frac{1}{2}[. \tag{22.268}$$

Notons qu'ici le fait d'être ouvert d'un côté et fermé de l'autre joue de façon essentielle (pour l'unicité aussi). Nous notons $z_2 = T^n * z_1$.

Supposons un instant que $|z_2| < 1$. Nous considérons l'élément

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{22.269}$$

qui fait

$$\Im(S * z) = \frac{\Im z}{|z|^2}. \tag{22.270}$$

Donc si $|z_2| < 1$ alors $\Im(S * z_2) > \Im(z_2)$, ce qui contredit la maximalité de $\Im(z_2)$ dans I_z. Nous en déduisons que $|z_2| \geqslant 1$. Nous en déduisons que $|z_2| \geqslant 1$.

Si $|z_2| > 1$, alors $z_2 \in D_1$ et c'est bon. Si $|z_2| = 1$, alors il faut encore un peu travailler. Si $z_2 \pm 1$ est à l'intérieur du disque, alors en agissant avec T ou T^{-1} nous retrouvons la même contradiction que précédemment. En écrivant $z_2 = e^{i\theta}$, nous devons donc avoir $2\cos(\theta) \leqslant 1$ ou encore $|\Re(z_2)| \leqslant \frac{1}{2}$. Donc si $\Re(z_2) \leqslant 0$ alors $z_2 \in D_2$.

Le dernier cas à traiter est $\Re(z_2) \in]0, \frac{1}{2}]$, c'est-à-dire $\theta \in [\frac{\pi}{3}, \frac{\pi}{2}[$. Dans ce cas l'action avec S ramène l'angle dans la bonne zone parce que $S * z = -\frac{1}{z}$ et donc $S * (\rho e^{-i\theta}) = -\frac{1}{\rho} e^{-i\theta}$.

20. Bien que cela ne soit pas indispensable pour la preuve, remarquons que I_z ne comprend qu'une quantité au plus dénombrable de valeurs. Le fait que, à z donné, la quantité $|cz+d|^2$ puisse être rendue aussi grande que l'on veut est évident. Donc I_z est borné vers le bas par zéro (qui n'est pas atteint, mais qui est une valeur d'adhérence).

Unicité Nous voulons à présent montrer que si $z \in D$, alors $A * z$ n'est plus dans D (sauf si $A = \pm\mathbb{1}$). Nous supposons que $z \in D$ et $A \in \mathrm{PSL}(2,\mathbb{Z})$ soient tels que $A * z \in D$, et nous prouvons qu'alors soit nous arrivons à une contradiction soit nous arrivons à $A = \mathbb{1}$. Pour cela nous allons décomposer en de nombreux cas.

(1) Nous commençons par $\Im(A * z) \geqslant \Im(z)$. Dans ce cas nous avons $|cz + d| \leqslant 1$ et en particulier $|c||\Im(z)| \leqslant 1$. Étant donné que le point de D qui a la partie imaginaire la plus petite est $-\frac{1}{2} + \frac{2}{\sqrt{3}}i$, nous trouvons $|c| \leqslant 2/\sqrt{3}$. Vu que c doit être entier, nous avons trois cas : $c = -1, 0, 1$.

(1a) Soit $c = 0$. Alors $A = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ et la condition de déterminant est $ad = 1$, ce qui signifie $a = d = 1$ (la possibilité $a = b = -1$ est « éliminée » le quotient par $\mathbb{Z}_2$ définissant $\mathrm{PSL}(2,\mathbb{Z})$). La matrice A doit alors être de la forme

$$A = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \tag{22.271}$$

et $A * z = z + b$. Si $z \in D$, alors le seul $z + b$ à être (peut-être) encore dans D est $b = 0$, mais alors A est l'identité.

(1b) Soit $c = 1$. Alors la condition $|cz + d| \leqslant 1$ nous donne trois possibilités[21] : $d = -1, 0, 1$.

i. Si $d = -1$, alors nous devons avoir $|z-1| \leqslant 1$. Il est instructif de faire un dessin, mais le point d'intersection entre les cercles $|z| = 1$ et $|z - 1| = 1$ est le point $\frac{1}{2} + \frac{\sqrt{3}}{2}i$, qui n'est pas dans D. Bref, il n'y a pas de points dans D vérifiant $|z - 1| \leqslant 1$.

ii. Si $d = 1$, alors (et c'est maintenant que la dissymétrie de D intervient) nous avons le point

$$z = -\frac{1}{2} + \frac{\sqrt{3}}{2}i \tag{22.272}$$

qui est dans D et qui vérifie $|z + 1| \leqslant 1$. Voyons à quoi ressemble la matrice A dans ce cas. Son déterminant est $a - b = 1$. Nous écrivons donc

$$A = \begin{pmatrix} b+1 & b \\ 1 & 1 \end{pmatrix}, \tag{22.273}$$

et en tenant compte du fait que $z\bar{z} = |z + 1| = 1$, nous calculons

$$A * z = \frac{(b+1)z + b}{z+1} \tag{22.274a}$$

$$= \frac{(bz + z + b)(\bar{z} + 1)}{|z+1|^2} \tag{22.274b}$$

$$= z + b + 1. \tag{22.274c}$$

La seule façon de ne pas quitter D est d'avoir $b = -1$, mais alors nous avons

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} \tag{22.275}$$

et $A * z = z$. Donc au final z est quand même le seul de son orbite à être dans D.

Notons au passage cette très intéressante propriété du point

$$z_0 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i. \tag{22.276}$$

C'est un point de qui vérifie $z_0 = A * z_0$ pour un élément non trivial A de $\mathrm{PSL}(2,\mathbb{Z})$. L'existence d'un tel élément est ce qui va nous coûter un peu de sueur pour prouver que $PSL(2,\mathbb{Z})$ est engendré par S et T.

21. Je ne rigolais pas quand je disais qu'on allait avoir de nombreux cas.

iii. Le cas $d = 0$ nous fait écrire $1 = \det A = -b$, donc $b = -1$ et

$$A = \begin{pmatrix} a & -1 \\ 1 & 0 \end{pmatrix}. \tag{22.277}$$

Nous avons alors $A * z = a - \frac{1}{z}$. De plus la condition $|z| \leqslant 1$ revient à $|z = 1|$. Pour les nombres complexes de module 1, l'opération $z \to -1/z$ est la symétrie autour de l'axe des imaginaires purs. Le seul à ne pas sortir de D est le fameux $z = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$, qui revient sur lui-même avec $a = -1$.

Nous passons à la possibilité $c = -1$. Dans ce cas la matrice est de la forme

$$A = \begin{pmatrix} a & b \\ -1 & d \end{pmatrix}, \tag{22.278}$$

et nous revenons au cas $c = 1$ en prenant $-A$ au lieu de A.

(2) Nous passons au cas $\Im(A * z) < \Im(z)$. Nous récrivons cette condition avec

$$\Im(A * z) < \Im\big(A^{-1} * (A * z)\big). \tag{22.279}$$

Si nous supposons que z et A sont tels que z et $A * z$ soient tous deux dans D, alors $z' = A * z$ est un élément de D tel que

$$\Im(z') < \Im(A^{-1} * z'). \tag{22.280}$$

Or nous avons vu qu'aucun élément de D vérifiant cette condition n'existait sans être trivial (celui qui ne bouge pas). Pour cela il suffit d'appliquer tout ce que nous venons de dire avec A^{-1} au lieu de A.

Quelques conclusions Après avoir passé tous les cas en revue, le fameux point $z_0 = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$ est l'unique point de D à accepter une matrice non triviale $A \in \mathrm{PSL}(2, \mathbb{Z})$ telle que $z_0 = A * z_0$. Nous remarquons aussi que tous les points de P sont ramenés dans D par une matrice obtenue comme produit de T, S, T^{-1} et S^{-1}.

□

Corolaire 22.94 ([423]).
Les matrices S et T génèrent le groupe modulaire au sens où toute matrice de $\mathrm{PSL}(2, \mathbb{Z})$ *s'écrit comme*

$$T^{m_1} S^{p_1} \dots T^{m_k} S^{p_k} \tag{22.281}$$

pour un certain k et des nombres $m_i, p_i \in \mathbb{Z}$. Autrement dit, $\mathrm{PSL}(2, \mathbb{Z}) = \mathrm{gr}(S, T)$.

Démonstration. Soit z, un point de D autre que z_0. Alors si $A \in \mathrm{PSL}(2, \mathbb{Z})$ est non trivial nous avons $A * z$ hors de D. Du coup, comme vu dans la démonstration du théorème 22.93, il existe $B \in \mathrm{gr}(S, T)$ tel que $B * (A * z) \in D$. Vu que D ne contient qu'un seul point de chaque orbite, nous avons

$$B * A * z = z, \tag{22.282}$$

et donc $BA = \pm\mathbb{1}$, ce qui prouve que[22] $A = B^{-1}$, c'est-à-dire que $A \in \mathrm{gr}(S, T)$. □

22. Dans $\mathrm{PSL}(2, \mathbb{Z})$, nous n'avons pas besoin de mettre $\pm$ parce qu'il est compris dans la définition.

Chapitre 23

Analyse vectorielle

23.1 Le théorème de Green

Soit un champ de vecteurs

$$F(x,y,z) = \begin{pmatrix} F_1(x,y,z) \\ F_2(x,y,z) \\ F_2(x,y,z) \end{pmatrix} \tag{23.1}$$

et un chemin $\sigma\colon [a,b] \to \mathbb{R}^3$ donné par

$$\sigma(t) = \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix}. \tag{23.2}$$

Nous avons défini la circulation de F le long de σ par

$$\begin{aligned} \int_\sigma F \cdot d\sigma &= \int_a^b F\big(\sigma(t)\big) \cdot \sigma'(t)dt \\ &= \int_a^b \Big[F_1\big(\sigma(t)\big)x'(t) + F_2\big(\sigma(t)\big)y'(y) + F_3\big(\sigma(t)\big)z'(t)\Big]dt \\ &= \int_\sigma F_1 dx + F_2 dy + F_3 dz. \end{aligned} \tag{23.3}$$

La dernière ligne est juste une notation compacte [1]. Elle sert à se souvenir qu'on va mettre x' à côté de F_1, y' à côté de F_2 et z' à côté de F_3. L'avantage de cette notation est qu'on peut écrire d'autres combinaisons.

Si f et g sont deux fonctions sur $\mathbb{R}^3$, nous pouvons écrire

$$\int_\sigma f dy + g dz. \tag{23.4}$$

Cela signifie

$$\int_a^b \Big[f\big(\sigma(t)\big)y'(t) + g\big(\sigma(t)\big)z'(t)\Big]dt. \tag{23.5}$$

Soit D une région du plan et σ, son contour que nous prenons, par convention [2], dans l'orientation trigonométrique, comme indiqué sur la figure 23.1. Nous supposons également que le domaine D n'a pas de trous intérieurs.

Nous notons par $\sigma = \partial D$ le bord de D, c'est-à-dire le contour dont nous venons de parler.

1. Il y aurai beaucoup de choses à dire là-dessus, mais la vie est trop courte pour parler de formes différentielles, et c'est dommage.
2. Il y aurait beaucoup de choses à dire sur ça aussi, mais...

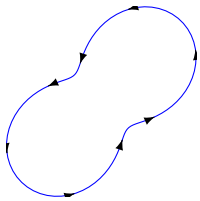


FIGURE 23.1 – Un contour avec son ortientation.

Théorème 23.1 (Théorème de Green)**.**
Soient $P, Q\colon D \to \mathbb{R}$ deux fonctions de classe C^1. Alors

$$\int_{\partial D} P dx + Q dy = \int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy. \tag{23.6}$$

Pour rappel, l'intégrale du membre de gauche signifie

$$\int_a^b \Big[P\big(\sigma(t)\big)\sigma'_x(t) + Q\big(\sigma(t)\big)\sigma'_y(t) \Big] dt. \tag{23.7}$$

Ce n'est d'ailleurs rien d'autre que l'intégrale du champ de vecteurs $\begin{pmatrix} P \\ Q \end{pmatrix}$.

Corolaire 23.2.
L'aire du domaine D est donnée par

$$A = \frac{1}{2} \int_{\partial D} (x dy - y dx). \tag{23.8}$$

Démonstration. L'intégrale $\int_{\partial D}(xdy - ydx)$ se traite avec le théorème de Green où l'on pose $P = -y$ et $Q = x$. Nous avons donc

$$\begin{aligned} \int_{\partial D} -y dx + x dy &= \int_D \left(\frac{\partial x}{\partial x} - \frac{\partial(-y)}{\partial y} \right) dx dy \\ &= \int_D 2\, dx dy. \end{aligned} \tag{23.9}$$

La dernière ligne est bien le double de la surface. □

Exemple 23.3.
Calculons (encore une fois) l'aire du disque de rayon R. Il s'agit de calculer l'intégrale

$$I = \frac{1}{2} \int_\sigma (x dt - y dx) \tag{23.10}$$

où σ est le cercle donné par

$$\sigma(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} R\cos(t) \\ R\sin(t) \end{pmatrix} \tag{23.11}$$

Le calcul est

$$\begin{aligned} I &= \frac{1}{2} \int_0^{2\pi} \underbrace{R\cos(\theta)}_{x} \underbrace{R\cos(\theta)}_{y'} - \underbrace{R\sin(\theta)}_{y} \underbrace{(-R\sin(\theta))}_{x'}\, d\theta \\ &= \frac{R^2}{2} \int_0^{\pi} d\theta \\ &= \pi R^2. \end{aligned} \tag{23.12}$$

△

Exemple 23.4.
Calculons l'aire de l'ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \leqslant 1 \tag{23.13}$$

dont le bord est donné par

$$\begin{cases} x(t) = a\cos(t) & (23.14a) \\ y(t) = b\sin(t). & (23.14b) \end{cases}$$

Le terme xdy devient $a\cos(t)b\cos(t) = ab\cos^2(t)$ et le terme ydx devient $b\sin(t)(-a\sin(t)) = -ab\sin^2(t)$. L'intégrale qui donne la surface est donc

$$\frac{1}{2}\int_{\partial D}(xdy - ydx) = \frac{1}{2}\int_0^{2\pi} ab = \pi ab. \tag{23.15}$$

△

Le théorème de Green peut être mis sous une autre forme.

Théorème 23.5 (Théorème de Green, forme vectorielle)**.**
Si G est un champ de vecteurs sur D, nous avons

$$\int_{\partial D} G \cdot d\sigma = \int_D (\nabla \times G) \cdot dS \tag{23.16}$$

où le second membre est le flux de $\nabla \times G$ sur la surface D.

Démonstration. Analysons le membre de droite. Nous savons que D est une surface dans le plan $\mathbb{R}^2$. Le vecteur normal à la surface est donc simplement le vecteur (constant) e_z. Le produit scalaire $(\nabla \times F) \cdot dS$ est donc $(\nabla \times F) \cdot e_z$ et se réduit à la troisième composante du rotationnel, c'est-à-dire

$$\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}. \tag{23.17}$$

Cela est bien le membre de droite de l'équation (23.6). Le membre de gauche de cette dernière est bien le membre de gauche de (23.16). □

Exemple 23.6.
Soit le champ de vecteurs $F(x, y) = \begin{pmatrix} xy^2 \\ y + x \end{pmatrix}$, et soit à calculer

$$\int_D \nabla \times F \cdot dS \tag{23.18}$$

où D est la région comprise entre les courbes $y = x^2$ et $y = x$ pour $x \geqslant 0$ (voir la figure 23.2).

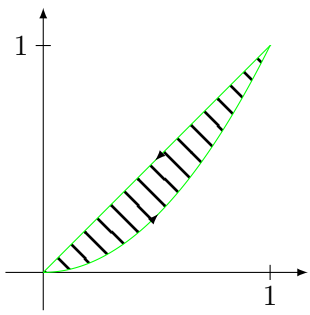


FIGURE 23.2 – Le contour d'intégration pour l'exemple 23.6.

Nous pouvons calculer cette intégrale directement en calculant le rotationnel de F :

$$\nabla \times F = \begin{pmatrix} 0 \\ 0 \\ 1-2xy \end{pmatrix}. \tag{23.19}$$

Par conséquent l'intégrale à effectuer est

$$I = \int_0^1 dx \int_{x^2}^x (1-2xy)dy = \frac{1}{12}. \tag{23.20}$$

```
----------------------------------------------------------------------
| Sage Version 4.6.1, Release Date: 2011-01-11                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: f(x,y)=1-2*x*y
sage: f.integrate(y,x**2,x).integrate(x,0,1)
(x, y) |--> 1/12
```

L'autre façon de calculer l'intégrale est d'utiliser le théorème de Green et de calculer la circulation de F le long de ∂D :

$$I = \int_{\partial D} F \cdot \sigma. \tag{23.21}$$

Le chemin $\sigma = \partial D$ est composé de la parabole $y = x^2$ et du segment de droite $x = y$. Attention : il faut respecter l'orientation. Nous avons

$$\sigma_1(t) = (t, t^2) \tag{23.22}$$

et

$$\sigma_2(t) = (1-t, 1-t). \tag{23.23}$$

Notez bien que le second chemin est $(1-t, 1-t)$ et non (t, t) parce qu'il faut le parcourir dans le bon sens (voir le dessin).

Commençons par le premier chemin :

$$\begin{aligned} \sigma_1(t) &= (t, t^2) \\ \sigma_1'(t) &= (1, 2t) \\ F\big(\sigma_1(t)\big) &= \begin{pmatrix} t^5 \\ t+t^2 \end{pmatrix}, \end{aligned} \tag{23.24}$$

et par conséquent

$$F\big(\sigma_1(t)\big) \cdot \sigma_1'(t) = t^5 + 2t^2 + 2t^3, \tag{23.25}$$

et le premier morceau de la circulation vaut

$$\int_{\sigma_1} F \cdot d\sigma_1 = \int_0^1 t^5 + 2t^2 + 2t^3 = \frac{4}{3}. \tag{23.26}$$

Pour le second chemin :

$$\begin{aligned} \sigma_2(t) &= (1-t, 1-t) \\ \sigma_2'(t) &= (-1, -1) \\ F\big(\sigma_2(t)\big) &= \begin{pmatrix} (1-t)^3 \\ 2(1-t) \end{pmatrix}. \end{aligned} \tag{23.27}$$

Par conséquent

$$F\big(\sigma_2(t)\big) \cdot \sigma_2(t) = -(1-t)^2 - 2(1-t). \tag{23.28}$$

Le second morceau de la circulation est par conséquent

$$\int_0^1 -(1-t)^2 - 2(1-t)dt = -\frac{5}{4}. \tag{23.29}$$

La circulation de F le long de σ est donc égale à

$$\frac{4}{3} - \frac{5}{4} = \frac{1}{12}. \tag{23.30}$$

Comme prévu, nous obtenons le même résultat. △

23.2 Théorème de la divergence dans le plan

23.2.1 La convention de sens de parcours

Soient D, un domaine dans le plan et un paramétrage

$$\begin{aligned} \sigma\colon [a,b] &\to \mathbb{R}^2 \\ t &\mapsto \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \end{aligned} \tag{23.31}$$

un paramétrage du bord ∂D de D. La normale à σ est perpendiculaire à la tangente, donc la normale extérieure de norme 1 vaut

$$n = \frac{\big(y'(t), -x'(t)\big)}{\sqrt{\big(x'(t)\big)^2 + \big(y'(t)\big)^2}} \quad \text{ou} \quad n- = \frac{\big(y'(t), -x'(t)\big)}{\sqrt{\big(x'(t)\big)^2 + \big(y'(t)\big)^2}}. \tag{23.32}$$

Comment faire le choix ?

Nous prenons comme convention que le sens *du chemin* doit être tel que le vecteur normal extérieur soit

$$n = \frac{\big(y'(t), -x'(t)\big)}{\sqrt{\big(x'(t)\big)^2 + \big(y'(t)\big)^2}}. \tag{23.33}$$

Donc si le chemin σ donne lieu à un vecteur n pointant vers l'intérieur, il faut utiliser le chemin qui va dans le sens contraire : $\tilde{\sigma}(t) = \sigma(1-t)$.

Les vecteurs tangents et normaux d'un contour sont dessinés sur la figure 23.3.

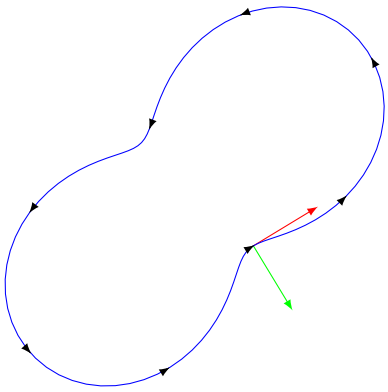


FIGURE 23.3 – Le champ de vecteurs tangents est dessiné en rouge tandis qu'en vert nous avons le champ de vecteurs normaux extérieurs.

23.2.2 Théorème de la divergence

Théorème 23.7 (Théorème de la divergence)**.**
Soit F un champ de vecteurs sur $\mathbb{R}^2$. Le flux de F à travers le bord de D est égal à l'intégrale de la divergence de F sur D. En formule :

$$\int_{\partial D} F \cdot n \, d\sigma = \int_D \nabla \cdot F \, dxdy. \tag{23.34}$$

Démonstration. Tant $F \cdot n$ que $\nabla \times F$ sont des fonctions. Le membre de gauche est donc l'intégrale d'une fonction sur un chemin et le membre de droite est l'intégrale d'une fonction sur une surface. Notre convention de sens de parcours du chemin permet d'écrire le produit scalaire $F \cdot n$ sous la forme suivante :

$$\begin{aligned} F \cdot n &= \frac{1}{\|\sigma'\|} \begin{pmatrix} F_x \\ F_y \end{pmatrix} \cdot \begin{pmatrix} y' \\ -x' \end{pmatrix} \\ &= \frac{1}{\|\sigma'\|} (F_x y' - F_y x') \\ &= \frac{1}{\|\sigma'\|} \begin{pmatrix} -F_y \\ F_x \end{pmatrix} \cdot \begin{pmatrix} x' \\ y' \end{pmatrix} \\ &= \frac{1}{\|\sigma'\|} \begin{pmatrix} -F_y \\ F_x \end{pmatrix} \cdot \sigma'. \end{aligned} \tag{23.35}$$

Par conséquent, la *fonction*

$$F \cdot n \tag{23.36}$$

est la même que la *fonction*

$$\frac{1}{\|\sigma'\|} \begin{pmatrix} -F_y \\ F_x \end{pmatrix} \cdot \sigma'. \tag{23.37}$$

L'intégrale de cette dernière fonction sur le chemin σ est

$$\begin{aligned} I &= \int_\sigma F \cdot n \\ &= \int_\sigma \frac{1}{\|\sigma'\|} \begin{pmatrix} -F_y \\ F_x \end{pmatrix} \cdot \sigma' \\ &= \int_a^b \frac{1}{\|\sigma'(t)\|} \begin{pmatrix} -F_y(\sigma(t)) \\ F_x(\sigma(t)) \end{pmatrix} \cdot \sigma'(t) \|\sigma'(t)\| dt \\ &= \int_a^b \begin{pmatrix} -F_y \\ F_x \end{pmatrix} \cdot \sigma'(t) dt. \end{aligned} \tag{23.38}$$

Cette dernière intégrale est la circulation du champ de vecteurs $\begin{pmatrix} -F_y \\ F_x \end{pmatrix}$ sur le chemin σ. Le théorème de Green 23.5 nous enseigne que la circulation le long d'un chemin est égale au flux du rotationnel à travers la surface. Par conséquent,

$$I = \int_D \left(\nabla \times \begin{pmatrix} -F_y \\ F_x \end{pmatrix} \right) \cdot dS = \int_D \nabla \cdot F \, dxdy \tag{23.39}$$

□

23.3 Théorème de Stokes

Nous nous mettons maintenant dans $\mathbb{R}^3$, et nous y considérons une surface paramétrée S donc le bord est ∂S.

Théorème 23.8 (Théorème de Stokes)**.**
Alors le flux du rotationnel de F à travers S est égal à la circulation de F le long du bord. En formule :

$$\int_S \nabla \times F \cdot dS = \int_{\partial S} F \cdot d\sigma. \tag{23.40}$$

Nous pouvons nous donner une idée du pourquoi ce théorème est vrai. D'abord, si la surface est plate, cela est exactement le théorème de Green 23.5. Supposons maintenant que le bord reste plat, mais que la surface se déforme un petit peu. Le chemin

$$\sigma(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \\ 0 \end{pmatrix} \tag{23.41}$$

est tout autant le bord du disque plat de rayon 1 que celui de la demi-sphère

$$\phi(x, y) = \begin{pmatrix} x \\ y \\ \sqrt{1 - x^2 - y^2} \end{pmatrix}. \tag{23.42}$$

Le champ de vecteur que nous considérons est $G = \nabla \times F$. Il a un certain flux à travers le disque plat, et ce plus est égal à la circulation de F sur σ. Quel est le flux de G à travers la demi-sphère ? Étant donné que $\nabla \cdot G = \nabla \cdot (\nabla \times F) = 0$, le champ de vecteurs G est incompressible, de telle façon que tout ce qui rentre dans la demi-sphère doit en sortir. Le flux de G à travers la demi-sphère doit par conséquent être égal à celui à travers le disque plat.

Exemple 23.9.
Soit C l'intersection entre le cylindre $x^2 + y^2 = 1$ et le plan $x + y + z = 1$. Calculer la circulation de

$$F(x, y, z) = \begin{pmatrix} -y^3 \\ x^3 \\ -z^3 \end{pmatrix} \tag{23.43}$$

le long de C.

Au lieu de calculer directement

$$\int_C F \cdot d\sigma, \tag{23.44}$$

nous allons calculer

$$\int_S \nabla \times F \cdot dS \tag{23.45}$$

où S est une surface dont C est le bord. Cette intégrale est à calculer avec la formule (19.267).

La première chose à faire est de trouver une surface dont le bord est C et en trouver un paramétrage ϕ. Le plus simple est de prendre le graphe du plan sur le cercle $x^2 + y^2 + 1$. Un paramétrage de cette surface est simplement

$$\begin{aligned} \phi \colon D &\to \mathbb{R}^3 \\ (x, y) &\mapsto \begin{pmatrix} x \\ y \\ 1 - x - y \end{pmatrix} \end{aligned} \tag{23.46}$$

où D est le disque de rayon 1. Étant donné que cela paramètre le plan $x + y + z - 1 = 0$, le vecteur normal est $n = e_x + e_y + z_z$. Nous pouvons cependant calculer ce vecteur normal en suivant la recette usuelle. D'abord les vecteurs tangents sont

$$\frac{\partial \phi}{\partial x} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad \frac{\partial \phi}{\partial y} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}. \tag{23.47}$$

Et le vecteur normal est donné par le produit vectoriel :

$$\begin{aligned} n &= \frac{\partial \phi}{\partial x} \times \frac{\partial \phi}{\partial y} \\ &= \begin{vmatrix} e_x & e_y & e_z \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{vmatrix} \\ &= e_x + e_y + z_z. \end{aligned} \tag{23.48}$$

Ensuite, le rotationnel de F est donné par

$$\nabla \times F = 3(x^2 + y^2)e_z. \tag{23.49}$$

Par conséquent,

$$\nabla \times F \cdot \left(\frac{\partial \phi}{\partial x} \times \frac{\partial \phi}{\partial y} \right) = 3(x^2 + y^2). \tag{23.50}$$

L'intégrale à calculer est donc

$$\begin{aligned} \int_S \nabla \times F \cdot dS &= \int_D (\nabla \times F)\big(\phi(x,y)\big) \cdot \left(\frac{\partial \phi}{\partial x} \times \frac{\partial \phi}{\partial y} \right) dxdy \\ &= 3 \int_D (x^2 + y^2) dxdy. \end{aligned} \tag{23.51}$$

Cette dernière intégrale est l'intégrale d'une fonction sur le disque de rayon 1. Elle s'effectue en passant aux coordonnées polaires :

$$3 \int_D (x^2 + y^2) dxdy = \int_0^{2\pi} d\theta \int_0^1 (r^2) r \, dr = \frac{3\pi}{2}. \tag{23.52}$$

△

23.4 Théorème de Gauss

Soit V une partie de $\mathbb{R}^3$ délimitée par une surface S sur laquelle nous considérons la normale extérieure. Soit F un champ de vecteurs sur $\mathbb{R}^3$.

Théorème 23.10 (Théorème de la divergence ou de Gauss)**.**
Le flux d'un champ de vecteur F à travers une surface fermée est égale à l'intégrale de la divergence sur le volume correspondant :

$$\int_{\partial V} F \cdot dS = \int_V \nabla \cdot F \, dxdydz. \tag{23.53}$$

Ce théorème signifie que la quantité de fluide qui s'accumule dans le volume (le flux est ce qui rentre moins ce qui sort) est égal à l'intégrale de $\nabla \cdot F$ sur le volume, alors que nous savons que, localement, la quantité $\nabla \cdot F(x, y, z)$ est la quantité de fluide qui s'accumule au point (x, y, z).

Remarque 23.11.
Ce théorème ne fonctionne qu'avec des surfaces fermées. Essayer de l'appliquer au calcul de flux à travers des surfaces ouvertes n'a pas de sens parce qu'une surface ouverte ne délimite pas un volume.

23.12.
La formule de la divergence peut être utilisée comme intégration par partie. Si u est une fonction et F un champ de vecteurs, $\nabla(uF) = \nabla(u) \cdot F + u\nabla \cdot F$ et alors

$$\int_{\partial V} uF \cdot n = \int V \nabla(uF) = \int_V u\nabla \cdot F + \int_V F \cdot \nabla u \tag{23.54}$$

où n est le champ de vecteurs normal extérieur à V. En remettant les termes dans un ordre qui ressemble plus à l'intégration par partie :

$$\int_V F \cdot \nabla u = \int_{\partial V} uF \cdot n - \int_V u \nabla F. \tag{23.55}$$

Exemple 23.13.
Calculer le flux du champ de vecteurs

$$F(x, y, z) = \begin{pmatrix} 2x \\ y^2 \\ z^2 \end{pmatrix} \tag{23.56}$$

à travers la sphère de rayon 1 centrée à l'origine. Nous utilisons le théorème de la divergence

$$\int_S F \cdot n \, dS = \int_B \nabla \cdot F \, dxdydz \tag{23.57}$$

où S est la sphère et B est la boule (la sphère pleine). La divergence de F se calcule :

$$\nabla \cdot F = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z} = 2 + 2x + 2y. \tag{23.58}$$

L'intégrale est donc en trois termes :

$$\begin{aligned} \int_B 2 = 2\text{Volume(B)} = \frac{8\pi}{3} \\ \int_B y \, dxdydz = 0 \\ \int_B z \, dxdydz = 0. \end{aligned} \tag{23.59}$$

△

Dans certains cas le théorème de Gauss permet de simplifier le calcul de l'intégrale d'une fonction sur une surface.

Exemple 23.14.
Soit à calculer l'intégrale

$$I = \int_{\partial B} (x^2 + y + z) dS, \tag{23.60}$$

c'est-à-dire l'intégrale de la fonction $x^2 + y + z$ sur la sphère. Le vecteur normal à la sphère est

$$n = xe_x + ye_y + ze_z. \tag{23.61}$$

Étant donné que nous sommes sur la sphère de rayon 1, ce vecteur est même normé. La fonction que nous regardons n'est rien d'autre que $F \cdot n$ avec

$$F = \begin{pmatrix} x \\ 1 \\ 1 \end{pmatrix}. \tag{23.62}$$

Nous pouvons donc simplement intégrer $\nabla \cdot F$ sur toute la boule :

$$I = \int_B \nabla \cdot F \, dxdydz = \int_B 1 \, dxdudz = \frac{4\pi}{3}. \tag{23.63}$$

△

23.5 Coordonnées curvilignes

23.5.1 Base locale

Nous connaissons déjà les coordonnées sphériques et cylindriques sur $\mathbb{R}^3$. Ce sont des systèmes « un peu courbes ». Il en existe bien entendu de nombreux autres (sinon ce ne serait pas drôle), et nous allons faire une étude un peu général de ces systèmes de **coordonnées curvilignes**. Des coordonnées curvilignes sur $\mathbb{R}^3$ est n'importe quel [3] système qui permet de repérer un point de $\mathbb{R}^3$ à partir de trois nombres.

Il s'agit donc d'un ensemble de trois applications

$$x_i \colon \mathbb{R}^3 \to \mathbb{R}. \tag{23.64}$$

Les coordonnées cylindriques sont

$$\begin{cases} x_1(r,\theta,z) = r\cos\theta & (23.65a) \\ x_2(r,\theta,z) = r\sin\theta & (23.65b) \\ x_3(r,\theta,z) = z & (23.65c) \end{cases}$$

Soit donc un système général $q = (q_1, q_2, q_3)$ et

$$M(q) = \begin{pmatrix} x_1(q) \\ x_2(q) \\ x_3(q) \end{pmatrix}. \tag{23.66}$$

Si nous fixons q_2 et q_3 et que nous laissons varier q_1, nous obtenons une courbe [4] dont nous pouvons considérer le vecteur vitesse, c'est-à-dire le vecteur tangent. En chaque point nous avons ainsi trois vecteurs

$$\frac{\partial M}{\partial q_i}(q). \tag{23.67}$$

Nous disons que le système de coordonnées curviligne est **orthogonal** si ces trois vecteurs sont orthogonaux. Dans la suite nous supposerons que c'est toujours le cas.

Nous posons

$$h_i = \left\| \frac{\partial M}{\partial q_i} \right\| \tag{23.68}$$

et nous considérons les trois vecteurs normés

$$e_i = h_i^{-1} \frac{\partial M}{\partial q_i}. \tag{23.69}$$

Les trois vecteurs $\{e_1, e_2, e_3\}$ forment une base orthonormée dite **base locale**. Ce sont des vecteurs liés [5] au point M.

23.5.2 Importance de l'orthogonalité

Nous avons dit que nous nous restreignons au cas où les vecteurs e_i sont orthogonaux. En termes de produits scalaires, cela signifie

$$e_i \cdot e_j = \delta_{ij}. \tag{23.70}$$

Nous en étudions maintenant quelques conséquences. L'équation (23.69) peut s'écrire plus explicitement sous la forme

$$e_i = \sum_k h_i^{-1} \frac{\partial x_k}{\partial q_i} 1_k. \tag{23.71}$$

3. Nous n'entrons pas dans les détails de régularité.
4. Dans le cas des sphériques, c'est une demi-droite horizontale d'angle q_2 et de hauteur q_3.
5. En géométrie différentielle on dira que ce sont des élément de l'espace tangent, mais c'est une toute autre histoire.

Notez que pour chaque k et i, la quantité $h_i^{-1}\frac{\partial x_k}{\partial q_i}$ est un simple nombre. Nous allons les mettre dans une matrice :

$$A_{ki} = h_i^{-1}\frac{\partial x_k}{\partial q_i}. \tag{23.72}$$

Cela nous donne le changement de base

$$e_i = \sum_k A_{ki} 1_k. \tag{23.73}$$

Le produit $e_i \cdot e_j$ s'écrit alors

$$\begin{aligned} e_i \cdot e_j &= \sum_{kl} A_{ki}A_{lj} \underbrace{1_k \cdot 1_l}_{=\delta_{kl}} \\ &= \sum_{kl} A_{ki}A_{lj}\delta_{kl} \\ &= \sum_k A_{ki}A_{kj} \\ &= \sum_k (A^T)_{ik}A_{kj}. \end{aligned} \tag{23.74}$$

Or cela doit valoir δ_{ij}. Par conséquent

$$A^T = A^{-1}. \tag{23.75}$$

Le fait que les coordonnées curvilignes considérées soient orthogonales s'exprime donc par la fait que la matrice de changement de base est une matrice orthogonale.

Cette circonstance nous permet d'inverser le changement de base (23.73) en multipliant cette équation par $(A^{-1})_{il}$ des deux côtés et en faisant la somme sur i :

$$\sum_i (A^{-1})_{il} e_i = \sum_{kl} \underbrace{A_{ki}(A^{-1})_{il}}_{=\delta_{kl}} 1_k, \tag{23.76}$$

par conséquent

$$\sum_i (A^T)_{il} e_i = 1_l, \tag{23.77}$$

et

$$1_l = \sum_i A_{li} e_i = \sum_i h_i^{-1}\frac{\partial x_l}{\partial q_i} e_i. \tag{23.78}$$

Armés de cette importante formule, nous pouvons exprimer les quantités que nous connaissons dans la base canonique en termes de la base locale.

Une autre conséquence du fait que e_1, e_2 et e_3 est une base orthonormée est que, éventuellement en réordonnant les vecteurs, on a

$$\begin{aligned} e_1 \times e_2 &= e_3 \\ e_2 \times e_3 &= e_1 \\ e_3 \times e_1 &= e_2 \end{aligned} \tag{23.79}$$

Ces trois relations s'écrivent en une seule avec

$$e_i \times e_j = \sum_k \epsilon_{ijk} e_k \tag{23.80}$$

où

$$\epsilon_{ijk} = \begin{cases} 0 & \text{si } i,j,k \text{ ne sont pas tous différents} \\ 1 & \text{si } ijk \text{ se ramène à 123 par un nombre pair de permutations} \\ -1 & \text{si } ijk \text{ se ramène à 123 par un nombre impair de permutations} \end{cases} \tag{23.81}$$

est le **symbole de Levi-Civita**. La formule du produit vectoriel peut également être utilisée à l'envers sous la forme

$$e_k = \frac{1}{2}\sum_{ij} \epsilon_{ijk}\, e_i \times e_j. \tag{23.82}$$

Le symbole de Levi-Civita possède de nombreuses formules. En voici certaines, facilement démontrables en considérant tous les cas :

$$\epsilon_{ijk}\epsilon_{ijl} = \delta_{kl}|\epsilon_{ijk}|. \tag{23.83}$$

Grâce au symboles de Levi-Civita, le produit mixte des vecteurs de base a une belle forme :

$$e_l \cdot (e_i \times e_j) = \sum_k \epsilon_{ijk} e_l \times e_k = \sum_k \epsilon_{ijk}\delta_{lk} = \epsilon_{ijl}. \tag{23.84}$$

23.5.3 Coordonnées polaires

Les coordonnées curvilignes polaires sont données par

$$M(r,\theta) = \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \end{pmatrix}, \tag{23.85}$$

et par conséquent

$$\frac{\partial M}{\partial r} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}, \quad \frac{\partial M}{\partial \theta} = \begin{pmatrix} -r\sin(\theta) \\ r\cos(\theta) \end{pmatrix}. \tag{23.86}$$

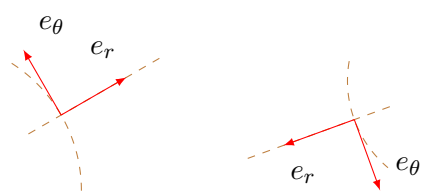
Nous avons les normes $h_r = 1$ et $h_\theta = r$, et donc les vecteurs de la base locale en (r,θ) sont

$$e_r = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} = \cos(\theta)e_x + \sin(\theta)e_y \tag{23.87}$$

ainsi que

$$e_\theta = \begin{pmatrix} -\sin(\theta) \\ \cos(\theta) \end{pmatrix} = -\sin(\theta)e_x + \cos(\theta)e_y. \tag{23.88}$$

Ces vecteurs sont représentés à la figure 23.4. Notez qu'il y en a une paire différente en chaque point.



(a) Base locale. (b) Base locale.

FIGURE 23.4 – En brun, les lignes que le point suivrait si on ne variait qu'une coordonnées polaires à la fois. Les vecteurs rouges sont les vecteurs e_r et e_θ.

23.5.4 Coordonnées cylindriques

Les coordonnées cylindriques sont les mêmes que les coordonnées polaires à part qu'il faut écrire

$$M(r,\theta,z) = \begin{pmatrix} r\cos(\theta) \\ r\sin(\theta) \\ z \end{pmatrix}, \tag{23.89}$$

et nous avons le vecteur de base supplémentaire

$$e_z = \frac{\partial M}{\partial z} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{23.90}$$

parce que $h_z = 1$.

23.5.5 Coordonnées sphériques

Les coordonnées curvilignes sphériques sont données par

$$M(\rho,\theta,\varphi) = \begin{pmatrix} \rho\sin(\theta)\cos(\varphi) \\ \rho\sin(\theta)\sin(\varphi) \\ \rho\cos(\theta) \end{pmatrix}, \tag{23.91}$$

dont les dérivées sont données par

$$\begin{aligned} \frac{\partial M}{\partial r} &= \begin{pmatrix} \sin(\theta)\cos(\varphi) \\ \sin(\theta)\sin(\varphi) \\ \cos(\theta) \end{pmatrix}, \qquad \frac{\partial M}{\partial \theta} = \begin{pmatrix} \rho\cos(\theta)\cos(\varphi) \\ \rho\cos(\theta)\sin(\varphi) \\ -\rho\sin(\theta) \end{pmatrix}, \\ \frac{\partial M}{\partial \varphi} &= \begin{pmatrix} -\rho\sin(\theta)\sin(\varphi) \\ \rho\sin(\theta)\cos(\varphi) \\ 0 \end{pmatrix} \end{aligned} \tag{23.92}$$

Les normes de ces vecteurs sont $h_\rho = 1$, $h_\theta = \rho$ et $h_\varphi = \rho\sin(\theta)$. Les vecteurs de la base locale en (ρ,θ,φ) sont donc

$$\begin{aligned} e_r &= \begin{pmatrix} \sin(\theta)\cos(\varphi) \\ \sin(\theta)\sin(\varphi) \\ \cos(\theta) \end{pmatrix}, \quad e_\theta = \begin{pmatrix} \cos(\theta)\cos(\varphi) \\ \cos(\theta)\sin(\varphi) \\ -\sin(\theta) \end{pmatrix}, \\ e_\varphi &= \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \\ 0 \end{pmatrix} \end{aligned} \tag{23.93}$$

23.5.6 Gradient en coordonnées curvilignes

Soit $(x,y,z) \mapsto f(x,y,z)$ une fonction sur $\mathbb{R}^3$. Nous pouvons la composer avec les coordonnées curvilignes q pour obtenir la fonction

$$\tilde{f}(q_1,q_2,q_3) = f\big(x_1(q),x_2(x),x_3(q)\big). \tag{23.94}$$

Nous disons que $\tilde{f}$ est l'expression de f dans les coordonnées q. Nous savons déjà comment calculer le gradient de f en coordonnées cartésiennes :

$$F(x,y,z) = \nabla f(x,y,z) = \begin{pmatrix} \partial_x f(x,y,z) \\ \partial_y f(x,y,z) \\ \partial_z f(x,y,z) \end{pmatrix}. \tag{23.95}$$

Cela est un vecteur lié au point (x,y,z). Notre objectif de bonheur dans la vie serait d'exprimer les coordonnées de ce vecteur dans la base $\{e_1,e_2,e_3\}$. En d'autres termes, nous voudrions trouver les nombres $\tilde{F}_1$, $\tilde{F}_2$ et $\tilde{F}_3$ tels que

$$F(x,y,z) = F\big(x(q),y(q),z(q)\big) = \tilde{F}_1 e_1 + \tilde{F}_2 e_2 + \tilde{F}_3 e_3. \tag{23.96}$$

Ces nombres seront des fonctions de (q_1,q_2,q_3).

Par définition,

$$\nabla f = \sum_l \frac{\partial f}{\partial x_l} 1_l. \tag{23.97}$$

En remplaçant 1_l par sa valeur en termes des e_i par la formule (23.78),

$$\begin{aligned}\nabla f &= \sum_l \frac{\partial f}{\partial x_l} 1_l \\ &= \sum_l \frac{\partial f}{\partial x_l} \sum_i h_i^{-1} \frac{\partial x_l}{\partial q_i} e_i \\ &= \sum_{il} \frac{1}{h_i} \frac{\partial f}{\partial x_l} \frac{\partial x_l}{\partial q_i} e_i \\ &= \sum_i \frac{1}{h_i} \frac{\partial \tilde{f}}{\partial q_i} e_i.\end{aligned} \tag{23.98}$$

Plus explicitement,

$$\nabla f\big(x(q), y(q), z(q)\big) = \sum_i \frac{1}{h_i(q)} \frac{\partial \tilde{f}}{\partial q_i}(q) e_i \tag{23.99}$$

où

$$h_i(q) = \left\| \frac{\partial M}{\partial q_i}(q) \right\|. \tag{23.100}$$

Le plus souvent nous n'allons pas noter explicitement la dépendance de h_i en q.

23.5.6.1 Coordonnées sphériques

Nous pouvons exprimer le gradient d'une fonction en coordonnées sphériques en utilisant la formule (23.99) :

$$\nabla \tilde{f}(\rho, \theta, \varphi) = \frac{\partial \tilde{f}}{\partial \rho} e_\rho + \frac{1}{\rho} \frac{\partial \tilde{f}}{\partial \theta} e_\theta + \frac{1}{\rho \sin(\theta)} \frac{\partial \tilde{f}}{\partial \varphi} r_\varphi. \tag{23.101}$$

Cette expression peut paraitre peu pratique parce que les vecteurs e_ρ, e_θ et e_φ eux-mêmes changent en chaque point. Elle est effectivement peu adaptée au dessin, mais elle est très pratique pour des fonctions ayant des symétries.

Exemple 23.15.
Le potentiel de la gravitation est la fonction

$$V(x, y, z) = \frac{1}{\sqrt{x^2 + y^2 + z^2}}. \tag{23.102}$$

En coordonnées sphériques elle s'écrit

$$\tilde{V}(\rho, \theta, \varphi) = \frac{1}{\rho}. \tag{23.103}$$

En voilà une fonction qu'elle est facile à dériver, contrairement à V ! En suivant la formule (23.101), nous avons immédiatement

$$\nabla \tilde{V} = -\frac{1}{\rho^2} e_\rho. \tag{23.104}$$

Nous voyons immédiatement que cela est un champ de vecteurs dont la norme diminue comme le carré de la distance à l'origine et qui est en permanence dirigé vers l'origine. △

23.5.7 Divergence en coordonnées curvilignes

Nous savons que

$$\nabla \tilde{f} = \sum_j \frac{1}{h_j} \frac{\partial \tilde{f}}{\partial q_j} e_j. \tag{23.105}$$

Nous pouvons en particulier considérer la fonction $f(q) = q_i$. De la même manière que nous avions noté x_i la fonction $x \mapsto x_i$, nous notons q_i la fonction $q \mapsto q_i$. Le gradient de cette fonction est donné par

$$\nabla q_i = \sum_j \frac{1}{h_j} \frac{\partial q_i}{\partial q_j} e_j, \tag{23.106}$$

mais $\frac{\partial q_i}{\partial q_j} = \delta_{ij}$, donc

$$\nabla q_i = \frac{e_i}{h_i}, \tag{23.107}$$

ou encore

$$e_i = h_i \nabla q_i. \tag{23.108}$$

Cela n'est pas étonnant : la direction dans laquelle la coordonnées q_i varie le plus est le vecteur e_i qui donne la tangente à la courbe obtenue lorsque *seul* q_i varie.

Commençons par calculer la divergence de e_i. En utilisant la formule (23.82),

$$\nabla \cdot e_k = \frac{1}{2} \sum_{ij} \epsilon_{ijk} \nabla \cdot (e_i \times e_j). \tag{23.109}$$

Nous avons, en utilisant les règles de Leibnitz de la proposition 11.424,

$$\begin{aligned}
\nabla \cdot (e_i \times e_j) &= \nabla \cdot (h_i \nabla q_i \times h_j \nabla q_j) \\
&= \nabla(h_i h_j) \cdot \big(\nabla q_i \times \nabla q_j\big) + h_i h_j \nabla \cdot \big(\nabla q_i \times \nabla q_j\big) \\
&= \nabla(h_i h_j) \cdot \big(\nabla q_i \times \nabla q_j\big) \\
&\quad + h_i h_j \nabla q_j \cdot \big(\underbrace{\nabla \times \nabla q_i}_{=0}\big) \\
&\quad + h_i h_j \nabla q_i \cdot \big(\underbrace{\nabla \times \nabla q_j}_{=0}\big)
\end{aligned} \tag{23.110}$$

Cela nous fait

$$\nabla \cdot e_k = \sum_{ij} \epsilon_{ijk} \frac{\nabla(h_i h_j)}{h_i h_j} \cdot (e_i \times e_j). \tag{23.111}$$

parce que $\nabla q_i = h_i^{-1} e_i$. Nous pouvons développer le gradient qui intervient :

$$\nabla(h_i h_j) = \sum_l \frac{1}{h_l} \frac{\partial}{\partial q_l} (h_i h_j) e_l. \tag{23.112}$$

Nous voyons donc arriver le produit mixte $e_l \cdot (e_i \times e_j)$. En utilisant la formule (23.84), cela s'exprime directement sous la forme ϵ_{ijl}.

Nous avons alors

$$\begin{aligned}
\nabla \cdot e_k &= \frac{1}{2} \sum_{ijl} \frac{1}{h_i h_j h_l} \frac{\partial}{\partial q_l} (h_i h_j) \epsilon_{ijk} \epsilon_{ijl} \\
&= \frac{1}{2} \sum_{ijl} \delta_{kl} |\epsilon_{ijk}| \frac{\partial}{\partial q_l} (h_i h_j) \\
&= \frac{1}{2} \sum_{ij} \frac{|\epsilon_{ijk}|}{h_i h_j h_k} \frac{\partial}{\partial q_k} (h_i h_j).
\end{aligned} \tag{23.113}$$

Par exemple,

$$\nabla \cdot e_1 = \frac{1}{h_1 h_2 h_3} \frac{\partial}{\partial q_1} (h_2 h_3). \tag{23.114}$$

Nous devons maintenant chercher le gradient d'un champ général

$$F(q) = \sum_k F_k(q) e_k. \tag{23.115}$$

La première chose à faire est d'utiliser la formule de Leibnitz :

$$\nabla \cdot F = \sum_k \nabla F_k(q) \cdot e_k + \sum_k F_k(q) \nabla \cdot e_k. \tag{23.116}$$

Afin d'alléger les notations, nous allons nous concentrer sur le terme numéro k et ne pas écrire la somme. Si i et j sont les nombres tels que $\epsilon_{ijk} = 1$, alors ce que la formule (23.113) signifie, c'est que

$$\nabla \cdot e_k = \frac{1}{h_1 h_2 h_3} \frac{\partial}{\partial q_k}(h_i h_j). \tag{23.117}$$

Nous savons déjà par la formule (23.99) que

$$\nabla F_k = \sum_l \frac{1}{h_l} \frac{\partial F_k}{\partial q_l} e_l, \tag{23.118}$$

par conséquent

$$\nabla F_k \cdot e_k = \sum_l \frac{1}{h_l} \frac{\partial F_k}{\partial q_l} \delta_{kl} = \frac{1}{h_k} \frac{\partial F_k}{\partial q_k}. \tag{23.119}$$

Pour obtenir cela nous avons utilisé le fait que $e_l \cdot e_k = \delta_{lk}$. Le terme numéro k de la somme (23.116) est donc

$$\frac{1}{h_k} \frac{\partial F_k}{\partial q_k} + \frac{F_k}{h_k h_i h_j} \frac{\partial (h_i h_j)}{\partial q_k} = \frac{1}{h_i h_j h_k} \frac{\partial (F_k h_i h_j)}{\partial q_k} \tag{23.120}$$

où il est entendu que i et j représentent les nombres tels que $\epsilon_{ijk} = 1$.

Au final, nous avons

$$\nabla \cdot F = \frac{1}{h_1 h_2 h_3} \sum_{ijk} |\epsilon_{ijk}| \frac{\partial (F_k h_i h_j)}{\partial q_k}. \tag{23.121}$$

Ici, la somme sur i et j consiste seulement à sélectionner les termes tels que i et j ne sont pas k. En écrivant la somme explicitement,

$$\nabla \cdot F = \frac{1}{h_1 h_2 h_3} \left[\frac{\partial}{\partial q_1}(F_1 h_2 h_3) + \frac{\partial}{\partial q_2}(F_2 h_1 h_3) + \frac{\partial}{\partial q_3}(F_3 h_1 h_2) \right]. \tag{23.122}$$

23.5.7.1 Coordonnées cylindriques

En coordonnées cylindriques, nous avons déjà vu que $h_r = 1$, $h_\theta = r$ et $h_z = 1$. La divergence est donc donnée par

$$\nabla \cdot F = \frac{1}{r} \left[\frac{\partial}{\partial r}(r F_r) + \frac{\partial}{\partial \theta}(F_\theta) + \frac{\partial}{\partial z}(r F_z) \right]. \tag{23.123}$$

Par exemple si

$$F(r, \theta, z) = r e_\theta + e_z, \tag{23.124}$$

nous avons

$$(\nabla \cdot F)(r, \theta, z) = \frac{1}{r} \left[\frac{\partial}{\partial \theta}(r) + \frac{\partial}{\partial z}(r) \right] = 0. \tag{23.125}$$

Cela est logique parce que $r e_\theta$ est à peu près le champ dont nous avons parlé dans l'exemple (11.421), qui était à divergence nulle. En réalité, le champ dont on parlait dans cet exemple était exactement $-e_\theta$. Le champ e_z est également à divergence nulle parce qu'il est constant.

23.5.7.2 Coordonnées sphériques

En coordonnées sphériques, nous avons $h_\rho = 1$, $h_\theta = r$ et $h_\varphi = r \sin\theta$, donc

$$\nabla \cdot F = \frac{1}{r^2 \sin\theta} \left[\frac{\partial}{\partial \rho}(\rho^2 \sin\theta F_\rho) + \frac{\partial}{\partial \theta}(\rho \sin\theta F_\theta) + \frac{\partial}{\partial \varphi}(\rho F_\varphi) \right]. \tag{23.126}$$

si $F(\rho, \theta, \varphi) = F_\rho e_\rho + F_\theta e_\theta + F_\varphi e_\varphi$.

23.5.8 Laplacien en coordonnées curvilignes orthogonales

Soit une fonction $f\colon \mathbb{R}^3 \to \mathbb{R}$. Le **laplacien** de f est donné par

$$\Delta f = \nabla \cdot (\nabla f). \tag{23.127}$$

En utilisant les formules données, nous avons

$$\Delta f = \frac{1}{h_1 h_2 h_3}\left[\frac{\partial}{\partial q_1}\left(\frac{h_2 h_3}{h_1}\frac{\partial f}{\partial q_1}\right) + \frac{\partial}{\partial q_2}\left(\frac{h_1 h_3}{h_2}\frac{\partial f}{\partial q_2}\right) + \frac{\partial}{\partial q_3}\left(\frac{h_1 h_2}{h_3}\frac{\partial f}{\partial q_3}\right)\right]. \tag{23.128}$$

Dans cette expression, la fonction f est donnée comme fonction de q_1, q_2 et q_3.

En coordonnées cylindriques, cela s'écrit

$$\begin{aligned}\Delta f &= \frac{1}{r}\left[\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{\partial}{\partial \theta}\left(\frac{1}{r}\frac{\partial f}{\partial \theta}\right) + \frac{\partial}{\partial z}\left(r\frac{\partial f}{\partial z}\right)\right]\\ &= \frac{\partial^2 f}{\partial r^2} + \frac{1}{r}\frac{\partial f}{\partial r} + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2}.\end{aligned} \tag{23.129}$$

Dans cette expression, f est fonction de r, θ et z.

En coordonnées sphériques, cela devient

$$\Delta f = \frac{1}{\rho^2 \sin\theta}\left[\frac{\partial}{\partial \rho}\left(\rho^2 \sin\theta \frac{\partial f}{\partial \rho}\right) + \frac{\partial}{\partial \theta}\left(\sin\theta\frac{\partial f}{\partial \theta}\right) + \frac{\partial}{\partial \varphi}\left(\frac{1}{\sin\theta}\frac{\partial f}{\partial \varphi}\right)\right]. \tag{23.130}$$

Dans cette expression, f est fonction de ρ, θ et φ.

23.5.9 Rotationnel en coordonnées curvilignes orthogonales

Nous voulons calculer le rotationnel de $F(q) = \sum_k F_k(q)e_k$. Pour cela nous commençons par écrire $e_k = h_k \nabla q_k$ et nous utilisons la formule de la proposition 11.424(3) avec $F_k h_k$ en guise de f :

$$\begin{aligned}\nabla \times F_k e_k &= \nabla \times (F_k h_k \nabla q_k)\\ &= F_k h_k \underbrace{\nabla \times (\nabla q_k)}_{=0} + \nabla(F_k h_k) \times \nabla q_k\\ &= \frac{1}{h_k}\nabla(F_k h_k) \times e_k.\end{aligned} \tag{23.131}$$

Nous utilisons à présent la formule (23.99) du gradient et le formule $e_j \times e_k = \sum_l \epsilon_{jkl} e_l$:

$$\begin{aligned}\nabla \times (F_k e_k) &= \sum_j \frac{1}{h_j h_k}\frac{\partial}{\partial q_j}(F_k h_k) e_j \times e_k\\ &= \sum_{jl} \frac{1}{h_j h_k}\epsilon_{jkl}\frac{\partial}{\partial q_j}(F_k h_k) e_l.\end{aligned} \tag{23.132}$$

Le rotationnel s'écrit donc

$$\nabla \times F = \sum_{jkl} \frac{1}{h_j h_k}\epsilon_{jkl}\frac{\partial}{\partial q_j}(F_k h_k) e_l. \tag{23.133}$$

Devant e_1 par exemple nous avons seulement les termes $j = 2$, $k = 3$ et $j = 3$, $k = 2$. Étant donné que $\epsilon_{231} = 1$ et $\epsilon_{321} = -1$, le coefficient de e_1 sera simplement

$$\frac{1}{h_2 h_3}\left(\frac{\partial}{\partial q_2}(F_3 h_3) - \frac{\partial}{\partial q_3}(F_2 h_2)\right). \tag{23.134}$$

La formule complète devient

$$\begin{aligned}\nabla \times \sum_k F_k e_k = &\frac{1}{h_2 h_3}\left(\frac{\partial}{\partial q_2}(F_3 h_3) - \frac{\partial}{\partial q_3}(F_2 h_2)\right)\\ &+ \frac{1}{h_1 h_3}\left(\frac{\partial}{\partial q_3}(F_1 h_1) - \frac{\partial}{\partial q_1}(F_3 h_3)\right)\\ &+ \frac{1}{h_2 h_1}\left(\frac{\partial}{\partial q_1}(F_2 h_2) - \frac{\partial}{\partial q_2}(F_1 h_1)\right).\end{aligned} \tag{23.135}$$

23.5.9.1 Coordonnées cylindriques

En utilisant $h_r = 1$, $h_\theta = r$ et $h_z = 1$, nous trouvons

$$\begin{aligned}\nabla \times (F_r e_r + F_\theta e_\theta + F_z e_z) = \frac{1}{r}\left(\frac{\partial F_z}{\partial \theta} - \frac{\partial (F_\theta r)}{\partial z}\right) e_r \\ + \left(\frac{\partial F_r}{\partial z} - \frac{\partial F_z}{\partial r}\right) e_\theta \\ + \left(\frac{\partial (F_\theta r)}{\partial r} - \frac{\partial F_r}{\partial \theta}\right) e_z.\end{aligned} \tag{23.136}$$

23.5.9.2 Coordonnées sphériques

En utilisant $h_\rho = 1$, $h_\theta = \rho$ et $h_\varphi = \rho \sin\theta$, nous trouvons

$$\begin{aligned}\nabla \times (F_\rho e_\rho + F_\theta e_\theta + F_\varphi e_\varphi) = \frac{1}{\rho \sin\theta}\left(\frac{\partial (F_\varphi) \sin\theta}{\partial \theta} - \frac{\partial F_\theta}{\partial \varphi}\right) e_\rho \\ + \frac{1}{\rho \sin\theta}\left(\frac{\partial F_\rho}{\partial \varphi} - \frac{\partial (F_\varphi \rho \sin\theta)}{\partial \rho}\right) e_\theta \\ + \frac{1}{\rho}\left(\frac{\partial F_\theta \rho}{\partial \rho} - \frac{\partial F_r}{\partial \theta}\right) e_\varphi.\end{aligned} \tag{23.137}$$

Note : dans le premier terme, il y a une simplification par ρ.

23.6 Les formules

23.6.1 Coordonnées polaires

Les vecteurs de base :

$$e_r = \begin{pmatrix}\cos(\theta) \\ \sin(\theta)\end{pmatrix} = \cos(\theta) e_x + \sin(\theta) e_y \tag{23.138a}$$

$$e_\theta = \begin{pmatrix}-\sin(\theta) \\ \cos(\theta)\end{pmatrix} = -\sin(\theta) e_x + \cos(\theta) e_y. \tag{23.138b}$$

Le gradient :

$$\nabla \tilde{f}(r,\theta) = \frac{\partial \tilde{f}}{\partial r}(r,\theta) e_r + \frac{1}{r}\frac{\partial \tilde{f}}{\partial \theta}(r,\theta) e_\theta. \tag{23.139}$$

La divergence :

$$\nabla \cdot F = \frac{1}{r}\left[\frac{\partial}{\partial r}(r F_r) + \frac{\partial}{\partial \theta}(F_\theta)\right]. \tag{23.140}$$

Le rotationnel :

$$\nabla \times (F_r e_r + F_\theta e_\theta) = \left(\frac{\partial (F_\theta r)}{\partial r} - \frac{\partial F_r}{\partial \theta}\right) e_z. \tag{23.141}$$

Notons que le rotationnel n'existe pas vraiment en deux dimensions. Ici nous avons vu le champ $F(r,\theta)$ comme un champ dans $\mathbb{R}^3$ ne dépendant pas de z et n'ayant pas de composante z. Le résultat est un rotationnel qui est dirigé selon l'axe z.

23.6.2 Coordonnées cylindriques

Les vecteurs de base : idem qu'en coordonnées polaires, et on ajoute e_z sans modifications.
Le gradient :

$$\nabla \tilde{f}(r,\theta,z) = \frac{\partial \tilde{f}}{\partial r}(r,\theta,z) e_r + \frac{1}{r}\frac{\partial \tilde{f}}{\partial \theta}(r,\theta,z) e_\theta + \frac{\partial \tilde{f}}{\partial z}(r,\theta,z) e_z. \tag{23.142}$$

La divergence :

$$\nabla \cdot F = \frac{1}{r}\left[\frac{\partial}{\partial r}(rF_r) + \frac{\partial}{\partial \theta}(F_\theta) + \frac{\partial}{\partial z}(rF_z)\right]. \tag{23.143}$$

Le rotationnel :

$$\begin{aligned}\nabla \times (F_r e_r + F_\theta e_\theta + F_z e_z) = \frac{1}{r}&\left(\frac{\partial F_z}{\partial \theta} - \frac{\partial (F_\theta r)}{\partial z}\right) e_r \\ &+ \left(\frac{\partial F_r}{\partial z} - \frac{\partial F_z}{\partial r}\right) e_\theta \\ &+ \left(\frac{\partial (F_\theta r)}{\partial r} - \frac{\partial F_r}{\partial \theta}\right) e_z.\end{aligned} \tag{23.144}$$

Note : les formules concernant les coordonnées polaires se réduisent de celles-ci en enlevant toutes les références à z.

23.6.3 Coordonnées sphériques

Les vecteurs de base :

$$\begin{aligned} e_r &= \begin{pmatrix}\sin(\theta)\cos(\varphi)\\ \sin(\theta)\sin(\varphi)\\ \cos(\theta)\end{pmatrix}, \quad e_\theta = \begin{pmatrix}\cos(\theta)\cos(\varphi)\\ \cos(\theta)\sin(\varphi)\\ -\sin(\theta)\end{pmatrix}, \\ e_\varphi &= \begin{pmatrix}-\sin(\varphi)\\ \cos(\varphi)\\ 0\end{pmatrix}\end{aligned} \tag{23.145}$$

Le gradient :

$$\nabla \tilde{f}(\rho, \theta, \varphi) = \frac{\partial \tilde{f}}{\partial \rho} e_\rho + \frac{1}{\rho}\frac{\partial \tilde{f}}{\partial \theta} e_\theta + \frac{1}{\rho \sin(\theta)}\frac{\partial \tilde{f}}{\partial \varphi} r_\varphi. \tag{23.146}$$

La divergence :

$$\nabla \cdot F = \frac{1}{\rho^2 \sin\theta}\left[\frac{\partial}{\partial \rho}(\rho^2 \sin\theta F_\rho) + \frac{\partial}{\partial \theta}(\rho \sin\theta F_\theta) + \frac{\partial}{\partial \varphi}(\rho F_\varphi)\right]. \tag{23.147}$$

Le rotationnel :

$$\begin{aligned}\nabla \times (F_\rho e_\rho + F_\theta e_\theta + F_\varphi e_\varphi) = \frac{1}{\rho \sin\theta}&\left(\frac{\partial (F_\varphi)\sin\theta}{\partial \theta} - \frac{\partial F_\theta}{\partial \varphi}\right) e_\rho \\ &+ \frac{1}{\rho \sin\theta}\left(\frac{\partial F_\rho}{\partial \varphi} - \frac{\partial (F_\varphi \rho \sin\theta)}{\partial \rho}\right) e_\theta \\ &+ \frac{1}{\rho}\left(\frac{\partial F_\theta \rho}{\partial \rho} - \frac{\partial F_r}{\partial \theta}\right) e_\varphi.\end{aligned} \tag{23.148}$$

Chapitre 24

Espaces de Hilbert

24.1 Espaces de Hilbert

Définition 24.1.
*Un espace **préhilbertien** est*

— *soit un espace vectoriel réel muni d'un produit scalaire*[1],

— *soit un espace vectoriel complexe muni d'un produit scalaire hermitien*[2].

*Un **espace de Hilbert** est un espace préhilbertien qui est complet*[3] *pour la norme induite par son produit (scalaire ou hermitien), c'est-à-dire tel que*

$$\|x\| = \sqrt{\langle x, x\rangle} \tag{24.1}$$

pour tout élément x.

Dans les deux cas nous considérons la topologie métrique dérivant du produit.

Dans les cas de dimension finie, les espaces vectoriels normés sont automatiquement complets par la proposition 7.298.

La différence entre un espace de Hilbert et un espace de Banach est que dans le cas d'un espace de Hilbert, nous demandons que la norme dérive d'un produit scalaire.

Proposition 24.2.
Si $\mathscr{H}$ est un espace de Hilbert réel, alors

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle. \tag{24.2}$$

Si $\mathscr{H}$ est un espace de Hilbert complexe alors

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\,\mathrm{Re}\langle x, y\rangle. \tag{24.3}$$

Dans les deux cas nous avons l'inégalité de Cauchy-Schwarz :

$$|\langle x, y\rangle| \leqslant \|x\|\|y\|. \tag{24.4}$$

Dans un espace vectoriel de dimension infinie, tous les opérateurs linéaires ne sont pas continus.

Exemple 24.3.
Soit un espace vectoriel V engendré par la base $\{e_k\}_{k\in\mathbb{N}}$ et l'application linéaire $T\colon V \to V$ donnée par

$$Te_k = ke_k. \tag{24.5}$$

1. Définition 7.186.
2. Définition 7.189.
3. Définition 7.275.

Nous allons montrer que l'image inverse de la boule unité ouverte $\mathcal{O}$ n'est pas ouverte. En effet si (ϵ_k) est une suite de réels strictement positifs tendant vers zéro, les vecteurs

$$a_k = \left(\frac{1}{k} + \epsilon_k\right) e_k \tag{24.6}$$

sont hors de $T^{-1}\mathcal{O}$ parce que

$$Ta_k = (1 + k\epsilon_k)e_k. \tag{24.7}$$

Mais la suite (a_k) converge vers 0 qui fait partie de $T^{-1}\mathcal{O}$. Donc le complémentaire de $T^{-1}\mathcal{O}$ n'est pas fermé, ce qui prouve que $T^{-1}\mathcal{O}$ n'est pas ouvert. △

Proposition 24.4 ([1]).
Soit un espace de Hilbert $\mathscr{H}$ et un espace vectoriel normé $(E, \|.\|)$ tel qu'il existe une bijection linéaire isométrique $\Phi\colon \mathscr{H} \to E$. Alors en posant

$$\langle \alpha, \beta \rangle_E = \langle \Phi^{-1}(\alpha), \Phi^{-1}(\beta) \rangle_{\mathscr{H}}, \tag{24.8}$$

l'espace E devient un espace de Hilbert.

Démonstration. Le fait que $\langle ., . \rangle_E$ soit un produit scalaire (hermitien) lorsque $\langle ., . \rangle_{\mathscr{H}}$ l'est est une conséquence de la linéarité de Φ. Il donne la norme parce que

$$\sqrt{\langle \alpha, \alpha \rangle} = \sqrt{\langle \Phi^{-1}(\alpha), \Phi^{-1}(\alpha) \rangle} = \|\Phi^{-1}(\alpha)\| = \|\alpha\| \tag{24.9}$$

parce que Φ^{-1} est une isométrie.

Pour voir que E est complet, soit une suite (α_k) de τ-Cauchy dans E. Nous devons prouver qu'elle converge. Nous posons $x_k = \Phi^{-1}(\alpha_k)$.

Commençons par voir que (x_k) est τ-Cauchy dans $\mathscr{H}$. Soit un voisinage V de 0 dans $\mathscr{H}$. Nous posons $U = \Phi(V)$. Vu que (α_k) est τ-Cauchy, il existe $N > 0$ tel que $\alpha_k - \alpha_l \in U$ dès que $k, l > N$. Du coup, si $k, l > N$ nous avons aussi $\Phi(\alpha_k - \alpha_l) \in \Phi(U) = V$.

Vu que $\mathscr{H}$ est complet, il existe $x \in \mathscr{H}$ tel que $x_k \xrightarrow{\mathscr{H}} x$. Montrons que $\alpha_k \xrightarrow{\mathscr{H}'} \Phi(x)$.

Soit $\epsilon > 0$ et $N > 0$ tel que $\|x_k - x\| < \epsilon$ dès que $k \geqslant N$. Alors nous avons aussi

$$\|\alpha_k - \Phi(x)\| = \|\Phi(x_k) - \Phi(x)\| = \|\Phi(x_k - x)\| = \|x_k - x\| < \epsilon \tag{24.10}$$

où nous avons utilisé le fait que Φ était linéaire et une isométrie. □

24.1.1 Sous-espace vectoriel fermé ? ? ?

Nous verrons que beaucoup de résultats demandent un sous-espace vectoriel *fermé*. Une question légitime est : est-ce qu'il existe des sous-espaces vectoriels qui ne soient pas fermés ? En dimension finie, tous les sous-espaces vectoriels sont fermés, mais cela n'est pas vrai en dimension infinie.

Soit en effet une partie libre infinie $A = \{v_i\}_{i\in\mathbb{N}}$ dans un espace de Hilbert $\mathscr{H}$. L'ensemble $\mathrm{Span}(A)$ des combinaisons linéaires d'éléments de A est un sous-espace vectoriel de $\mathscr{H}$, mais il n'est pas fermé.

En effet, supposons pour simplifier les notations que $\|v_i\| = 1$ pour tout i. Alors nous considérons la combinaison

$$a = \sum_{k=1}^{\infty} \alpha_k v_k \tag{24.11}$$

où les α_k sont suffisamment décroissants pour assurer les convergences [4]. Le vecteur a n'est pas dans $\mathrm{Span}(A)$, mais la suite $a_n = \sum_{k=1}^{n} \alpha_k v_k$ est dans $\mathrm{Span}(A)$ et converge vers a : $a_n \xrightarrow{\mathscr{H}} a$. En effet,

$$\|a_n - a\| = \|\sum_{k=n+1}^{\infty} \alpha_k v_k\| \leqslant \sum_{k=n+1}^{\infty} |\alpha_k| \xrightarrow{n\to\infty} 0. \tag{24.12}$$

4. Par exemple $\alpha_k = 1/k^2$ si les v_k sont orthonormaux.

Vous voulez des détails sur la dernière limite ? Vu que la somme $\sum_k |\alpha_k|$ converge, la suite des sommes de queues de suites [5] converge vers zéro :

$$\lim_{n\to\infty} \sum_{k=n+1}^{\infty} |\alpha_k| = 0. \tag{24.13}$$

24.2 Théorème de la projection

Théorème 24.5 (Projection sur partie fermée convexe[424, 425])**.**
Soit $\mathscr{H}$ un espace de Hilbert, $x \in \mathscr{H}$, et C un sous-ensemble fermé convexe de $\mathscr{H}$.

(1) Les deux conditions suivantes sur $y \in \mathscr{H}$ sont équivalentes :

(1a) $\|x - y\| = \inf\{\|x - z\| \text{ tel que } z \in C\}$,

(1b) pour tout $z \in C$, $\mathrm{Re}\langle x - y, z - y\rangle \leqslant 0$.

(2) Il existe un unique $y \in \mathscr{H}$, noté $y = \mathtt{proj}_C(x)$ vérifiant ces conditions.

Démonstration. Nous commençons par prouver l'existence et l'unicité d'un élément dans C vérifiant la première condition. Ensuite nous verrons l'équivalence.

Nous nommons d l'infimum en question de la première condition.

Existence Soit (y_n) une suite dans C telle que

$$\lim_{n\to\infty} \|x - y_n\| = \inf\{\|x - y\| \text{ tel que } z \in C\} = d. \tag{24.14}$$

Nous allons montrer que cette suite peut être choisie de Cauchy. Elle convergera donc dans $\mathscr{H}$ parce que ce dernier est complet. Mais C étant supposé fermé dans $\mathscr{H}$, la limite appartiendra à C. Soient $r, s \in \mathbb{N}$. D'abord nous avons

$$\|y_r - y_s\|^2 = \langle y_r - y_s + x - x, y_r - y_s + x - x\rangle \tag{24.15a}$$
$$= \|y_r - x\|^2 + \|y_s - x\|^2 - 2\langle y_r - x, y_s - x\rangle. \tag{24.15b}$$

Ensuite,

$$4\left\|\frac{y_r + y_s}{2} - x\right\|^2 = \langle y_r + y_s - 2x, y_r - y_s - 2x\rangle \tag{24.16a}$$
$$= \|y_r - x\|^2 + \|y_s - x\|^2 + 2\langle y_r - x, y_s - x\rangle. \tag{24.16b}$$

Si nous égalisons les valeurs de $2\langle y_r - x, y_s - x\rangle$ nous trouvons

$$\|y_r - y_s\|^2 = -4\left\|\frac{y_r + y_s}{2} - x\right\|^2 + 2\|y_r - x\|^2 + 2\|y_s - x\|^2. \tag{24.17}$$

La distance infimum étant d, nous pouvons choisir y_n de telle façon à avoir

$$\|y_n - x\| \leqslant d + \frac{1}{n}. \tag{24.18}$$

D'autre part étant donné que C est convexe, $(y_r + y_s)/2$ est dans C et nous avons

$$\left\|\frac{y_r + y_s}{2} - x\right\| \leqslant d. \tag{24.19}$$

En mettant ces majorations dans (24.17) nous trouvons

$$\|y_r - y_s\|^2 \leqslant -4d + 2\left(d + \frac{1}{r}\right) + 2\left(d + \frac{1}{s}\right) = \frac{1}{r} + \frac{1}{s}. \tag{24.20}$$

5. On se comprend hein.

La suite (y_n) est donc de Cauchy et la limite est un élément de C. Prouvons que cet élément y réalise l'infimum. Pour cela nous avons les inégalités

$$d \leqslant \|x - y\| \leqslant \|x - y_n\| + \|y_n - y\|. \tag{24.21}$$

En prenant le limite $n \to \infty$ nous trouvons

$$d \leqslant \|x - y\| \leqslant d. \tag{24.22}$$

Unicité Même preuve que pour le théorème en dimension finie 11.232.

(1a)⇒ (1b) Même preuve que pour le théorème en dimension finie 11.232.

(1b)⇒ (1a) Même preuve que pour le théorème en dimension finie 11.232.

□

Proposition 24.6.
Soit C une partie convexe et fermée de l'espace de Hilbert $\mathscr{H}$. Alors pour tout $x_1, x_2 \in \mathscr{H}$ nous avons

$$\| \mathtt{proj}_C(x_1) - \mathtt{proj}_C(x_2)\| \leqslant \|x_1 - x_2\|. \tag{24.23}$$

En particulier la projection est une application continue.

Démonstration. Nous posons $y_1 = \mathtt{proj}_C(x_1)$ et $y_2 = \mathtt{proj}_C(x_2)$. Par la partie (1b) du théorème 24.5, nous avons, pour tout $z, z' \in C$ les inégalités

$$\mathrm{Re}\langle x_1 - y_1, z - y_1\rangle \leqslant 0 \tag{24.24a}$$
$$\mathrm{Re}\langle x_2 - y_2, z' - y_2\rangle \leqslant 0. \tag{24.24b}$$

En prenant $z = y_2$ et $z' = y_1$ et en sommant nous trouvons

$$\mathrm{Re}\langle (x_1 - y_1) + (y_2 - x_2), y_2 - y_1\rangle \leqslant 0. \tag{24.25}$$

Nous pouvons maintenant calculer

$$\begin{aligned}
\|y_1 - y_2\|^2 &= \mathrm{Re}\,\|y_1 - y_2\| \\
&= \mathrm{Re}\langle y_1 - y_2, (y_1 - x_1) + x_1 - x_2 + (x_2 - y_2)\rangle \\
&= \mathrm{Re}\langle y_1 - y_2, x_1 - x_2\rangle + \mathrm{Re}\langle y_2 - y_1, (x_1 - y_1) + (y_2 - x_2)\rangle \\
&\leqslant \langle y_1 - y_2, x_1 - x_2\rangle \\
&\leqslant \|\langle y_1 - y_2, x_1 - x_2\rangle\| \\
&=\leqslant \|y_1 - y_2\|\|x_1 - x_2\|.
\end{aligned} \tag{24.26}$$

En simplifiant par $\|y_1 - y_2\|$ [6] nous trouvons le résultat

$$\|y_1 - y_2\| \leqslant \|x_1 - x_2\|. \tag{24.27}$$

□

Théorème 24.7 (Projection orthogonale)**.**
Soit $\mathscr{H}$ un espace de Hilbert et K, un sous-espace vectoriel fermé non réduit à $\{0\}$ et $x \in \mathscr{H}$. L'élément $y = \mathtt{proj}_K(x)$ est l'unique élément de K tel que

$$x - y \in K^\perp. \tag{24.28}$$

De plus l'application $x \mapsto \mathtt{proj}_K(x)$ est linéaire, continue et de norme 1.

L'élément y ainsi définit est la **projection orthogonale** de x sur K et sera noté $\mathtt{proj}_K(f)$.

6. Si c'est nul, alors la preuve est évidente.

Démonstration. La continuité de la projection est donné par la proposition 24.6.

Soit z un élément de K tel que $\langle z-x, a\rangle = 0$ pour tout $a \in K$. Nous avons

$$\|x-a\|^2 = \|z-x\|^2 + \|a-z\|^2 + 2\underbrace{\langle z-x, a-z\rangle}_{=0} \tag{24.29a}$$

$$\geqslant \|z-x\|^2. \tag{24.29b}$$

Le produit scalaire est nul parce que $a - z \in K$. La distance $\|z-x\|$ est donc bien la plus petite distance entre x et les éléments de K.

Dans l'autre sens, nous supposons que $y \in K$ minimise la distance à x dans K. Par hypothèse pour tout a et pour tout $\lambda \in \mathbb{R}$, la différence

$$\|(y+\lambda a) - x\|^2 - \|y-x\|^2 \tag{24.30}$$

est positive. En développant les produits scalaires nous trouvons la conditions suivante

$$\lambda^2\|a\|^2 + 2\lambda\langle a, y-x\rangle \geqslant 0 \tag{24.31}$$

qui doit être vraie pour tout $\lambda \in \mathbb{R}$. En tant que polynôme du second degré en λ, cela n'aura pas deux racines réelles distinctes uniquement si $\langle a, y-x\rangle = 0$.

Nous montrons maintenant la linéarité de la projection orthogonale. Soient $x_1, x_2 \in \mathscr{H}$. L'élément $y = \mathtt{proj}_K x_1 + \mathtt{proj}_K x_2$ satisfait à la condition d'orthogonalité : pour tout $z \in K$,

$$\langle x_1 + x_2 - \mathtt{proj}_K x_1 - \mathtt{proj}_K x_2, z\rangle = \langle x_1 - \mathtt{proj}_K x_1, z\rangle + \langle x_2 - \mathtt{proj}_K x_2, z\rangle = 0. \tag{24.32}$$

Étant donné que K est un sous-espace vectoriel, la condition de minimalité est automatiquement vérifiée (seconde partie du théorème 24.5).

En ce qui concerne la norme opérateur de $\mathtt{proj}_K$, la décomposition de $x \in \mathscr{H}$ en composantes dans K et $K^\perp$ est

$$x = x + (x - \mathtt{proj}_K x). \tag{24.33}$$

Étant deux parties orthogonales nous avons

$$\|\mathtt{proj}_K x\|^2 = \|x\|^2 - \|x - \mathtt{proj}_K x\|^2. \tag{24.34}$$

En prenant $\|x\| = 1$ nous trouvons $\|\mathtt{proj}_K x\|^2 \leqslant 1$ et par conséquent $\|\mathtt{proj}_K\| \leqslant 1$. Mais d'autre part en prenant $x \in K$ nous avons automatiquement $\|\mathtt{proj}_K\| \geqslant 1$. □

Proposition 24.8.
Soit $\mathscr{H} = L^2(\Omega, \mathcal{A}, \mu)$, $\mathcal{F}$ une sous tribu de $\mathcal{A}$ et K l'ensemble de fonctions $\mathcal{F}$-mesurables dans $L^2(\Omega, \mathcal{A}, \mu)$. Si $f \in L^2(\Omega, \mathcal{A}, \mu)$ est positive, alors $\mathtt{proj}_K f$ est positive (presque partout).

Démonstration. L'ensemble $A = \{\mathtt{proj}_K f < 0\}$ est dans $\mathcal{F}$. En effet

$$A = (\mathtt{proj}_K f)^{-1}\big(]-\infty, 0[\big) \tag{24.35}$$

alors que, par construction, $\mathtt{proj}_K f$ est $\mathcal{F}$-mesurable. La fonction indicatrice $\mathbb{1}_A$ est alors $\mathcal{F}$-mesurable (c'est-à-dire $1_A \in K$) et nous avons

$$0 \leqslant \int_\Omega f\mathbb{1}_A = \int_\Omega \mathtt{proj}_K f\mathbb{1}_A \leqslant 0. \tag{24.36}$$

Étant donné que nous avons supposé $f \geqslant 0$ nous avons alors $\mu(A) = 0$. D'où le fait que $\mathtt{proj}_K f$ est presque partout positive. □

24.3 Systèmes orthogonaux et bases

Dans cette partie nous noterons $\mathbb{K}$ le corps de base de l'espace $\mathscr{H}$. Seuls deux cas sont envisagés : $\mathbb{K} = \mathbb{C}$ ou $\mathbb{K} = \mathbb{R}$.

Pour chaque $x \in \mathscr{H}$ nous considérons l'application

$$\begin{aligned} \Phi_y \colon \mathscr{H} &\to \mathbb{K} \\ x &\mapsto \langle x, y \rangle. \end{aligned} \tag{24.37}$$

Ce sont des applications continues.

24.3.1 Orthogonal d'une partie

Définition 24.9.
*Soit une partie A de l'espace de Hilbert $\mathscr{H}$. L'**orthogonal** de A est l'ensemble*

$$A^\perp = \{v \in \mathscr{H} \text{ tel que } \langle v, x \rangle = 0 \forall x \in A\}. \tag{24.38}$$

Proposition 24.10.
Si $\mathscr{H}$ est une préhilbert et si $A \subset \mathscr{H}$, alors l'ensemble $A^\perp$ est un sous-espace fermé de $\mathscr{H}$.

Démonstration. L'application Φ_x définie en (24.37) est continue pour chaque $x \in \mathscr{H}$ et par conséquent l'ensemble $\ker \Phi_x = \Phi_x^{-1}(\{0\})$ est fermé. L'ensemble

$$A^\perp = \bigcap_{x \in A} \ker \Phi_x \tag{24.39}$$

est donc fermé. □

Proposition 24.11 (wikipédia)**.**
Soit V un espace vectoriel normé et une décomposition en somme directe $V = F \oplus G$. Alors les trois conditions suivantes sont équivalentes.

(1) L'isomorphisme naturel $F \times G \to v$ est un homéomorphisme.

(2) F et G sont fermés et la restriction à G de la projection $V \to V/F$ est un homéomorphisme.

(3) F et G sont fermés et la projection de E sur F est continue.

Lorsqu'une décomposition en somme directe vérifie la proposition 24.11, nous disons que la décomposition est **topologique**.

Théorème 24.12.
Soit $\mathscr{H}$ un espace de Hilbert et F un sous-espace fermé de $\mathscr{H}$. Alors

(1) Nous avons la décomposition

$$\mathscr{H} = F \oplus F^\perp. \tag{24.40}$$

(2) La projection sur F par rapport à la somme directe (24.40) est la projection $\mathtt{proj}_F$ du théorème de projection.

(3) La décomposition (24.40) est topologique.

Proposition 24.13.
Si F est un sous-espace de l'espace de Hilbert $\mathscr{H}$ alors on a $F^{\perp\perp} = \bar{F}$.

Démonstration. Nous savons par la proposition 24.10 que $F^\perp$ est fermé, par conséquent le théorème 24.12 donne la somme directe

$$\mathscr{H} = F^\perp \oplus F^{\perp\perp}. \tag{24.41}$$

Mais $\bar{F}$ étant également fermé nous avons la somme directe

$$\mathscr{H} = \bar{F} \oplus (\bar{F})^{\perp}. \tag{24.42}$$

Montrons que $(\bar{F})^{\perp} = F^{\perp}$. En effet si $x \in F^{\perp}$ et si $y \in \bar{F}$, alors il existe une suite y_n dans F qui converge vers y. Pour chaque n nous avons $\langle x, y_n \rangle = 0$ et dons $\langle x, y \rangle = 0$ par continuité du produit scalaire.

Nous avons donc

$$\mathscr{H} = F^{\perp} \oplus F^{\perp\perp} = \bar{F} \oplus F^{\perp}. \tag{24.43}$$

Mais $\bar{F} \subset F^{\perp\perp}$ (prendre une suite). Les espaces $\bar{F}$ et $F^{\perp\perp}$ étant tous deux des supplémentaires de $F^{\perp}$, nous déduisons qu'ils doivent être égaux. □

Proposition 24.14.
Soit $\mathscr{H}$ un espace de Hilbert et F un sous-espace vectoriel de $\mathscr{H}$. Alors F est dense si et seulement si $F^{\perp} = \{0\}$.

Démonstration. Nous savons que

$$\mathscr{H} = F^{\perp} \oplus \bar{F}. \tag{24.44}$$

Donc nous avons $H = \bar{F}$ si et seulement si $F^{\perp} = \{0\}$. □

Pour vérifier si un sous-espace vectoriel d'un espace de Hilbert est dense, il suffit donc de montrer que son orthogonal est réduit à zéro.

24.3.2 Dual, théorème de représentation de Riesz

Lemme 24.15.
L'application linéaire

$$\begin{aligned} \Phi \colon \mathscr{H} &\to \mathcal{L}(\mathscr{H}, \mathbb{K}) \\ y &\mapsto \Phi_y \end{aligned} \tag{24.45}$$

est une isométrie : nous avons $\|\Phi_y\| = \|y\|$.

De plus pour chaque y, l'application Φ_y est continue.

Démonstration. En utilisant la définition de la norme opérateur et l'inégalité de Cauchy-Schwarz,

$$\|\Phi_y\| = \sup_{\|x\|=1} \|\Phi_y(x)\| = \sup |\langle x, y \rangle| \leqslant \sup \|x\|\|y\| = \|y\|. \tag{24.46}$$

Par conséquent $\|\Phi_y\| \leqslant \|y\|$. Mais d'autre part le fait que $\Phi_y(y) = \|y\|^2$ montre que $\|\Phi_y\| \geqslant \|y\|$.

En ce qui concerne la continuité de Φ_y, elle est garantie par le fait que c'est une application linéaire bornée via la proposition 10.18. □

Définition 24.16.
*Le **dual** de l'espace de Hilbert $\mathscr{H}$ est l'ensemble*

$$\mathscr{H}' = \{f \colon \mathscr{H} \to \mathbb{K} \text{ linéaire et continue}\}. \tag{24.47}$$

Notons que dans le contexte des espaces de Hilbert nous demandons la continuité des éléments du dual parce qu'elle n'est pas automatique par la linéarité dans les cas de dimension infinie. En principe nous devrions préciser dual *topologique*, mais nous ne le ferons pas systématiquement lorsque le contexte parle clairement de topologie (ce qui est le cas lorsqu'on parle d'espaces de Hilbert). De temps en temps le dual *algébrique* d'un espace est noté E^* ; dans ce cas la continuité n'est pas demandée.

Théorème 24.17 (Théorème de représentation de Riesz, thème 34)**.**
Soit un espace de Hilbert $\mathscr{H}$ sur le corps $\mathbb{K}$ ($\mathbb{R}$ ou $\mathbb{C}$). L'application

$$\begin{aligned}\Phi\colon \mathscr{H} &\to \mathscr{H}'\\ y &\mapsto \Phi_y\end{aligned} \tag{24.48}$$

est une bijection isométrique.

Démonstration. Nous savons du lemme 24.15 que Φ est une isométrie. Nous devons seulement montrer que Φ est surjective. L'application nulle est dans l'image de Φ. Soit $f \in \mathscr{H}'$ non nulle. Par continuité nous savons que $F = \ker(f)$ est fermé, donc

$$\mathscr{H} = \ker(f) \oplus (\ker f)^{\perp} \tag{24.49}$$

par le théorème 24.12.

Une base adaptée Nous considérons une base de $\mathscr{H}$ adaptée à la décomposition 24.49, c'est-à-dire

— $(u_s)_{s\in S}$ une base de $\ker f$,

— $(v_t)_{t\in T}$ une base de $(\ker f)^{\perp}$.

$(\ker f)^{\perp} \cap \ker(f) = \{0\}$ Si $v \in (\ker f)^{\perp}$ alors pour tout $w \in \ker(f)$ nous avons $\langle v, w\rangle = 0$. En particulier si v lui-même est dans $\ker(f)$ alors $\langle v, v\rangle = 0$ et $v = 0$.

Une base encore plus adaptée Si v est non nul dans $\in (\ker f)^{\perp}$, alors $f(v) \neq 0$ (sinon il serait dans l'intersection entre $\ker(f)$ et $(\ker f)^{\perp}$). Mais $f(v) \neq 0$ dans $\mathbb{K}$ implique que $f(\mathbb{K}v) = \mathbb{K}$, c'est-à-dire que $\{f(v)\}$ est une base de $\text{Image}(f)$ (l'image de f est $\mathbb{K}$).

Le théorème du rang (théorème 4.39) assure alors que

$$\{v\} \cup \{u_s\}_{s\in S} \tag{24.50}$$

est une base de $\mathscr{H}$ avec $v \in (\ker f)^{\perp}$ et $u_s \in \ker f$. Nous choisissons v pour avoir $\|v\| = 1$.

Existence Nous pouvons maintenant prouver l'existence de y tel que $\Phi_y = f$. Prouvons qu'en posant $y = \overline{f(v)}v$ nous avons $\Phi_y = f$. Pour cela, un peu de calcul :

$$\Phi_y(v) = \langle v, y\rangle = f(v)\langle v, v\rangle = f(v) \tag{24.51a}$$

$$\Phi_y(u_s) = \langle u_s, y\rangle = 0. \tag{24.51b}$$

Par conséquent Φ_y et f coïncident sur une base de $\mathscr{H}$.

Unicité En ce qui concerne l'unicité, d'abord si $\Phi_y = f$ alors nous devons avoir $y \in (\ker f)^{\perp}$ et par conséquent $y = \lambda v$ pour un certain $\lambda \in \mathbb{K}$. Nous avons alors

$$\Phi_y(v) = \bar{\lambda}\langle v, v\rangle = \bar{\lambda}. \tag{24.52}$$

Pour que cela soit égal à $f(v)$, nous fixons $\lambda = \overline{f(v)}$.

□

Notons que nous avons réellement utilisé le théorème du rang pour l'unicité. Si nous ne demandions pas l'unicité, alors nous n'avions pas besoin du fait que $\dim(\ker f)^{\perp} = 1$, et nous aurons donc pu parler de formes plus générales à valeurs dans $\mathbb{K}^n$.

24.3.3 Séparabilité

Définition 24.18.
Un espace topologique est ***séparable*** *s'il possède une partie dénombrable dense.*

Définition 24.19 ([426])**.**
Si E est un espace vectoriel normé nous disons que Δ est une partie ***totale*** *de E si* $\mathrm{Span}(\Delta)$ *est dense dans E. Attention : nous rappelons que* $\mathrm{Span}(\Delta)$ *est l'ensemble des combinaisons linéaires* finies *d'éléments de Δ.*

Proposition 24.20.
Un espace vectoriel normé est séparable si et seulement s'il possède une partie totale dénombrable.

Démonstration. Si Δ est une partie dénombrable de E, alors le $\mathbb{Q}$-espace vectoriel $\mathrm{Span}_{\mathbb{Q}}\,\Delta$ est dénombrable, et sa fermeture est la même que celle de $\mathrm{Span}_{\mathbb{K}}\,\Delta$. □

Définition 24.21.
Une famille $(u_i)_{i\in I}$ d'éléments de $\mathscr{H}$ indicée par un ensemble quelconque I est un ***système orthonormé*** *si*

(1) $\|u_i\| = 1$ pour tout $i \in I$,

(2) $u_i \perp u_j$ pour tout $i \neq j$.

Notons que si $(u_n)_{n\in\mathbb{N}}$ est un système orthonormé dénombrable alors en utilisant les formules de la proposition 24.2, nous avons

$$\Big\|\sum_{k=1}^{n} \xi_k u_k\Big\|^2 = \sum_{k=1}^{n} |\xi_k|^2. \tag{24.53}$$

Exemple 24.22.
Dans l'ensemble $L^2(a,b)$ avec $b-a=L$ l'ensemble des fonctions

$$e_n(t) = \frac{1}{\sqrt{b-a}}\exp(\frac{2\pi}{L}int) \tag{24.54}$$

avec $n \in \mathbb{Z}$ forme un système orthonormé. En effet

$$\langle e_n, e_n\rangle = \frac{1}{b-a}\int_a^b \exp(\frac{2\pi}{L}int)\exp(-\frac{2\pi}{L}int) = 1 \tag{24.55}$$

et

$$\langle e_n, e_m\rangle = \frac{1}{b-a}\int_a^b e^{\frac{2\pi}{L}(n-m)t}dt = 0. \tag{24.56}$$

Dans cette intégrale nous utilisons le fait que $b = a + (b-a)$ pour simplifier les expressions en cours de calcul.

La famille (24.54) est le **système trigonométrique** de $L^2(a,b)$. On en parle aussi dans [388]. △

Proposition 24.23.
Une partie orthonormée est libre.

Démonstration. Soit $(u_i)_{i\in I}$ une famille orthonormée de $\mathscr{H}$ et une combinaison linéaire *finie* nulle :

$$\sum_{k=1}^{n} a_k u_k = 0. \tag{24.57}$$

Nous développons la somme en utilisant les formules de la proposition 24.2 :

$$0 = \|\sum_k a_k u_k\|^2 = \sum_k |a_k|^2\|u_k\|^2 + 2\sum_{k<l}\langle a_k u_k, a_l u_l\rangle. \tag{24.58}$$

La famille étant orthonormée, les choses se simplifient en

$$\sum_k |a_k|^2 = 0, \tag{24.59}$$

ce qui signifie que $a_k = 0$ pour tout k. □

Avant de continuer nous devons définir comment nous calculons des sommes sur des ensembles quelconques. Si I est un ensemble et si pour chaque $i \in I$ nous avons un nombre réel positif a_i, alors nous définissons

$$\sum_{i\in I} a_i = \sup_{\substack{J\subset I\\ J \text{ fini}}} \sum_{j\in J} a_j. \tag{24.60}$$

Cela est discuté dans la section 10.5.

Proposition 24.24 (Inégalités de Bessel)**.**
Soit $\mathscr{H}$ un préhilbert. Si $(u_i)_{i\in I}$ est un système orthonormé et si $x \in \mathscr{H}$, alors

$$\sum_{i\in I} \big|\langle x, u_i\rangle\big|^2 \leqslant \|x\|^2. \tag{24.61}$$

Démonstration. Les éléments de la somme étant des réels positifs, la notion de somme à utiliser est celle de la définition 10.91.

Posons $c_i(x) = \langle x, u_i\rangle$. Pour toute partie finie $J \subset I$ nous avons

$$0 \leqslant \Big\|x - \sum_{j\in J} c_j(x)u_j\Big\|^2 = \|x\|^2 - 2\operatorname{Re}\sum_j \langle x, c_j(x)u_j\rangle + \sum_j |c_j(x)|^2. \tag{24.62}$$

Mais en tenant compte du fait que

$$\langle x, c_j(x)u_j\rangle = \overline{c_j(x)}\langle x, u_j\rangle = |c_j(x)|^2, \tag{24.63}$$

nous restons avec

$$\|x - \sum_{j\in J} c_j(x)u_j\| = \|x\|^2 - \sum_{j\in J} |c_j(x)|^2. \tag{24.64}$$

Finalement,

$$\sum_{j\in J} |c_j(x)|^2 \leqslant \|x\|^2. \tag{24.65}$$

Ayant cette inégalité pour toute partie finie de I, nous l'avons encore pour le supremum. □

Proposition 24.25.
Soit $\mathscr{H}$ un préhilbert et une famille orthonormé $(u_i)_{i\in I}$. Si

$$x = \sum_{i\in I} \xi_i u_i \tag{24.66}$$

alors $\xi_i = \langle x, u_i\rangle$.

Démonstration. Nous appliquons l'application Φ_{u_k} du théorème de représentation de Riesz[7] à l'équation (24.66).

Si I est dénombrable, alors permuter Φ_{u_k} avec la somme consiste à invoquer la continuité, et permuter la limite des sommes partielles avec Φ_{u_k} (l'application Φ_{u_k} est continue parce qu'isométrique).

Sinon, il faut utiliser la proposition 10.102. Il faut donc montrer que la famille $\Phi_{u_k}\big(\xi_i u_i\big)$ est sommable. Cela est fort vrai parce que cette famille ne contient en réalité qu'en seul élément non nul, celui avec $i = k$, qui vaut ξ_k. Au final nous avons :

$$\langle x, u_k\rangle = \sum_{i\in I} \Phi_{u_k}(\xi_i u_i) = \xi_k. \tag{24.67}$$

□

7. Théorème 24.17.

24.3.4 Base hilbertienne

Définition 24.26.
*Une **base orthonormée** est une famille dénombrable orthonormé et totale*[8]*. Cela sera souvent aussi appelé une **base hilbertienne**.*

24.27.
Cette définition demande quelques remarques.

(1) La notion de base hilbertienne pas la même notion de base qu'en algèbre. En effet pour avoir une base algébrique d'un espace vectoriel, nous demandons que les éléments soient des combinaisons linéaires *finies* des éléments de la base, tandis qu'ici en demandant que la partie soit totale nous demandons simplement que les combinaisons linéaires finies soient denses.

(2) Nous allons voir qu'un espace de Hilbert est généré par les sommes *infinies* de vecteurs d'une base hilbertienne avec des coefficients qui forment une suite dans ℓ^2.

(3) Nous ne demandons pas que la famille soit libre ? La belle affaire. Une famille orthogonale est toujours libre, proposition 24.23.

Lemme 24.28 ([427])**.**
Si $\mathcal{E}$ est une base hilbertienne[9] *de l'espace de Hilbert $\mathscr{H}$ et si $x \in \mathscr{H}$ alors l'ensemble*

$$\{e \in \mathcal{E} \text{ tel que } \langle x, e\rangle \neq 0\} \tag{24.68}$$

est au plus dénombrable.

Démonstration. Notons qu'ici, $\mathscr{H}$ n'est pas supposé séparable. Nous savons par l'inégalité de Bessel (24.61) que

$$\sum_{e\in\mathcal{E}} |\langle x, e\rangle|^2 \leqslant \|x\|^2. \tag{24.69}$$

Donc si $\epsilon > 0$ est donné, l'ensemble $\{e \in \mathcal{E} \text{ tel que } |\langle x, e\rangle| > \epsilon\}$ est fini. Or

$$\{e \in \mathcal{E} \text{ tel que } \langle x, e\rangle \neq 0\} = \{e \in \mathcal{E} \text{ tel que } |\langle x, e\rangle| > 0\} = \bigcup_{n=1}^{\infty} \{e \in \mathcal{E} \text{ tel que } |\langle x, e\rangle| > \frac{1}{n}\}. \tag{24.70}$$

Bref, cet ensemble est une union dénombrable d'ensembles finis. Il est donc dénombrable. □

Remarque 24.29.
Le lemme 24.28 ne signifie pas que la base $\mathcal{E}$ doive être dénombrable. Il signifie seulement que pour chaque x séparément, seule une partie dénombrable de $\mathcal{E}$ est nécessaire.

Corolaire 24.30 ([427])**.**
Si un espace de Hilbert possède une base hilbertienne dénombrable, alors toutes ses bases hilbertiennes sont dénombrables.

Démonstration. Soient $\mathcal{E}$ et $\mathcal{F}$ des bases hilbertiennes de l'espace de Hilbert $\mathscr{H}$, en supposant que $\mathcal{E}$ soit dénombrable. Pour chaque $f \in \mathcal{F}$ nous avons $\langle f, e\rangle \neq 0$ pour au moins un $e \in \mathcal{E}$, sinon en vertu de la décomposition (24.66) de f dans la base $\mathcal{E}$, nous aurions $f = 0$. Nous avons donc

$$\mathcal{F} \subset \bigcup_{e\in\mathcal{E}} \{f \in \mathcal{F} \text{ tel que } \langle f, e\rangle \neq 0\}, \tag{24.71}$$

alors que le lemme 24.28 indique que chacun des ensembles de l'union est au plus dénombrable. La partie $\mathcal{F}$ est donc une union dénombrable d'ensemble dénombrables. Elle est dénombrable. □

8. Définition 24.19
9. Définition 24.26.

Remarque 24.31.
En travaillant un peu plus sur la notion de cardinalité, le corolaire 24.30 indique que toutes les bases hilbertiennes ont même cardinalité. En effet en laissant tomber l'hypothèse de dénombrabilité sur $\mathcal{E}$, l'inclusion (24.71) donne que $\mathcal{F}$ est une union de $\mathrm{Card}(\mathcal{E})$ ensembles dénombrables et est alors de cardinalité $\mathrm{Card}(\mathcal{E})$.

Définition 24.32.
Une partie orthonormale B est ***maximale*** *si le seul $x \in \mathscr{H}$ vérifiant $\langle x, b\rangle = 0$ pour tout $b \in B$ est $x = 0$.*

Lemme 24.33 ([1, 428])**.**
Tout espace de Hilbert possède une partie orthonormale maximale.

Démonstration. Soit $\mathscr{H}$ un espace de Hilbert et $\mathcal{O}$, l'ensemble de parties orthonormales de $\mathscr{H}$, ordonné [10] par l'inclusion. Cet ensemble est inductif. En effet soit une partie totalement ordonnée $\{A_i\}_{i\in I}$ de $\mathcal{O}$: chaque A_i est une partie orthonormale de $\mathscr{H}$, et de plus pour $i, j \in I$ nous avons soit $A_i \subset A_j$ soit $A_j \subset A_i$. Nous pouvons considérer

$$A = \bigcup_{i\in I} A_i. \tag{24.72}$$

Cela est encore une partie orthonormé de $\mathscr{H}$ parce que si $x, y \in A$, alors il existe $i, j \in I$ tels que $x \in A_i$ et $y \in A_j$. Vu que $\{A_i\}_{i\in I}$ est totalement ordonné nous supposons pour fixer les idées que $A_i \subset A_j$. Alors x et y sont dans A_j qui est une partie orthonormée ; nous en déduisons que $\langle x, y\rangle = 0$ et donc que A est une partie orthonormée. C'est-à-dire : $A \in \mathcal{O}$. Par ailleurs, A est un majorant de $\{A_i\}_{i\in I}$ pour l'inclusion parce que $A_i \subset A$ pour tout i.

Nous avons prouvé que $\mathcal{O}$ est un ensemble inductif. Le lemme de Zorn 1.15 nous dit alors que $\mathcal{O}$ possède un maximum. Ce maximum est une partie orthonormale inclue dans aucune autre partie orthonormale. C'est-à-dire que ce maximum est une partie orthonormale maximale au sens de la définition 24.32. □

Proposition 24.34 ([428])**.**
Si un espace de Hilbert possède une partie orthonormale maximale dénombrable, alors toutes les parties orthonormales sont dénombrables (ou finies).

Démonstration. Soit B une partie orthonormale maximale, et A une partie orthonormale. Pour chaque $b \in B$ nous notons

$$A(b) = \{a \in A \text{ tel que } \langle a, b\rangle \neq 0\} \cup \{0\}. \tag{24.73}$$

Un élément de A qui ne serait dans aucun des $A(b)$ serait perpendiculaire à tous les éléments de B, et serait donc l'élément nul. Mais l'élément nul est dans tous les $A(b)$; donc nous avons

$$A \subset \bigcup_{b\in B} A(b). \tag{24.74}$$

Or par l'inégalité de Bessel, l'ensemble $A(b)$ est dénombrable. Par conséquent A est inclus dans une union dénombrable d'ensembles dénombrables ; A est dénombrable. □

Le fait que tout espace de Hilbert possède une base hilbertienne est vrai. Nous allons démontrer ce résultat d'abord pour les espaces séparables et ensuite, indépendamment, en général. Si vous vous la sentez de maîtriser la proposition 24.36, vous pouvez sauter la 24.35.

Proposition 24.35.
Tout espace de Hilbert séparable possède une base hilbertienne.

10. Définition 1.7.

Démonstration. Vu que nous supposons avoir un espace de Hilbert séparable, il possède une partie totale dénombrable par la proposition 24.20. Soit $(v_n)_{n\in\mathbb{N}}$ une telle partie. Quitte à supprimer les v_i qui sont combinaisons linéaires des précédents, nous pouvons supposer que cette partie est libre. Nous considérons l'espace vectoriel

$$F_n = \mathrm{Span}\{v_1, \dots, v_n\}. \tag{24.75}$$

Sur F_n nous pouvons appliquer un procédé de Gram-Schmidt pour construire une base orthonormée $\{u_1, \dots, u_n\}$ de F_n au sens usuel. En considérant F_{n+1} et en recommençant, les vecteurs $u_1, \dots, u_n$ ne changent pas, mais nous obtenons un vecteur u_{n+1}.

Nous construisons ainsi une suite (u_n) qui est alors orthonormée au sens des espaces de Hilbert. Nous devons encore prouver qu'il s'agit d'un ensemble total. Cela est simplement dû au fait que tout élément de $\mathrm{Span}\{v_n\}$ est contenu dans $\mathrm{Span}\{u_n\}$ parce que Span ne considère que des combinaisons linéaires finies. □

Proposition 24.36 ([428]).
À propos de parties orthonormales maximales.

(1) *Une partie d'un espace de Hilbert est orthonormale maximale si et seulement si elle est une base hilbertienne.*

(2) *Tout espace de Hilbert possède une base hilbertienne.*

Démonstration. Soit une partie orthonormale maximale B et la fermeture de son espace engendré : $F = \overline{\mathrm{Span}(B)}$. Pour que B soit une base, nous devons démontrer que $F = \mathscr{H}$. Pour cela nous considérons $x \in \mathscr{H}$ et nous utilisons le théorème de projection orthogonale 24.7 pour mentionner le fait que

$$x - \mathtt{proj}_F(x) \perp F \tag{24.76}$$

Cela dit que $x - \mathtt{proj}_F(x)$ est un vecteur orthogonal en particulier à tous les éléments de B ; par maximalité nous avons $x - \mathtt{proj}_F(x) = 0$, c'est-à-dire $x = \mathtt{proj}_F(x)$ ou encore $x \in F$. Cela prouve que $F = \mathscr{H}$.

Note : pour être pointilleux, nous aurions dû travailler non avec $x - \mathtt{proj}_F(x)$, mais avec le vecteur normalisé à 1.

En ce qui concerne le second point, nous invoquons le lemme 24.33 pour dire que tout espace de Hilbert possède une partie orthonormale maximale. Ensuite la première partie de cette proposition nous dit que cette dernière est une base hilbertienne. □

Voici un petit résumé de ce que nous avons vu en termes de dénombrabilité et séparabilité.

Théorème 24.37.
Pour un espace de Hilbert, les choses suivantes sont équivalentes.

(1) *L'espace est séparable*[11].

(2) *L'espace possède au moins une base hilbertienne*[12] *dénombrable.*

(3) *Toutes les bases hilbertiennes sont dénombrables.*

(4) *Toute partie libre est dénombrable.*

Démonstration. Plein de résultats à citer ...

(1) implique (2) est la proposition 24.35.

(2) implique (3) est la proposition 24.30.

(3) implique (4) Le procédé de Gram-Schmidt met en bijection une partie libre avec une partie orthonormale (qui engendre le même espace, mais c'est une autre affaire). Or lorsque l'espace de Hilbert possède une base dénombrable, toutes les parties orthonormales sont dénombrables par la proposition 24.34.

11. Définition 24.18.
12. Définition 24.26.

(4) implique (1) La proposition 24.36(2) dit que tout espace de Hilbert possède des bases hilbertiennes. Une telle base est forcément une partie libre parce que toute famille orthonormale est libre (proposition 24.23), et donc dénombrable par hypothèse. À ce point nous avons montré que notre espace de Hilbert possédait une base hilbertienne dénombrable. Cela implique qu'il est séparable par la proposition 24.20.

□

Remarque 24.38.
À mon avis il doit exister un théorème de complétion de base hilbertienne disant que si on a une famille orthonormée, alors elle se prolonge en base. Utilisant cela, nous trouvons une nouvelle démonstration de la proposition 24.24 en disant que la somme sur la « partie de base » est plus petite que la somme sur la « base complète ».

24.39.
Vu que nous n'avons l'intention de ne travailler qu'avec des espaces de Hilbert séparables et que toutes leurs bases sont dénombrables, nous n'allons travailler qu'avec des bases dénombrables, et donc des systèmes orthonormés dénombrables. Nous allons conventionnellement les indicer par $\mathbb{N}$.

La proposition suivante explique que la notion de projection est compatible avec la décomposition d'un vecteur dans un système orthonormé.

Proposition 24.40.
Soit $(u_k)_{k\geqslant 1}$ un système orthonormé d'un préhilbert $\mathscr{H}$. Soient

$$x = \sum_{k=1}^{\infty} \xi_k u_k \tag{24.77}$$

et

$$F = \operatorname{Span}\{u_1, \dots, u_n\}. \tag{24.78}$$

Alors

$$\mathtt{proj}_F(x) = \sum_{k=1}^{n} \xi_k u_k. \tag{24.79}$$

Démonstration. Nous allons dans un premier temps montrer que

$$y = x - \sum_{k=1}^{n} \xi_k u_k \tag{24.80}$$

est dans $F^\perp$. Pour cela nous calculons

$$\langle x - \sum_{k=1}^{n} \xi_k u_k, u_j, \rangle = \langle \sum_{k=n+1}^{\infty} \xi_k u_k, u_j \rangle = 0 \tag{24.81}$$

où nous avons utilisé la continuité du produit scalaire pour permuter la somme (infinie) et le produit. Étant donné que $y \in F^\perp$ nous avons $\mathtt{proj}_F\, y = 0$ par le point (2) du théorème 24.12.

D'autre part $\mathtt{proj}_F\, y$ peut être calculé selon

$$\mathtt{proj}_F\, y = \mathtt{proj}_F\, x - \sum_{k=1}^{n} \xi_k\, \mathtt{proj}_F\, u_k \tag{24.82}$$

tandis que $\mathtt{proj}_F\, u_k = u_k$ lorsque $1 \leqslant k \leqslant n$. Par conséquent l'annulation de $\mathtt{proj}_F\, y$ donne

$$\mathtt{proj}_F\, x = \sum_{k=1}^{n} \xi_k u_k, \tag{24.83}$$

donc le résultat. □

Théorème 24.41 (Meilleur approximation)**.**
Soit $\{u_i\}_{i\in I}$ une famille orthonormé de l'espace de Hilbert $\mathscr{H}$. Alors pour tout $x \in \mathscr{H}$ et pour tout J fini dans I et pour toute famille de nombres complexes $(a_j)_{j\in J}$ nous avons

$$\| \sum_{j\in J} \langle x, u_j\rangle u_j - x\| \leqslant \| \sum_{j\in J} a_j u_j - x\|. \tag{24.84}$$

Ce théorème exprime le fait que les nombres $\langle x, u_i\rangle$ sont les meilleurs coefficients à mettre devant les u_i pour approximer x.

Corolaire 24.42.
Soit $\{u_i\}_{i\in I}$ une famille orthonormé de $\mathscr{H}$. Pour tout $x \in \mathscr{H}$ et pour toutes parties finies J, K de I avec $J \subset K$ nous avons

$$\| \sum_{j\in K} \langle x, u_j\rangle u_j - x\| \leqslant \| \sum_{j\in J} \langle x, u_j\rangle u_j - x\|. \tag{24.85}$$

Ce corolaire exprime le fait que plus on prend de termes de la forme $\langle , u_i\rangle u_i$, mieux c'est.

Proposition 24.43.
Soit $\mathscr{H}$ un espace de Hilbert et (u_n) un système orthonormé dans $\mathscr{H}$. Si $(\xi_n)_{n\in\mathbb{N}}$ est une suite dans ℓ^2 alors la série

$$\sum_{n=1}^{\infty} \xi_n u_n \tag{24.86}$$

converge dans $\mathscr{H}$.

Autrement dit l'application

$$\begin{aligned} S\colon \mathscr{H} &\to \ell^2 \\ x &\mapsto \big(\langle x, u_n\rangle\big)_{n\geqslant 1} \end{aligned} \tag{24.87}$$

est surjective.

Démonstration. Nous allons montrer que la série $\sum_{n=1}^{\infty} \xi_n u_n$ est de Cauchy, c'est-à-dire que la limite

$$\lim_{n\to\infty} \| \sum_{k=n}^{n+p} \xi_k u_k\| = 0 \tag{24.88}$$

est uniforme en p. Cela est un corolaire de la formule (24.53) parce que

$$\| \sum_{k=n}^{n+p} \xi_k u_k\|^2 = \sum_{k=n}^{n+p} |x_k|^2. \tag{24.89}$$

Mais si (ξ_n) est dans ℓ^2, pour tout ϵ, il existe N tel que si $n > N$ alors le membre de droite est inférieur à ϵ indépendamment de p. □

24.3.5 Décomposition dans une base hilbertienne

Étant donné une base hilbertienne de $\mathscr{H}$, nous notons

$$c_k(x) = \langle x, u_k\rangle. \tag{24.90}$$

Dans le théorème suivant (et d'ailleurs partout), les sommes sur I sont prises au sens de la définition 10.92.

Théorème 24.44 (Décomposition dans une base orthogonale)**.**
Soit $\mathscr{H}$ un espace de Hilbert séparable $\{u_i\}_{i\in I}$ une base orthonormée (I est un ensemble dénombrable quelconque).

(1) Pour tout $x \in \mathscr{H}$ nous avons

$$x = \sum_{i \in I} \langle x, u_i \rangle u_i \tag{24.91}$$

où la somme est prise au sens de la définition 10.92. En particulier, la somme converge de façon commutative.

(2) Si $\{e_i\}_{i \in I}$ est une famille orthonormée qui satisfait la décomposition (24.91) pour tout $x \in \mathscr{H}$ alors $\{e_i\}$ est une base hilbertienne.

(3) Nous avons l'identité de ***Plancherel***

$$\|x\|^2 = \sum_{i \in I} |\langle x, u_i \rangle|^2. \tag{24.92}$$

Le point (5) nous indiquera que cette égalité est en fait suffisante pour dire que nous avons une base.

(4) Nous avons l'identité de ***Parseval***

$$\langle x, y \rangle = \sum_{i \in I} \langle x, u_i \rangle \overline{\langle y, u_i \rangle}. \tag{24.93}$$

(5) Si $\{e_i\}$ est une famille de vecteurs unitaires vérifiant l'identité de Plancherel, alors c'est une base hilbertienne.

(6) Si $\{u_i\}_{i \in I}$ est une base hilbertienne, la suite $n \mapsto |\langle x, e_n \rangle|$ appartient à $\ell^2(I)$.

Démonstration. (1) Étant donné que le système $\{u_i\}_{i \in I}$ est total, nous pouvons considérer une suite de combinaisons linéaires finies des u_i qui converge vers x. Nous écrivons

$$x_n = \sum_{x \in J_n} a_{n,k} u_j \tag{24.94}$$

et $x_n \to x$ dans $\mathscr{H}$. Les ensembles J_n sont des sous-ensembles finis de I. Nous pouvons les choisir de telle sorte que $J_n \subset J_{n+1}$ et $\bigcup_{n \in \mathbb{N}} J_n = I$. Ce choix correspond à éventuellement prendre $a_{n,j} = 0$ pour toutes les valeurs de j « en trop ».

Soit $\epsilon > 0$ et N tel que $\|x_n - x\| < \epsilon$ pour tout $n \geqslant N$. Nous allons montrer que pour tout J fini tel que $J_N \subset J$ nous avons $\| \sum_{j \in J} \langle x, u_j \rangle u_j - x \| < \epsilon$. Étant donné que

$$\sum_{j \in J_N} \langle x, u_j \rangle u_j = \sum_{j \in J} a_j u_j \tag{24.95}$$

avec

$$a_j = \begin{cases} \langle x, u_j \rangle & \text{si } j \in J_N \\ 0 & \text{sinon,} \end{cases} \tag{24.96}$$

le théorème de meilleure approximation 24.41 nous enseigne que

$$\| \sum_{j \in J} \langle x, u_j \rangle u_j - x \| \leqslant \| \sum_{j \in J_N} \langle x, u_j \rangle u_j - x \| < \epsilon \tag{24.97}$$

par conséquent la somme $\sum_{i \in I} \langle x, u_i \rangle u_i$ converge vers x au sens général, et en particulier commutativement.

Les sommes étant commutatives (en particulier $x = \sum_{i \in I} \langle x, u_i \rangle u_i$), et les bases hilbertiennes étant dénombrables, nous ne perdons aucune généralité en ne considérant que des bases indexées par $\mathbb{N}$.

(2) Nous devons montrer que l'ensemble des combinaisons linéaires finies est dense dans $\mathscr{H}$. Par hypothèse, pour tout ϵ, il existe un ensemble fini J tel que

$$\| \sum_{j \in J} \langle x, u_j \rangle u_j - x \| < \epsilon. \tag{24.98}$$

Cela prouve la densité dont nous avions besoin.

(3) La norme étant une fonction continue, elle commute avec les sommes infinies, de telle sorte que l'égalité de Plancherel donne

$$\|x\|^2 = \|\sum_{i\in I}\langle x, u_i\rangle u_i\|^2. \tag{24.99}$$

Le système des u_i étant orthonormé,

$$\|x\|^2 = \|\sum_{i\in I}\langle x, u_i\rangle u_i\|^2 = \sum_{i\in I}|\langle x, u_i\rangle|^2. \tag{24.100}$$

(4) Au tour de Parseval. Nous commençons par prouver que la somme du membre de droite converge. En utilisant l'inégalité $|zz'| \leqslant |z|^2 + |z'|^2$ (valable pour $z, z' \in \mathbb{C}$) et Plancherel, nous avons

$$\begin{aligned}\sum_{i\in I}|\langle x, u_i\rangle\overline{\langle y, u_i\rangle}| &\leqslant \sum_{i\in I}|\langle x, u_i\rangle|^2 + |\langle y, u_i\rangle|^2\\ &\leqslant \|x\|^2 + \|y\|^2.\end{aligned} \tag{24.101}$$

Nous en déduisons que la famille $\langle x, u_i\rangle\overline{\langle y, u_i\rangle}$ est (commutativement) sommable en utilisant la proposition 10.101. Par ailleurs nous savons que

$$x = \sum_{i\in I} c_i(x)u_i \tag{24.102a}$$

$$y = \sum_{i\in I} c_i(y)u_i, \tag{24.102b}$$

et le produit scalaire étant une forme bilinéaire continue,

$$\langle x, y\rangle = \sum_{i\in I}\sum_{j\in I}\langle c_i(x)u_i, \overline{c_j(y)u_j}\rangle \tag{24.103a}$$

$$= \sum_{i\in I}\sum_{j\in J} c_i(x)\overline{c_j(y)}\delta_{ij} \tag{24.103b}$$

$$= \sum_{i\in I} c_i(x)\overline{c_j(y)}. \tag{24.103c}$$

(5) Nous utilisons l'égalité de Plancherel avec $x = u_j$:

$$\|u_j\|^2 = \|u_j\| + \sum_{i\in I\setminus\{j\}}|\langle u_j, u_i\rangle|^2. \tag{24.104}$$

Par conséquent $\langle u_j, u_i\rangle = 0$ dès que $i \neq j$. Cela prouve que le système $\{u_i\}_{i\in I}$ est orthonormé. Nous devons encore prouver que le système est total. Pour cela nous repartons de l'équation (24.64) que nous avions déduites dans la démonstration de l'inégalité de Bessel :

$$\|x - \sum_{j\in J} c_j(x)u_j\| = \|x\|^2 - \sum_{j\in J}|c_j(x)|^2. \tag{24.105}$$

Par hypothèse le membre de droite peut être rendu aussi petit que l'on veut en prenant J grand (mais fini) dans I. Le membre de gauche indique alors que le système $\{u_i\}_{i\in I}$ est total.

(6) L'identité de Plancherel signifie entre autres que si $x \in \mathscr{H}$ alors $\sum_{i\in I}|\langle x, u_i\rangle|^2$ converge. Du coup la suite $(\langle x, u_i\rangle)_{i\in I}$ est dans $\ell^2(I)$.

□

Remarque 24.45.
Nous avons décidé d'indexer les bases hilbertiennes par $\mathbb{N}$; cela est légitime parce que les sommes sont commutatives. Il ne faut cependant pas perdre de vue qu'en pratique l'ensemble naturel avec lequel on indexe une base est parfois $\mathbb{Z}$. Un tel cas est donné par la base trigonométrique de L^2. Indexer cette dernière par $\mathbb{N}$ plutôt que par $\mathbb{Z}$ serait une contorsion inutile.

Remarque 24.46.
L'égalité de Parseval est la raison pour laquelle les physiciens écrivent souvent

$$\mathrm{Id} = \sum_{n=1}^{\infty} |u_n\rangle\langle u_n| \tag{24.106}$$

dans les livres de mécanique quantique par exemple. Dans certains[429], nous lisons même

$$\int_{-\infty}^{+\infty} dq|q\rangle\langle q| = \hat{I}. \tag{24.107}$$

Notons que ces personnes travaillent avec un espace de Hilbert dont la base n'est pas dénombrable. Pour dire que la physique, ça n'utilise pas des mathématiques pour rire !

Remarque 24.47.
Par définition une base orthonormée est donc une partie dénombrable dont l'espace vectoriel engendré est dense. Un espace de Hilbert possédant une base orthonormée est donc séparable. C'est ce fait qui nous pousse à ne considérer que des espaces de Hilbert séparables ; nous n'allons donc pas étudier ce qu'il se passerait par exemple en considérant l'espace vectoriel librement engendré par les éléments de $\mathbb{R}$.

Exemple 24.48.
L'identité de Parseval (24.93), dans le cas de l'espace des fonctions continues périodiques de période 2π signifie qu'en posant

$$c_n(f) = \frac{1}{2\pi}\int_0^{2\pi} f(s)e^{-ins}, \tag{24.108}$$

nous avons

$$\frac{1}{2\pi}\int_0^{2\pi} |f(s)|^2 = \sum_{n=-\infty}^{\infty} |c_n(f)|^2. \tag{24.109}$$

△

Corolaire 24.49.
Soit $\mathscr{H}$ un espace de Hilbert et (u_n) une base orthonormée. L'application

$$\begin{aligned} S\colon \mathscr{H} &\to \ell^2 \\ x &\mapsto \big(\langle x, u_n\rangle\big)_{n\in\mathbb{N}} \end{aligned} \tag{24.110}$$

est un isomorphisme d'espaces de Hilbert.

De plus l'isomorphisme réciproque est

$$\begin{aligned} S^{-1}\colon \ell^2 &\to \mathscr{H} \\ (\xi_n) &\mapsto \sum_{n=1}^{\infty} \xi_n u_n. \end{aligned} \tag{24.111}$$

Démonstration. Nous devons prouver que l'application est bijective et qu'elle vérifie

$$S(x) \cdot S(y) = \langle x, y\rangle \tag{24.112}$$

où le point dénote le produit dans ℓ^2.

Pour la surjectivité, si $(\xi_n) \in \ell^2$ alors nous savons que la somme $\sum_n \xi_n u_n$ converge par la proposition 24.43 et par conséquent (ξ_n) est l'image par S de ce vecteur de $\mathscr{H}$.

Pour l'injectivité, si $S(x) = S(y)$ alors

$$x = \sum_n \langle x, u_n\rangle u_n = \sum_n \langle y, u_n\rangle u_n = y \tag{24.113}$$

en utilisant la décomposition (24.91).

Le fait que S soit une isométrie est contenu dans Parseval. □

Proposition 24.50 ([1]).
Soit un espace de Hilbert séparable $\mathscr{H}$, un sous-espace vectoriel fermé V et une base orthonormée $\{b_i\}_{i\in I}$ de $\mathscr{H}$. En posant $v_i = \texttt{proj}_V(b_i)$ alors

$$C = \overline{\text{Span}\{b_i - v_i\}} \tag{24.114}$$

est l'orthogonal de V.

Démonstration. Juste pour rappel, lorsque nous écrivons $v_i = \texttt{proj}_V(b_i)$, nous parlons de la projection orthogonale du théorème 24.7. Donc tous les vecteurs $b_i - v_i$ sont dans $V^\perp$. En passant aux limites, $C \subset V^\perp$.

Soit $x \in V^\perp$ que nous décomposons dans la base $\{b_i\}_{i\in I}$ comme $x = \sum_{i\in I} x_i b_i$. Posons $c_i = b_i - v_i = b_i - \texttt{proj}_V(b_i)$. Alors d'une part

$$0 = \texttt{proj}_V(x) = \sum_i x_i \,\texttt{proj}_V(b_i) = \sum_i x_i(b_i - c_i). \tag{24.115}$$

Nous avons utilisé la continuité de $\texttt{proj}_V$ pour permuter avec la somme. D'autre part,

$$x = \sum_i x_i b_i = \sum_i x_i(b_i - c_i + c_i) = \underbrace{\sum_i x_i(b_i - c_i)}_{=0} + \sum_i x_i c_i \in C. \tag{24.116}$$

Notons que pour la dernière appartenance, il est important de prendre la fermeture pour définir C. □

Proposition 24.51 ([1, 430]).
Toute partie orthonormée d'un espace de Hilbert séparable se prolonge en une base hilbertienne.

Démonstration. Soit une partie orthonormée $\{u_i\}_{i\in I}$ de l'espace de Hilbert $\mathscr{H}$. Nous mentionnons que cette partie est libre et donc, par le théorème 24.37, dénombrable [13]. Montrons pour commencer que la partie $V = \overline{\text{Span}\{u_i\}_{i\in I}}$ est un sous-espace vectoriel fermé de $\mathscr{H}$.

Vectoriel Soit $v, w \in V$ et $\epsilon > 0$. Il existe $a \in \text{Span}\{u_i\}$ tel que $\|a - v\| < \epsilon$ et $b \in \text{Span}\{u_i\}$ tel que $\|b - w\| < \epsilon$. Dans ce cas,

$$\|v + w - (a + b)\| \leqslant 2\epsilon. \tag{24.117}$$

Cela prouve que $v + w \in V$. Nous procédons de même pour λv.

Fermé Par construction.

L'orthogonal [14] de V est un sous-espace vectoriel de $\mathscr{H}$, et nous pouvons donc en considérer une base hilbertienne $C = \{c_\alpha\}_{\alpha\in A}$. Nous prétendons que $C \cup U$ est une base hilbertienne [15] de $\mathscr{H}$.

$C \cup U$ est libre Bing ! Il ne faut pas le démontrer : ça ne fait pas partie de la définition d'une base hilbertienne.

$C \cup U$ est générateur Bang ! Il ne faut pas le démontrer : ça ne fait pas partie de la définition d'une base hilbertienne.

$C \cup U$ est orthogonal Ah, voilà quelque chose à démontrer. Nous devons vérifier que les produits sont nuls. Soient $u, v \in U$ et $a, b \in C$. Nous avons :

- $\langle u, v\rangle = 0$ par hypothèse.
- $\langle u, a\rangle = 0$ parce que les éléments de C sont orthogonaux à V et que $u_i \in V$.
- $\langle a, b\rangle = 0$ parce que C est une base hilbertienne de $V^\perp$.

13. Avec un peu de mauvaise foi, vous pouvez quand même dire que cela n'implique pas que I lui-même soit dénombrable, mais vous pouvez le supposer pour fixer les idées.
14. Voir la définition 24.9 et la proposition 24.50.
15. Définition 24.26.

$C \cup U$ est dénombrable L'ensemble C est dénombrable parce que c'est une base hilbertienne. Quant à $\{u_i\}_{i\in I}$, nous avons déjà mentionné le fait qu'il doive être dénombrable. L'union deux parties dénombrables est dénombrable.

$C \cup U$ est total Nous devons prouver que $\overline{\mathrm{Span}(C \cup U)} = \mathscr{H}$ parce qu'il y a bien la fermeture qui intervient dans la définition 24.19. Pour cela nous utilisons la proposition 24.14. Si $x \in \mathscr{H}$ alors nous avons

$$x = \mathtt{proj}_V(x) + \big(x - \mathtt{proj}_V(x)\big) = \sum_{i\in I} x_i u_i + \sum_{\alpha\in A} x_\alpha c_\alpha \tag{24.118}$$

pour des coefficients x_i et x_α. Notons que I et A sont deux ensembles différents. Aucun des x_α n'est un des x_i, ni inversement. Supposons que $x \in \mathrm{Span}(C \cup U)^\perp$; alors

$$0 = \langle x, u_j\rangle = \sum_{i\in I} x_i \underbrace{\langle u_i, u_j\rangle}_{=\delta_{ij}} + \sum_{\alpha\in A} \underbrace{\langle c_\alpha, u_j\rangle}_{=0} = x_j. \tag{24.119}$$

Donc $x_j = 0$. En faisant de même avec $0 = \langle x, c_\beta\rangle = x_\beta$ nous déduisons $x = 0$.

□

Proposition 24.52 ([1]).
Si $\mathscr{H}$ est un espace de Hilbert séparable de base hilbertienne $\{e_i\}_{i\in I}$, alors $\mathscr{H}'$ est un espace de Hilbert séparable dont une base hilbertienne est donnée par les formes

$$\begin{aligned} \alpha_i \colon \mathscr{H} &\to \mathbb{C} \\ e_j &\mapsto \delta_{ij}. \end{aligned} \tag{24.120}$$

Démonstration. C'est la proposition 24.4 qui fait tout. □

24.3.6 Digression sur les normes opérateurs

Le théorème 24.44 nous indique que si $\{u_i\}_{i\in I}$ est une base hilbertienne, alors pour tout $x \in \mathscr{H}$ nous avons

$$\mathtt{proj}_{u_i}(x) = \langle x, u_i\rangle u_i, \tag{24.121}$$

et donc

$$\sum_{i\in I} \mathtt{proj}_{u_i} x = x. \tag{24.122}$$

Nous ne pouvons cependant pas conclure que

$$\sum_{i\in I} \mathtt{proj}_{u_i} = \mathrm{Id} \tag{24.123}$$

au sens de la norme opérateur de la définition 10.3. En effet en prenant $I = \mathbb{N}$, l'égalité (24.123) demanderait d'avoir

$$\lim_{N\to\infty} \left\| \sum_{i=1}^{N} \mathtt{proj}_{u_i} - \mathrm{Id} \right\|_\infty = 0, \tag{24.124}$$

or pour tout N, le vecteur u_{N+1} réalise

$$\sum_{i=1}^{N} \mathtt{proj}_{u_i}(u_{N+1}) - u_{N+1} = -u_{N+1}. \tag{24.125}$$

Par conséquent pour tout N nous avons

$$\sup_{\|x\|=1} \left\| \sum_{i=1}^{N} \mathtt{proj}_{u_i} x - x \right\| \geqslant 1. \tag{24.126}$$

Nous ne pouvons donc pas dire

$$\sum_{n=1}^{\infty} \mathtt{proj}_{u_i} = \mathrm{Id} \tag{24.127}$$

au sens de la norme opérateur.

Nous avons cependant la convergence au sens faible.

Proposition 24.53.
Soit $\mathscr{H}$ un espace de Hilbert et $\{u_i\}_{i\in\mathbb{N}}$ une base hilbertienne de $\mathscr{H}$. Au sens de la topologie faible sur l'espace des opérateurs nous avons

$$\sum_{i=1}^{\infty} \mathtt{proj}_{u_i} = \mathrm{Id}\,. \tag{24.128}$$

Démonstration. Pour chaque $N \in \mathbb{N}$ et $x \in \mathscr{H}$, en vertu de la décomposition (24.91) nous avons

$$\sum_{i=1}^{N} \mathtt{proj}_{u_i}(x) - x = \sim_{i=N+1}^{\infty} \langle x, u_i\rangle u_i. \tag{24.129}$$

Par l'orthonormalité de la base nous avons

$$\sum_{i=N+1}^{\infty} \|\langle x, u_i\rangle u_i\| = \sum_{i=N+1}^{\infty} |\langle x, u_i\rangle|, \tag{24.130}$$

dont la limite $N \to \infty$ est zéro étant donné que la suite $i \mapsto |\langle x, u_i\rangle|$ est dans $\ell^2(\mathbb{R})$ par le théorème 24.44. □

24.3.7 Applications linéaires et continuité

Nous avons déjà vu dans l'exemple 10.28 que la fonction

$$\begin{aligned} f\colon H &\to H \\ e_k &\mapsto k e_k \end{aligned} \tag{24.131}$$

n'était pas continue en zéro alors qu'elle est linéaire. Nous allons maintenant voir qu'elle est un contre-exemple à la proposition 11.244. Calculons les dérivées partielles :

$$\frac{\partial f}{\partial e_j}(x) = \frac{d}{dt}\Big[f(x + te_j)\Big]_{t=0} = \frac{d}{dt}\Big[\sum_k k(x_k + t\delta_{jk})e_k\Big]_{t=0} = je_j. \tag{24.132}$$

où nous avons permuté la somme et la dérivée en considérant la suite de fonctions $f_k(t) = k(x_k + tu_k)e_k$. Donc $\frac{\partial f}{\partial e_j}(x)$ existe et est continue sur un voisinage de $x = 0$ (c'est même constant). Nous savons pourtant que la fonction f n'est pas différentiable en zéro parce que non continue.

L'endroit qui coince dans la preuve de la proposition 11.244 est l'introduction des « contres-termes » dans l'équation (11.624). En effet les contre-termes à ajouter seraient

$$l(x) = \frac{d}{dt}\Big[f\big(a + s(x - a)\big)\Big]_{t=0} \tag{24.133}$$

qui ici serait

$$l\big(\sum_k x_k e_k\big) = \frac{d}{dt}\Big[f\big(s\sum_k x_k e_k\big)\Big]_{t=0} = \sum_k x_k e_k, \tag{24.134}$$

dont la convergence est plus que douteuse.

Notons que les dérivées directionnelles n'existent pas toutes, loin s'en faut : si $u \in H$ nous avons

$$\frac{\partial f}{\partial u}(x) = \frac{d}{dt}\Big[f(x + tu)\Big]_{t=0} = \frac{d}{dt}\Big[\sum_k k(x_k + tu_k)e_k\Big]_{t=0} = \sum_k k u_k e_k \tag{24.135}$$

La convergence de la dernière somme n'est pas garantie pour tout u.

Exemple 24.54.
Soit un espace de Hilbert $\mathscr{H}$. L'application

$$\begin{aligned}\phi\colon \mathscr{H} &\to \mathbb{C}\\ e_i &\mapsto 1\end{aligned} \tag{24.136}$$

n'est pas continue.

Nous considérons les vecteurs $y_n = \sum_{i=1}^n e_i$ et nous posons $x_n = y_n/\|y_n\|$. Nous avons

$$\|y_n\|^2 = \langle y_n, y_n\rangle = \sum_{i,j}^n \langle e_i, e_j\rangle = n, \tag{24.137}$$

donc $\|y_n\| = \sqrt{n}$, et

$$\phi(x_n) = \frac{n}{\sqrt{n}} = \sqrt{n}. \tag{24.138}$$

L'application ϕ n'est donc pas bornée et pas continue non plus. △

Proposition 24.55.
Soient deux espaces de Hilbert séparables[16] $\mathscr{H}_1$ *et* $\mathscr{H}_2$*. Soit une base orthonormée* $\{e_i\}_{i\in I}$ *où* I *est dénombrable. Si* $\Phi\colon \mathscr{H}_1 \to \mathscr{H}_2$ *est une bijection linéaire continue, alors*

$$\Phi\big(\sum_{i\in I} x_i e_i\big) = \sum_{i\in I} x_i \Phi(e_i). \tag{24.139}$$

Démonstration. Soit $\epsilon > 0$. Vu que $\sum_{i\in I} x_i e_i$ est une somme convergente[17], il existe une partie finie J de I telle que pour tout K fini contenant J,

$$\|\sum_{j\in K} x_j e_j - x\| < \epsilon. \tag{24.140}$$

Vu que Φ est une isométrie linéaire, nous avons

$$\|\sum_{j\in K} x_j \Phi(e_j) - \Phi(x)\| < \epsilon, \tag{24.141}$$

et donc bien l'égalité

$$\sum_{i\in I} x_i \Phi(e_j) = \Phi(x). \tag{24.142}$$

□

La proposition suivante donne une formule pour l'inverse du Φ du théorème de représentation de Riesz 24.17 lorsque nous avons une base hilbertienne.

Proposition 24.56.
Soit un espace de Hilbert séparable $\mathscr{H}$*. Si* $\phi \in \mathscr{H}'$*, alors*

$$\Phi^{-1}(\phi) = \sum_{i\in I} \phi(e_i) e_i \tag{24.143}$$

dès que $\{e_i\}_{i\in I}$ *est une base hilbertienne de* $\mathscr{H}$*.*

Démonstration. Nous considérons la base des $\{\alpha_i\}_{i\in I}$ de $\mathscr{H}'$. Remarquez au passage que $\alpha_i = \Phi_{e_i}$. Nous avons $\phi = \sum_{i\in I} \phi_i \alpha_i$ pour certains $\phi_i \in \mathbb{C}$ (qui valent $\phi_i = \phi(e_i)$, ça a son importance pour la suite). En utilisant la proposition 24.55 pour permuter Φ et la somme nous avons le calcul suivant :

$$\phi = \sum_{i\in I} \phi_i \alpha_i = \sum_{i\in I} \phi_i \Phi(e_i) = \sum_{i\in I} \Phi(\phi_i e_i) = \Phi\big(\sum_{i\in I} \phi_i e_i\big) = \Phi(\sum_{i\in I} \phi(e_i) e_i). \tag{24.144}$$

En résulté :

$$\phi = \Phi(\sum_{i\in I} \phi(e_i) e_i), \tag{24.145}$$

et donc la formule (24.143). □

16. Donc à base dénombrable, voir 24.37.
17. Définition 10.92.

24.4 Théorème de Kochen-Specker

Le théorème suivant est central en mécanique quantique. La démonstration provient de [431] et de Wikipédia. Nous allons démontrer complètement le théorème seulement pour les espaces de Hilbert de dimension plus grande ou égale à 4.

Théorème 24.57 (Kochen-Specker[431])**.**
Soit $\mathscr{H}$ un espace de Hilbert de dimension plus grande ou égale à 3. Une fonction v sur l'ensemble des opérateurs de $\mathscr{H}$ ne peut pas satisfaire aux conditions suivantes :

(1) Si A et B sont compatibles, alors $v(A+B) = v(A) + v(B)$,

(2) Si A et B sont compatibles, alors $v(AB) = v(A)v(B)$.

Ici nous disons que deux opérateurs sont ***compatibles*** *lorsqu'ils possèdent une base hilbertienne commune de vecteurs propres.*

Démonstration. Soit $\{u_n\}_{n\in\mathbb{N}}$ une base hilbertienne de $\mathscr{H}$. Nous notons $\mathtt{proj}_i$ l'opérateur de projection sur l'espace (fermé) engendré par u_i. Ce sont des opérateurs compatibles deux à deux parce que la base $\{u_n\}_{n\in\mathbb{N}}$ est une base commune de vecteurs propres [18].

D'abord nous devons avoir $v(\mathbb{1}) = 1$. En effet pour tout opérateur A, nous avons

$$v(A) = v(A\mathbb{1}) = v(A)v(\mathbb{1}). \tag{24.146}$$

Pour peu que $v(A) \neq 0$, cela nous fait $v(\mathbb{1}) = 1$.

En vertu du théorème 24.44, un vecteur $x \in \mathscr{H}$ se décompose en $x = \sum_n \langle x, u_n\rangle u_n$, et nous avons

$$\mathtt{proj}_i\, x = \langle x, u_i\rangle u_i. \tag{24.147}$$

En effet le théorème de la projection orthogonale 24.7 nous enseigne que $\mathtt{proj}_i\, x$ serait l'unique vecteur de la forme λu_i tel que $\lambda u_i - x \perp u_i$. Il est facile de vérifier que le vecteur proposé par (24.147) vérifie cette propriété.

Une conséquence est que

$$\left(\sum_{i=0}^{\infty} \mathtt{proj}_i\right)(x) = x. \tag{24.148}$$

Par conséquent, par hypothèse du théorème nous devons avoir

$$\sum_i v(\mathtt{proj}_i) = v(\mathbb{1}) = 1. \tag{24.149}$$

Étant donné que les projections sont idempotentes,

$$v(\mathtt{proj}_i) = v(\mathtt{proj}_i^2) = v(\mathtt{proj}_i)^2 \tag{24.150}$$

et donc $v(\mathtt{proj}_i)$ doit valoir zéro ou un. Mais la relation (24.149) donne une forte contrainte sur le choix de 0 et de 1. En effet, parmi les $v(\mathtt{proj}_i)$, un et un seul doit valoir 1, les autres doivent valoir 0.

Refaisant le raisonnement pour une autre base orthonormale hilbertienne, nous trouvons que les valeurs de v sur les opérateurs de projection sur les différentes directions doivent être choisies de telle façon que tout choix de base hilbertienne orthogonale contienne exactement un 1 et le reste de zéros.

Nous voudrions maintenant insister sur un point. Le problème de déterminer de façon cohérente les valeurs 0 ou 1 pour tous les $v(\mathtt{proj}_i)$ revient à attacher 0 ou 1 à tous les *rayons* de $\mathscr{H}$ de façon que toute base orthogonale de $\mathscr{H}$ contienne exactement un 1. Un rayon est une direction, c'est-à-dire une classe d'équivalence $x \sim \lambda x$. Si nous décidons de nommer « blanc » les rayons attachés à

18. Pour les besoins de la physique, nous remarquons que ces opérateurs sont des opérateurs hermitiens qui commutent, mais ça ne joue pas ici.

la valeur 0 et « noirs » ceux attachés à la valeur 1, le problème se réduit à colorer la boule unité de façon compatible.

Soit $\{u_i\}_{i\in\mathbb{N}}$ une base orthogonale de $\mathscr{H}$ numérotée de telle sorte que $v(u_0) = 1$ et $v(u_k) = 0$ pour $k \neq 0$. Nous allons maintenant nous particulariser au cas de dimension supérieure ou égale à 4. Si R est une rotation dans le plan $\mathrm{Span}\{u_0, u_1, u_2, u_3\}$, alors l'ensemble

$$\{Ru_0, Ru_1, Ru_2, Ru_3, u_k\}_{k\geqslant 3} \tag{24.151}$$

est encore une base orthogonale de $\mathscr{H}$ et nous avons encore $v(u_k) = 0$ pour $k \geqslant 4$. Par conséquent un et un seul des vecteurs Ru_0, Ru_1, Ru_2 ou Ru_3 est colorié en noir ; les trois autres étant blancs. Le problème est maintenant complètement réduit à la dimension 4. Note : pour réduire à la dimension 3, on procède de même, mais pour conclure, il faut travailler plus.

Nous allons construire 9 base orthogonales de $\mathbb{R}^4$ à partir de 18 vecteurs, chacun arrivant dans exactement deux des bases. Ils sont donnés dans le tableau suivant :

0	1	0	0	0	0	1	1	1	−1	1	−1	−1	1	1	1	1	1
0	0	1	0	0	1	1	1	1	−1	−1	1	1	1	−1	1	1	−1
0	0	0	1	0	0	1	−1	1	−1	1	1	1	1	−1	1	1	1
0	1	0	0	1	0	1	−1	−1	1	1	1	1	−1	1	1	−1	1
1	0	1	0	1	1	1	0	1	1	1	0	1	0	1	0	1	−1
1	0	0	1	0	0	−1	0	0	0	0	−1	0	1	1	0	0	0
1	0	1	0	1	−1	0	1	0	0	0	1	0	1	0	1	0	0
−1	0	0	−1	0	0	0	−1	1	1	−1	0	−1	0	0	−1	1	1

Chaque case de ce table représente un rayon de $\mathbb{R}^4$; il y en a 18 différents, chacun écris deux fois. Une simple vérification montre que chaque colonne est un système orthogonal. La preuve du théorème de Kochen-Specker revient à montrer que nous ne pouvons pas colorier ce tableau de façon cohérente. En effet, étant donné que chaque vecteur est écrit deux fois, le tableau doit contenir un nombre pair de cases blanches et un nombre pair de cases noires.

Par ailleurs chaque colonne étant un système orthogonal, chaque colonne contient exactement une case noire ; il y a donc exactement neuf cases noires dans le tableau, ce qui est impossible.

□

24.5 Théorème de Lax-Milgram

Définition 24.58.
Une forme bilinéaire $a\colon V \times V \to \mathbb{R}$ sur un espace vectoriel normé V est ***coercitive*** *s'il existe $\alpha > 0$ tel que $a(u, u) \geqslant \alpha\|u\|^2$ pour tout $u \in V$.*

Théorème 24.59 (Lax-Milgram[432])**.**
Soit un espace de Hilbert réel V muni de différentes choses.

(1) L'application linéaire $L\colon V \to \mathbb{R}$ qui est bornée sur V. Nous notons $\|L\|$ sa norme.

(2) La forme bilinéaire continue a sur $V \times V$.

(3) La forme a est coercitive[19]*.*

Alors

(1) Il existe $M > 0$ tel que $|a(u, v)| \leqslant M\|u\|\|v\|$ pour tout $u, v \in V$.

(2) Le problème qui consiste à chercher $u \in V$ tel que $a(u, v) = L(v)$ pour tout $v \in V$ admet une unique solution. De plus cette solution vérifie l'inégalité

$$\|u\| \leqslant \frac{\|L\|}{\alpha} \tag{24.152}$$

Démonstration. Évacuons deux points faciles avant de commencer les choses sérieuses. D'abord L est continue par la proposition 10.18. Ensuite, l'existence du M est donnée par la proposition 10.35.

Maintenant nous commençons.

19. Définition 24.58.

Reformulation en équation linéaire La forme L est continue et donc dans le dual V' ; le théorème de Riesz 24.17 nous donne donc $f \in V$ tel que

$$L(v) = \langle f, v \rangle \tag{24.153}$$

pour tout $v \in V$. De plus si $w \in V$ est fixé, l'application $b_w \colon v \mapsto a(w, v)$ est linéaire et bornée parce que

$$\|b_w\| = \sup_{\|v\|=1} |b_w(v)| = \sup_{\|v\|=1} |a(w, v)| \leqslant M\|w\|\|v\| = M\|w\|. \tag{24.154}$$

Encore une fois, b_w étant continue et linéaire, elle est dans V' et Riesz nous fournit un élément $A(w) \in V$ tel que

$$b_w(v) = \langle A(w), v \rangle \tag{24.155}$$

pour tout $v \in V$.

Le problème variationnel $a(u, v) = L(v)$ est équivalent à $\langle A(u), v \rangle = \langle f, v \rangle$. L'ensemble des solutions de cette dernière est égal à l'ensemble des solutions de l'équation

$$A(u) = f. \tag{24.156}$$

A est linéaire Soient $\alpha, \beta \in \mathbb{R}$ et $w, z \in V$. Nous avons pour tout $v \in V$:

$$\begin{aligned}
\langle A(\alpha w + \beta z), v \rangle &= a(\alpha w + \beta z, v) && (24.157a)\\
&= \alpha a(w, v) + \beta a(z, v) && (24.157b)\\
&= \alpha \langle A(w), v \rangle + \beta \langle A(z), v \rangle && (24.157c)\\
&= \langle \alpha A(w) + \beta A(z), v \rangle. && (24.157d)
\end{aligned}$$

Étant donné que nous avons égalité pour tout $v \in V$ nous en déduisons que $A(\alpha w + \beta z) = \alpha A(w) + \beta A(z)$, ce qui signifie que A est linéaire.

Une autre propriété de A Nous déduisons une majoration de $\|A(v)\|^2$ lorsque ce n'est pas nul. Pour ce faire,

$$\begin{aligned}
\|A(v)\|^2 a &= \langle A(v), A(v) \rangle && (24.158a)\\
&= a(v, A(v)) && (24.158b)\\
&\leqslant M\|v\|\|A(v)\|. && (24.158c)
\end{aligned}$$

En simplifiant, $\|A(v)\| \leqslant M\|v\|$. Et donc

$$\|A(v)\|^2 \leqslant M^2\|v\|^2. \tag{24.159}$$

Une contraction Nous allons choisir une valeur de $\rho > 0$ telle que l'application

$$T \colon w \mapsto w - \rho\big(A(w) - f\big) \tag{24.160}$$

soit une contraction[20]. Nous avons $T(w) - T(w') = w - w' - \rho\big(A(w - w')\big)$ et donc

$$\begin{aligned}
\|T(w) - T(w')\|^2 &= \|w - w'\|^2 + \rho^2\|A(w - w')\|^2 - 2\rho\langle A(w - w'), w - w' \rangle && (24.161a)\\
&= \|w - w'\|^2 + \rho^2\|A(w - w')\|^2 - 2\rho a(w - w', w - w'). && (24.161b)
\end{aligned}$$

Vu que le dernier terme arrive avec un signe moins, pour majorer l'expression, il faut minorer ce terme, c'est-à-dire utiliser $a(w - w', w - w") \geqslant \alpha\|w - w - w'\|$. Et en même temps nous utilisons (24.159) pour le second terme. Au final pour pouvons factoriser $\|w - w'\|$ et

$$\|T(w) - T(w')\| \leqslant \|w - w'\|\big(1 + \rho^2 M^2 - 2\rho\alpha\big). \tag{24.162}$$

Pout que T soit contractante, il faut $0 < P(x) < 1$ avec $P(x) = M^2x^2 - 2\alpha x + 1$. Le minimum de ce polynôme est obtenu en $x = \frac{\alpha}{M^2}$ (la formule du $x_{min} = -b/2a$) et vaut $1 - \frac{\alpha^2}{M^2} < 1$. Vu que par ailleurs $\lim_{x\to\infty} P(x) = +\infty$, et que ce polynôme passe par au moins une valeur strictement inférieure à 1, nous savons qu'il existe un x tel que $0 < P(x) < 1$. En donnant à ρ cette valeur, l'application T est une contraction.

20. Définition 16.33.

Point fixe et conclusion L'ensemble des solutions du problème (24.156) est égal à l'ensemble des points fixes de $T(v) = v - \rho\big(A(v) - f\big)$.

L'application $T\colon V \to V$ est contractante et V est métrique et complet. Ergo le théorème de point fixe de Picard 16.34 s'applique et il existe un unique point fixe $u \in V$ pour l'application T. Ce point fixe est l'unique solution de notre problème initial.

La majoration Nous savons que pour tout $v \in V$, la relation $a(u,v) = L(v)$ est vérifiée. En particulier pour $v = u$ nous avons

$$a(u,u) = L(u). \tag{24.163}$$

D'un côté nous utilisons $a(u,u) \geqslant \alpha\|u\|^2$ et de l'autre, $L(u) \leqslant \|L\|\|u\|$:

$$\alpha\|u\|^2 \leqslant \|L\|\|u\| \tag{24.164}$$

et donc

$$\|u\| \leqslant \frac{\|L\|}{\alpha}. \tag{24.165}$$

Notons que $L(u)$ et $a(u,u)$ sont positifs.

□

Exemple 24.60 ([433]).
La borne $\|u\| \leqslant \|L\|/\alpha$ est optimale au sens où il existe des cas d'égalité. En effet nous pouvons considérer

$$a(u,v) = s\langle u,v\rangle \tag{24.166}$$

et

$$L(v) = \langle f,v\rangle \tag{24.167}$$

pour un certain $s > 0$ et $f \in V$. L'application L n'est autre que Φ_f dont nous avons abondamment parlé autour du théorème de représentation de Riesz 24.17. Le lemme 24.15 nous apprend que L est une application borné (et donc continue) de norme $\|L\|_{V'} = \|f\|_V$.

L'application $a\colon V \times V \to \mathbb{R}$ est continue par la proposition 10.39. Elle est coercive parce que

$$a(u,u) = s\|u\|^2. \tag{24.168}$$

Les hypothèses du théorème de Lax-Milgram 24.59 sont réunies et il existe un unique élément $u \in V$ tel que $a(u,v) = L(v)$ pour tout v. Cet élément n'est autre que

$$u = \frac{1}{s}f. \tag{24.169}$$

Nous avons alors $\|u\| = \frac{1}{s}\|f\|$.

Mais s est la constante de coercivité de a et $\|f\|$ est la norme de L. Cette dernière affirmation est contenue dans le fait que Φ est une isométrique dans le lemme 24.15. En reprenant les notations du théorème, nous avons l'égalité

$$\|u\| = \frac{\|L\|}{\alpha}, \tag{24.170}$$

ce qui signifie que l'inégalité du théorème est dans un certain sens optimale. △

Proposition 24.61.
Soit un espace de Hilbert réel V et une application $a\colon V \times V \to \mathbb{R}$ coercive et continue.

L'application $\phi\colon V' \to V$ qui à L fait correspondre l'unique solution u de $a(u,v) = L(v)$ est

(1) bien définie,

(2) linéaire,

(3) continue.

Démonstration. Le fait que ϕ soit bien définie est le théorème de Lax-Milgram 24.59.

L'application ϕ dont nous parlons ici est une application entre deux espaces normés. Soient $L_1, L_2 \in V'$; nous définissons $u_1 = \phi(L_1)$ et $u_2 = \phi(L_2)$. Nous avons

$$a(u_1, v) = L_1(v) \tag{24.171a}$$
$$a(u_2, v) = L_2(v) \tag{24.171b}$$

pour tout $v \in V$. Par bilinéarité de a nous avons aussi

$$a(u_1 + u_2, v) = (L_1 + L_2)(v). \tag{24.172}$$

Vu que $L_1 + L_2$ est également bornée, $u_1 + u_2 = \phi(L_1 + L_2)$. Pour le même type de raisons, si $\lambda \in \mathbb{R}$,

$$a(\lambda u, v) = (\lambda L)(v), \tag{24.173}$$

c'est-à-dire que $\phi(\lambda L) = \lambda\phi(L)$.

Donc ϕ est linéaire. Pour la continuité, il suffit de s'assurer que ϕ est bornée. Pour cela nous utilisons la majoration (24.152) et la définition de la norme d'une application linéaire :

$$\|\phi\| = \sup_{\|L\|=1} \|\phi(L)\| \leqslant \sup_{\|L\|=1} \frac{1}{\alpha}\|L\| = \frac{1}{\alpha}. \tag{24.174}$$

Donc ϕ est bornée et par voie de conséquence continue (proposition 10.18). □

Un problème est bien posé au sens de Hadamard si la solution est unique et dépend continuement des données ; nous ne rentrerons pas plus dans les détails pour l'instant : nous en reparlerons dans la définition 32.14 mais sachez qu'il existe en fin de compte autant de Hadamard qu'il y a de contextes dans lequel le mot « problème » a un sens.

Ce que dit la proposition 24.61 est que, a étant donné, le problème qui consiste à résoudre $a(u, v) = L(v)$ est bien posé au sens de Hadamard, au moins par rapport à L.

Théorème 24.62 (Lax-Milgram version symétrique[432])**.**
Nous considérons les mêmes hypothèses que celles du théorème de Lax-Milgram, c'est-à-dire un espace de Hilbert réel V muni de différentes choses.

(1) L'application linéaire $L\colon V \to \mathbb{R}$ qui est bornée sur V. Nous notons C sa norme.

(2) La forme bilinéaire continue a sur $V \times V$. Nous considérons $M > 0$ tel que $|a(u, v)| \leqslant M\|u\|\|v\|$ pour tout $u, v \in V$.

(3) La forme a est coercitive.

Nous supposons de plus que a est symétrique.

Alors un élément $u \in V$ est tel que $a(u, w) = l(w)$ pour tout $w \in V$ si et seulement si elle minimise la ***fonctionnelle d'énergie***

$$J(w) = \frac{1}{2}a(w, w) - l(w). \tag{24.175}$$

Démonstration. Nous séparons les deux sens.

$\Rightarrow$ Soit u, un élément vérifiant $a(u, w) = l(w)$ pour tout $w \in V$. Soit aussi un élément quelconque $w \in V$, et montrons que $J(u + w) \geqslant J(u)$ (tout élément de V peut être écrit sous la forme $u + w$). Nous avons :

$$J(u + w) = \frac{1}{2}\big(a(u, u) + a(u, w) + a(w, u) + a(w, w)\big) - l(u) - l(w) \tag{24.176a}$$
$$= J(u) + \underbrace{a(u, w) - l(w)}_{=0} + \frac{1}{2}a(w, w) \tag{24.176b}$$
$$= J(u) + \frac{1}{2}a(w, w) \tag{24.176c}$$

Vu que a est coercive, le second terme est positif et nous avons

$$J(u+w) \geqslant J(u), \tag{24.177}$$

ce qu'il fallait.

$\Leftarrow$ Soit $u \in V$, un élément minimisant la fonctionnelle J. Nous fixons $w \in V$ et considérons la fonction $g\colon \mathbb{R} \to \mathbb{R}$ définie par

$$g(\epsilon) = J(u + \epsilon w). \tag{24.178}$$

En développant un peu et en regroupant les termes,

$$g(\epsilon) = J(u) + \epsilon\big(a(u,w) - l(w)\big) + \frac{\epsilon^2}{2}a(w,w). \tag{24.179}$$

Cela est une fonction éminemment continue et dérivable ; en réalité c'est un bête polynôme de degré deux. Vu que u minimise J, pour tout $\epsilon \neq 0$ nous avons $g(\epsilon) \geqslant g(0)$ ou encore : $\epsilon = 0$ est minimum local (et même global) de g. Le polynôme (24.179) prend son minimum en $\epsilon = 0$ si et seulement si

$$a(u,w) - l(w) = 0. \tag{24.180}$$

Vous ne me croyez pas ? Faites $g'(\epsilon) = 0$ ou bien reprenez la formule du $-b/2a$ pour le sommet d'une parabole, en tenant compte que $a(w,w) \neq 0$. Notons qu'ici encore le fait que a soit coercive joue parce que c'est cela qui nous permet d'affirmer que la parabole a un minimum et non un maximum.

$\square$

Chapitre 25

Analyse complexe

25.1 Fonctions holomorphes

La dérivée complexe est discutée à la section 11.26, et la définition d'une fonction holomorphe est 11.248.

25.1.1 Équations de Cauchy-Riemann

Notons que la formule (11.643) donne un **développement limité** pour les fonctions holomorphes. Si f est holomorphe en z_0 alors si z est dans un voisinage de z_0, il existe une fonction $s\colon \mathbb{R} \to \mathbb{C}$ telle que $\lim_{t\to 0} s(t)/t = 0$ et

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + s(|z - z_0|). \tag{25.1}$$

Nous introduisons les opérateurs

$$\frac{\partial}{\partial z} = \partial = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \tag{25.2a}$$

$$\frac{\partial}{\partial \bar{z}} = \bar{\partial} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) \tag{25.2b}$$

Si f est une fonction $\mathbb{C}$-dérivable représentée par la fonction $F = P + iQ$, les équations de Cauchy-Riemann signifient que $\Delta P = \Delta Q = 0$, c'est-à-dire que la fonction f a des composantes harmoniques.

Théorème 25.1.
Si $f \in C^1(\Omega)$ alors nous avons équivalence des faits suivants :

(1) f est holomorphe sur Ω,

(2) f vérifie $\partial_{\bar{z}} f = 0$.

Proposition 25.2.
Une application $f\colon \Omega \to \mathbb{C}$ est $\mathbb{C}$-dérivable sur Ω si et seulement si elle est différentiable et

$$\begin{cases} \dfrac{\partial u}{\partial x} = \dfrac{\partial v}{\partial y} & \text{(25.3a)} \\ \dfrac{\partial u}{\partial y} = -\dfrac{\partial v}{\partial x} & \text{(25.3b)} \end{cases}$$

où $f(x + iy) = u(x, y) + iv(x, y)$. Ces équations se notent de façon plus compacte

$$\frac{\partial f}{\partial \bar{z}} = 0. \tag{25.4}$$

Les équations (25.3) sont les équations de **Cauchy-Riemann**.

Démonstration. La différentielle de $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ est donnée par la matrice

$$T = \begin{pmatrix} \partial_x u(a) & \partial_y u(a) \\ \partial_x v(a) & \partial_y v(a) \end{pmatrix}. \tag{25.5}$$

Cette matrice est une similitude si et seulement si les équations de Cauchy-Riemann sont satisfaites. En effet si $1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ et $i = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, la matrice T est une similitude (écrivons $\alpha + i\beta$ son coefficient) si

$$\begin{cases} T(1) = \alpha + i\beta & \text{(25.6a)} \\ T(i) = -\beta + i\alpha, & \text{(25.6b)} \end{cases}$$

c'est-à-dire

$$T = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}. \tag{25.7}$$

Identifier cette matrice à (25.5) fournit le résultat annoncé. □

Proposition 25.3.
Si $f\colon \mathbb{C} \to \mathbb{C}$ est holomorphe, alors nous avons

$$df_{z_0} = (\partial_z f)(z_0) \tag{25.8}$$

au sens où l'opérateur linéaire $df_{z_0}\colon \mathbb{C} \to \mathbb{C}$ est l'opération de multiplication par le nombre complexe $(\partial_z f)(z_0)$.

Démonstration. Soit $f(x+iy) = f_1(x,y) + if_2(x,y)$ une fonction holomorphe [1]. Les fonctions réelles f_1 et f_2 sont assujetties aux équations de Cauchy-Riemann de la proposition 25.2 :

$$\begin{cases} \partial_x f_1 = \partial_y f_2 & \text{(25.9a)} \\ \partial_x f_2 = -\partial_y f_1. & \text{(25.9b)} \end{cases}$$

Nous avons, en recourant à un petit abus de notation entre $f_i\colon \mathbb{R}^2 \to \mathbb{R}$ et $f_i\colon \mathbb{C} \to \mathbb{R}$:

$$\begin{aligned} df_{z_0}(u) &= \frac{d}{dt}\Big[f(z_0 + tu)\Big]_{t=0} && \text{(25.10a)} \\ &= \frac{d}{dt}\Big[f_1(z_0 + tu) + if_2(z_0 + tu)\Big]_{t=0} && \text{(25.10b)} \\ &= \partial_x f_1 u_1 + \partial_y f_1 u_2 + i\big(\partial_x f_2 u_1 + \partial_y f_2 u_2\big) && \text{(25.10c)} \\ &= \begin{pmatrix} \partial_x f_1 & \partial_y f_1 \\ \partial_x f_2 & \partial_y f_2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} && \text{(25.10d)} \\ &= \begin{pmatrix} \partial_x f_1 & \partial_y f_1 \\ -\partial_y f_1 & \partial_y f_2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}. && \text{(25.10e)} \end{aligned}$$

En utilisant le lemme 11.251 nous reconnaissons la matrice de multiplication par le nombre $\partial_x f_1 - i\partial_y f_1$. Or justement,

$$\partial_z f = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) f = \frac{1}{2}\big(\partial_x f_1 + i\partial_x f_2 - i\partial_y f_1 + \partial_y f_2\big), \tag{25.11}$$

qui se réduit à $\partial_x f_1 - i\partial_y f_1$ lorsque nous y appliquons les équations de Cauchy-Riemann. □

1. Définition 11.248.

25.1.2 Intégrales sur des chemins fermés

Lemme 25.4.
Si g est une fonction continue dans un ouvert $\Omega \subset \mathbb{C}$ et si g admet une primitive complexe sur Ω alors

$$\int_\gamma g(z)dz = 0 \tag{25.12}$$

pour tout chemin fermé γ de classe C^1 contenu dans Ω.

Démonstration. Nommons G une primitive de g. Par définition,

$$\begin{aligned}
\int_\gamma g &= \int_\gamma G' && (25.13a)\\
&= \int_0^1 G'\big(\gamma(t)\big)\gamma'(t)dt && (25.13b)\\
&= \int_0^1 (G\circ g\gamma)'(t)dt && (25.13c)\\
&= G(\gamma(1)) - G\big(\gamma(0)\big) && (25.13d)\\
&= 0 && (25.13e)
\end{aligned}$$

parce que le chemin est fermé : $\gamma(0) = \gamma(1)$. □

Lemme 25.5 (Goursat[349])**.**
Soit Ω un ouvert dans $\mathbb{C}$ et f une fonction continue sur Ω, holomorphe sur Ω moins éventuellement un point (nommé $z_1 \in \Omega$). Soit T, un triangle[2] *fermé inclus dans Ω. Alors nous avons*

$$\int_{\partial T} f(z)dz = 0. \tag{25.14}$$

Démonstration. Nous notons $\gamma = \partial T$. Dans la suite nous allons définir une suite de triangles $T^{(n)}$ et nous noterons $\gamma_n = \partial T^{(n)}$ avec une orientation que nous allons expliquer. Pour commencer nous posons $T^{(0)} = T$ et $\gamma_0 = \partial T^{(0)}$.

Nous considérons le cas $z_1 \notin T$, et nous posons

$$c = l(\gamma)^{-2}|\int_\gamma f|. \tag{25.15}$$

Notre objectif est de montrer que $c = 0$. Soit A, B, C les trois sommes du triangle ; nous divisons le triangle de la façons suivante. D'abord nous considérons les points A', B, C' respectivement milieux de BC, AC et AB. En traçant le triangle $A'B'C'$, nous construisons quatre triangles que nous nommons $T_i^{(0)}$. Le théorème de Thalès assure que le périmètre de chacun des quatre triangles est la moitié du périmètre du grand triangle T.

Sur T nous choisissons l'orientation ABC. De façon à être « compatible », nous choisissons les orientations $AC'B'$, $BA'C'$ et $A'CB'$. La somme de ces trois triangles donne T plus le triangle $A'C'B'$. Par conséquent nous choisissons sur le triangle central l'orientation (inverse) $AB'C'$ de façon à avoir

$$\int_\gamma f = \sum_{i=1}^4 \int_{\partial T_i^{(0)}} f. \tag{25.16}$$

Cela implique que pour au moins un des quatre triangles (disons $T_k^{(0)}$ pour fixer les idées) nous ayons

$$\int_{\partial T_k^{(0)}} f \geqslant \frac{1}{4}\int_{\partial T^{(0)}} f \tag{25.17}$$

2. Nous considérons ici le triangle « plein ».

Nous notons $T^{(1)}$ ce triangle. Comme noté précédemment nous avons

$$l(\partial T^{(1)}) = \frac{1}{2} l(\partial T^{(0)}), \tag{25.18}$$

et donc

$$l(\gamma_1)^{-2} | \int_{\gamma_1} |f = 4 l(\gamma_0)^{-2} | \int_{\gamma_1} f | \geqslant 4 l(\gamma_0)^{-2} \frac{1}{4} | \int_{\gamma_0} f | = c. \tag{25.19}$$

En répétant le procédé nous construisons une suite de triangles $T^{(n)}$ qui satisfont toujours

$$l(\partial T^{(n)}) = \frac{1}{2^n} l(\partial T^{(0)}). \tag{25.20}$$

Ces triangles forment une suite de fermés emboités dont le diamètre tend vers zéro. Leur intersection contient donc exactement un point (lemme 16.119) que nous nommons z_0 (et qui appartient évidemment à Ω). Étant donné que f est holomorphe nous utilisons le développement limité (25.1) autour de z_0 :

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + s(|z - z_0|)(z - z_0) \tag{25.21}$$

avec $\lim_{t \to 0} s(t) = 0$. Nous posons $g(z) = f(z_0) + f'(z_0)(z - z_0)$ et nous considérons $\epsilon > 0$. Soit $\alpha > 0$ tel que

$$|f(z) - g(z)| < \epsilon |z - z_0| \tag{25.22}$$

pour tout $|z - z_0| < \alpha$. Le α à choisir pour obtenir cet effet est celui qui donne $s(|z - z_0|) < \epsilon$. Soit $N \in \mathbb{N}$ tel que $l(\gamma_n) < \alpha$ pour tout $n > N$. D'autre part, deux points dans un triangle sont toujours à distance moindre que la longueur d'un côté, donc pour tout $z \in T^{(n)}$ nous avons $|z - z_0| < \alpha$ et par conséquent pour tout z dans $T^{(n)}$ nous avons

$$|f(z) - g(z)| < \epsilon |z - z_0|. \tag{25.23}$$

Notons que la fonction g est une dérivée : c'est la dérivée de la fonction

$$G(z) = z f(z_0) + \frac{1}{2} f'(z_0)(z - z_0)^2. \tag{25.24}$$

Par conséquent nous avons

$$\int_{\gamma_n} g = 0 \tag{25.25}$$

par le lemme 25.4. Nous avons donc

$$\begin{aligned} | \int_{\gamma_n} f | &= | \int_{\gamma_n} (f - g) | && (25.26a) \\ &\leqslant l(\gamma_n) \max\{|f(z) - g(z)| \text{ tel que } z \in T^{(n)}\} && (25.26b) \\ &\leqslant \epsilon l(\gamma_n)^2, && (25.26c) \end{aligned}$$

et par conséquent

$$c \leqslant l(\gamma_n)^{-2} | \int_{\gamma_n} f | \leqslant \epsilon, \tag{25.27}$$

ce qui signifie que $c = 0$ parce que ϵ est arbitraire. Nous avons donc prouvé le lemme de Goursat dans le cas où le point de non holomorphie z_1 est en dehors de T.

Si z_1 est sur un côté, disons sur le côté AB, alors nous considérons un vecteur $v \in \mathbb{C}$ tel que $T_\epsilon = T + \epsilon v$ ne contienne z_1 pour aucun ϵ. Le vecteur $v = z_1 - C$ fait par exemple l'affaire. En vertu du point précédent nous avons

$$\int_{\partial T_\epsilon} f = 0 \tag{25.28}$$

pour tout $\epsilon > 0$. Étant donné que la fonction f est continue (y compris en z_1), l'intégrale sur ∂T est également nulle.

Si maintenant le point z_1 est à l'intérieur de T nous décomposons T en trois triangles ayant z_1 comme sommet commun. Si nous considérons les orientations Az_1C, ABz_1 et BCz_1, alors nous avons

$$\int_T f = \int_{Az_1C} f + \int_{ABz_1} f + \int_{BCz_1} f, \tag{25.29}$$

alors que par le point précédent les trois intégrales du membre de droite sont nulles. □

Proposition 25.6 ([349]).
Soient Ω un ouvert étoilé et f une fonction holomorphe sur Ω sauf éventuellement en un point z_1 où f est seulement continue. Alors si γ est un chemin fermé dans Ω, nous avons

$$\int_\gamma f = 0. \tag{25.30}$$

Proposition 25.7.
Si $f(z) = \sum_n a_n z^n$ a pour rayon de convergence R, alors f est $\mathbb{C}$-dérivable et nous pouvons dériver terme à terme dans la boule ouverte $B(0, R)$.

Démonstration. Cela est exactement la proposition 14.25. □

25.1.3 Lacets, indice et homotopie

Proposition-définition 25.8.
Soit γ un chemin fermé[3] *dans $\mathbb{C}$. L'**indice** de la courbe γ est la fonction*

$$\begin{aligned} \mathrm{Ind}_\gamma \colon \mathbb{C}\backslash\gamma &\to \mathbb{Z} \\ z &\mapsto \frac{1}{2\pi i}\int_\gamma \frac{d\omega}{\omega - z}. \end{aligned} \tag{25.31}$$

*Un chemin continu et fermé (au sens $\gamma(1) = \gamma(0)$) est un **lacet**.*

(1) La fonction Ind_γ est continue et prend effectivement des valeurs entières.

(2) La fonction indice est constante sur chaque composante connexe[4] *de $\mathbb{C}\backslash\gamma$ et est nulle sur la composante non bornée.*

Le second point est en partie la proposition 7.44.

Définition 25.9.
*Si γ_1 et γ_2 sont deux lacets en $x_0 \in X$ (un espace topologique), une **équivalence d'homotopie** est une application $f \colon [0,1] \times [0,1] \to X$ telle que*

(1) $f(0,t) = \gamma_1(t)$ pour tout t ;

(2) $f(1,t) = \gamma_1(t)$ pour tout t ;

(3) pour chaque $t \in [0,1]$, l'application $s \mapsto f(s,t)$ est continue ;

(4) pour chaque $s \in [0,1]$, l'application $t \mapsto f(s,t)$ est un lacet basé en x_0.

Exemple 25.10.
Si γ est un cercle de centre $z_0 \in \mathbb{C}$ et de rayon r, alors

$$\mathrm{Ind}_\gamma(z) = \begin{cases} 2\pi i & \text{si } z \in B(z_0, r) \\ 0 & \text{sinon.} \end{cases} \tag{25.32}$$

La seconde ligne provient directement de la proposition 25.8. Pour la première, le cercle γ se paramètre par

$$\gamma(\theta) = z_0 + re^{i\theta}, \tag{25.33}$$

3. Par abus de langage, nous désignerons par γ à la fois le chemin et son image.
4. Définition 7.43.

et l'intégrale vaut

$$\int_\gamma \frac{d\omega}{\omega - z_0} = \int_0^{2\pi} \frac{1}{re^{i\theta}} ire^{i\theta} d\theta = 2\pi i. \tag{25.34}$$

L'indice de ce chemin va évidemment jouer un rôle particulier dans la suite. △

Théorème 25.11 (Cauchy, version homotopique[434])**.**
Soient Ω un ouvert de $\mathbb{C}$ et f une fonction holomorphe sur Ω. Si γ_1 et γ_2 sont deux lacets homotopes de classe C^1 dans Ω, alors

$$\int_{\gamma_1} f(z)dz = \int_{\gamma_2} f(z)dz. \tag{25.35}$$

Corolaire 25.12 ([434])**.**
Soient $a \in \mathbb{C}$ ainsi que deux chemins γ_1 et γ_2 homotopes dans $\mathbb{C}\backslash\{a\}$. Alors $\mathrm{Int}(\gamma_1, a) = \mathrm{Ind}(\gamma_2, a)$.

Il y a aussi des choses sur l'indice dans [349].

25.1.4 Théorème de Cauchy et analycité

Cette sous-section veut prouver le théorème de Cauchy. Comme d'habitude, une référence qui ne peut pas rater est [349].

Théorème 25.13 (Formule de Cauchy)**.**
Soient Ω ouvert dans $\mathbb{C}$, $z_0 \in \Omega$ et f une fonction holomorphe sur Ω. Soit $r > 0$ tel que $B(z_0, r) \subset \Omega$. Alors pour tout $z \in B(z_0, r)$ nous avons

$$f(z) = \frac{1}{2\pi i} \int_{\partial B(z_0,r)} \frac{f(\omega)}{\omega - z} d\omega. \tag{25.36}$$

Démonstration. Soit $z \in B(z_0, r)$. Considérons la fonction

$$g(\omega) = \begin{cases} \frac{f(\omega)-f(z)}{\omega - z} & \text{si } \omega \neq z \\ f'(z) & \text{si } \omega = z. \end{cases} \tag{25.37}$$

Cette fonction est holomorphe sur $B(z_0, r)\backslash\{z\}$ et continue en z. Elle vérifie donc la proposition 25.6 et nous avons

$$\int_\gamma g = 0 \tag{25.38}$$

où γ est le cercle de centre z_0 et de rayon r. Nous avons donc

$$0 = \int_\gamma \frac{f(\omega)}{\omega - z} - \int_\gamma \frac{f(z)}{\omega - z}, \tag{25.39}$$

et ayant déjà calculé la seconde intégrale dans l'exemple 25.10 nous en déduisons

$$\int_\gamma \frac{f(\omega)}{\omega - z} d\omega = 2\pi i f(z), \tag{25.40}$$

ce qu'il fallait. □

Théorème 25.14.
Soient Ω ouvert dans $\mathbb{C}$ et f holomorphe sur Ω. Soient encore $z_0 \in \Omega$ et r_0 tels que $B(z_0, r_0) \subset \Omega$. Alors :

(1) Sur $B(z_0, r_0)$, la fonction f s'écrit

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n. \tag{25.41}$$

(2) Nous avons

$$a_n = \frac{f^{(n)}(z_0)}{n!} = \frac{1}{2\pi i}\int_\gamma \frac{f(\omega)}{(\omega - z_0)^{n+1}}d\omega \tag{25.42}$$

où $\gamma = \partial B(z_0, r)$ *avec* $|z - z_0| < r < r_0$.

(3) En particulier f *est infiniment dérivable.*

Démonstration. Soit $r > 0$ tel que $|z - z_0| < r < r_0$. La formule de Cauchy (théorème 25.13) nous dit que

$$f(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(\omega)}{\omega - z}d\omega \tag{25.43}$$

où $\gamma = \partial B(z_0, r)$. Nous pouvons paramétrer ce chemin par $\omega = z_0 + re^{i\theta}$ et $\theta \in [0, 2\pi]$. Nous avons

$$f(z) = \frac{1}{2\pi i}\int_0^{2\pi} \frac{f(z_0 + re^{i\theta})}{z_0 + re^{i\theta} - z} rie^{i\theta}d\theta \tag{25.44a}$$

$$= \frac{1}{2\pi}\int_0^{2\pi} \frac{f(z_0 + re^{i\theta})}{1 - e^{-i\theta}(z - z_0)/r}d\theta. \tag{25.44b}$$

Nous pouvons développer l'intégrante en puissance de $(z - z_0)$ en utilisant la formule 14.366. Ici le rôle de x est tenu par

$$e^{-i\theta}(z - z_0)/r \tag{25.45}$$

dont le module est bien plus petit que 1, par hypothèse sur r. Nous avons donc

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} \sum_{n=0}^{\infty} f(z_0 + re^{i\theta})e^{-in\theta}r^{-n}(z - z_0)^n d\theta. \tag{25.46}$$

L'art est maintenant de permuter la somme et l'intégrale. Pour cela nous remarquons que ce qui se trouve dans la somme est majoré en module par

$$M\left|\frac{z - z_0}{r}\right|^n \tag{25.47}$$

où M est le maximum de $|f|$ sur γ. La borne (25.47) ne dépend pas de θ ; par conséquent la convergence de la somme est uniforme en θ par le critère de Weierstrass (théorème 11.309). Le théorème 14.1 s'applique[5] et nous pouvons permuter la somme avec l'intégrale.

Ce que nous trouvons est que

$$f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n \tag{25.48}$$

où

$$a_n = \frac{1}{2\pi}\int_0^{2\pi} f(z_0 + re^{i\theta})e^{-in\theta}r^{-n}d\theta = \frac{1}{2\pi i}\int_\gamma \frac{f(\omega)}{(\omega - z_0)^{n+1}}. \tag{25.49}$$

Cette formule est valable pour $|z - z_0| < r$. Sur cette boule, la fonction est donc une série entière. Le théorème de Taylor 14.102 nous permet donc d'affirmer que f est partout infiniment continument dérivable (parce que en chaque point on a un voisinage sur lequel c'est vrai), et d'identifier les coefficients (qui, eux, ne sont valables que localement) sous la forme

$$a_n = \frac{f^{(n)}(z_0)}{n!}. \tag{25.50}$$

□

5. Étant donné que nous savions déjà que la somme était une fonction intégrable, nous sommes loin d'avoir utilisé toute la puissance du théorème.

Corolaire 25.15.
Soit f une fonction continue sur un ouvert Ω telle que pour toute boule $B(a,r)$ contenue dans Ω, nous ayons

$$f(a) = \frac{1}{2\pi i}\int_{\partial B(a,r)} \frac{f(\xi)}{\xi - a} d\xi. \tag{25.51}$$

Alors f est holomorphe.

Démonstration. Il suffit de recopier la démonstration du théorème 25.14 pour savoir que f se développe en série de puissances et est donc en particulier dérivable. □

Le fait qu'une fonction holomorphe soit C^∞ comme dit dans la proposition 25.14 permet de démonter un résultat de dérivation sous l'intégrale, qui dépend de pouvoir majorer la différentielle.

Proposition 25.16.
Soit une fonction continue $g\colon \mathbb{R}\times\mathbb{C}\to\mathbb{C}$. Nous supposons que pour tout t, la fonction $z\mapsto g(t,z)$ est $\mathbb{C}$-dérivable (définition 11.247) et différentiable. Soit B compact dans $\mathbb{R}$ et la fonction

$$G(z) = \int_B g(t,z)dt. \tag{25.52}$$

que nous supposons exister pour tout z.

Alors

$$G'(z) = \int_B g'(t,z)dt \tag{25.53}$$

où le prime réfère à la $\mathbb{C}$-dérivée par rapport à la variable z à t fixé.

Démonstration. Nous fixons $z\in\mathbb{C}$ et nous considérons la suite de fonctions

$$g_i(t) = \frac{g(t,z+\epsilon_i) - g(t,z)}{\epsilon_i} \tag{25.54}$$

où ϵ_i est une suite de nombres complexes tendant vers zéro ($\epsilon_i \xrightarrow{\mathbb{C}} 0$). Si la limite existe et ne dépend pas de la suite choisie, alors $\lim_{i\to\infty} g_i(t) = g'(t,z)$. Et vu que g est supposée dérivable, c'est le cas.

Nous avons aussi, par linéarité de l'intégrale :

$$G'(z) = \lim_{i\to\infty}\int_B g_i(t)dt. \tag{25.55}$$

La difficulté est de permuter la limite et l'intégrale. Pour cela nous allons utiliser la convergence dominée de Lebesgue (théorème 13.184). Afin de majorer $|g_i(t)|$ par une fonction intégrable en t (uniformément en i), nous exploitons le théorème des accroissements finis, théorème 10.145. En notant dg la différentielle de g par rapport à z à t fixé, pour chaque t et chaque i nous avons

$$|g(t,z+\epsilon_i) - g(t,z)| \leqslant \sup_{\xi\in[z,z+\epsilon_i]} \|dg_\xi\|\|\epsilon_i\|. \tag{25.56}$$

Vu que z est fixé et que ξ est dans le compact $[z,z+\epsilon_i]$ et que dg est continue (parce que la $\mathbb{C}$-dérivabilité implique la continuité de la différentielle parce que nous avons l'analycité par le théorème 25.14), nous pouvons majorer $\|dg_\xi\|$ par une constante $M_i(z)$ qui dépend à priori de i et de z.

Heureusement, nous pouvons prendre a fortiori le supremum sur $\overline{B(z,|\epsilon_i|)}$ (qui est tout autant compact) et supposer que $|\epsilon_i|$ est strictement décroissante ; de toutes façons, il y a un maximum parce que $|\epsilon_i|\to 0$. Dans ce cas, il suffit de prendre le supremum de $\|dg_\xi\|$ pour $\xi\in\overline{B(z,|\epsilon_1|)}$ et ça contente tout le monde.

Quoi qu'il en soit nous avons une constante $M(z)$ telle que

$$|g(t,z+\epsilon_i) - g(t,z)| \leqslant M(z)\|\epsilon_i\| \tag{25.57}$$

et donc $|g_i(t)| \leqslant M(z)$. La constante (par rapport à t) $M(z)$ est évidemment intégrable sur le compact B et nous pouvons permuter la limite avec l'intégrale :

$$G'(z) = \lim_{i\to\infty} \int_B g_i(t)dt = \int_B \lim_{i\to\infty} g_i(t)dt = \int_B g'(t,z)dt. \tag{25.58}$$

□

Proposition 25.17.
Une fonction continue f est holomorphe si et seulement si la 1-forme différentielle $f(z)dz$ est localement exacte.

Démonstration. Si f est holomorphe, alors nous avons vu que f était différentiable et que $df_z = f(z)dz$ par la formule 11.649.

Dans le sens inverse, supposons que $f(z)dz$ est localement exacte, et soit F telle que $dF = f(z)dz$. Ce que nous allons faire est montrer que la dérivée de F existe et vaut f. En effet, la définition de la différentielle nous dit que

$$\lim_{h\to 0} \left| \frac{F(z+h) - F(z) - dF_z(h)}{h} \right| = 0. \tag{25.59}$$

La limite vaut évidemment encore zéro si nous enlevons les modules :

$$0 = \lim_{h\to 0} \frac{F(z+h) - F(z) - f(z)h}{h} \tag{25.60a}$$

$$= \lim_{h\to 0} \frac{F(z+h) - F(z)}{h} - f(z). \tag{25.60b}$$

Donc $F' = f$. Cela montre que F est $\mathbb{C}$-dérivable et donc holomorphe. En conséquence du théorème 25.14, la fonction F est infiniment dérivable et f l'est alors aussi. La fonction f est donc holomorphe[6]. □

25.1.5 Théorème de Brouwer en dimension 2

Pour d'autres versions du théorème de Brouwer, voir la sous-section 19.5.1.

Théorème 25.18 (Brouwer en dimension 2[54])**.**
Soit $\mathcal{B}$ la boule unité fermée de $\mathbb{R}^2$. Alors toute application continue de $\mathcal{B}$ dans elle-même admet un point fixe.

Démonstration. Supposons que la fonction $f \in C^0(\mathcal{B},\mathcal{B})$ n'admette pas de points fixes sur $\mathcal{B} = \overline{B(0,1)}$. Pour $x \in \mathcal{B}$ nous notons $g(x)$ l'intersection entre $\partial\mathcal{B}$ et la demi-droite allant de $f(x)$ vers x. C'est bien parce que f n'a pas de points fixes que g est bien définie.

En reprenant le même début de la preuve de la proposition 19.30 nous savons que la fonction

$$\begin{aligned} g\colon \overline{B(0,1)} &\to \partial B(0,1) \\ x &\mapsto \lambda(x)\big(x - f(x)\big) + f(x) \end{aligned} \tag{25.61}$$

est continue. De plus $g(x) = x$ sur $\partial B(0,1)$. Nous allons montrer qu'une telle fonction[7] ne peut pas exister.

Pour $s \in [0,1]$ nous paramétrons le cercle $\partial B(0,s)$ par

$$\begin{aligned} x_s\colon [0,1] &\to \partial B(0,1) \\ t &\mapsto \big(s\cos(2\pi t), s\sin(2\pi t)\big). \end{aligned} \tag{25.62}$$

6. Dire que la dérivée d'une fonction holomorphe est holomorphe est un raisonnement classique.
7. Qui est nommée *rétraction* de la sphère sur elle-même.

Ensuite nous considérons les chemins

$$\begin{aligned}\gamma_s\colon [0,1] &\to \partial B(0,s)\\ t &\mapsto g\circ x_s.\end{aligned}\tag{25.63}$$

L'application γ_s est continue et $\gamma_s(0)=\gamma_s(1)$. Les chemins γ_s sont des lacets ; nous nous intéressons maintenant à l'indice au point 0 de γ_0 et γ_1. D'une part $\gamma_0(t)=g(0)$ (lacet constant) et $\gamma_1(t)=e^{2i\pi t}$ (parce que $g(x)=x$ sur le bord). Nous avons donc, en utilisant l'indice de la définition 25.8,

$$\mathrm{Ind}_{\gamma_0}(0)=\frac{1}{2\pi i}\mathrm{Ind}_{\gamma_0}\frac{d\omega}{\omega}=\frac{1}{2\pi i}\int_0^1\frac{\gamma_0'(t)}{\gamma_0(t)}dt=0,\tag{25.64}$$

alors que

$$\mathrm{Ind}_{\gamma_1}(0)=\frac{1}{2\pi i}\int_0^1\frac{2i\pi e^{2i\pi t}}{e^{2i\pi t}}dt=1.\tag{25.65}$$

Nous considérons l'homotopie

$$\begin{aligned}\gamma\colon [0,1]\times[0,1] &\to \overline{B(0,1)}\\ (s,t) &\mapsto \gamma_s(t)=(g\circ x_s)(t).\end{aligned}\tag{25.66}$$

Nous avons $g(0)\neq 0$ parce que g prend ses valeurs sur le bord. Vu que c'est une équivalence d'homotopie[8] entre γ_1 et γ_2, les indices devraient être égaux par le corolaire 25.12. □

25.1.6 Principe des zéros isolés

Lemme 25.19.
Si f est une fonction holomorphe[9] sur le compact K, alors il existe une fonction polynôme P_f et une fonction holomorphe h_f ne s'annulant pas sur K telles que $f=h_fP_f$.

Démonstration. Soit une fonction f vérifiant les conditions. Si f est identiquement nulle, alors il suffit de prendre $P_f=0$ et c'est fait. Nous supposons donc que f n'est pas identiquement nulle.

Quantité finie de racines D'abord f ne peut s'annuler qu'un nombre fini de fois sur K. Sinon, on pourrait considérer une suite des racines[10] de f dans K. Vu qu'une suite dans un compact contient une sous-suite convergente (théorème 7.103), la fonction f aurait un point d'accumulation de racines. Alors le principe des zéros isolés (théorème 16.126) nous donne un ouvert sur lequel f est nulle et donc le corolaire 16.127 nous dit que f est identiquement nulle.

Autour d'une racine Bref, la fonction f possède un nombre fini de racines sur K. Soit z_0 l'un d'eux.

Par le théorème 25.14(1), nous avons, sur un voisinage de z_0 :

$$f(z)=\sum_{n=0}^{\infty}a_n(z-z_0)^n.\tag{25.67}$$

En particulier, $0=f(z_0)=a_0$. Donc $a_0=0$. Soit k, le plus petit naturel pour lequel $a_k\neq 0$. Nous avons

$$f(z)=(z-z_0)^kg(z)\tag{25.68}$$

avec $g(z)=\sum_{n=0}^{\infty}a_{k+n}(z-z_0)^n$.. Vu que $a_k\neq 0$ nous avons $g(z_0)\neq 0$. Montrons à présent que g est holomorphe sur un voisinage de z_0. Vu que la série définissant g est une sous-série d'une série convergente sur un voisinage, elle converge sur un voisinage et la proposition 25.7 nous dit que g est $\mathbb{C}$-dérivable. C'est-à-dire holomorphe par définition.

8. Définition 25.9
9. Définition 11.249.
10. Notez l'utilisation de la proposition 1.31 que je vous invite à ne pas considérer comme une trivialité absolue.

Autour de toutes les racines Soient (z_i) les racines (en nombre fini). Pour chaque i nous avons une boule $B(z_i, r_i)$ sur laquelle $f = P_i g_i$ où P_i est un polynôme de la forme $(z - z_i)^k$ et g_i est holomorphe sur $B(z_i, r_i)$. Nous définissons la fonction suivante :

$$h(z) = \begin{cases} \dfrac{f(z)}{\prod_k P_k(z)} & \text{si } z \neq z_i \\ \dfrac{g_i}{\prod_{k\neq i} P_k(z)} & \text{si } z = z_i. \end{cases} \tag{25.69}$$

Cette fonction ne s'annule jamais. Mais est-elle holomorphe ?

Si $z \neq z_i$ (sous-entendu : pour tout i), alors sur un voisinage, $h = f/\prod P_k$ qui est un quotient de fonctions holomorphes dont le dénominateur ne s'annule pas. Elle est donc holomorphe sur ce voisinage par le lemme 11.256.

Pour les autres notons que pour tout $z \in B(z_i, r_i)$,

$$h = \frac{g_i}{\prod_{k\neq i} P_k}. \tag{25.70}$$

Cela est encore un quotient dont le dénominateur ne s'annule pas [11].

La réponse Nous avons, pour tout $z \in K$:

$$f(z) = h(z) \prod_k P_k(z). \tag{25.71}$$

□

Afin de détendre l'atmosphère, nous allons laisser tomber l'analyse quelques instants et prouver un résultat d'algèbre.

Proposition 25.20 ([1, 435]).
L'anneau des fonctions holomorphes sur un compact [12] *donné de* $\mathbb{C}$ *est principal* [13].

Démonstration. Nous nommons A l'ensemble des fonctions holomorphes sur le compact K, et J un anneau de A.

A est un anneau Le point délicat de la définition 1.73 est le fait que la somme et le produit d'éléments de A sont des éléments de A parce que les résultats type « la somme de deux fonctions holomorphes est holomorphes » sont valides sur des ouverts alors que nous sommes ici sur un compact. Soient f et g dans A ; nous nommons Ω_f et Ω_g des ouverts contenant K tels que f est holomorphe sur Ω_f et g sur Ω_g.

L'ensemble $\Omega_f \cap \omega_g$ est un ouvert (intersection d'ouverts) contenant K et sur lequel f et g sont holomorphes. Donc $f + g$ et fg y sont holomophes.

Engendré par des polynômes Pour chaque $f \in J$ nous écrivons $f = P_f h_f$ en vertu de la décomposition donnée par le lemme 25.19. Vu que h_f ne s'annule pas, $1/h_f$ est encore holomorphe sur K et nous déduisons que $P_f = f/h_f$ est dans J. La partie

$$S = \{P_f \text{ tel que } f \in J\} \tag{25.72}$$

est génératrice de J parce que, par construction, tous les éléments de J sont des produits d'éléments de S par des fonctions holomorphes sur K (donc, des éléments de A). Mais tous les éléments de S sont dans J, donc $(S) = J$.

11. Nous avons choisi les r_i de telle sorte que les boules ne s'intersectent pas.
12. Être holomorphe sur un compact signifie qu'il existe une extension holomorphe à un ouvert contenant le compact.
13. Définition 3.118

Un polynôme pour tous les engendrer Soit M, l'idéal de $\mathbb{C}[X]$ engendré par S. Attention : J est l'idéal de A engendré par S. Mais l'idéal de $\mathbb{C}[X]$ engendré par S est peut-être autre chose. Vu que $\mathbb{C}$ est un corps, le lemme 3.178 dit que $\mathbb{C}[X]$ est principal. Donc M est un idéal principal de $\mathbb{C}[X]$ et nous avons un polynôme $p \in \mathbb{C}[X]$ tel que

$$M = \mathbb{C}[X]p. \tag{25.73}$$

Si vous avez compris le chausse trappe, vous saurez pourquoi il faut écrire $M = \mathbb{C}[X]p$ et non utiliser l'écriture plus simple « $M = (p)$ ».

$A\mathbb{C}[X] = A$ L'inclusion $A \subset A\mathbb{C}[X]$ est dûe au fait que $1 \in \mathbb{C}[X]$, et l'autre inclusion est le fait que $\mathbb{C}[X] \subset A$ alors que A est un anneau.

Suite des opérations Nous avons :

$$J = AS \subset A\mathbb{C}[X]p. \tag{25.74}$$

Voilà une inclusion de montrée. Reste à faire l'autre.

Vu que $p \in J$ nous avons aussi $Ap \subset J$. Et donc

$$A\mathbb{C}[X]p = Ap \subset J. \tag{25.75}$$

Avec ces deux inclusions, $J = A\mathbb{C}[X]p = Ap$. Donc J est engendré par un seul élément et est principal.

□

25.1.7 Prolongement de fonctions holomorphes

Proposition 25.21.
Soient Ω un ouvert de $\mathbb{C}$ et $f\colon \Omega \to \mathbb{C}$ une fonction holomorphe sur $\Omega\backslash\{a\}$ ($a \in \Omega$). Nous supposons qu'il existe $r > 0$ tel que f est bornée sur $B(a,r) \cap \Omega$. Alors f se prolonge en une fonction holomorphe sur Ω.

Le théorème de prolongement de Riemann 25.35 donnera plus d'informations.

Démonstration. Nous définissons la fonction $g\colon \Omega \to \mathbb{C}$ par

$$g(z) = \begin{cases} (z-a)f(z) & \text{si } z \neq a \\ 0 & \text{si } z = a. \end{cases} \tag{25.76}$$

Sur $\Omega\backslash\{a\}$, la fonction g est holomorphe (produit de fonctions holomorphes), et elle est continue en a. Par conséquent elle est holomorphe sur Ω. Nous la développons en série entière sur une boule $B(a,r)$:

$$g(z) = \sum_{n=0}^{\infty} c_n (z-a)^n. \tag{25.77}$$

Nous avons $g(a) = c_0 = 0$. Nous posons

$$\varphi(z) = \sum_{n=0}^{\infty} c_{n+1} (z-a)^n. \tag{25.78}$$

Si $z \neq a$, alors $\varphi(z) = f(a)$ parce que $\varphi(z) = g(z)/(z-a)$. Mais φ est continue en a, et donc holomorphe en a.

La fonction φ est par conséquent un prolongement holomorphe de f en a. □

25.1.8 Théorème de Runge

Le théorème que nous allons prouver n'est en réalité qu'une partie de ce qui est usuellement appelle le théorème de Runge.

Théorème 25.22 (Théorème de Runge)**.**
Soit K, un compact de $\mathbb{C}$ tel que $\complement K$ soit connexe. Si $a \in \complement K$ alors la fonction

$$\varphi_a(z) = \frac{1}{z-a} \tag{25.79}$$

est limite uniforme de polynômes sur K.

Démonstration. Nous considérons $P(K)$, l'adhérence des polynômes sur K pour la norme uniforme (sur K). Nous devons montrer que pour tout $a \in \complement K$, la fonction φ_a est dans $P(K)$. Pour cela nous considérons l'ensemble

$$A = \{a \in \complement K \text{ tel que } \varphi_a \in P(K)\} \tag{25.80}$$

et nous allons montrer qu'il est à la fois non vide, ouvert et fermé dans le connexe $\complement K$.

Je répète : nous allons prouver l'ouverture et la fermeture *pour la topologie de* $\complement K$. Nous n'allons pas prouver que A est un ouvert de $\mathbb{C}$. Ce qui sera par conséquent prouvé est que $A = \complement K$.

Non vide Soit $R = \sup_{z\in K} |z|$ et $a \in \complement K$ tel que $|a| > R$. Nous avons

$$\varphi_a(z) = \frac{1}{a}\frac{1}{\frac{z}{a}-1} = -\frac{1}{a}\frac{1}{1-\frac{z}{a}} = -\frac{1}{a}\sum_{k=0}^{\infty}\left(\frac{z}{a}\right)^k = \sum_{k=0}^{\infty}\frac{z^k}{a^{k+1}}. \tag{25.81}$$

Ici la convergence de la série et sa limite sont assurées par le fait que $|z/a| < 1$ par choix de R et a. La suite de polynômes

$$P_n(z) = \sum_{k=0}^{n}\frac{z^k}{a^{k+1}} \tag{25.82}$$

converge uniformément sur $B(0,R)$ et en particulier sur K. Donc $P_n \to \varphi_a$.

Fermé Nous allons montrer que la fermeture de A (dans $\complement K$) est inclue dans A, et donc qu'elle est égale à A et donc que A est fermé. Par le lemme 7.17, la fermeture de A dans $\complement K$ est l'ensemble $\bar{A} \cap \complement K$ où $\bar{A}$ est la fermeture de A au sens usuel.

Bref, soit $a \in \bar{A} \cap \complement K$, et montrons que $\varphi_a \in \overline{P(K)}$. Vu que $P(K)$ est déjà une fermeture, nous aurons en fait $\varphi_a \in P(K)$ et donc $a \in A$, ce qui signifierait que $\bar{A} \cap \complement A = A$ et donc que A est fermé.

Au travail.

Soit $(a_n) \in A$ une suite convergente vers a. Soit aussi $d = d(a,K)$; on a $d > 0$ parce que K est compact et a est hors de a alors le complémentaire de K est ouvert. Nous choisissons en plus la suite a_n pour avoir $|a_n - a| < \frac{d}{2}$; au pire on prend la queue de suite. Soit $z \in K$; nous avons

$$|\varphi_{a_n}(z) - \varphi_a(z)| = \left|\frac{1}{z-a_n} - \frac{1}{z-a}\right| = \left|\frac{a_n - a}{(z-a_n)(z-a)}\right|. \tag{25.83}$$

Vu que $a_n \in B(a, \frac{d}{2})$ et que $z \in K$ et $d = d(a,K)$ nous avons $|a_n - z| \geqslant \frac{d}{2}$; et aussi $|a - z| \geqslant \frac{d}{2}$. Nous pouvons donc majorer (25.83) par

$$|\varphi_{a_n}(z) - \varphi_a(z)| \leqslant 2\frac{|a_n - a|}{d^2}. \tag{25.84}$$

Donc nous avons

$$\|\varphi_a - \varphi_{a_n}\|_K \leqslant 2\frac{|a_n - a|}{d^2} \to 0 \tag{25.85}$$

où la norme $\|.\|_K$ est la norme supremum sur K. Donc $a \in \overline{P(K)} = P(K)$ et A est fermé.

Ouvert Vu que K est compact, il est fermé et donc $\complement K$ est ouvert. Par conséquent, ainsi que précisé dans l'exemple 7.18, les ouverts de $\complement K$ sont les ouverts de $\mathbb{C}$ contenus dans $\complement K$. Afin de prouver que A est ouvert, nous prenons $a \in A$ et nous cherchons une boule (au sens de $\mathbb{C}$) autour de a qui serait incluse dans A.

Soit donc $h \in \mathbb{C}$ « petit » dans un sens que nous allons préciser plus tard. Encore une fois nous posons $d = d(a, K)$. Nous avons

$$\varphi_{a+h}(z) = \frac{1}{z-a-h} = \frac{1}{z-a}\frac{1}{1-\frac{h}{z-a}} = \sum_{k=0}^{\infty} \frac{h^k}{(z-a)^{k+1}}. \tag{25.86}$$

Déjà ici nous demandons $h < \sup_{z\in K} |z-a|$. Puisque $|z-a| > d$, nous avons alors

$$|\varphi_{a+h}(z)| \leqslant \sum_{k=0}^{\infty} \frac{h^k}{d^{k+1}} < \infty. \tag{25.87}$$

Cela pour dire que la somme à droite de (25.86) converge bien pourvu que h soit bien petit. Nous pouvons donc poursuivre :

$$\varphi_{a+h}(z) = \sum_{k=0}^{\infty} \frac{h^k}{(z-a)^{k+1}} = \sum_{k=0}^{\infty} h^k \varphi_a(z)^{k+1}. \tag{25.88}$$

Nous montrons maintenant que la convergence de la somme (25.88) est en réalité uniforme en z. En effet

$$\Big|\varphi_{a+h}(z) - \sum_{k=0}^{N} h^k \varphi_a(z)^{k+1}\Big| = \Big|\sum_{k=N+1}^{\infty} h^k \varphi_a(z)^{k+1}\Big| \tag{25.89a}$$

$$\leqslant \sum_{k=N+1}^{\infty} |h|^k |\varphi_a(z)|^{k+1}. \tag{25.89b}$$

Étant donné que φ_a est continue sur le compact K, elle y est majorée en module ; on peut même être plus précis :

$$|\varphi_a(z)| = \frac{1}{|z-a|} \leqslant \frac{1}{d}. \tag{25.90}$$

Nous pouvons donc écrire

$$\Big|\varphi_{a+h}(z) - \sum_{k=0}^{N} h^k \varphi_a(z)^{k+1}\Big| \leqslant \frac{1}{d} \sum_{k=N+1}^{\infty} \Big|\frac{h}{d}\Big|^k. \tag{25.91}$$

Étant donné que la somme $\sum_{k=0}^{\infty} |h/d|^k$ converge, la limite $N \to \infty$ est nulle et nous avons

$$\lim_{N\to\infty} \|\varphi_{a+h} - \sum_{k=0}^{N} h^k \varphi_a^{k+1}\|_K = 0. \tag{25.92}$$

Pour avoir $\varphi_{a+h} \in P(K)$, il faut encore savoir si les fonctions φ_a^k sont dans $P(K)$ pour tout k. Dans ce cas pour chaque N la somme sera encore dans $P(K)$ et φ_{a+h} sera limite uniforme d'éléments de $P(K)$.

Par hypothèse, $\varphi_a \in P(K)$; soit P_n une suite de polynômes qui converge uniformément vers φ_a. Nous allons montrer qu'alors la suite de polynômes P_n^k converge uniformément vers φ_a^k. Soit n tel que $\|P_n - \varphi_a\|_K \leqslant \epsilon$ et utilisons le produit remarquable

$$a^k - b^k = (a-b) \sum_{i=0}^{k-1} a^i b^{k-1-i} \tag{25.93}$$

pour obtenir

$$|P_n(z)^k - \varphi_a(z)^k| \leqslant |P_n(z) - \varphi_a(z)| \sum_{i=0}^{k-1} |P_n(z)^i \varphi_a(z)^{k-1-i}|. \tag{25.94}$$

Vu que P_n et φ_a sont continues sur le compact K, on peut majorer la somme par une constante M, et il restera

$$|P_n(z)^k - \varphi_a(z)^k| \leqslant M|P_n(z) - \varphi_a(z)|, \tag{25.95}$$

ou encore

$$\|P_n^k - \varphi_a^k\| \leqslant M\epsilon. \tag{25.96}$$

Cela prouve que $\varphi_a^k \in P(K)$ et donc que φ_{a+h} est limite uniforme (sur K) d'éléments de $P(K)$ et donc fait partie de $P(K)$ lui aussi.

Ceci achève de prouver que A est ouvert dans $\complement K$.

Conclusion L'ensemble A est non vide, ouvert et fermé dans $\complement K$, donc il est égal à $\complement K$. Le théorème est ainsi démontré.

□

25.2 Intégrales de fonctions holomorphes

Nous commençons par le lemme technique.

Lemme 25.23 ([349]).
Soit f une fonction holomorphe sur $B(z_0, r_0)$. Pour tout $z \in B(z_0, r)$ (avec $r < r_0$) nous avons

$$|f'(z)| \leqslant \frac{r}{\big(r - |z - z_0|\big)^2} \max\big\{f(z_0 + re^{i\theta})\big\}_{\theta\in\mathbb{R}}. \tag{25.97}$$

Démonstration. Par translation nous pouvons supposer que $z_0 = 0$. Étant donné que f est holomorphe, elle admet un développement en séries entières

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \tag{25.98}$$

et nous notons $M = \max\{f(z) \text{ tel que } z \in \overline{B(0, r)}\}$. Nous avons $r^n|a_n| \leqslant M$. Par conséquent

$$|f'(z)| = \left|\sum_{n=1}^{\infty} n a_n z^{n-1}\right| \tag{25.99a}$$

$$\leqslant \frac{1}{r} \sum r^n |a_n| n \left(\frac{|z|}{r}\right)^{n-1} \tag{25.99b}$$

$$< \frac{M}{r} \sum n \left(\frac{|z|}{r}\right)^{n-1}. \tag{25.99c}$$

À ce point nous devons utiliser la série de l'exemple 14.110. Nous avons alors

$$|f'(z)| \leqslant \frac{M}{r} \frac{1}{\left(1 - \frac{|z|}{r}\right)^2} = \frac{Mr}{(r - |z|)^2}. \tag{25.100}$$

□

Théorème 25.24 (Holomorphie sous l'intégrale[349]).
Soit un espace mesuré (Ω, μ), un ouvert A dans $\mathbb{C}$ et une fonction $f\colon A \times \Omega \to \mathbb{C}$. Nous voulons étudier la fonction

$$F(z) = \int_\Omega f(z, \omega) d\mu(\omega) \tag{25.101}$$

pour tout $z \in A$. Nous supposons que

(1) la fonction $f(., \omega)$ est holomorphe sur A pour chaque ω.

(2) La fonction $f(z, .)$ est mesurable sur (Ω, μ).

(3) Pour tout compact $K \subset A$, il existe une fonction $g_K\colon \Omega \to \mathbb{R}$ telle que $|f(z, \omega)| \leqslant g_K(\omega)$ et telle que

$$\int_\Omega g_K(\omega) d\mu(\omega) \tag{25.102}$$

existe.

Alors la fonction F est holomorphe et

$$F'(z) = \int_\Omega \frac{\partial f}{\partial z}(z, \omega) d\mu(\omega). \tag{25.103}$$

Démonstration. Soient $z_0 \in A$ et $r > 0$ tels que $K = \overline{B(z_0, r)} \subset A$. Pour chaque $\omega \in \Omega$ nous considérons la fonction

$$\begin{aligned} f_\omega\colon \overline{B(z_0, r)} &\to \mathbb{C} \\ z &\mapsto f(z, \omega). \end{aligned} \tag{25.104}$$

Étant donné que $\overline{B(z_0, r)}$ est compacte, la fonction $|f_\omega|$ est majorée par un nombre que nous notons $f_K(\omega)$ qui est indépendant de z (pour autant que $z \in K$). Nous désignons par $S(z_0, r)$ la frontière de la boule $B(z_0, r)$. Étant donné que la majoration est valable sur $\overline{B(z_0, r)}$, nous avons en particulier

$$|f_\omega(z)| \leqslant f_K(\omega) \tag{25.105}$$

pour tout $z \in S$. En utilisant la lemme 25.23 nous avons

$$|f'_\omega(z)| \leqslant \frac{r}{(r - |z - z_0|)^2} \max\{f(z_0 + re^{i\theta})\}_{\theta \in \mathbb{R}} \tag{25.106a}$$

$$\leqslant \frac{r f_K(\omega)}{(r - |z - z_0|)^2}. \tag{25.106b}$$

Cette majoration est valable pour tout $z \in B(z_0, r)$. Si nous supposons de plus que $z \in B(z_0, r/2)$ nous avons

$$|f'(z)| \leqslant \frac{r f_K(\omega)}{\left(r - \frac{r}{2}\right)^2} = \frac{4}{r} f_K(\omega). \tag{25.107}$$

Étant donné que la boule $B(z_0, r/2)$ est convexe, la fonction f_ω est Lipschitz et pour tout $h \in \mathbb{C}$ tel que $|h| < r/2$ nous avons

$$\left| \frac{f_\omega(z_0 + h) - f_\omega(z_0)}{h} \right| \leqslant \frac{4 f_K(\omega)}{r}. \tag{25.108}$$

Soit maintenant une suite (h_n) qui converge vers 0 dans $\mathbb{C}$. Nous considérons la suite de fonctions correspondantes

$$g_n(\omega) = \frac{f(z_0 + h_n, \omega) - f(z_0, \omega)}{h_n}. \tag{25.109}$$

Cette suite de fonctions vérifie la convergence ponctuelle

$$g_n(\omega) \to \frac{\partial f}{\partial z}(z_0, \omega). \tag{25.110}$$

De plus g_n est une fonction (de ω) dominée par $\frac{4f_K}{r}$ qui est intégrable. Par conséquent le théorème de la convergence dominée[14] nous indique que

$$\int_\Omega g_n(\omega)d\mu(\omega) \to \int_\Omega \frac{\partial f}{\partial z}(z_0,\omega)d\mu(\omega), \tag{25.111}$$

tandis que

$$F'(z) = \lim_{n\to\infty} \frac{F(z_0+h_n)-F(z_0)}{h_n} = \lim_{n\to\infty}\int_\Omega g_N(\omega)d\mu(\omega). \tag{25.112}$$

□

Corolaire 25.25.
Si f est une fonction holomorphe sur l'ouvert ξ contenant la fermeture de la boule $B(z_0,r)$, alors pour tout z dans $B(z_0,\rho)$ ($\rho<r$) les dérivées de f s'expriment pas la formule suivante :

$$f^{(k)}(z) = \frac{1}{2\pi i}\int_{\partial B(z_0,r)} \frac{f(\xi)}{(\xi-z)^{k+1}}d\xi. \tag{25.113}$$

Démonstration. Nous faisons par récurrence.

Pour la dérivée première Nous appliquons le théorème 25.24 à la fonction

$$g(z,\xi) = \frac{f(\xi)}{\xi-z} \tag{25.114}$$

avec $\xi = \partial B(z_0,r)$ et $A = B(z_0,\rho)$ avec $\rho<r$. Étant donné que f est holomorphe, elle est continue et donc bornée sur tout compact $K\subset A$ par une constante M (qui dépend du compact choisi). D'autre part, nous avons toujours $|\xi-z|>r-\rho$ et donc

$$|g(z,\xi)| \leqslant \frac{M}{r-\rho}. \tag{25.115}$$

La fonction constante $g_K = \frac{M}{r-\rho}$ est évidemment intégrable. Le théorème conclut que f est holomorphe (cela, nous le savions déjà[15]), et

$$f'(z) = \frac{1}{2i\pi}\int_{\partial B}\frac{f(\xi)}{(\xi-z)^2}d\xi. \tag{25.116}$$

Les dérivées suivantes Pour la récurrence[436] nous supposons que

$$f^{(k)}(z) = \frac{k!}{2i\pi}\int_{\partial B(z_0,r)}\frac{f(\xi)}{(\xi-z)^{k+1}}d\xi, \tag{25.117}$$

et nous tentons de calculer $f^{(k+1)}(z)$. Pour cela nous paramétrons l'intégrale de façon très usuelle :

$$f^{(k)}(z) = \frac{k!}{2\pi i}\int_0^{2\pi}\frac{f(re^{it})}{(re^{it}-z)}ire^{it}dt. \tag{25.118}$$

Nous permettons de permuter la $\mathbb{C}$-dérivation (par rapport à z) et l'intégrale en vertu de la proposition 25.16 appliquée à la fonction

$$g(t,z) = \frac{f(re^{it})}{(re^{it}-z)^k+1}ire^{it}. \tag{25.119}$$

Cela donne

$$f^{(k+1)}(z) = \frac{k!}{2i\pi}\int_0^{2\pi} f(re^{it})ire^{it}\frac{k+1}{(re^{it}-z)^{k+1}}dt = \frac{(k+1)!}{2i\pi}\int_{\partial B}\frac{f(\xi)}{(\xi-z)^{k+2}}d\xi. \tag{25.120}$$

14. Lebesgue, théorème 13.184.
15. Et cela fournit une preuve alternative à la réciproque du théorème de Cauchy : une fonction continue qui vérifie la formule de Cauchy est holomorphe.

□

Théorème 25.26.
Si f est une fonction holomorphe sur le disque ouvert $B(z_0, R)$ alors

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(z_0)}{n!}(z - z_0)^n \tag{25.121}$$

et cette série converge uniformément sur tout compact.

Démonstration. Sans perte de généralité nous supposons que $z_0 = 0$. La formule de Cauchy (théorème 25.13) fournit, pour $z \in B(0, R)$,

$$f(z) = \frac{1}{2\pi i}\int_{\partial B} \frac{f(\xi)}{\xi - z}d\xi = \frac{1}{2\pi i}\int_{\partial B} \frac{f(\xi)}{1 - (z/\xi)}\frac{d\xi}{\xi}. \tag{25.122}$$

En particulier notons que $z \in B(0, R)$ alors que ξ est sur le bord de cette boule ouverte. Donc $|\xi| > |z|$ pour tous les ξ et z qui interviennent. Nous utilisons la série géométrique

$$\frac{1}{1 - (z/\xi)} = \sum_{n=0}^{\infty} \left(\frac{z}{\xi}\right)^n. \tag{25.123}$$

Permuter une intégrale et une somme En utilisant la mesure de comptage[16] sur $\mathbb{N}$ (qui est σ-finie), nous pouvons écrire

$$\int_{\partial B} \sum_{n=0}^{\infty} \frac{z^n f(\xi)}{\xi^{n+1}}d\xi = \int_{\partial B} \left(\int_{\mathbb{N}} g(\xi, n)dm(n)\right) d\xi \tag{25.124}$$

où

$$\begin{aligned} g\colon \partial B \times \mathbb{N} &\to \mathbb{C} \\ (\xi, n) &\mapsto \frac{z^n f(\xi)}{\xi^{n+1}}. \end{aligned} \tag{25.125}$$

Nous allons permuter les intégrales en utilisant le théorème de Fubini, selon la procédure décrite en 13.263. Nous commençons par l'intégrale sur $\mathbb{N}$:

$$\int_{\mathbb{N}} |g(n, \xi)| = |\frac{f(\xi)}{\xi}| \sum_{n\in\mathbb{N}} |\frac{z}{\xi}|^n = \frac{1}{R}|f(\xi)|\frac{1}{1 - |z|/R}. \tag{25.126}$$

Ici nous avons utilisé $|\xi| = R$. Notons que z est fixé depuis longtemps à l'intérieur de la boule de rayon R de telle sorte que $|z/\xi|$ est une constante strictement inférieure à 1.

L'intégrale sur $\xi \in \partial B$ n'a pas à être effectuée explicitement : nous nous contentons de prouver qu'elle est finie. La fonction f est continue sur le compact ∂B. Cela parce que B est une boule fermée dans l'ouvert Ω sur lequel f est continue. Au final l'expression à droite de (25.126) est bornée sur le compact ∂B et son intégrale donne un nombre fini.

Tout ceci pour invoquer le corolaire 13.261 qui nous indique que $g \in L^1(\mathbb{N} \times \partial B)$.

Une fois g intégrable sur l'espace produit $\mathbb{N} \times \partial B$, nous pouvons utiliser Fubini 13.262 pour permuter les intégrales.

Une fois la somme et l'intégrale permutées, nous avons

$$f(z) = \frac{1}{2\pi i}\sum_{n=0}^{\infty}\int_{\partial B} \frac{z^n f(\xi)}{\xi^{n+1}} = \sum_{n=0}^{\infty}\left(\frac{1}{2\pi i}\int_{\partial B}\frac{f(\xi)}{\xi^{n+1}}\right) z^n. \tag{25.127}$$

Nous devons maintenant montrer que ce qui se trouve dans la grande parenthèse vaut $f^{(n)}(0)/n!$. Cela est immédiat en comparant avec la formule (25.113).

□

16. Définition 13.230.

Proposition 25.27 (Morera [437]).
Soit Ω ouvert dans $\mathbb{C}$ et f continue. Si

$$\int_{\partial T} f = 0 \tag{25.128}$$

pour tout triangle (plein) T contenu dans Ω, alors f est holomorphe sur Ω.

Démonstration. Il est suffisant de prouver que f est holomorphe sur toute boule ouverte $B(a,r)$ inclue dans Ω. Nous posons, pour tout $z \in B(a,r)$,

$$F(z) = \int_{[p,z]} f, \tag{25.129}$$

et nous considérons le chemin triangulaire $a \to z \to z+h \to a$ où $h \in \mathbb{C}$ est choisi assez petit pour que $z+h \in B(a,r)$. L'intégrale sur le triangle étant nulle, nous avons

$$0 = \int_{a\to z} f + \int_{z\to z+h} f + \int_{z+h\to a} f, \tag{25.130}$$

c'est-à-dire

$$F(z+h) - F(z) = \int_{z\to z+h} f. \tag{25.131}$$

En paramétrant le chemin par $z+th$ avec $t \in [0,1]$, et en tenant compte de la remarque 19.44,

$$F'(z) = \lim_{h\to 0} \frac{F(z+h)-F(z)}{h} \tag{25.132a}$$

$$= \lim_{h\to 0} \frac{1}{h} \int_0^1 f(z+th)hdt, \tag{25.132b}$$

ce qui prouve que F est dérivable et $F' = f$. Par définition (11.248), F est holomorphe, et donc C^∞ par le théorème 25.14. Du coup f est également C^∞ et donc en particulier holomorphe. □

25.2.1 Mesure de Radon

Définition 25.28.
*Une **mesure de Radon** sur un compact K de $\mathbb{C}$ est une forme linéaire continue sur $C(K)$. Si μ est une mesure de Radon, on définit la **transformée de Cauchy** de μ par*

$$\begin{aligned} \hat{\mu}\colon \mathbb{C}\backslash K &\to \mathbb{C} \\ z &\mapsto -\frac{1}{\pi}\mu\left(\frac{1}{\xi - z}\right). \end{aligned} \tag{25.133}$$

Théorème 25.29.
Si μ est une mesure de Radon sur K alors $\hat{\mu}$ est infiniment $\mathbb{C}$-dérivable sur $\Omega = \mathbb{C}\backslash K$ et nous avons

$$\hat{\mu}^{(n)}(z) = -\frac{n!}{\pi}\mu\left(\frac{1}{(\xi - z)^{n+1}}\right). \tag{25.134}$$

Cette théorie permet de fournir une démonstration plus technologique du corolaire 25.25.

Lemme 25.30.
Si f est holomorphe sur Ω et si B est une boule fermée dans Ω alors pour tout $z \in \mathrm{Int}(B)$ nous avons

$$f^{(k)}(z) = \frac{k!}{2i\pi} \int_{\partial B} \frac{f(\xi)}{(\xi - z)^{k+1}} d\xi. \tag{25.135}$$

Démonstration. Appliquer le théorème 25.29 à la mesure de Radon

$$\mu(\phi) = \int_{\partial B} \phi(\xi) d\xi. \tag{25.136}$$

□

Tout ce petit monde à propos de la mesure de Radon permet également de redémontrer que

$$\left(\frac{1}{2\pi i}\int_{\partial B}\frac{f(\xi)}{\xi^{n+1}}\right) = f^{(n)}(0)/n!, \tag{25.137}$$

comme nous l'avons déjà fait autour de l'équation (25.127). Nous utilisons le théorème de Radon 25.29 à la mesure

$$\mu(\phi) = \int_{\partial B}\phi(\xi)d\xi. \tag{25.138}$$

La transformée de Cauchy est

$$\hat{\mu}(z) = -\frac{1}{\pi}\mu\left(\frac{1}{\xi - z}\right) = -\frac{1}{\pi}\int_{\partial B}\frac{1}{\xi - z}d\xi, \tag{25.139}$$

et le théorème assure que

$$\hat{\mu}^{(n)}(z) = -\frac{n!}{\pi}\mu\left(\frac{1}{(\xi - z)^{n+1}}\right) = -\frac{n!}{\pi}\int_{\partial B}\frac{1}{(\xi - z)^{n+1}}d\xi. \tag{25.140}$$

En comparant la formule (25.139) avec la formule de Cauchy nous voyons que $\hat{\mu}(z) = -2if(z)$. Par conséquent

$$f^{(n)}(z) = -\frac{1}{2i}\hat{\mu}^{(n)}(z) = \frac{n!}{2\pi i}\int_{\partial B}\frac{1}{(\xi - z)^{n+1}}d\xi, \tag{25.141}$$

et

$$f^{(n)}(0) = \frac{n!}{2\pi i}\int_{\partial B}\frac{1}{\xi^{n+1}}d\xi. \tag{25.142}$$

25.3 Conditions équivalentes à l'holomorphie

Nous nous proposons de lister les conditions que nous avons vues être équivalentes à l'holomorphie.

Théorème 25.31.
Soient Ω un ouvert de $\mathbb{C}$ et $f\colon \Omega \to \mathbb{C}$ une fonction continue. Les conditions suivantes sont équivalentes.

(1) f est holomorphe.

(2) Pour tout triangle (plein) T contenu dans Ω, $\int_T f = 0$.

(3) f est $\mathbb{C}$-dérivable.

(4) f est C^∞

(5) $\frac{\partial f}{\partial \bar{z}} = 0$; ce sont les équations de Cauchy-Riemann.

(6) La 1-forme différentielle $f(z)dz$ est localement exacte.

(7) Pour toute boule $B(a,r)$ contenue dans Ω nous avons

$$f(a) = \frac{1}{2\pi i}\int_{\partial B(a,r)}\frac{f(z)}{z - a}dz. \tag{25.143}$$

La fonction f est holomorphe en z_0 si et seulement si il existe un voisinage $B(z_0,r)$ de z_0 et des nombres a_k tels que sur la boule,

$$f(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n. \tag{25.144}$$

Dans ce cas, f est holomorphe sur toute la boule.

Démonstration. (1) implique (2) est le lemme de Goursat 25.5. (2) implique (1) est le théorème de Morera 25.27.

(3) est la définition de l'holomorphie, définition 11.248.

(4) implique (1) est un a fortiori sur la définition. (1) implique (4) est contenu dans le théorème de développement en série entière 25.14.

L'équivalence entre (5) et l'holomorphie est le théorème 25.2.

L'équivalence entre (6) et (1) est la proposition 25.17.

L'équivalence entre (1) et (7) est d'une part le théorème 25.14 et d'autre part le corolaire 25.15.

En ce qui concerne la dernière affirmation, si f est holomorphe en z_0, alors le théorème 25.14(1) donne la série. Si au contraire nous avons la série, la proposition 25.7 nous donne le résultat. □

25.4 Singularités, pôles et méromorphe

Définition 25.32.
*Si f est holomorphe sur un ouvert Ω, alors une **singularité** de f est un point isolé du bord de Ω.*

*(1) La singularité est **effaçable** si la fonction f s'y prolonge en une fonction holomorphe.*

*(2) La singularité Z est un **pôle** d'ordre k de f si elle n'est pas effaçable et si la fonction $z \mapsto (z-Z)^k f(z)$ se prolonge en une fonction holomorphe en Z.*

Exemple 25.33.
La fonction

$$z \mapsto \frac{\sin(z)}{z} \tag{25.145}$$

n'est pas définie en $z = 0$, mais elle s'y prolonge en une fonction continue en posant $f(0) = 1$. △

Proposition 25.34.
Une singularité de f est un pôle si et seulement si

$$\lim_{z\to Z} f(z) = \infty. \tag{25.146}$$

Le théorème suivant compète la proposition 25.21.

Théorème 25.35 (Prolongement de Riemann)**.**
Soient $f\colon \Omega \to \mathbb{C}$ et Z une singularité de f. Nous avons équivalence de

(1) la singularité Z est effaçable ;

(2) f possède un prolongement continue en Z ;

(3) il existe un voisinage épointé de Z sur lequel f est bornée ;

(4) $\lim_{z\to Z}(z-Z)f(z) = 0$.

Définition 25.36 (Fonction méromorphe[438])**.**
*Soient $\mathcal{U}$ un ouvert de $\mathbb{C}$ et $\{p_i\}$ une suite de points dans $\mathcal{U}$ sans points d'accumulation (éventuellement il y a un nombre fini de p_i). Si la fonction f est holomorphe sur $\mathcal{U}\backslash\{p_i\}$ et si chaque p_i est un point régulier ou un pôle de f, alors nous disons que f est **méromorphe** sur $\mathcal{U}$.*

Proposition 25.37.
Soient Ω un ouvert de $\mathbb{C}$ et $f_n\colon \Omega \to \mathbb{C}$ une suite de fonctions telles que pour tout compact K de Ω il existe $N_K \geqslant 0$ tel que

(1) f_n n'a pas de pôle dans K dès que $n \geqslant N_K$;

(2) la série $\sum_{n\geqslant N_K} f_n$ converge uniformément sur K.

Alors

(1) La fonction

$$f(z) = \sum_{n=0}^{\infty} f_n(z) \tag{25.147}$$

est méromorphe sur Ω et ses pôles sont l'union de ceux des f_n.

(2) Nous pouvons permuter la somme et la dérivée :

$$f'(z) = \sum_{n=0}^{\infty} f_n'(z). \tag{25.148}$$

Théorème 25.38 (Série de Laurent)**.**
Soient C une couronne de rayons $r_1 < r_2$ centrée en zéro et une fonction f holomorphe dans cette couronne. Alors nous avons la ***série de Laurent***

$$f(z) = \sum_{n\in\mathbb{Z}} a_n z^n. \tag{25.149}$$

(1) Cette série converge uniformément sur tout compact de C.

(2) Les coefficients sont donnés par

$$a_n = \frac{1}{2\pi i}\int_\gamma \frac{f(z)}{z^{n+1}}dz \tag{25.150}$$

où γ est un cercle centré en zéro.

(3) Ce développement en série est unique.

(4) La valeur des a_n ne dépend pas du choix du rayon du cercle γ.

25.5 Fonctions d'Euler

Théorème 25.39 (Prolongement méromorphe de la fonction Γ d'Euler[54])**.**
Nous considérons la formule

$$\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt. \tag{25.151}$$

Alors

(1) Cette formule définit une fonction holomorphe sur

$$\mathcal{P} = \{z \in \mathbb{C} \text{ tel que } \Re(z) > 0\}. \tag{25.152}$$

(2) La fonction $\Gamma\colon \mathcal{P} \to \mathbb{C}$ admet un unique prolongement méromorphe sur $\mathbb{C}$, lequel a des pôles sur les entiers négatifs.

Démonstration. **Holomorphie sous l'intégrale** Pour étudier l'holomorphie de la fonction Γ sur $\mathcal{P}$ nous utilisons le théorème 25.24.

Nous considérons la fonction

$$\begin{aligned} g\colon \mathcal{P}\times\mathbb{R}^+ &\to \mathbb{C}\\ (z,t) &\mapsto e^{-t}z^{z-1} \end{aligned} \tag{25.153}$$

et nous commençons par montrer que c'est holomorphe en z pour chaque $t > 0$ fixé. Nous le vérifions par le critère de $\partial_{\bar{z}f=0}$ [17] et en nous souvenant que $t^i = e^{\ln(t^i)} = e^{i\ln(t)}$. Nous obtenons rapidement que

$$\frac{\partial g}{\partial \bar{z}} = 0. \tag{25.154}$$

Le fait que la fonction $t \mapsto g(z,t)$ soit mesurable pour tout z est d'accord.

17. Théorème 25.2.

Et enfin soit K compact dans $\mathcal{P}$. Il faut trouver une fonction $g_K(t)$ intégrable sur $[0,\infty[$ telle que pour tout $z \in K$ et $t \in [0,\infty[$ nous ayons $|f(z,t) \leqslant g(t)|$. Pour cela nous majorons séparément les parties $t \in]0,1[$ et $t \geqslant 1$.
Soit donc K compact dans $\mathcal{P}$; nous posons $M = \max_{z\in K} \Re(z)$ et $\epsilon = \min_{z\in K} \Re(z)$.
Si $t \in]0,1[$ alors nous avons

$$e^{-t}t^{z-1} = e^{-t}e^{(z-1)\ln(t)}, \tag{25.155}$$

de telle façon à que que

$$\begin{aligned}
|e^{-t}t^{z-1}| &\leqslant |e^{(x-1+iy)\ln(t)}| && (25.156a)\\
&= |e^{(\Re(z)-1)\ln(t)}| && (25.156b)\\
&= |t^{\Re(z)-1}| && (25.156c)\\
&\leqslant |t^{\epsilon-1}| && (25.156d)\\
&= \frac{1}{t^{1-\epsilon}}. && (25.156e)
\end{aligned}$$

Cette dernière fonction est intégrable sur $]0,1[$.
Nous considérons maintenant $t \geqslant 1$. Dans ce cas nous avons

$$|e^{-t}z^{z-1}| = e^{-t}t^{\Re(z)-1} \leqslant e^{-t}t^{M-1}. \tag{25.157}$$

Cette dernière fonction est un produit d'une exponentielle décroissante avec un polynôme. C'est donc intégrable entre 1 et l'infini.
La fonction g_K que nous considérons est donc

$$g_K(t) = \begin{cases} \frac{1}{t^{1-\epsilon}} & \text{si } t < 1\\ \text{borné} & \text{si } 1 \leqslant t \leqslant b\\ e^{-t}t^{M-1} & \text{si } t > b. \end{cases} \tag{25.158}$$

Cela est une fonction intégrable sur $]0\infty,[$ et qui majore f uniformément en z sur le compact K de $\mathcal{P}$. Le théorème 25.24 nous permet donc de conclure que

$$\Gamma(z) = \int_0^\infty f(z,t)dt \tag{25.159}$$

est holomorphe en z sur $\mathcal{P}$ et que

$$\Gamma'(z) = \int_0^\infty \frac{\partial f}{\partial z}(z,t)dt. \tag{25.160}$$

En deux morceaux Nous passons maintenant à la seconde partie du théorème. Pour $z \in \mathcal{P}$ nous coupons l'intégrale en deux :

$$\Gamma(z) = \int_0^1 e^{-t}t^{z-1}dt + \int_1^\infty e^{-t}t^{z-1}dt \tag{25.161}$$

Première partie Nous commençons par parler de la première partie : $\int_0^1 e^{-t}t^{z-1}dt$ dans laquelle nous voulons utiliser le développement en série de l'exponentielle e^{-t}. Nous devons donc traiter

$$\int_0^1 \sum_{n=0}^\infty \frac{(-1)^n}{n!}t^{n+z-1}dt. \tag{25.162}$$

Nous allons permuter la somme avec l'intégrale à l'aide du théorème de Fubini 13.262 en posant la fonction

$$g(n,t) = \frac{(-1)^n}{n!}t^{n+z-1} \tag{25.163}$$

et en considérant le produit entre la mesure de Lebesgue sur $\mathbb{C}$ et la mesure de comptage sur $\mathbb{N}$, c'est-à-dire que nous étudions

$$\int_0^1 \int_{\mathbb{N}} g(n,t)dndt. \tag{25.164}$$

Pour permuter il suffit de prouver que $|g|$ est intégrable pour la mesure produit, c'est-à-dire que

$$\int_0^1 \int_{\mathbb{N}} \left| \frac{(-1)^n}{n!} t^{n+z-1} \right| < \infty. \tag{25.165}$$

Nous avons $|t^z = t^{\Re(z)}|$, donc

$$\sum_{n=0}^{\infty} \left| \frac{t^{n+z-1}}{n!} \right| = t^{\Re(z)-1} \sum_{n=0}^{\infty} \frac{t^n}{n!} = t^{\Re(z)-1} e^t. \tag{25.166}$$

Étant donné que nous avons fixé $z \in \mathcal{P}$, nous avons $\Re(z) - 1 > -1$ et donc $t^{\Re(z)-1}$ est intégrable entre 0 et 1. La partie e^t se majore sur $[0,1]$ par une constante quelconque. Nous avons donc payé le droit d'inverser la somme et l'intégrale :

$$\int_0^1 e^{-t} t^{z-1} dt = \sum_{n=0}^{\infty} \int_0^1 \frac{(-1)^n}{n!} t^{n+z-1} dt = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} [t^{n+z}]_0^1 = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(n+z)}. \tag{25.167}$$

Nous avons donc l'intéressante formule suivante, valable pour tout $z \in \mathcal{P}$:

$$\Gamma(z) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(n-z)} + \int_1^{\infty} e^{-t} t^{z-1} dt. \tag{25.168}$$

Prolongation de la première partie Nous voudrions montrer maintenant que la fonction

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{n!(n-z)} \tag{25.169}$$

est méromorphe sur $\mathbb{C}$ avec des pôles en les entiers négatifs. Pour cela nous considérons la suite de fonctions

$$f_n(z) = \frac{(-1)^n}{n!(z+n)} \tag{25.170}$$

et nous allons utiliser la proposition 25.37. Si $n \geqslant 0$, la fonction f_n est méromorphe sur $\mathbb{C}$ avec un pôle simple en $z = -n$. Soit K compact de $\mathbb{C}$ et N_K tel que $K \subset \overline{B(0, N_K)}$. Pour $n \geqslant N_K + 1$, la fonction f_n n'a pas de pôle dans K et de plus pour tout $z \in K$ nous avons

$$|z+n| = |z-(-z)| \geqslant \big|n - |z|\big| \geqslant n - |z| \geqslant n - N_K, \tag{25.171}$$

et par conséquent

$$|f_n(z)| \leqslant \frac{1}{n!(n-N)}, \tag{25.172}$$

ou pour le dire de façon plus snob :

$$\|f_n\|_{\infty,K} \leqslant \frac{1}{n!(n-N)}, \tag{25.173}$$

dont la série converge. Cela signifie que la série $\sum_{n \geqslant N} f_n$ converge normalement [18] sur K, donc la fonction

$$f(z) = \sum_{n=0}^{\infty} f_n(z) \tag{25.174}$$

est une fonction méromorphe dont les pôles sont ceux des f_n, c'est-à-dire les entiers négatifs (proposition 25.37).

18. Définition 10.52.

La seconde partie Nous allons à présent prouver que la fonction

$$g(z) = \int_1^\infty e^{-t} t^{z-1} dt \tag{25.175}$$

est holomorphe sur $\mathbb{C}$. Pour cela nous considérons la fonction de deux variables $f(z,t) = e^{-t}t^{z-1}$ et nous utilisons le théorème d'holomorphie sous l'intégrale 25.24. D'abord pour z_0 fixé dans $\mathbb{C}$ nous avons

$$\int_1^\infty |e^{-t}t^{z_0-1}| \leqslant \int_1^\infty e^{-t}t^{\Re(z_0)-1}dt, \tag{25.176}$$

donc l'intégrale converge parce que c'est polynôme contre exponentielle. Par ailleurs pour chaque t_0 fixé sur $[0,\infty[$, la fonction $z \mapsto e^{-t_0}t_0^{z-1}$ est holomorphe sur $\mathbb{C}$ comme en témoigne le calcul suivant :

$$\frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) t_0^{x+iy-1} = 0. \tag{25.177}$$

Et enfin si K est compact dans $\mathbb{C}$ nous avons

$$|f(z,t)| = |e^{-t}t^{z-1}| = e^{-t}|t^{\Re(z)-1}| \leqslant e^{-t}t^{M-1} \tag{25.178}$$

où $M = \max_{z\in K} \Re(z)$. Nous en déduisons que la fonction

$$z \mapsto \int_1^\infty e^{-t}t^{z-1}dt \tag{25.179}$$

est une fonction holomorphe sur $\mathbb{C}$.

Conclusion Au final nous avons prouvé que la fonction Γ d'Euler admet le prolongement méromorphe sur $\mathbb{C}$ donné par

$$\Gamma(z) = \sum_{n=0}^\infty \frac{(-1)^n}{n!(z+n)} + \int_1^\infty e^{-t}t^{z-1}dt. \tag{25.180}$$

□

25.5.1 Euler et factorielle

Proposition 25.40.
Nous avons la formule $\Gamma(n) = (n-1)!$ *pour tout* $n \in \mathbb{N}$.

Démonstration. Nous partons de la formule

$$\Gamma(n) = \int_0^\infty e^{-t}t^{n-1}dt \tag{25.181}$$

que nous intégrons par partie en posant

$$\begin{aligned} u &= t^{n-1} & u' &= (n-1)t^{n-1} \\ v &= e^{-t} & v' &= -e^{-t}. \end{aligned} \tag{25.182}$$

Les termes au bord s'annulent (ici il y a un passage à la limite qui n'est pas écrit) et nous trouvons

$$\Gamma(n) = \int_0^\infty (n-1)e^{-t}t^{n-2}dt = (n-1)\Gamma(n-1). \tag{25.183}$$

Pour conclure il suffit de remarquer que

$$\Gamma(1) = \int_0^\infty = -[e^{-t}]_0^\infty = 1. \tag{25.184}$$

□

25.6 Partition d'un entier en parts fixées

Proposition 25.41 ([54]).
Soient $a_1, \dots, a_k \in \mathbb{N}^*$ des entiers premiers entre eux dans leur ensemble. Pour $n \geqslant 1$ nous posons

$$u_n = \text{Card}\left\{(x_1, \dots, x_k) \in \mathbb{N}^* \text{ tel que } \sum_{i=1}^{k} a_i x_i = n\right\}, \tag{25.185}$$

et $u_0 = 1$.

Alors nous avons l'équivalence de suite (pour $n \to \infty$) :

$$u_n \sim \frac{1}{a_1 \dots a_k} \frac{n^{k-1}}{(k-1)!}. \tag{25.186}$$

Démonstration. Pour chacun des $i \in \{1, \dots, k\}$ nous considérons la série entière

$$\sum_{x=0}^{\infty} z^{xa_i} = \sum_k (z^{a_i})^x. \tag{25.187}$$

Étant donné que $|z^{a_i}| < 1$ si et seulement si $|z < 1|$, cette série a un rayon de convergence égal à 1. Nous allons calculer le produit de Cauchy de ces k séries, en nous souvenant que le théorème 14.15 nous assure que la série résultante aura un rayon de convergence au moins égal à 1 et vaudra le produit des différentes séries.

Le coefficient de z^n dans cette série vaut

$$\sum_{\substack{x \in \mathbb{N}^k \\ \sum x_i a_i = n}} 1 = u_n \tag{25.188}$$

parce que dans chacune des séries, le coefficient de tous les z^{xa_i} est 1. Nous définissons la fonction

$$f(z) = \sum_{n=0}^{\infty} u_n z^n = \prod_{i=1}^{k} \left(\sum_{x=0}^{\infty} z^{xa_i} \right) = \prod_{i=1}^{k} \frac{1}{1 - z^{a_i}}. \tag{25.189}$$

La fonction f existe sur $|z| < 1$ parce que nous venons de voir qu'elle peut s'exprimer comme un produit de Cauchy ; et la dernière égalité est simplement la somme de la série harmonique. D'autre part la fonction f est la série génératrice de la suite (u_n).

Nous sommes en présence d'une fonction ayant des pôles aux racines $a_1, \dots, a_k$e de l'unité. Étant donné que 1 est une racine de l'unité de tous les ordres, le pôle en $z = 1$ est de multiplicité k. Les autres pôles sont de multiplicité strictement inférieure ; en effet soit $\omega \in \mathbb{C}$ tel que $\omega^{a_i} = 1$ pour tout i. Alors Bezout [19] nous donne des entiers $v_i \in \mathbb{Z}$ tels que $\sum_i v_i a_i = 1$. Alors nous avons

$$\omega = \omega^{\sum v_i a_i} = \prod_{i=1}^{k} (\omega^{a_i})^{v_i} = 1. \tag{25.190}$$

Donc nous voyons que 1 est le seul à être racine de tous les ordres en même temps. Nous notons

$$P = \{\omega_1, \dots, \omega_p\} \tag{25.191}$$

l'ensemble des pôles avec $\omega_1 = 1$. Par ailleurs la fonction f est une fraction rationnelle dont nous connaissons les racines du dénominateur (ce sont les ω_i) et à peu près leurs ordres. Nous utilisons le truc de la décomposition en fractions simples en séparant le terme de puissance k qui n'existe que pour la racine $\omega_1 = 1$:

$$f(z) = \frac{\alpha}{(1-z)^k} + \sum_{i=1}^{p} \sum_{j=1}^{k-1} \frac{c_{ij}}{(\omega_i - z)^j}. \tag{25.192}$$

19. Théorème 3.13.

Ce développement est valable pour tout $|z| < 1$. Nous considérons maintenant $\omega \in P$ et $j \in \mathbb{N}$ et nous étudions la fonction

$$g(z) = \frac{1}{\omega - z}. \tag{25.193}$$

Un rapide calcul (par exemple par récurrence) montre que

$$g^{(k)}(z) = \frac{k!}{(\omega - z)^{k+1}}, \tag{25.194}$$

et étant donné que $|\omega| = 1$ nous pouvons écrire la série

$$\frac{1}{\omega - z} = \sum_{k=0}^{\infty} \frac{z^k}{\omega^{k+1}}, \tag{25.195}$$

valable pour $|z| < 1$. Ce qui nous intéresse, c'est d'exprimer une série pour $1/(\omega - z)^j$; et voyant (25.194), nous voyons qu'il suffit de calculer les dérivées de la série de g. Nous dériver terme à terme à l'intérieur du rayon de convergence. Avec quelques abus d'écriture, et en utilisant la bête formule (14.79) nous avons [20]

$$\begin{aligned}
\frac{1}{(\omega - z)^j} &= \frac{g^{(j-1)}(z)}{(j-1)!} && (25.196a)\\
&= \frac{1}{(j-1)!}\left(\frac{1}{\omega - z}\right)^{(j-1)} && (25.196b)\\
&= \frac{1}{(j-1)!}\sum_{k=0}^{\infty}\frac{1}{\omega^{k+1}}(z^k)^{(j-1)} && (25.196c)\\
&= \sum_{k=j-\infty}\frac{1}{(j-1)!}\frac{1}{\omega^{k+1}}\frac{k!}{(k-j+1)!}z^{k-j+1} && (25.196d)\\
&= \sum_{n=0}^{\infty}\frac{1}{\omega^{n+j}}\frac{(n+j-1)!}{n!(j-1)!}z^n && (25.196e)\\
&= \sum_{n=0}^{\infty}\frac{1}{\omega^{n+j}}\binom{n+j-1}{n}z^n. && (25.196f)
\end{aligned}$$

Nous pouvons utiliser cela pour récrire la formule (25.192) de façon considérablement plus compliquée :

$$f(z) = \alpha\sum_{n=0}^{\infty}\binom{n+j-1}{n}z^n + \sum_{i=1}^{p}\sum_{j=1}^{k-1}\sum_{n=0}^{\infty}c_{ij}\binom{n+j-1}{n}\frac{z^n}{\omega_i^{n+j}}. \tag{25.197}$$

Mais nous savons que ce f est la série génératrice de la suite (u_n) et que nous pouvons donc utiliser la formule (14.354) pour exprimer les nombres u_l : u_l est simplement le coefficient de z^l divisé par $l!$. C'est-à-dire

$$u_l = \alpha\binom{l+k-1}{l} + \sum_{i=1}^{p}\sum_{j=1}^{k-1}c_{ij}\binom{l+j-1}{l}\frac{1}{\omega_i^{l+j}}. \tag{25.198}$$

Notre boulot est d'examiner le comportement de cela lorsque $l \to \infty$, c'est-à-dire regarder quels sont les puissances de l en présence. Notons que

En ce qui concerne le premier terme, la puissance dominante dans le coefficient binomial est l^{k-1}. Dans les autres termes [21], c'est l^{j-1} qui est de degré moins grand. Donc le comportement de u_l en termes de l est

$$u_l \sim \alpha\frac{l^{k-1}}{(k-1)!}. \tag{25.199}$$

20. À ce niveau j'ai pas exactement le même coefficient binomial que dans [54], mais je n'exclus absolument pas que ce soit moi qui me trompe. Écrivez-moi si vous pouvez infirmer ou confirmer l'erreur. Quoi qu'il en soit, cela ne change pas le résultat asymptotique que nous cherchons.

21. Attention : les termes $i = 1$ ont $\omega_1 = 1$ et il n'est donc pas possible de conclure simplement en disant que $\omega_i^{l-j} \to 0$ pour $l \to \infty$; bien que cela soit vrai pour tous les $i \neq 1$.

Il nous reste à voir ce que vaut α. Pour cela nous repartons de l'expression 25.189 que nous écrivons sous la forme

$$(1-z)^k f(z) = \prod_{i=1}^{k} \frac{1-z}{1-z^{a_i}}. \tag{25.200}$$

Nous reconnaissons l'inverse d'une somme harmonique partielle :

$$(1-z)^k f(z) = \prod_{i=1}^{k} \frac{1}{1+z+z^2+\cdots+z^{a_i-1}}. \tag{25.201}$$

Par ailleurs, nous ne savons pas si $f(1)$ existe parce que son rayon de convergence n'est que de 1 ; et nous savons même qu'elle n'existe pas (parce que ce serait la somme des u_n). Mais nous savons aussi que le pôle de plus grande multiplicité de f est en $z = 1$ et est de multiplicité k. Donc $(1-z)^k f(z)$ devrait converger pour $z \to 1$. Pour tout $|z < 1|$ nous avons

$$(1-z)^k f(z) = \alpha + \sum_{i=1}^{p} \sum_{j=1}^{k-1} c_{ij} \frac{(1-z)^k}{(\omega_i - z)^j}. \tag{25.202}$$

Lorsque $z \to 1$, tous les termes des sommes tendent vers zéro, y compris ceux avec $i = 1$ parce que $j < k$. Il reste donc

$$\lim_{z\to 0} (1-z)^k f(z) = \alpha. \tag{25.203}$$

En calculant la même limite avec (25.201) nous trouvons

$$\lim_{z\to 1} (1-z)^k f(z) = \lim_{z\to 1} \prod_{i=1}^{k} \frac{1}{1+z+z^2+\cdots+z^{a_i-1}} = \frac{1}{a_1 \dots a_k}. \tag{25.204}$$

Donc

$$\alpha = \frac{1}{a_1 \dots a_k}, \tag{25.205}$$

et le résultat est prouvé.

□

25.7 Exponentielle et logarithme complexe

25.7.1 Propriétés de l'exponentielle

Proposition 25.42.
Soit $z \in \mathbb{C}$ fixé. La fonction

$$\begin{aligned} E\colon \mathbb{R} &\to \mathbb{C} \\ t &\mapsto e^{tz} \end{aligned} \tag{25.206}$$

est C^∞, sa dérivée est

$$E'(t) = ze^{tz}. \tag{25.207}$$

La fonction E est développable en série entière (voir définition 14.100) sur $\mathbb{R}$ en $t = 0$ et

$$e^{tz} = \sum_{n=0}^{\infty} \frac{z^n}{n!} t^n. \tag{25.208}$$

Démonstration. Nous fixons $z \in \mathbb{C}$. Par définition 17.5, la série suivante est e^{tz} :

$$f(t) = \sum_{n=0}^{\infty} \frac{z^n}{n!} t^n. \tag{25.209}$$

Cette série a un rayon de convergence infini et la fonction f est donc C^∞ sur $\mathbb{R}$. Nous pouvons la dériver terme à terme :

$$f'(t) = \sum_{n=1}^{\infty} \frac{z^n}{n!} n t^{n-1} = z \sum_{n=1}^{\infty} \frac{z^{n-1}}{(n-1)!} t^{n-1} = z e^{tz}. \tag{25.210}$$

□

Théorème 25.43.
La fonction exponentielle vérifie les propriétés suivantes.

(1) exp *est holomorphe*[22].

(2) $(e^z)' = e^z$.

(3) L'exponentielle est développable en série entière,

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} \tag{25.211}$$

et la série converge normalement sur tout compact de $\mathbb{C}$.

Démonstration. En tant que application $E\colon \mathbb{R}^2 \to \mathbb{C}$, la fonction

$$E(x, y) = e^x(\cos y + i \sin y) \tag{25.212}$$

est C^∞. De plus nous avons

$$\frac{\partial E}{\partial x}(x, y) = e^{x+iy} = E(x, y) \tag{25.213a}$$

$$\frac{\partial E}{\partial y}(x, y) = iE(x, y), \tag{25.213b}$$

et par conséquent la fonction E vérifie les équations de Cauchy-Riemann.

Si r est fixé, par le critère d'Abel appliqué à la suite $r/n!$ nous savons que la série $\sum z^n/n!$ converge normalement sur le compact $B(0, r)$. □

25.7.2 Intégrale de Fresnel

Nous allons calculer l'**intégrale de Fresnel**

$$\int_0^\infty e^{-ix^2} dx = \frac{\sqrt{\pi}}{2} e^{-i\pi/4} \tag{25.214}$$

en suivant la démarche présentée par Wikipédia[439]. Nous commençons par prouver que l'intégrale est convergente en nous contentant de justifier la convergence de

$$\int_0^\infty \sin(x^2) dx. \tag{25.215}$$

Pour chaque $a > 0$ fixé, l'intégrale $\int_0^a \sin(x^2)dx$ ne pose pas de problèmes. Le lemme 19.192 nous permet de passer à la limite ; nous devons donc seulement calculer

$$\lim_{b\to\infty} \int_a^b \sin(x^2) dx \tag{25.216}$$

où a est une constante strictement positive. Nous effectuons une intégration par partie en posant

$$u = \frac{1}{x} \qquad u' = -\frac{1}{x^2} \tag{25.217a}$$

$$v' = x \sin(x) \qquad v = \frac{1 - \cos(x)}{2}. \tag{25.217b}$$

22. Définition 11.248.

Notons que la primitive v a été choisie pour avoir $v(0) = 0$. Nous avons

$$\int_a^b \sin(x^2)dx = \left[\frac{1-\cos(x^2)}{2x}\right]_a^b - \int_a^b \frac{\cos(x^2)-1}{2x^2}dx \tag{25.218}$$

Pour le premier terme nous avons

$$\lim_{b\to\infty}\left[\frac{1-\cos(x^2)}{2x}\right]_a^b = \lim_{b\to\infty}\frac{1-\cos(b^2)}{2b} - \frac{1-\cos(a^2)}{2a} = -\frac{1-\cos(a^2)}{2a}. \tag{25.219}$$

C'est borné. Pour le second terme de (25.218), la fonction

$$\frac{\cos(x^2)-1}{2x^2} \tag{25.220}$$

est majorée par la fonction $1/x^2$ qui est intégrable entre a et ∞.

Nous allons calculer l'intégrale demandée en passant par la fonction

$$f(x) = e^{-z^2} \tag{25.221}$$

définie sur le plan complexe. Nous l'intégrons sur le chemin $\gamma = \gamma_1 + \gamma_2 - \gamma_3$ indiqué à la figure 25.1.

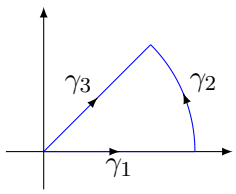


FIGURE 25.1 – Chemin d'intégration pour l'intégrale de Fresnel

Ces chemins sont donnés par

$$\begin{aligned}\gamma_1 \colon [0,R] &\to \mathbb{C}\\ t &\mapsto t,\end{aligned} \tag{25.222}$$

$$\begin{aligned}\gamma_2 \colon [0,\frac{\pi}{4}] &\to \mathbb{C}\\ t &\mapsto Re^{it},\end{aligned} \tag{25.223}$$

$$\begin{aligned}\gamma_3 \colon [0,R] &\to \mathbb{C}\\ t &\mapsto te^{i\pi/4}.\end{aligned} \tag{25.224}$$

Tout d'abord la fonction f est bien holomorphe par le critère du théorème 25.2. Le calcul de $\frac{\partial f}{\partial \bar{z}}$ se fait simplement en posant $f(x,y) = e^{-(x+iy)^2}$. Le calcul est usuel :

```
----------------------------------------------------------------------
| Sage Version 4.8, Release Date: 2012-01-20                         |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: f(x,y)=exp(-(x+I*y)**2)
sage: A=f.diff(x)+I*f.diff(y)
sage: A.simplify_full()
(x, y) |--> 0
```

Nous avons donc

$$0 = \int_\gamma f = \underbrace{\int_0^R e^{-t^2}dt}_{I_1(R)} + \underbrace{\int_0^{\pi/4} e^{-R^2e^{2it}}Rie^{it}dt}_{I_2(R)} + \underbrace{\int_0^R e^{-t^2e^{i\pi/2}}e^{i\pi/4}dt}_{I_3(R)}. \tag{25.225}$$

L'intégrale est nulle pour tout R en vertu de la proposition 25.6. L'intégrale I_1 est une gaussienne et nous avons

$$\lim_{R\to\infty} I_1(R) = \frac{\sqrt{\pi}}{2} \tag{25.226}$$

par l'exemple 13.265. Nous montrons maintenant que $\lim_{R\to\infty} |I_2(R)| = 0$ [23]. D'abord nous majorons en prenant la norme puis nous effectuons le changement de variables $u = 2t$:

$$|I_2(R)| \leqslant \int_0^{\pi/4} Re^{-R^2\cos(2t)}dt \tag{25.227a}$$

$$= \frac{R}{2}\int_0^{\pi/2} e^{-R^2\cos(u)}du. \tag{25.227b}$$

Nous savons que le graphe du cosinus est concave : il reste au dessus de la droite que joint $(0,1)$ à $(\frac{\pi}{2},0)$. Du coup $\cos(u) \geqslant 1 - \frac{2}{\pi}u$ et par conséquent

$$e^{-R^2\cos(u)} \leqslant e^{-R^2(1-\frac{2}{\pi}u)} = e^{R^2(\frac{2}{\pi}u-1)}. \tag{25.228}$$

Nous effectuons l'intégrale

$$|I_2(R)| \leqslant \frac{R}{2}\int_0^{\pi/2} e^{-R^2}e^{\frac{2R^2}{\pi}u}du \tag{25.229a}$$

$$= \frac{R}{2}e^{-R^2}\left[\frac{\pi}{2R^2}e^{2R^2u/\pi}\right]_0^{\pi/2} \tag{25.229b}$$

$$= \frac{\pi}{4R} - \frac{\pi e^{-R^2}}{4R}, \tag{25.229c}$$

et nous avons bien $\lim_{R\to\infty} |I_2(R)| = 0$. Nous passons à la troisième intégrale. En tenant compte que $e^{i\pi/2} = i$, nous avons

$$I_3(R) = -\int_0^R e^{-\gamma_3(t)^2}e^{i\pi/4}dt \tag{25.230a}$$

$$= -\frac{1+i}{\sqrt{2}}\int_0^R e^{-t^2}e^{2i\pi/4} \tag{25.230b}$$

$$= -\frac{1+i}{\sqrt{2}}\int_0^R e^{-it^2}. \tag{25.230c}$$

En passant à la limite $R \to 0$, de l'équation (25.225) il ne reste que

$$0 = \frac{\sqrt{2}}{2} - \frac{1+i}{\sqrt{2}}\int_0^\infty e^{-it^2}dt, \tag{25.231}$$

ce qui signifie que

$$\int_0^\infty e^{-it^2}dt = \frac{\sqrt{2\pi}}{2(1+i)} = \frac{\sqrt{\pi}}{2}e^{-i\pi/4}. \tag{25.232}$$

25.7.3 Logarithme complexe

25.7.3.1 La fonction argument

Nous savons la définition 17.5 de l'exponentielle complexe.

Définition 25.44.
*Un **logarithme** de $\alpha \in \mathbb{C}$ est une solution de l'équation $e^z = \alpha$.*

Notons bien que cela définit *un* logarithme, et non *le* logarithme.

23. Il y a moyen de démontrer cela via le lemme de Jordan[440]. Nous donnons ici une démonstration moins technologique.

Lemme 25.45.
Si z_1 et z_2 sont des logarithmes de α alors il existe $k \in \mathbb{Z}$ tel que $z_1 = z_2 + 2ik\pi$.

Démonstration. Nous commençons par déterminer les logarithmes de $\alpha = 1$. Nous avons besoin de $e^{a+bi} = 1$ $(a, b \in \mathbb{R})$. Nous avons

$$e^a e^{bi} = 1, \tag{25.233}$$

et en prenant la norme nous trouvons $|e^a| = 1$, ce qui donne $a = 0$. Ensuite $e^{bi} = 1$, qui signifie $b = 2k\pi$. Les logarithmes de 1 sont donc les nombres de la forme $2ik\pi$.

Soient maintenant z_1 et z_2 des logarithmes de α. Alors $e^{z_1} = e^{z_2}$, donc[24] $e^{z_1 - z_2} = 1$, ce qui signifie que $z_1 - z_2$ est un logarithme de 1. Donc il existe un $k \in \mathbb{Z}$ tel que $z_1 - z_2 = 2ik\pi$. □

Remarque 25.46.
Jusqu'ici nous n'avons pas donné de conditions donnant l'existence d'un logarithme. Nous avons seulement supposé des existences et donné des propriétés sur ces hypothétiques objets.

Définition 25.47 ([441])**.**
*Si $z \in \mathbb{C}^*$ nous définissons la **valeur principale** de son argument le nombre $\theta \in]-\pi, \pi]$ tel que*

$$z = |z|e^{i\theta} \tag{25.234}$$

Nous le notons $\arg(z)$.

25.48.
Il ne faut pas se ruer sur $\arg(x + iy) = \arctan(y/x)$. Pour rappel, la fonction arctan a été définie dans le théorème 17.30, et elle prend ses valeurs dans $]-\pi/2, \pi/2[$. La formule $v(x, y) = \arctan(y/x)$ n'est donc valable que pour $x > 0$. Les valeurs sont :

$$\arg(x + iy) = \begin{cases} \arctan(y/x) & \text{si } x > 0 \\ \pi + \arctan(y/x) & \text{si } x < 0 \text{ et } y \geqslant 0 \\ -\pi + \arctan(y/x) & \text{si } x < 0 \text{ et } y < 0 \\ \frac{\pi}{2} & \text{si } x = 0 \text{ et } y > 0 \\ \frac{-\pi}{2} & \text{si } x = 0 \text{ et } y < 0. \end{cases} \tag{25.235}$$

Pour $x > 0$ nous avons $\arg(x + iy) = \arctan(y/x)$ parce que justement la fonction arctan prend ses valeurs en particulier entre $-\pi$ et π. Pour $x < 0$ et $y > 0$ nous avons $\arg(x+iy) = \pi + \arctan(y/x)$ (dans ce cas, $\arctan(y/x) < 0$) et si $x < 0$, $y < 0$ nous avons $\arg(x + iy) = -\pi + \arctan(y/x)$.

25.49 (Les dérivées partielles de la fonction argument)**.**
Vu que nous en aurons besoin plusieurs fois, nous calculons maintenant les dérivées partielles de la fonction

$$\begin{aligned} \varphi\colon \mathbb{R}^2 &\to \mathbb{R} \\ (x, y) &\mapsto \arg(x + iy). \end{aligned} \tag{25.236}$$

Nous commençons par la dérivée $\partial_x \varphi(x, y)$. Et il y a de nombreux cas à séparer.

<u>$x > 0$</u> Nous avons

$$\frac{\partial \varphi}{\partial x}(x, y) = \lim_{\epsilon \to 0} \frac{\arctan(y/(x + \epsilon)) - \arctan(y/x)}{\epsilon}, \tag{25.237}$$

qui n'est autre que la dérivée de la fonction $x \mapsto \arctan(y/x)$. Nous pouvons la calculer facilement avec le théorème 17.30(2) :

$$\frac{\partial \varphi}{\partial x}(x, y) = -\frac{y}{x^2 + y^2}. \tag{25.238}$$

24. C'est facile de dire « donc ». Il faut surtout citer la proposition 17.8(2).

$\underline{x < 0}$ Nous avons

$$\frac{\partial\varphi}{\partial x}(x,y) = \lim_{\epsilon\to 0}\frac{\pm\pi + \arctan(y/(x+\epsilon)) - \big(\pm\pi + \arctan(y/x)\big)}{\epsilon} \tag{25.239}$$

où les signes $\pm$ dépendent du signe de y. De toutes façons, les termes en π se simplifient et le calcul est le même que celui du cas $x > 0$. Encore une fois nous avons

$$\frac{\partial\varphi}{\partial x}(x,y) = -\frac{y}{x^2+y^2}. \tag{25.240}$$

$\underline{x = 0}$ Nous devons calculer

$$\frac{\partial\varphi}{\partial x}(0,y) = \lim_{\epsilon\to 0}\frac{\arg(\epsilon+iy) - \arg(iy)}{\epsilon}. \tag{25.241}$$

Il y a quatre cas d'après les signes de ϵ (séparer limite à gauche et à droite) et y.

Si $\epsilon > 0$ et $y > 0$ alors nous avons à faire le calcul

$$\lim_{\epsilon\to 0^+}\frac{\arctan(y/\epsilon) - \pi/2}{\epsilon} \tag{25.242}$$

qui se traite par la règle de l'Hospital. Cela donne $-1/y$.

Les trois autres cas ne se distinguent que par des constantes au numérateur, lesquelles disparaissent en appliquant la règle de l'Hospital[25]. Au final,

$$\frac{\partial\varphi}{\partial x}(0,y) = -\frac{1}{y}. \tag{25.243}$$

Nous avons calculé jusqu'ici :

$$\frac{\partial\varphi}{\partial x}(x,y) = \frac{-y}{x^2+y^2} \tag{25.244}$$

pour tout $(x,y) \in \mathbb{R}^2\backslash\{(0,0)\}$. En particulier vous avez noté que cette dérivée partielle est continue sur $\mathbb{R}^2\backslash\{(0,0)\}$.

Nous calculons à présent la dérivée partielle par rapport à y :

$$\frac{\partial\varphi}{\partial y}(x,y) = \lim_{\epsilon\to 0}\frac{\arg(x+iy+i\epsilon) - \arg(x+iy)}{\epsilon}. \tag{25.245}$$

$\underline{x > 0}$ Nous avons à calculer

$$\lim_{\epsilon\to 0}\frac{\arctan\frac{y+\epsilon}{x} - \arctan\frac{y}{x}}{\epsilon}, \tag{25.246}$$

qui n'est autre que la dérivée de la fonction $t \mapsto \arctan\frac{t}{x}$ en $t = y$. Résultat :

$$\frac{\partial\varphi}{\partial y}(x,y) = \frac{x}{x^2+y^2}. \tag{25.247}$$

$\underline{x < 0 \textbf{ et } y \neq 0}$ Le calcul à faire est :

$$\lim_{\epsilon\to 0}\frac{\pm\pi + \arctan\frac{y+\epsilon}{x} - \big(\pm\pi + \arctan\frac{y}{x}\big)}{\epsilon} \tag{25.248}$$

Une chose importante à remarquer est que dans le calcul de la limite nous pouvons supposer que y et $y+\epsilon$ aient le même signe, quelle que soit la valeur et le signe de ϵ (assez petit). C'est pour cela que les deux termes $\pm\pi$ arrivent avec le même signe des deux côtés de la différence, et se simplifient. Nous tombons sur une limite déjà faite et

$$\frac{\partial\varphi}{\partial y}(x,y) = \frac{x}{x^2+y^2} \tag{25.249}$$

25. Nonobstant le fait que ces constantes se mettent bien pour avoir un vrai cas d'indétermination 0/0, sinon la règle de l'Hospital ne s'applique pas.

$x < 0$ **et** $y = 0$ Vu que $x < 0$ nous avons $\arg(x) = \pi$ et nous devons calculer

$$\lim_{\epsilon \to 0} \frac{\arg(x + i\epsilon) - \pi}{\epsilon}. \tag{25.250}$$

La limite $\epsilon \to 0^+$ est classique et donne $1/x$.

Mais la limite $\epsilon \to 0^-$ n'existe pas :

$$\lim_{\epsilon \to 0^-} \frac{-\pi + \arctan(\epsilon/x) - \pi}{\epsilon} \tag{25.251}$$

n'existe pas.

Donc

$$\frac{\partial \varphi}{\partial y}(x, 0) \tag{25.252}$$

n'existe pas pour $x < 0$.

$x = 0$ **et** $y \neq 0$ Le calcul est immédiat

$$\lim_{\epsilon \to 0} \frac{\arg(iy + i\epsilon) - \arg(iy)}{\epsilon} = 0, \tag{25.253}$$

donc

$$\frac{\partial \varphi}{\partial y}(0, y) = 0. \tag{25.254}$$

En ce qui concerne la continuité, nous avons que $\partial_y \varphi$ est continue partout sauf sur la demi-droite $\{(x, 0) \text{ tel que } x \leqslant 0\}$ où elle n'existe pas.

25.7.3.2 Une définition possible du logarithme

Définition 25.50.
Nous définissons la fonction ***logarithme*** *par*

$$\begin{aligned} \ln \colon \mathbb{C}^* &\to \mathbb{C} \\ z &\mapsto \ln\big(|z|\big) + i \arg(z) \end{aligned} \tag{25.255}$$

où le $\ln$ *à droite est le logarithme usuel sur* $\mathbb{R}^+$.

Remarque 25.51.
Cette fonction généralise le logarithme déjà vu sur $]0, \infty[\subset \mathbb{R}$. En effet pour des valeurs de z dans cette partie nous avons $\arg(z) = 0$ et $|z| = z$.

Lemme 25.52.
Le nombre $\ln(z)$ *est un logarithme de* z.

Démonstration. Nous avons

$$e^{\ln(z)} = e^{\ln|z|} e^{i \arg(z)} = |z| e^{i \arg(z)} = z. \tag{25.256}$$

Nous avons utilisé le fait que $e^{\ln(x)} = x$ pour $x \in \mathbb{R}^+$ et $|z| e^{i \arg(z)} = z$ par définition de la fonction arg. □

Notons que si on avait pris d'autres conventions pour définir arg, nous aurions eu d'autres définitions possibles de ln.

Exemple 25.53.
Nous avons

$$\ln(-1) = \ln(1) + i \arg(-1). \tag{25.257}$$

Mais $\ln(1) = 0$ et $\arg(-1) = \pi$ (et non $-\pi$), donc

$$\ln(-1) = i\pi. \tag{25.258}$$

C'est cette définition du logarithme qui est prise par Sage, et c'est cela qui lui permet de donner la primitive de $1/x$ comme $\ln(x)$ et non $\ln(|x|)$, parce que Sage connaît les logarithmes de nombres réels négatifs :

```

 SageMath version 7.3, Release Date: 2016-08-04
 Type "notebook()" for the browser-based notebook interface.
 Type "help()" for help.

sage: ln(-1)
I*pi
sage: f(x)=1/x
sage: f.integrate(x)
x |--> log(x)
```

tex/sage/sageSnip010.sage

△

Nous avons jusqu'ici défini une fonction sur $\mathbb{C}^*$ qui fait correspondre à chaque nombre complexe un de ses logarithmes. Il reste quelques questions à régler :

— Est-ce que cette fonction est continue ? Holomorphe ? (réponses : non et non)

— Si non, est-ce qu'il y avait moyen de trouver une définition plus efficace ? (réponse : non)

Lemme 25.54.
La fonction ln *n'est pas continue sur* $]-\infty, 0]$.

Démonstration. Attention à bien comprendre l'énoncé. La fonction

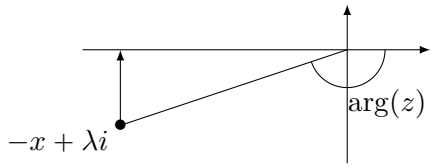
$$\begin{aligned} f\colon]-\infty, 0[&\to \mathbb{C} \\ x &\mapsto \ln(x) \end{aligned} \tag{25.259}$$

est continue. D'ailleurs c'est $\ln(x) = \ln(|x|) + i\pi$. Ce dont il est question dans l'énoncé, c'est de la fonction ln vue comme fonction sur $\mathbb{C}^*$.

Soit $x > 0$ dans $\mathbb{R}$; nous avons

$$\ln(-x) = \ln(x) + i\pi. \tag{25.260}$$

Cependant $\lim_{\substack{\lambda\to 0^-\\ \lambda\in\mathbb{R}}} \ln(-x + \lambda i)$ va valoir $\ln(|x| - i\pi)$. En effet lorsque $\lambda < 0$ est petit, l'argument de $-x + \lambda i$ se rapproche de $-\pi$ (et non de π).



Donc

$$\lim_{\substack{\lambda\to 0^-\\ \lambda\in\mathbb{R}}} \ln(-x + \lambda i) = \lim \ln(|x + \lambda i|) + i \arg(-x + \lambda i) = \ln(|x|) - i\pi. \tag{25.261}$$

Nous n'avons donc pas continuité de la fonction logarithme comme fonction sur $\mathbb{C}^*$. □

Théorème 25.55.
La restriction

$$\ln\colon \mathbb{C}\backslash]-\infty,0] \to \mathbb{C} \tag{25.262}$$

est holomorphe.

Démonstration. Nous allons utiliser la proposition 11.254 et considérer la fonction

$$\begin{aligned} F\colon S &\to \mathbb{R}^2 \\ (x,y) &\mapsto \big(\ln(|x+iy|), \arg(x+iy)\big) \end{aligned} \tag{25.263}$$

où $S = \mathbb{R}^2\backslash\{(x,0) \text{ tel que } x \leqslant 0\}$. Nous devons vérifier que F est différentiable et que sa différentielle en un point de S est une similitude.

Nous posons

$$u(x,y) = \ln\left(\sqrt{x^2+y^2}\right) \tag{25.264}$$

et

$$v(x,y) = \arg(x+iy). \tag{25.265}$$

Les dérivées partielles de u ne sont pas très compliquées :

```
sage: var('x,y')
(x, y)
sage: u(x,y)=ln(sqrt(x**2+y**2))
sage: u.diff(x)
(x, y) |--> x/(x^2 + y^2)
```

tex/sage/sageSnip011.sage

c'est-à-dire

$$\frac{\partial u}{\partial x} = \frac{x}{x^2+y^2} \tag{25.266a}$$

$$\frac{\partial u}{\partial y} = \frac{y}{x^2+y^2}. \tag{25.266b}$$

Pour celles de v par contre, il faut se poser des questions, par exemples résister à la tentation d'écrire $v(x,y) = \arctan(y/x)$ et lire 25.48.

Nous avons déjà calculé les dérivées partielles de v dans 25.49, et nous avons vu qu'elles étaient continues sur $\mathbb{R}^2$ privé de la demi-droite.

Vu que les dérivées partielles sont continues, le théorème 11.243 nous dit que F est différentiable. La matrice de la différentielle est alors la matrice des dérivées partielles

$$\begin{pmatrix} \frac{x}{x^2+y^2} & \frac{y}{x^2+y^2} \\ \frac{-y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{pmatrix}, \tag{25.267}$$

qui a la forme requise (11.636) pour que la proposition 11.254 nous assure que ln soit $\mathbb{C}$-dérivable, c'est-à-dire holomorphe. □

25.7.3.3 Pas plus de continuité

Bon. La fonction logarithme que nous avons définie est holomorphe sur $\mathbb{C}^*$ privé d'une demi-droite U. Et elle n'est pas continue sur U ; elle y est cependant continue « par le haut ». Pouvons-nous faire mieux ? Nous allons maintenant prouver quelques résultats d'impossibilité de faire mieux que holomorphe partout sauf une partie pas si petite que ça.

Proposition 25.56.
Il n'existe pas de fonctions continues $f\colon \mathbb{C}^* \to \mathbb{C}$ telle que $e^{f(z)} = z$ pour tout $z \in \mathbb{C}^*$.

Démonstration. Pour tout z, le nombre $f(z)$ est un logarithme de z. Or $\ln(z)$ en est également un. Donc par le lemme 25.45

$$f(z) = \ln(z) + 2ik(z)\pi \tag{25.268}$$

pour une certaine fonction $k\colon \mathbb{C}^* \to \mathbb{Z}$. Sur le domaine d'holomorphie de ln, les fonctions ln et f étant continues, la fonction k l'est aussi. Mais une fonction continue à valeurs dans $\mathbb{Z}$ est constante (son domaine est connexe).

Il existe donc $k \in \mathbb{Z}$ tel que

$$f(z) = \ln(z) + 2ik\pi \tag{25.269}$$

au moins pour tout $z \in \mathbb{C}^*\backslash U$. Une telle fonction ne peut pas être continue sur U parce que ln ne l'est pas. □

Ok. Pas continue sur tout $\mathbb{C}$. Mais continue sur un peu plus que $\mathbb{C}$ privé de toute une demi-droite ? La proposition suivante répond que bof.

Proposition 25.57.
Soit Ω un ouvert de $\mathbb{C}$ contenant $S(0,r)$ (le cercle centré en 0 et de rayon $r > 0$). Il n'existe pas de fonction continue $f\colon \Omega \to \mathbb{C}$ telle que $e^{f(z)} = z$ pour tout $z \in \Omega$.

Démonstration. Encore une fois, pour tout $z \in \Omega$ nous avons

$$f(z) = \ln(z) + 2i\pi k(z) \tag{25.270}$$

pour une certaine fonction $k\colon \Omega \to \mathbb{Z}$. Sur $\Omega\backslash U$, la fonction ln est continue et k doit également l'être. Donc k est constante sur les composantes connexes de $\Omega\backslash U$.

Vu que $S(0,r)$ est compact, on peut le recouvrir par un nombre fini de boules centrées en des points de $S(0,r)$. En prenant le minimum des rayons de ces boules, nous voyons que Ω contient une couronne

$$\{z \in \mathbb{C} \text{ tel que } r - \delta \leqslant |z| \leqslant r + \delta\}. \tag{25.271}$$

Soit le point $x_0 = -r$. C'est un point de Ω contenu dans U. Nous allons prouver que $B(x_0,\delta)\backslash U$ est dans une seule composante connexe de Ω.

Soit un point $z_1 \in B(x_0,\delta)$ situé au-dessus de U, et z_2 un point de $B(x_0,\delta)$ situé en dessous de U. Le cercle $S(0,r)$ coupe $B(x_0,\delta)$ en deux points : un au-dessus et un en-dessous de U. On peut lier z_1 au point de « sortie » supérieur de $S(0,r)$ en restant dans $B(x_0,\delta)$; ce point est ensuite relié en suivant le cercle au point d'entrée inférieur du cercle dans $B(x_0,\delta)$. Ce dernier point est lié à z_2 par un chemin restant dans la boule.

Tout cela pour dire que z_1 et z_2 sont dans la même composante connexe de Ω et que $k(z_1) = k(z_2)$. Il existe donc $k \in \mathbb{Z}$ tel que

$$f(z) = \ln(z) + 2ik\pi \tag{25.272}$$

sur $B(x_0,\delta)\backslash U$. Une telle fonction f ne peut pas être continue. □

25.7.3.4 Pas d'unicité : autres déterminations de l'argument

25.58.
Nous avons pris la fonction d'argument $\arg\colon \mathbb{C} \to]-\pi,\pi]$. Il y en a évidemment beaucoup d'autres de possibles. Par exemple pour $\alpha \in \mathbb{R}$ nous pouvons considérer

$$\arg_{\alpha^+}\colon \mathbb{C} \to]\alpha, \alpha + 2\pi] \tag{25.273}$$

ou

$$\arg_{\alpha^-}\colon \mathbb{C} \to [\alpha, \alpha + 2\pi[. \tag{25.274}$$

En posant

$$\ln_{\alpha^\pm}(z) = \ln(|z|) + i\arg_{\alpha^\pm}(z) \tag{25.275}$$

nous avons une fonction réciproque de l'exponentielle définie sur $\mathbb{C}^*$ et holomorphe sur $\mathbb{C}^*$ privé d'une demi-droite D_α (dépendante de la valeur de α).

La différence entre $\ln_{\alpha^+}$ et $\ln_{\alpha^-}$ est seulement la valeur sur la demi-droite de non-holomorphie. L'une sera semi-continue d'un côté et l'autre, de l'autre côté.

Remarque 25.59.
La fonction $\arg_{0^-}$ a déjà été utilisée en 17.4.2 pour écrire un inverse de la fonction

$$\begin{aligned}\varphi\colon [0,2\pi[&\to S^1\\ t &\mapsto e^{it}.\end{aligned} \tag{25.276}$$

Définition 25.60 ([442])**.**
*Soit un ouvert $\Omega \subset \mathbb{C}^*$. Nous disons que la fonction $f\colon \Omega \to \mathbb{C}$ est une **détermination** sur Ω si elle est continue et vérifie*

$$e^{f(z)} = z \tag{25.277}$$

pour tout $z \in \mathbb{C}$.

Les différents résultats vus jusqu'ici montrent qu'il n'existe pas de détermination du logarithme sur $\mathbb{C}^*$.

Définition 25.61.
*La **détermination principale** du logarithme est la restriction de notre logarithme 25.50*

$$\begin{aligned}\ln\colon \mathbb{C}^* &\to \mathbb{C}\\ z &\mapsto \ln(|z|) + i\arg(z)\end{aligned} \tag{25.278}$$

à l'ouvert $\mathbb{C}^*\backslash U$ où U est la partie $\Re(z) \leqslant 0$ de $\mathbb{C}$.

Remarque 25.62.
Beaucoup de sources[443] ne définissent pas $\ln_{\alpha^\pm}$ sur la droite D_α. C'est-à-dire qu'ils notent $\ln_\alpha$ notre fonction $\ln_{\alpha^+}$ restreinte à $\mathbb{C}^*\backslash D_\alpha$. Dans ce cas, les fonctions $\ln_{\alpha^+}$ et $\ln_{\alpha^-}$ sont identiques[26].

Cette remarque est importante parce que certains vont vous dire « le logarithme n'est pas définit sur la demi-droite » ; de leur point de vue, la fonction que nous avons définie est une prolongation (non continue) à U du logarithme, qui est continu.

(1) Certaines personnes pourraient vous dire que notre logarithme « n'est pas bien définit parce que si on fait le tour dans un sens ou dans l'autre nous n'obtenons pas la même valeur pour $\ln(z)$ lorsque z est sur U ». Et cela avec des arguments aussi forts que « 2π et 0, c'est le même point ».
Nous préférons être bien clairs[27] sur ce point : notre fonction ln est parfaitement définie sur $\mathbb{C}^*$ et 2π n'est pas la même chose que zéro. En particulier $\arg(e^{2i\pi}) = 0$ et $\arg(e^{-i\pi}) = \pi$ et non $-\pi$.

(2) Il n'en reste pas moins que Sage donne $\ln(-1) = I\pi$ et que nous avons choisi de faire de même, parce que le Frido n'est pas un cours d'agrégation, mais un texte qui donne quelques éléments de mathématique dans le but d'utiliser Sage efficacement.

(3) Tout ceci pour dire que si vous utilisez ce livre pour l'agrégation, vous devriez sérieusement considérer l'option de ne pas donner du logarithme la définition donnée ici, mais bien sa restriction.

En fait notre logarithme est maximum pour la propriété « être une réciproque de l'exponentielle » alors que beaucoup de monde préfère avoir une fonction maximale pour la propriété « être réciproque de l'exponentielle tout en étant continue ».

De toutes les fonctions ayant le droit de vouloir être appelée « logarithme », celle que nous avons choisie (un peu arbitrairement) pour s'appeler « logarithme » et accaparer de la notation « ln » est $\ln_{\pi^+}$. Elle est d'une certaine manière celle qui arrive le plus naturellement.

26. Cela n'est pas tout à fait évident ; vous devriez y penser.
27. Est-ce qu'il faut vraiment un pluriel ici ?

En effet si nous pensons au logarithme népérien $\ln\colon \mathbb{R}^+ \to \mathbb{R}$ que nous voulons prolonger sur $\mathbb{R}$, nous devons poser

$$\ln(-x) = \ln(-1) + \ln(x) \tag{25.279}$$

pour $x > 0$. Que peut valoir $\ln(-1)$? Il doit vérifier $e^{\ln(-1)} = -1$. La première valeur qui nous tombe sous la main est $\ln(-1) = \pi$. Bien entendu, d'autres possibilités étaient possibles, comme $\ln(-1) = 2017\pi$ par exemple.

25.7.3.5 Pas d'unicité : développement en série

Pour $z_0 \in \mathbb{C}^*$ nous pouvons écrire un développement en série de la réciproque de l'exponentielle autour de z_0. La fonction ainsi définie est holomorphe sur la boule $B(z_0, |z_0|)$ et diverge en dehors de cette boule.

Voilà encore une fonction « logarithme » pour chaque point de $\mathbb{C}^*$. Nous nommons $\ln_{z_0}$ la fonction

$$\ln_{z_0}\colon B(z_0, |z_0|) \to \mathbb{C} \tag{25.280}$$

donnée par la série.

En général nous n'avons pas $\ln_{z_1} = \ln_{z_2}$ sur l'intersection des disques de convergence. Si c'était le cas, de proche en proche nous pourrions construire une fonction continue réciproque du logarithme sur $\mathbb{C}^*$, ce qui est impossible.

25.7.3.6 Pas d'unicité : laquelle choisir ?

Bon. Pour chaque demi-droite D nous avons une détermination du logarithme sur $\mathbb{C}^*\backslash D$. Et pour tout $z_0 \in \mathbb{C}^*$ nous en avons une sur $B(z_0, |z_0|)$.

En pratique, quel logarithme choisir ? Cela dépend du problème.

Si vous avez besoin ou envie de travailler avec des série entières, le mieux est de choisir une détermination donnée par un développement autour d'un point bien choisi par rapport à votre problème.

Si vous avez surtout besoin d'holomorphie, et que vous en avez besoin sur un grand domaine, vous devriez choisir une détermination sur un des ensembles $\mathbb{C}^*\backslash D_\alpha$ en choisissant α de telle sorte que la demi-droite maudite ne passe pas par la zone sur laquelle vous travaillez.

Dans tous les cas, vous devez préciser très explicitement la détermination choisie. Dans ce texte, sauf mention du contraire, nous utiliserons la détermination principale, et même son extension (non continue) à $\mathbb{C}^*$. Lorsque nous aurions besoin d'holomorphie, nous préciserons que nous considérons la restriction.

25.7.3.7 Logarithme comme primitive

Tout le monde sait que le logarithme $\ln\colon \mathbb{R}^+ \to \mathbb{R}$ est une primitive de la fonction $x \mapsto 1/x$. Qu'en est-il dans le cas complexe ? Tout d'abord précisons que nous ne comptons pas encore parler d'intégrale sur $\mathbb{C}$, mais seulement d'intégrales sur $\mathbb{R}$ d'une fonction à valeur complexes.

Proposition 25.63.
Si $z \in \mathbb{C}$ alors

$$\int \frac{1}{x+z} dx = \ln(x+z) \tag{25.281}$$

Démonstration. Il est important de comprendre que la formule (25.281) est un abus de notation pour dire que si nous considérons la fonction

$$\begin{aligned} \varphi\colon \mathbb{R} &\to \mathbb{C} \\ x &\mapsto \ln(x+z) \end{aligned} \tag{25.282}$$

alors nous avons $\varphi'(x) = \frac{1}{x+z}$. Ici la dérivation est une dérivation sur $\mathbb{R}$ et l'intégrale est une intégrale sur $\mathbb{R}$, c'est-à-dire « composante par composantes ». La fonction φ se décompose en

partie réelle et imaginaire qui sont à dériver séparément :

$$\varphi(x) = \ln(|x+z|) + i\arg(x+z). \tag{25.283}$$

Si z est imaginaire pur Nous posons $z = \lambda i$ avec $\lambda \in \mathbb{R}^*$. D'abord nous avons

$$\frac{1}{x+\lambda i} = \frac{x}{x^2+\lambda^2} - i\frac{\lambda}{x^2+\lambda^2}. \tag{25.284}$$

La partie réelle de $\varphi(x)$ est

$$\varphi_1(x) = \ln\left(\sqrt{x^2+\lambda^2}\right), \tag{25.285}$$

dont la dérivée est

$$\varphi_1'(x) = \frac{x}{x^2+\lambda^2}, \tag{25.286}$$

qui correspond bien à la partie réelle de $\frac{1}{x+\lambda i}$.
En ce qui concerne la partie imaginaire, $\varphi_2(x) = \arg(x+\lambda i)$, et sa dérivée n'est rien d'autre que la dérivée partielle par rapport à x de la fonction argument, déjà calculée en (25.244) :

$$\varphi_2'(x) = \frac{-\lambda}{x^2+\lambda}. \tag{25.287}$$

Cela est bien la partie imaginaire de $\frac{1}{x+\lambda i}$.
Notons que nous n'avons pas de problèmes sur la demi-droite des réels négatifs parce que nous ne considérons au final que la dérivée partielle par rapport à x de la fonction argument, laquelle existe et est continue, même sur cette partie.

Pour z quelconque Soit $z = s + \lambda i$ avec $s, \lambda \in \mathbb{R}$. En posant $\varphi_0(x) = \ln(x+\lambda i)$ nous avons $\varphi(x) = \varphi_0(x+s)$ et donc

$$\varphi'(x) = \varphi_0'(x+s) = \frac{1}{x+s+\lambda i} = \frac{1}{x+z}. \tag{25.288}$$

Tout va bien.

□

Exemple 25.64.
Un petit calcul d'intégrale, que nous avions déjà faite dans l'exemple 19.96 (avec la méthode de Rothstein-Trager). En passant par une décomposition en fractions simples :

$$\int \frac{1}{x^3+x} = \int \left(\frac{1}{x} - \frac{1/2}{x-i} - \frac{1/2}{x+i}\right) \tag{25.289a}$$

$$= \ln(x) - \frac{1}{2}\ln(x-i) - \frac{1}{2}\ln(x+i) \tag{25.289b}$$

$$= \ln(x) - \frac{1}{2}\ln(x^2+1). \tag{25.289c}$$

Attention aux justifications. Il n'est pas vrai en général dans le cas de nombres complexes a et b que $\ln(ab) = \ln(a) + \ln(b)$. En effet, pour la partie réelle, ça passe parce que $|ab| = |a||b|$. Mais en ce qui concerne la partie imaginaire,

$$\arg(ab) \neq \arg(a) + \arg(b) \tag{25.290}$$

lorsque la somme dépasse les bornes de $]-\pi, \pi]$. Le passage à (25.289c) fonctionne parce que dans le cas particulier des nombres $x+i$ et $x-i$, les arguments se somment à zéro : $\arg(x+i)+\arg(x-i) = 0$. △

25.8 Théorème de Weierstrass

Théorème 25.65 (Théorème de Weierstrass[444]).
Soit (f_n) une suite de fonctions holomorphes sur un ouvert Ω de $\mathbb{C}$ que nous supposons converger uniformément sur tout compact vers f. Alors f est holomorphe sur Ω et pour tout k nous avons

$$f_n^{(k)} \to f^{(k)} \tag{25.291}$$

uniformément sur tout compact.

Dit en peu de mots, la limite uniforme d'une suite de fonctions holomorphes est holomorphe, et on peut permuter la limite avec la dérivation.

Démonstration. Chacune des fonctions f_n étant holomorphes, si $a \in \Omega$ et r est tel que $B(a,r) \subset \Omega$, nous avons par la formule de Cauchy 25.13 :

$$f_n(z) = \frac{1}{2\pi i} \int_{\partial B(a,r)} \frac{f_n(\xi)}{\xi - z} d\xi \tag{25.292}$$

pour tout z dans un boule $B(a,\rho)$ incluse dans $B(a,r)$. Étant donné que le cercle ∂B est compact, elle y est majorée par une constante M. Montrons que de plus nous pouvons choisir M de telle façon à avoir $|f_n(\xi)| \leqslant M$ pour tout n et tout ξ en même temps. D'abord nous utilisons la continuité de la limite f sur le compact ∂B pour poser $A = \max_{z\in\partial B} |f(z)|$. Ensuite nous considérons un $\epsilon > 0$ et N tel que $| f_n - f\|_{\partial B} \leqslant \epsilon$ pour tout $n \geqslant N$. Nous savons maintenant que

$$\{|f_n(\xi)| \text{ tel que } n \geqslant N, \xi \in \partial B\} \tag{25.293}$$

est majoré par $A + \epsilon$. Nous posons enfin

$$B = \max_{n\leqslant N} \max_{\xi\in\partial B} |f_n(z)|, \tag{25.294}$$

et alors le nombre $M = \max\{A + \epsilon, B\}$ majore $|f_n(\xi)|$ pour tout n et tout $\xi \in \partial B$.

De plus pour tout $\xi \in \partial B$ et pour tout z dans la petite boule, nous avons $|\xi - z| > r - \rho$, donc la fonction dans l'intégrale est majorée par une constante ne dépendant ni de n ni de ξ. Nous pouvons donc permuter l'intégrale et la limite sur n :

$$f(z) = \frac{1}{2i\pi} \int_{\partial B} \frac{f(\xi)}{\xi - z}. \tag{25.295}$$

Cela implique que la fonction f est holomorphe par le corolaire 25.15.

Nous voudrions maintenant parler des dérivées des f_n et de f. Pour cela nous voulons permuter l'intégrale et les dérivées, ce qui est fait au corolaire 25.25 :

$$f_n^{(k)} = \frac{1}{2\pi i} \int_{\partial B(z_0,r)} \frac{f(\omega)}{(\omega - z)^{k+1}} d\omega. \tag{25.296}$$

Nous voulons la convergence sur tout compact contenu dans l'ouvert Ω. Pour ce faire, nous allons considérer un compact $K \subset \Omega$ et prouver la convergence uniforme dans toute boule de la forme $B(z_0,r)$ avec $z_0 \in K$ et $B(z_0,r) \subset \Omega$. Pour chaque tel couple (z_0,r), nous aurons un $N_{(z_0,r)} \in \mathbb{N}$ tel que si $n \geqslant N_{(z_0,r)}$,

$$\|f_n^{(k)} - f^{(k)}\|_{B(z_0,r)} \leqslant \epsilon. \tag{25.297}$$

Vu que ces boules $B(z_0,r)$ forment un recouvrement de K par des ouverts, nous pouvons en retirer un sous-recouvrement fini et prendre, comme N, le maximum des $N_{(z_0,r)}$ correspondants. Pour ce N nous aurons

$$\|f_n^{(k)} - f^{(k)}\|_K \leqslant \epsilon. \tag{25.298}$$

Au travail !

Pour $z \in B(z_0, r)$ nous considérons $r' > r$ tel que $B(z_0, r') \subset \Omega$ et nous avons

$$|f_n^{(k)}(z) - f^{(k)}(z)| = \left| \frac{1}{2\pi i} \int_{\partial B(z_0,r')} \frac{f_n(\xi) - f(\xi)}{(\xi - z)^{k+1}} d\xi \right| \tag{25.299a}$$

$$\leqslant \frac{1}{2\pi} \int_{\partial B(z_0,r')} \frac{|f_n(\xi) - f(\xi)|}{|r - r'|^{k+1}} d\xi. \tag{25.299b}$$

Nous avons pris ce r' de telle manière que $|\xi - z|$ soit borné par le bas par $|r - r'|$; sinon la majoration que nous venons de faire ne marche pas. Étant donné que $f_n \to f$ uniformément, nous pouvons considérer n assez grand pour que le numérateur soit plus petit que ϵ indépendamment de ξ et de z. Donc pour un n assez grand,

$$|f_n^{(k)}(z) - f^{(k)}(z)| \leqslant \frac{\epsilon}{2\pi} \frac{2\pi r'}{|r - r'|^{k+1}} \tag{25.300}$$

pour tout $z \in B(z_0, r)$. Donc nous avons convergence uniforme $f_n^{(k)} \to f^{(k)}$ sur cette boule. Par l'argument de compacité donné plus haut, nous avons la convergence uniforme sur tout compact. □

Chapitre 26

Analyse fonctionnelle

Théorème 26.1 (Théorème d'isomorphisme de Banach)**.**
Une application linéaire continue et bijective entre deux espaces de Banach est un homéomorphisme.

26.1 Théorème d'Ascoli

Définition 26.2.
Une partie A d'un espace topologique X est ***relativement compacte*** *dans X si sa fermeture est compacte.*

Proposition 26.3 ([252])**.**
Soient E et F deux espaces vectoriels normés sur $\mathbb{R}$ ou $\mathbb{C}$ et une application $f \in \mathcal{L}(E, F)$. Les propriétés suivantes sont équivalentes.

(1) L'image d'un borné de E par f est relativement compact dans F.

(2) L'image par f de la boule unité fermée est relativement compacte dans F.

(3) Si (x_n) est une suite bornée dans E, alors nous pouvons en extraire une sous-suite $(x_{\varphi(n)})$ telle que $\big(fx_{\varphi(n)}\big)$ converge dans F.

Définition 26.4.
Une application vérifiant les conditions équivalentes de la proposition 26.3 est dite ***compacte****.*

Théorème 26.5 (Théorème d'Ascoli[445])**.**
Soient un espace topologique compact K et un espace métrique (E, d). Nous considérons la topologie uniforme sur $C(K, E)$. Une partie A de $C(K, E)$ est relativement compacte si et seulement si les deux conditions suivantes sont remplies :

(1) A est équicontinu[1]*,*

(2) $\forall x \in K$, l'ensemble $\{f(x)$ tel que $f \in A\}$ est relativement compact dans E.

Corolaire 26.6.
Soient $\big(E, (p_l)\big)$ et $\big(F, (q_k)\big)$ deux espaces localement compacts munis de semi-normes. Nous supposons que E est métrisable et complet. Soit $(T_j)_{j\in\mathbb{N}}$ une suite d'application linéaires $E \to F$ telles que pour tout $x \in E$ il existe $\alpha_x \in F$ tel que

$$T_j x \xrightarrow{F} \alpha_x. \tag{26.1}$$

Si nous posons $Tx = \alpha_x$ alors

(1) l'application T est linéaire et continue,

1. Définition 7.321.

(2) pour tout compact K dans E et pour tout k nous avons

$$\lim_{j\to\infty} \sup_{x\in K} q_k(T_j x - Tx) = 0, \tag{26.2}$$

(3) si $x_j \to x$ dans E alors

$$T_j x_j \to Tx \tag{26.3}$$

dans F.

La version suivante du théorème de Banach-Steinhaus est énoncée de façon *ad hoc* pour fonctionner avec l'espace $\mathscr{D}(K)$ des fonctions de classe C^∞ à support dans le compact K. Un énoncé un peu plus fort est donné dans le cadre des espaces de Fréchet dans [153].

Théorème 26.7 (Banach-Steinhaus avec des semi-normes)**.**
Soit (E, d) un espace vectoriel métrique complet dont la topologie est également[2] *donnée par une famille $\mathcal{P}$ de semi-normes. Soit $\{T_\alpha\}_{\alpha\in A}$ une famille d'applications linéaires continues $T_\alpha\colon E \to \mathbb{R}$ telles que pour tout $x \in E$ nous ayons*

$$\sup_{\alpha\in A} \big|T_\alpha(x)\big| < \infty. \tag{26.4}$$

Alors il existe une constante $C > 0$ et un sous-ensemble fini $J \subset \mathcal{P}$ tels que pour tout $x \in E$ nous ayons

$$\big|T_\alpha(x)\big| \leqslant C \max_{j\in J} p_j(x). \tag{26.5}$$

Démonstration. Pour chaque $k \in \mathbb{N}^*$ nous posons

$$\Omega_k = \{x \in E \text{ tel que } \sup_{\alpha\in A} \big|T_\alpha(x)\big| > k\}. \tag{26.6}$$

Ces ensembles sont des ouverts (pour la même raison que dans la preuve du théorème 10.84) et leur union est E en entier parce que par hypothèse $\sup_{\alpha\in A} \big|T_\alpha(x)\big| < \infty$.

Si les Ω_k étaient tous dense, le théorème de Baire 7.339 nous dit que leur intersection est dense également ; elle est donc non vide et si $x_0 \in \bigcap_{k\in\mathbb{N}} \Omega_k$ nous avons

$$\sup_{\alpha\in A} \big|T_\alpha(x_0)\big| = \infty, \tag{26.7}$$

ce qui contredirait l'hypothèse. Donc les Ω_k ne sont pas tous denses. Soit $k_0 \in \mathbb{N}^*$ tel que Ω_{k_0} n'est pas dense dans E. Il existe donc $x_0 \in E$ et un ouvert autour de x_0 n'intersectant pas Ω_{k_0}.

Nous jouons à présent sur la topologie de E. L'ouvert dont il est question est un d-ouvert et donc un $\mathcal{P}$-ouvert, lequel contient une $\mathcal{P}$-boule ouverte. Cette dernière boule n'est pas spécialement une d-boule, mais c'est un d-ouvert.

Il existe dont J fini dans $\mathcal{P}$ et $\rho > 0$ tels que $B_J(x_0, \rho) \cap \Omega_{k_0} = \varnothing$. Donc pour tout $y \in B_J(x_0, \rho)$ nous avons

$$\sup_{\alpha\in A} \big|T_\alpha(y)\big| \leqslant k_0. \tag{26.8}$$

Si maintenant $y \in B(0, \rho)$, nous avons $y = (y + x_0) - x_0$ et donc

$$\begin{aligned}
\big|T_\alpha(y)\big| &= \big|T_\alpha(y + x_0) - T_\alpha(x_0)\big| && \text{(26.9a)}\\
&\leqslant \big|T_\alpha(y + x_0)\big| + \big|T_\alpha(x_0)\big| && \text{(26.9b)}\\
&\leqslant k_0 + C && \text{(26.9c)}
\end{aligned}$$

où nous avons posé $C = \sup_{\alpha\in A} \big|T_\alpha(x_0)\big|$. En normalisant, sur la boule $B_J(0, 1)$ nous avons

$$\big|T_\alpha(y)\big| \leqslant \rho^{-1}(k_0 + C). \tag{26.10}$$

2. Au sens où les ouverts sont les mêmes.

Enfin su $x \in E$ nous avons

$$\frac{x}{\max_{j\in J} p_j(x)} \in B_J(0,1) \tag{26.11}$$

et donc

$$\left|T_\alpha\left(\frac{x}{\max_{j\in J}}p_j(x)\right)\right| \leqslant \rho^{-1}(k_0+C). \tag{26.12}$$

Utilisant encore la linéarité de T_α nous trouvons ce que nous devions trouver :

$$\left|T_\alpha(x)\right| \leqslant \max_{j\in J} p_j(x)\rho^{-1}(k_0+C) \tag{26.13}$$

à redéfinition près de $\rho^{-1}(k_0+C)$ en C. □

26.2 Espaces de Lebesgue L^p

26.2.1 Généralités

Définition 26.8.
*Soit $(\Omega, \mathcal{F}, \mu)$ un espace mesuré. Deux fonctions à valeurs complexes f et g sur cet espaces sont dites **équivalentes** et nous notons $f \sim g$ si elles sont μ-presque partout égales. Nous notons $[f]$ la classe de f pour cette relation.*

Nous allons souvent noter f indifféremment pour la fonction f et un des représentants de la classe de f. Toutefois, lorsque la distinction sera importante, nous essayerons de faire faire l'effort de distinguer la fonction f de sa classe $[f]$.

Lemme 26.9.
Une classe contient au maximum une seule fonction continue.

Démonstration. Soient deux fonctions continues f_1 et f_2 avec $f_1(a) \neq f_2(a)$. Si $|f_1(a) - f_2(a)| = \delta$ alors il existe un ϵ tel que $|f_1(x) - f_1(a)| < \delta$ pour tout $x \in B(a,\epsilon)$. En particulier $f_1 \neq f_2$ sur $B(a,\epsilon)$. Cette dernière boule est de mesure de Lebesgue non nulle ; ergo f_1 et f_2 ne sont pas dans la même classe. □

Définition 26.10.
Pour $p \geqslant 1$, nous introduisons l'opération

$$\|f\|_p = \left(\int_\Omega |f(x)|^p d\mu(x)\right)^{1/p} \tag{26.14}$$

et nous notons $\mathcal{L}^p(\Omega,\mu)$ l'ensemble des fonctions mesurables sur Ω telles que $\|f\|_p < \infty$.

26.11.
Le fait que f doive être mesurable pour être dans L^p est bien dans la définition de L^p, et non une propriété qui pourrait être déduite de la finitude de l'intégrale (26.14).

Soit $\Omega = \mathbb{R}$ muni de ses boréliens ou même de sa tribu de Lebesgue et de la mesure de Lebesgue. Si A est une partie bornée non mesurable (exemple 13.141), alors nous considérons un borné mesurable K contenant A et la fonction

$$f(x) = \begin{cases} 1 & \text{si } x \in A \\ -1 & \text{si } x \in K\backslash A \\ 0 & \text{sinon.} \end{cases} \tag{26.15}$$

Nous avons alors $|f| = \mathbb{1}_K$. Vu que K est borné, $\int_{\mathbb{R}} |f| = \mu(K)$ ne présente pas de problèmes. Donc f serait un élément de $L^1(\mathbb{R})$ sans être mesurable.

Pour éviter cela, nous incluons la mesurabilité dans la définition de $\mathcal{L}^p$.

Lemme 26.12 ([1]).
L'ensemble $\mathcal{L}^p$ est un espace vectoriel sur $\mathbb{C}$.

Démonstration. Pour rappel, nous ne considérons les choses que pour $p \geqslant 1$. Le fait que si $f \in L^p$, alors $\lambda f \in L^p$ est évident. Ce qui est moins immédiat, c'est le fait que $f + g \in L^p$ lorsque f et g sont dans L^p. Cela découle d'une part du fait que la fonction $\varphi\colon x \mapsto x^p$ est convexe sur les positifs (lemme 16.85), de telle sorte que

$$\varphi\left(\frac{|a| + |b|}{2}\right) \leqslant \frac{\varphi(|a|) + \varphi(|b|)}{2}, \tag{26.16}$$

ou encore

$$(|a| + |b|)^p \leqslant 2^{p-1}(|a|^p + |b|^p) \tag{26.17}$$

Et d'autre part, nous savons que pour $z_1, z_2 \in \mathbb{C}$, $|z_1 + z_2| \leqslant |z_1| + |z_2|$ (proposition 1.168(5)). Donc

$$\begin{aligned}
\int_\Omega |f + g|^p d\mu &= \int_\Omega |f(\omega) + g(\omega)|^p d\mu(\omega) && (26.18a)\\
&\leqslant \int_\Omega \big(|f(\omega)| + |g(\omega)|\big)^p d\mu(\omega) && (26.18b)\\
&\leqslant 2^{p-1}\big(|f(\omega)|^p + |g(\omega)|^p\big) d\mu(\omega) && (26.18c)\\
&= 2^{p-1}\big(\|f\|_p^p + |g|_p^p\big). && (26.18d)
\end{aligned}$$

Donc si f et g sont dans $\mathcal{L}^p$, alors $f + g$ est dans $\mathcal{L}^p$. □

26.13.
Il est à noter que nous ne considérons que des valeurs $p \geqslant 1$, précisément parce que la fonction $x \mapsto |x|^p$ n'est pas convexe lorsque $p < 1$.

Dans le même ordre d'idées, si $p \geqslant 1$, alors le $q \in \mathbb{R}$ tel que

$$\frac{1}{p} + \frac{1}{q} = 1 \tag{26.19}$$

est également $q \geqslant 1$. Cela est important pour un certain nombres de théorèmes qui vont venir, en particulier l'inégalité de Hölder (26.59).

Si vous en voulez à propos de $0 < p < 1$, vous pouvez lire [446].

26.14.
L'opération $f \mapsto \|f\|_p$ n'est pas une norme sur $\mathcal{L}^p$ parce que pour f presque partout nulle, nous avons $|f|_p = 0$. Il y a donc des fonctions non nulles sur lesquelles $\|.\|_p$ s'annule.

26.15.
Soit un espace mesuré non complet[3] $(\Omega, \mathcal{A}, \mu)$. Il existe une partie N de mesure nulle et $A \subset N$ non mesurable. Considérons

$$f(x) = \begin{cases} 2 & \text{si } x \in A \\ 1 & \text{si } x \in N \backslash A \\ 0 & \text{sinon.} \end{cases} \tag{26.20}$$

Cette fonction est non nulle exactement sur N. Donc $f \sim 0$. Mais f n'est pas mesurable parce que $f^{-1}(2) = A$ n'est pas mesurable.

Il est donc possible d'être dans la classe d'équivalence d'une fonction mesurable sans être mesurable. Ceci est cependant un détail (presque) sans importances pour deux raisons.

— La mesure de Lebesgue est complète par définition (définition 13.127) si nous considérons bien la tribu de Lebesgue et non seulement les boréliens.

3. Définition 13.56.

— Dans le cas des espaces de Lebesgue $L^p(\Omega, \mathcal{A}, \mu)$, il s'agit d'un quotient de $\mathcal{L}^p$ qui ne contient que des fonctions mesurables. Donc dans l'étude de L^p, tous les représentants sont mesurables, même si $(\Omega, \mathcal{A}, \mu)$ n'est pas complet.

Lemme 26.16.
Si $f \in \mathcal{L}^p(\Omega)$ et $f \sim g$, alors $g \in \mathcal{L}^p(\Omega)$ et $\|f\|_p = \|g\|_p$.

Démonstration. Soit $h(x) = |g(x)|^p - |f(x)|^p$; c'est une fonction par hypothèse presque partout nulle et donc intégrable sur Ω ; son intégrale y vaut zéro. Nous avons

$$\int_\Omega |f(x)|^p d\mu(x) = \int_\Omega \Big(|f(x)|^p + h(x)\Big) d\mu(x) = \int_\omega |g(x)|^p d\mu(x). \tag{26.21}$$

Cela prouve que la dernière intégrale existe et vaut la même chose que la première. □

Nous pouvons donc considérer la norme $|.|_p$ comme une norme sur l'ensemble des classes plutôt que sur l'ensemble des fonctions. Nous notons L^p l'ensemble des classes des fonctions de $\mathcal{L}^p$. Cet espace est muni de la norme

$$\|[f]\|_p = \|f\|_p, \tag{26.22}$$

formule qui ne dépend pas du représentant par le lemme 26.16.

Proposition 26.17.
Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$.

(1) L'ensemble des classes $L^p(\Omega, \mathcal{A}, \mu)$ est un espace vectoriel.

(2) La formule

$$\|[f]\|_p = \left(\int_\Omega |f(x)|^p d\mu(x)\right)^{1/p} \tag{26.23}$$

défini une norme[4] *sur $L^p(\Omega, \mathcal{A}, \mu)$.*

Démonstration. En deux parties.

Espace vectoriel Le lemme 26.12 dit que $\mathcal{L}^p$ est un espace vectoriel. Une structure d'espace vectoriel est donnée en posant

$$[f] + [g] = [f + g] \tag{26.24}$$

et

$$\lambda[f] = [\lambda f] \tag{26.25}$$

qui sont deux définitions correctes parce qu'elles ne dépendent pas du choix du représentant. De plus le lemme 26.12 dit que $f + g \in \mathcal{L}^p$ dès que $f, g \in \mathcal{L}^p$. Donc $[f + g] \in L^p$ dès que $[f], [g] \in \text{Ł}^p$.

Norme Pour être une norme, il faut vérifier les trois propriétés de la définition 7.113.

D'abord, si $\|[f]\|_p = 0$, nous avons

$$\int_\Omega |f(x)|^p d\mu(x) = 0, \tag{26.26}$$

ce qui par le lemme 13.179 implique que $|f(x)|^p = 0$ pour presque tout x. Ou encore $f \sim 0$, c'est-à-dire $[f] = [0]$ au niveau des classes.

Ensuite pour $\lambda \in \mathbb{C}$ et $f \in L^p$ nous avons

$$\|\lambda f\|_p = \left(\int_\Omega |\lambda f(\omega)|^p d\mu(\omega)\right)^{1/p} \tag{26.27a}$$

$$= \left(\int_\Omega |\lambda|^p |f(\omega)|^p d\mu(\omega)\right)^{1/p} \tag{26.27b}$$

$$= |\lambda| \|f\|_p. \tag{26.27c}$$

4. Définition 7.113.

Et enfin, en suivant le calcul (26.18) nous avons

$$\|f+g\|_p \leqslant \|f\|_p + \|g\|_p. \tag{26.28}$$

□

26.18.
À partir de maintenant $\big(L^p(\Omega,\mu), \|.\|_p\big)$ est un espace métrique avec toute la topologie qui va avec.

Dans la suite nous n'allons pas toujours écrire $[f]$ pour la classe de f. Par abus de notations nous allons souvent parler de $f \in L^p$ comme si c'était une fonction.

De même nous notons $L^p(\Omega)$ ou $L^p(\Omega,\mu)$ ou $L^p(\Omega,\mathcal{A},\mu)$ d'après ce sur quoi nous voulons insister. Mais seule la dernière notation est parfaitement correcte.

26.2.2 Un peu de convergence de suites

Proposition 26.19 ([447]).
Soit $1 \leqslant p \leqslant \infty$ et supposons que la suite $[f_n]$ dans $L^p(\Omega,\mathcal{F},\mu)$ converge vers $[f]$ au sens L^p. Alors il existe une sous-suite (h_n) qui converge ponctuellement μ-presque partout vers f.

Démonstration. Si $p=\infty$ nous sommes en train de parler de la convergence uniforme et il ne faut même pas prendre ni de sous-suite ni de « presque partout ».

Supposons que $1 \leqslant p < \infty$. Nous considérons une sous-suite $[h_n]$ de $[f_n]$ telle que

$$\|[h_j]-[f]\|_p < 2^{-j}, \tag{26.29}$$

puis nous posons $u_k(x) = |h_k(x)-f(x)|^p$. Notons que ce u_k est une vraie fonction, pas une classe. Et en plus c'est une fonction positive. Nous avons

$$\int_\Omega u_k d\mu = \int_\omega |h_k(x)-f(x)|^p d\mu(x) = \|h_k - f\|_p^p \leqslant 2^{-kp}. \tag{26.30}$$

Vu que u_k est une fonction positive la suite des sommes partielles de $\sum_k u_k$ est croissante et vérifie donc le théorème de la convergence monotone 13.160 :

$$\int_\Omega \left(\sum_{k=0}^\infty u_k(x)\right) d\mu(x) = \sum_{k=0}^\infty \int_\Omega u_k(x) d\mu(x) \leqslant \sum_{k=0}^\infty 2^{-kp} < \infty. \tag{26.31}$$

Le fait que l'intégrale de la fonction $\sum_k u_k$ est finie implique que cette fonction est finie μ-presque partout. Donc le terme général tend vers zéro presque partout, c'est-à-dire

$$|h_k(x)-f(x)|^p \to 0. \tag{26.32}$$

Cela signifie que $h_k \to f$ presque partout ponctuellement. □

Est-ce qu'on peut faire mieux que la convergence ponctuelle presque partout d'une sous-suite ? En tout cas on ne peut pas espérer grand chose comme convergence pour la suite elle-même, comme le montre l'exemple suivant.

Exemple 26.20.
Nous allons montrer une suite de fonctions qui converge vers zéro dans $L^p[0,1]$ (avec $p<\infty$) mais qui ne converge ponctuellement pour *aucun* point. Cet exemple provient de bibmath.net.

Nous construisons la suite de fonctions par paquets. Le premier paquet est formé de la fonction constante 1.

Le second paquet est formé de deux fonctions. La première est $\mathbb{1}_{[0,1/2]}$ et la seconde $\mathbb{1}_{[1/2,1]}$.

Plus généralement le paquet numéro k est constitué des k fonctions $\mathbb{1}_{[i/k,(i+1)/k]}$ avec $i = 0,\dots,k-1$.

Vu que les fonctions du paquet numéro k ont pour norme $\|f\|_p = \frac{1}{k}$, nous avons évidemment $f_n \to 0$ dans L^p. Il est par contre visible que chaque paquet passe en revue tous les points de $[0,1]$. Donc pour tout x et pour tout N, il existe (même une infinité) $n > N$ tel que $f_n(x) = 1$. Il n'y a donc convergence ponctuelle nulle part. △

La proposition suivante est une espèce de convergence dominée de Lebesgue pour L^p.

Proposition 26.21.
Soit $f \in L^p(\Omega)$ avec $1 \leqslant p < \infty$ et (f_n) une suite de fonctions convergeant ponctuellement vers f et telle que $|f_n| \leqslant |f|$. Alors $f_n \xrightarrow{L^p} f$.

Démonstration. Nous avons immédiatement $|f_n(x)|^p \leqslant |f(x)|^p$, de telle sorte que le théorème de la convergence dominée implique que $f_n \in L^p$. La convergence dominée donne aussi que $\|f_n\|_p \to \|f\|_p$, mais cela ne nous intéresse pas ici.

Nous posons $h_n(x) = |f_n(x) - f(x)|$. En reprenant la formule de majoration (26.17) et en tenant compte du fait que $|f_n(x)| \leqslant |f(x)|$, nous avons

$$h_n(x) \leqslant 2^{p-1}\big(|f_n(x)|^p + |f(x)|^p\big) \leqslant 2^p |f(x)|^p, \tag{26.33}$$

ce qui prouve que $|h_n|$ est uniformément (en n) majorée par une fonction intégrable, donc h_n est intégrable et on peut permuter la limite et l'intégrale (théorème de la convergence dominée 13.184) :

$$\lim_{n\to\infty} \|f_n - f\|_p^p = \lim_{n\to\infty} \int_{\mathbb{R}^d} |f_n(x) - f(x)|^p dx = \int_{\mathbb{R}^d} \lim_{n\to\infty} h_n(x) dx = 0. \tag{26.34}$$

□

Proposition 26.22 ([1]).
Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$ et une fonction mesurable $f\colon \Omega \to \mathbb{C}$ telle que

$$\int_\Omega f \mathbb{1}_A d\mu = 0 \tag{26.35}$$

pour toute partie mesurable A de mesure finie. Alors $[f] = 0$, c'est-à-dire que f est non nulle uniquement sur une partie de mesure nulle.

Démonstration. Nous rappelons que dire que φ est intégrable signifie que $\mathrm{Re}(f)^+$, $\mathrm{Re}(f)^-$, $\mathrm{Im}(f)^+$ et $\mathrm{Im}(f)^-$ sont intégrables.

Les parties

$$\{x \in \Omega \text{ tel que } \mathrm{Re}(f)^+(x) \in [k, k+1[\}_{k=1,\dots} \tag{26.36}$$

et

$$\{x \in \Omega \text{ tel que } \mathrm{Re}(f)^+(x) \in [\frac{1}{k+1}, \frac{1}{k}[\}_{k=1,\dots} \tag{26.37}$$

sont mesurables en tant qu'images inverses de mesurables par la fonction mesurable $\mathrm{Re}(f)^+$.

Nous notons $\{A_i\}_{i\in\mathbb{N}}$ une énumération quelconque de ces parties. L'important est que

$$\bigcup_{i\in\mathbb{N}} A_i = \{x \in \Omega \text{ tel que } \mathrm{Re}(f)^+(x) \neq 0\} \tag{26.38}$$

et que pour chaque i, il existe $\alpha_i > 0$ tel que $\mathrm{Re}(f)^+(x) > \alpha_i$ pour tout x dans A_i.

Par hypothèse nous avons $\int_\Omega f \mathbb{1}_{A_i} d\mu = 0$. En particulier,

$$0 = \int_\Omega \mathrm{Re}(f\mathbb{1}_{A_i})^+ = \int_\Omega \mathrm{Re}(f)^+(x) \mathbb{1}_{A_i}(x) d\mu(x). \tag{26.39}$$

Mais pour tout $x \in A_i$ nous avons $\mathrm{Re}(f)^+(x) > \alpha_i > 0$, donc

$$0 = \int_\Omega \mathrm{Re}(f)^+(x) \mathbb{1}_{A_i}(x) d\mu(x) \geqslant \alpha_i \int_\Omega \mathbb{1}_{A_i} = \alpha_i \mu(A_i) \geqslant 0. \tag{26.40}$$

Vu cet encadrement par zéro, nous avons $\alpha_i \mu(A_i) = 0$ et donc $\mu(A_i) = 0$ pour tout i.

Nous en déduisons, par union dénombrable de parties de mesures nulles, que $\mathrm{Re}(f)^+$ est non nulle seulement sur une partie de mesure nulle.

Le même raisonnement pour $\mathrm{Re}(f)^-$, $\mathrm{Im}(f)^+$ et $\mathrm{Im}(f)^-$ donne que f est non nulle sur une partie de mesure nulle. Donc $f = 0$ au sens des classes dans L^p. □

26.2.3 L'espace L^∞

Il n'est pas possible de définir le supremum d'une fonction définie à ensemble de mesure nulle près parce que toute classe contient des fonctions qui peuvent être arbitrairement grandes en n'importe que point. Nous cherchons alors à définir une notion de supremum qui ne tient pas compte des ensembles de mesure nulle.

Définition 26.23.
Soit $f\colon \Omega \to \mathbb{C}$. Un nombre M est un ***majorant essentiel*** *de f si*

$$\mu\big(|f(x)| \leqslant M\big) = 0. \tag{26.41}$$

Nous posons alors

$$N_\infty(f) = \inf\{M \text{ tel que } |f(x)| \leqslant M \text{ presque partout}\}. \tag{26.42}$$

Cela revient à prendre le supremum à ensemble de mesure nulle près.

Définition 26.24.
Nous définissons alors les espaces de Lebesgue correspondants :

$$\mathcal{L}^\infty(\Omega) = \{f\colon \Omega \to \mathbb{C} \text{ tel que } N_\infty(f) < \infty\}, \tag{26.43}$$

et L^∞ en est le quotient usuel.

26.25.
Le point sur les notations telles qu'elles devraient être respectées :

(1) Si f est une fonction, $N_\infty(f)$ est son supremum essentiel.

(2) Si f est une fonction, $\|f\|_\infty$ est sa norme supremum. Ce n'est pas la même chose que $N_\infty(f)$.

(3) Si $[f]$ est une classe de fonctions (pour l'égalité presque partout), alors $\|[f]\|_{L^\infty}$ est la norme de cette classe dans L^∞, c'est-à-dire $\|[f]\|_{L^\infty} = N_\infty(f)$ où f est un représentant.

Le point sur les abus tolérables :

(1) Si f est une fonction, on peut écrire $\|f\|_{L^\infty}$ pour $N_\infty(f)$. Attention toutefois que $N_\infty(f)$ peut valoir ∞, alors que les éléments de L^∞ sont sélectionnés pour être les classes des fonctions telles que $N_\infty(f) < \infty$. Donc il est parfois possible, lorsque f est une fonction, de parler de $\|f\|_{L^\infty}$ alors que $[f]$ n'est pas un élément de L^∞.

(2) Si $[f]$ est une classe, on peut écrire $N_\infty([f])$ pour $\|[f]\|_{L^\infty}$.

(3) Noter f la classe de la fonction f. Attention qu'alors, écrire $\|f\|_\infty$ n'a pas de sens.

Le point sur les abus intolérable :

(1) Si f est une fonction, noter $\|f\|_\infty$ pour $N_\infty(f)$.

(2) Si $[f]$ est une classe de fonctions, noter $\|[f]\|_\infty$ pour $\|[f]\|_{L^\infty}$.

26.26.
Tout ceci pour dire que si f_k sont des fonctions et si f est une fonction, nous avons la convergence

$$f_k \xrightarrow{L^\infty} f \tag{26.44}$$

si et seulement si $N_\infty(f_k - f) \to 0$. Cette convergence (qui se sert des abus tolérables de notations) est équivalente à la convergence

$$[f_k] \xrightarrow{L^\infty} [f]. \tag{26.45}$$

Mais ce n'est pas du tout équivalent à $\|f_k - f\|_\infty \to 0$. Tout au plus, il est vrai que si $\alpha \in L^\infty$ (donc α est une classe), il existe un représentant $f \in \alpha$ tel que $\|f\|_\infty = \alpha_{L^\infty}$.

Lemme 26.27 ([1]).
Si f_k est une suite de fonctions telle que $\|f_k - f\|_\infty \to 0$, alors

$$f_k \xrightarrow{L^\infty} f \tag{26.46}$$

où la convergence signifie $N_\infty(f_k - f) \to 0$.

Démonstration. Si g est une fonction nous avons toujours $N_\infty(g) \leqslant \|g\|_\infty$. Donc

$$N_\infty(f_k - f) \leqslant \|f_k - f\|_\infty \to 0. \tag{26.47}$$

□

26.28.
Attention toutefois que ce lemme ne signifie pas que si $\|f_k - f\|_\infty \to 0$, alors $[f_k] \xrightarrow{L^\infty} [f]$ parce que nous pourrions avoir $N_\infty(f_k) = \infty$ et alors $[f_k]$ n'est pas un élément de L^∞.

Cela arrive par exemple pour $f_k(x) = x$ et $f(x) = x$. Nous avons $\|f_k - f\|_\infty = 0$ pour tout k, alors que ni f_k ni f ne donnent lieu à un élément de L^∞.

Exemple 26.29.
Même si g est bornée, nous n'avons pas spécialement $\|g\|_\infty = N_\infty(g)$. Par exemple

$$\begin{aligned} g\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} 1 & \text{si } x = 0 \\ 0 & \text{sinon .} \end{cases} \end{aligned} \tag{26.48}$$

Cette fonction vérifie $\|g\|_\infty = 1$ mais $N_\infty(g) = 0$. △

26.2.4 Quelques identifications

Il est intuitivement clair que ce qui peut arriver à une fonction en un seul point ne va pas influencer la fonction lorsqu'elle est vue dans L^p. En tout cas lorsqu'on considère des mesures pour lesquelles les singletons sont de mesure nulle, et c'est bien le cas de la mesure de Lebesgue. Il est peut-être intuitivement moins clair que l'on peut non seulement modifier le comportement d'une fonction en un point, mais également modifier l'ensemble de base. En voici un exemple.

Proposition 26.30.
Nous avons les égalités suivantes d'espaces

$$L^p\big(]0, 2\pi[\big) = L^p\big([0, 2\pi]\big) = L^p(S^1) \tag{26.49}$$

au sens où il existe des bijections isométriques de l'un à l'autre. Ici nous sous-entendons la mesure de Lebesgue partout[5].

Démonstration. Voici une application bien définie où le crochet dénote la prise de classe :

$$\begin{aligned} \psi\colon L^p\big(]0, 2\pi[\big) &\to L^p\big([0, 2\pi]\big) \\ [f] &\mapsto \text{la classe de } f_e(x) = \begin{cases} f(x) & \text{si } x \in]0, 2\pi[\\ 0 & \text{si } x = 0 \text{ ou } x = 2\pi. \end{cases} \end{aligned} \tag{26.50}$$

Injective Si $[f] = [g]$ dans $L^p\big(]0, 2\pi[\big)$ alors $f_e(x) = g_e(x)$ pour tout $x \in]0, 2\pi[$ sauf une partie de mesure nulle. L'union de cette partie avec $\{0, 2\pi\}$ est encore de mesure nulle dans $[0, 2\pi]$. Les images par ψ sont donc égales dans $L^p\big([0, 2\pi]\big)$.

5. Vu que la mesure de Lebesgue est définie pour $\mathbb{R}^d$ munie de sa tribu des boréliens (complétée), vous êtes en droit de vous demander quelle est la tribu et la mesure que nous considérons sur le cercle S^1.

Surjective Un élément de $L^p\big([0,2\pi]\big)$ est l'image de sa restriction ... ou plutôt l'image de la classe de la restriction d'un quelconque de ses représentants.

Isométrie L'intégrale qui donne la norme sur L^p ne change pas selon que nous ajoutions ou non les bornes au domaine d'intégration.

De la même manière nous avons

$$L^p\big([0,2\pi[\big) = L^p\big([0,2\pi]\big). \tag{26.51}$$

En ce qui concerne l'identification avec $L^p(S^1)$, il faut passer par l'isométrie $\varphi\colon [0,2\pi[\to S^1$ donnée par $\varphi(t) = e^{it}$, et être heureux que ce soit bien une isométrie parce qu'il faudra l'utiliser pour un changement de variables pour montrer que

$$\int_0^{2\pi} f(t)dt = \int_{S^1} (f\circ\varphi^{-1})(z)dz. \tag{26.52}$$

□

26.2.5 Inégalité de Jensen, Hölder et de Minkowski

Proposition 26.31 (Inégalité de Jensen[298])**.**
Soit un espace mesuré de probabilité[6] $(\Omega,\mathcal{A},\mu)$ *ainsi qu'une fonction convexe* $f\colon\mathbb{R}\to\mathbb{R}$ *et une application* $\alpha\colon\Omega\to\mathbb{R}$ *tels que* α *et* $f\circ\alpha$ *soient intégrables sur* Ω*. Alors*

$$f\Big(\int_\Omega \alpha\, d\mu\Big) \leqslant \int_\Omega (f\circ\alpha)d\mu. \tag{26.53}$$

Démonstration. Soient $a\in\mathbb{R}$ et c_a le nombre donné par la proposition 16.88 : pour tout $\omega\in\Omega$ nous avons

$$f\big(\alpha(\omega)\big) - f(a) \geqslant c_a\big(\alpha(\omega)-a\big). \tag{26.54}$$

Cela est en particulier vrai pour $a = \int_\Omega \alpha\, d\mu$. Nous intégrons l'inégalité (26.54) sur Ω en nous souvenant que $\int d\mu = 1$:

$$\int_\Omega (f\circ\alpha)d\mu - \int_\Omega f(a)d\mu \geqslant c_a\Big(\int_\Omega \alpha - \int_\Omega a\Big) \tag{26.55a}$$

$$\int_\Omega (f\circ\alpha)d\mu - f(a) \geqslant c_a(a-a) \tag{26.55b}$$

$$f(a) \leqslant \int_\Omega (f\circ\alpha)d\mu. \tag{26.55c}$$

Cette dernière inégalité est celle que nous devions prouver. □

Corolaire 26.32.
Soit un espace mesuré de probabilité $(\Omega,\mathcal{A},\mu)$ *et une application* $\alpha\in L^1(\Omega,\mu)$ *et* $\alpha\in L^p(\Omega)$ *avec* $1\leqslant p<+\infty$*. Alors*

$$\Big|\int_\Omega \alpha(s)d\mu(s)\Big| \leqslant \|\alpha\|_p. \tag{26.56}$$

Démonstration. Il suffi d'utiliser l'inégalité de Jensen sur la fonction convexe $f(x)=|x|^p$. Nous avons alors

$$\Big|\int_\Omega \alpha(s)d\mu(s)\Big|^p \leqslant \int_\Omega |\alpha(s)|^p d\mu(s), \tag{26.57}$$

c'est-à-dire

$$\Big|\int_\Omega \alpha(s)d\mu(s)\Big| \leqslant \Big[\int_\Omega |\alpha(s)|^p d\mu(s)\Big]^{1/p} = \|\alpha\|_p \tag{26.58}$$

où ma norme $\|.\|_p$ est prise au sens de la mesure μ. □

6. C'est-à-dire que $\int_\Omega d\mu = 1$.

Proposition 26.33 (Inégalité de Hölder)**.**
Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré et $1 \leqslant p$, $q \leqslant \infty$ satisfaisant $\frac{1}{p} + \frac{1}{q} = 1$. Si $f \in L^p(\Omega)$, $g \in L^q(\Omega)$, alors le produit fg est dans $L^1(\Omega)$ et nous avons

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q. \tag{26.59}$$

Plus généralement si $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$ alors

$$\|fg\|_r \leqslant \|f\|_p \|g\|_q \tag{26.60}$$

Ce qui est appelé « inégalité de Hölder » est généralement l'inéquation (26.59).

Démonstration. Nous allons prouver l'inégalité (26.60). D'abord nous supposons $\|g\|_q = 1$ et nous posons

$$A = \{x \in \Omega \text{ tel que } |g(x)| > 0\}. \tag{26.61}$$

Hors de A, les intégrales que nous allons écrire sont nulles. Nous avons

$$\|fg\|_r^p = \Big| \int_A |f|^r |g|^{r-q} |g|^q \Big|^{p/r}, \tag{26.62}$$

et le coup tordu est de considérer cette intégrale comme étant une intégrale par rapport à la mesure $\nu = |g|^q d\mu$ qui a la propriété d'être une mesure de probabilité par hypothèse sur g. Nous pouvons alors utiliser l'inégalité de Jensen [7] parce que $p/r > 1$, ce qui fait de $x \mapsto |x|^{p/r}$ une fonction convexe. Nous avons alors

$$\|fg\|_r^p \leqslant \int_A \big(|f|^r |g|^{r-q}\big)^{p/r} |g|^q d\mu \tag{26.63a}$$

$$= \int_A |f|^p |g|^{p(r-q)/r} |g|^q d\mu \tag{26.63b}$$

La puissance de $|g|$ dans cette expression est : $q + \frac{p(r-q)}{r} = 0$ parce que $p(q-r) = rq$. Nous avons alors montré que

$$\|fg\|_r^p \leqslant \int_A |f|^p d\mu \leqslant \|f\|_p^p. \tag{26.64}$$

La dernière inégalité est le fait que le domaine A n'est pas tout le domaine Ω.

Si maintenant $\|g\|_q \neq 1$ alors nous calculons

$$\|fg\|_r = \|g\|_q \|f \frac{g}{\|g\|_q}\|_r \leqslant \|g\|_q \|f\|_p \tag{26.65}$$

en appliquant la première partie à la fonction $\frac{g}{\|g\|_q}$ qui est de norme 1. □

Remarque 26.34.
Dans le cas d'un espace de probabilité, la fonction constante $g = 1$ appartient à $L^p(\Omega)$. En prenant $p = q = 2$ nous obtenons

$$\|f\|_1 \leqslant \|f\|_2. \tag{26.66}$$

Lemme 26.35.
Lorsque I est borné nous avons $L^2(I) \subset L^1(I)$. Si I n'est pas borné alors $L^2(I) \subset L^1_{loc}(I)$.

Démonstration. En effet si I est borné, alors la fonction constante 1 est dans $L^2(I)$ et l'inégalité de Hölder 26.33 nous dit que le produit $1u$ est dans $L^1(I)$.

Si I n'est pas borné, nous refaisons le même raisonnement sur un compact K de I. □

7. Proposition 26.31.

Corolaire 26.36 ([448]).
Soit l'espace $L^2(I)$ avec $I =]0,1[$ avec la mesure de Lebesgue. Si $u_n \in L^2$ converge vers u dans L^2 alors nous pouvons permuter l'intégrale et la limite :

$$\lim_{n\to\infty} \int_I u_n = \int_I u. \tag{26.67}$$

Démonstration. Nous considérons la forme linéaire

$$\begin{aligned} T\colon L^2(I) &\to \mathbb{R} \\ u &\mapsto \int_I u. \end{aligned} \tag{26.68}$$

Elle est bien définie par l'inégalité de Hölder $\|fg\|_1 \leqslant \|f\|_2\|g\|_2$ appliqué à $g(x) = 1$ qui vérifie $\|g\|_2 = 1$. Nous avons aussi

$$T(u) \leqslant \int_I |u| \leqslant \|u\|_1 \leqslant \|u\|_2 \tag{26.69}$$

où la dernière inégalité est celle de Hölder 26.33. Bref, T est continue. Cela signifie que si $u_n \xrightarrow{L^2(I)} u$ alors $T(u_n) = T(u)$. Cela est l'égalité demandée. □

Proposition 26.37 (Inégalité de Minkowski[449, 450]).
Si $1 \leqslant p < \infty$ et si $f, g \in L^p(\Omega, \mathcal{A}, \mu)$ alors

(1) $\|f + g\|_p \leqslant \|f\|_p + \|g\|_p$
(2) Il y a égalité si et seulement si les vecteurs $f(x)$ et $g(x)$ sont presque partout colinéaires : il existe α, β tels que $\alpha f + \beta g = 0$ presque partout.
(3) Si $f(x,y)$ est mesurable sur l'espace produit $\big(X \times Y, \mu \otimes \nu\big)$ et si $p \geqslant 1$, alors

$$\Big\| x \mapsto \int_Y f(x,y) d\nu(y) \Big\|_p \leqslant \int_Y \|f_y\|_p d\nu(y) \tag{26.70}$$

où $f_y(x) = f(x,y)$.

La partie (3) est une généralisation de l'inégalité triangulaire (c'est-à-dire du point (1)) dans le cas où nous n'avons pas une somme de deux fonctions mais d'une infinité paramétrée par $y \in Y$. Elle sera le plus souvent utilisée sous la forme déballée :

$$\Big[\int_X \Big(\int_Y |f(x,y)| d\nu(y)\Big)^p d\mu(x)\Big]^{1/p} \leqslant \int_Y \Big(\int_X |f(x,y)|^p d\mu(x)\Big)^{1/p} d\nu(y). \tag{26.71}$$

26.2.6 Ni inclusions ni inégalités

Aucun espace $L^p(\mathbb{R})$ n'est inclus dans aucun autre ni aucune norme n'est plus grande qu'une autre (sur les intersections). Nous verrons cependant en la proposition 26.38 que de telles inclusions et inégalités sont possibles pour $L^p\big([0,1]\big)$.

Nous allons donner des exemples de tout ça en supposant $p < q$ et en nous appuyant lourdement sur les intégrales de $\frac{1}{x^\alpha}$ étudiées par la proposition 13.251.

$L^p \nsubseteq L^q$ La fonction

$$f(x) = \begin{cases} \frac{1}{x^{1/q}} & \text{si } 0 < x < 1 \\ 0 & \text{sinon} \end{cases} \tag{26.72}$$

est dans L^p mais pas dans L^q. En effet

$$\|f\|_p^p = \int_0^1 \frac{1}{x^{p/q}} dx < \infty \tag{26.73}$$

parce que $p < q$ et $p/q < 1$. Par contre

$$\|f\|_q^q = \int_0^1 \frac{1}{x} dx = \infty. \tag{26.74}$$

$L^q \not\subseteq L^p$ La fonction

$$f(x) = \begin{cases} \frac{1}{x^{1/p}} & \text{si } x > 1 \\ 0 & \text{sinon} \end{cases} \tag{26.75}$$

est dans L^q mais pas dans L^p. En effet

$$\|f\|_p^p = \int_1^\infty \frac{1}{x} = \infty \tag{26.76}$$

alors que

$$\|f\|_q^q = \int_1^\infty \frac{1}{x^{q/p}} dx < \infty. \tag{26.77}$$

Exemple de $\|f\|_p > \|f\|_q$ La fonction

$$f(x) = \begin{cases} 1 & \text{si } x \in [0,2] \\ 0 & \text{sinon.} \end{cases} \tag{26.78}$$

Nous avons

$$\|f\|_p = 2^{1/p} \tag{26.79a}$$
$$\|f\|_q = 2^{1/q}. \tag{26.79b}$$

Mais comme $p < q$ donc $\|f\|_p > \|f\|_q$.

Exemple de $\|f\|_p < \|f\|_q$ La fonction

$$f(x) = \begin{cases} 1 & \text{si } x \in [0,\frac{1}{2}] \\ 0 & \text{sinon.} \end{cases} \tag{26.80}$$

Alors

$$\|f\|_p = \frac{1}{2^{1/p}} \tag{26.81a}$$
$$\|f\|_q = \frac{1}{2^{1/q}} \tag{26.81b}$$

et donc $\|f\|_p < \|f\|_q$.

Ces exemples donnent un exemple de fonction f telle que $\|f\|_p < \|f\|_q$ pour tout espace $L^p(I)$ et $L^q(I)$ avec $I \subset \mathbb{R}$. Par contre l'exemple $\|f\|_p > \|f\|_q$ ne fonctionne que si la taille de I est plus grande que 1. Et pour cause : il y a des inclusions si I est borné.

Proposition 26.38 ([6]).
Inclusions et inégalités dans le cas d'un ensemble de mesure finie.

(1) Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré fini et $1 \leqslant p \leqslant +\infty$. Alors $L^q(\Omega) \subset L^p(\Omega)$ dès que $p \leqslant q$.

(2) Si $1 < p < 2$ et si $f \in L^2\big([0,1]\big)$ alors $\|f\|_p \leqslant \|f\|_2$.

Démonstration. Pour la simplicité des notations nous allons noter L^p pour $L^p(\Omega)$, et pareillement pour L^q. Soit $f \in L^q$. Nous posons

$$A = \{x \in [0,1] \text{ tel que } |f(x)| \geqslant 1\}. \tag{26.82}$$

Étant donné que $p \leqslant q$ nous avons $|f|^p \leqslant |f|^q$ sur A ; par conséquent $\int_A |f|^p$ converge parce que $\int_A |f|^q$ converge.

L'ensemble A^c est évidemment borné (complémentaire dans Ω) et sur A^c nous avons $|f(x)| \leqslant 1$ et donc $|f|^p \leqslant 1$. L'intégrale $\int_{A^c} |f|^p$ converge donc également.

Au final $\int_\Omega |f|^p$ converge et $f \in L^p$.

Soit à présent $f \in L^2$; par le premier point nous avons immédiatement $f \in L^2 \cap L^p$. Soit aussi $r \in \mathbb{R}$ tel que $\frac{1}{2/p} + \frac{1}{r} = 1$. Nous avons $|f|^p \in L^{2/p}$, et vu que nous sommes sur un domaine borné, $1 \in L^r$. Nous écrivons l'inégalité de Hölder (26.59) avec ces fonctions. D'une part

$$\|f\|_1 = \||f|^p\|_1 = \|f\|_p^p. \tag{26.83}$$

D'autre part

$$\||f|^p\|_{2/p} = \left(\int |f|^2\right)^{p/2} = \|f\|_2^p. \tag{26.84}$$

Donc $\|f\|_p^p \leqslant \|f\|_2^p$, ce qui prouve l'assertion (2) parce que $p > 1$. □

Remarque 26.39.
Nous n'avons cependant pas $L^2\big([0,1]\big) = L^p\big([0,1]\big)$ parce que l'exemple (26.72) fonctionne encore :

$$f(x) = \frac{1}{\sqrt{x}} \tag{26.85}$$

pour $x \in [0,1]$ donne bien

$$\|f\|_2 = \int_0^1 \frac{1}{x} = \infty \tag{26.86}$$

et $\|f\|_p = \int_0^1 \frac{1}{x^{p/2}} < \infty$ parce que $1 < p < 2$.

26.2.7 Complétude

Théorème 26.40 ([451, 452])**.**
Pour $1 \leqslant p < \infty$, l'espace $L^p(\Omega, \mathcal{A}, \mu)$ est complet.

Démonstration. Soit $(f_n)_{n\in\mathbb{N}}$ une suite de Cauchy dans L^p. Pour tout i, il existe $N_i \in \mathbb{N}$ tel que $\|f_p - f_q\|_p \leqslant 2^{-i}$ pour tout $p, q \geqslant N_i$. Nous considérons la sous-suite $g_i = f_{N_i}$, de telle sorte qu'en particulier

$$\|g_i - g_{i-1}\|_p \leqslant 2^{-i}. \tag{26.87}$$

Pour chaque j nous considérons la somme télescopique

$$g_j = g_0 + \sum_{i=1}^{j} (g_i - g_{i-1}) \tag{26.88}$$

et l'inégalité

$$|g_j| \leqslant |g_0| + \sum_{i=1}^{j} |g_i - g_{i-1}|. \tag{26.89}$$

Nous allons noter

$$h_j = |g_0| + \sum_{i=1}^{j} |g_i - g_{i-1}|. \tag{26.90}$$

La suite de fonctions (h_j) ainsi définie est une suite croissante de fonctions positive qui converge donc (ponctuellement) vers une fonction h qui peut éventuellement valoir l'infini en certains points. Par continuité de la fonction $x \mapsto x^p$ nous avons

$$\lim_{j\to\infty} h_j^p = h^p, \tag{26.91}$$

puis par le théorème de la convergence monotone (théorème 13.160) nous avons

$$\lim_{j\to\infty} \int_\Omega h_j^p d\mu = \int_\Omega h^p d\mu. \tag{26.92}$$

Utilisant à présent la continuité de la fonction $x \mapsto x^{1/p}$ nous trouvons

$$\lim_{j\to\infty}\left(\int h_j^p\right)^{1/p} = \left(\int |h|^p\right)^{1/p}. \tag{26.93}$$

Nous avons donc déjà montré que

$$\lim_{j\to\infty}\|h_j\|_p = \left(\int |h|^p\right)^{1/p} \tag{26.94}$$

où, encore une fois, rien ne garantit à ce stade que l'intégrale à droite soit un nombre fini. En utilisant l'inégalité de Minkowski (proposition 26.37) et l'inégalité (26.87) nous trouvons

$$\|h_j\|_p \leqslant \|g_0\|_p + \sum_{i=1}^{j}\|g_i - g_{i-1}\|_p \leqslant \|g_0\|_p + 1. \tag{26.95}$$

En passant à la limite,

$$\left(\int |h|^p\right)^{1/p} = \lim_{j\to\infty}\|h_j\|_p \leqslant \|g_0\|_p + 1 < \infty. \tag{26.96}$$

Par conséquent $\int |h|^p$ est finie et

$$h \in L^p(\Omega, \mathcal{A}, \mu). \tag{26.97}$$

En particulier, l'intégrale $\int h$ est finie (parce que $p \geqslant 1$) et donc que $h(x) < \infty$ pour presque tout $x \in \Omega$.

Nous savons que $h(x)$ est la limite des sommes partielles (26.90), en particulier la série

$$\sum_{j=1}^{\infty}|g_i - g_{i-1}| \tag{26.98}$$

converge ponctuellement. En vertu du corolaire 13.186, la série de terme général $g_i - g_{i-1}$ converge ponctuellement. La suite g_i converge donc vers une fonction que nous notons g. Par ailleurs la suite g_i est dominée par $h \in L^p$, le théorème de la convergence dominée (théorème 13.184) implique que

$$\lim_{j\to\infty}\|g_j - g\|_p = 0. \tag{26.99}$$

Nous allons maintenant prouver que $\lim_{n\to\infty\|f_n - g\|_p} = 0$. Soit $\epsilon > 0$. Pour tout n et i nous avons

$$\|f_n - g\|_p = \|f_n - f_{N_i} + f_{N_i} - g\|_p \leqslant \|f_n - f_{N_i}\|_p + \|f_{N_i} - g\|_p. \tag{26.100}$$

Pour rappel, $f_{N_i} = g_i$. Si i et n sont suffisamment grands nous pouvons obtenir que chacun des deux termes est plus petit que $\epsilon/2$.

Il nous reste à prouver que $g \in L^p(\Omega, \mathcal{A}, \mu)$. Nous avons déjà vu (équation (26.97)) que $h \in L^p$, mais $|g_i| \leqslant h^p$, par conséquent $g \in L^p$.

Nous avons donc montré que la suite de Cauchy (f_n) converge vers une fonction de L^p, ce qui signifie que L^p est complet. □

Théorème 26.41 (Fischer-Riesz[54]).
Soit un ouvert Ω de $\mathbb{R}^n$ et $p \in [1, \infty]$. Alors

(1) Toute suite convergente dans $L^p(\Omega)$ admet une sous-suite convergente presque partout sur Ω.

(2) La sous-suite donnée en (1) est dominée par un élément de $L^p(\Omega)$.

(3) L'espace $L^p(\Omega)$ est de Banach.

Démonstration. Le cas $p = \infty$ est à séparer des autres valeurs de p parce qu'on y parle de norme uniforme, et aucune sous-suite n'est à considérer.

Cas $p = \infty$. Nous commençons par prouver dans le cas $p = \infty$. Soit (f_n) une suite de Cauchy dans $L^\infty(\Omega)$, ou plus précisément une suite de représentants d'éléments de L^p. Pour tout $k \geqslant 1$, il existe $N_k \geqslant 0$ tel que si $m, n \geqslant N_k$, on a

$$\|f_m - f_n\|_\infty \leqslant \frac{1}{k}. \tag{26.101}$$

En particulier, il existe un ensemble de mesure nulle E_k sur lequel

$$|f_m(x) - f_n(x)| \leqslant \frac{1}{k}. \tag{26.102}$$

D'après le lemme 13.27, la partie $E = \bigcup_{k\in\mathbb{N}} E_k$, est encore de mesure nulle. En résumé, nous avons un N_k tel que si $m, n \geqslant N_k$, alors

$$|f_n(x) - f_m(x)| \leqslant \frac{1}{k} \tag{26.103}$$

pour tout x hors de E. Donc pour chaque $x \in \Omega\backslash E$, la suite $n \mapsto f_n(x)$ est de Cauchy dans $\mathbb{R}$ et converge donc. Cela défini donc une fonction

$$\begin{aligned} f\colon \Omega\backslash E &\to \mathbb{R} \\ x &\mapsto \lim_{n\to\infty} f_n(x). \end{aligned} \tag{26.104}$$

Cela prouve le point (1) : la convergence ponctuelle.
En passant à la limite $n \to \infty$ dans l'équation 26.103 et tenant compte que cette majoration tient pour presque tout x dans Ω, nous trouvons

$$\|f - f_n\|_\infty \leqslant \frac{1}{k}. \tag{26.105}$$

Donc non seulement f est dans L^∞, mais en plus la suite (f_n) converge vers f au sens L^∞, c'est-à-dire uniformément. Cela prouve le point (3). En ce qui concerne le point (2), la suite f_n est entièrement (à partir d'un certain point) dominée par la fonction $1 + |f|$ qui est dans L^p.

Cas $p < \infty$. Toute suite convergente étant de Cauchy, nous considérons une suite de Cauchy (f_n) dans $L^p(\Omega)$ et ce sera suffisant pour travailler sur le premier point. Pour montrer qu'une suite de Cauchy converge, il est suffisant de montrer qu'une sous-suite converge. Soit $\varphi\colon \mathbb{N} \to \mathbb{N}$ une fonction strictement croissante telle que pour tout $n \geqslant 1$ nous ayons

$$\|f_{\varphi(n+1)} - f_{\varphi(n)}\|_p \leqslant \frac{1}{2^n}. \tag{26.106}$$

Pour créer la fonction φ, il est suffisant de prendre le N_k donné par la condition de Cauchy pour $\epsilon = 1/2^k$ et de considérer la fonction définie par récurrence par $\varphi(1) = N_1$ et $\varphi(n+1) > \max\{N_n, \varphi(n-1)\}$. Ensuite nous considérons la fonction

$$g_n(x) = \sum_{k=1}^{n} |f_{\varphi(k+1)}(x) - f_{\varphi(k)}(x)|. \tag{26.107}$$

Notons que pour écrire cela nous avons considéré des représentants f_k qui sont alors des fonctions à l'ancienne. Étant donné que g_n est une somme de fonctions dans L^p, c'est une fonction L^p, comme nous pouvons le constater en calculant sa norme :

$$\|g_n\|_p \leqslant \sum_{k=1}^{n} \|f_{\varphi(k+1)} - f_{\varphi(k)}\|_p \leqslant \sum_{k=1}^{n} \frac{1}{2^k} \leqslant \sum_{k=1}^{\infty} \frac{1}{2^k} = 1. \tag{26.108}$$

Étant donné que tous les termes de la somme définissant g_n sont positifs, la suite (g_n) est croissante. Mais elle est bornée en norme L^p et donc sujette à obéir au théorème de Beppo-Levi 13.160 sur la convergence monotone. Il existe donc une fonction $g \in L^p(\Omega)$ telle que $g_n \to g$ presque partout.

Soit un $x \in \Omega$ pour lequel $g_n(x) \to g(x)$; alors pour tout $n \geqslant 2$ et $\forall q \geqslant 0$,

$$|f_{\varphi(n+q)}(x) - f_{\varphi(n)}(x)| = \left| f_{\varphi(n+q)}(x) - \sum_{k=1}^{q-1} f_{\varphi(n+k)}(x) - \sum_{k=1}^{q-1} f_{\varphi(n+k)}(x) - f_{\varphi(n)}(x) \right| \tag{26.109a}$$

$$= \left| \sum_{k=1}^{q} f_{\varphi(n+k)} - \sum_{k=1}^{q} f_{\varphi(n+k-1)}(x) \right| \tag{26.109b}$$

$$\leqslant \sum_{k=1}^{q} \left| f_{\varphi(n+k)}(x) - f_{\varphi(n+k-1)}(x) \right| \tag{26.109c}$$

$$= g_{n+q+1}(x) - g_{n+1}(x) \tag{26.109d}$$

$$\leqslant g(x) - g_{n-1}(x). \tag{26.109e}$$

Nous prenons la limite $n \to \infty$; la dernière expression tend vers zéro et donc

$$|f_{\varphi(n+q)}(x) - f_{\varphi(n)}(x)| \to 0 \tag{26.110}$$

pour tout q. Donc pour presque tout $x \in \Omega$, la suite $n \mapsto f_{\varphi(n)}(x)$ est de Cauchy dans $\mathbb{R}$ et donc y converge vers un nombre que nous nommons $f(x)$. Cela définit une fonction

$$\begin{aligned} f\colon \Omega\backslash E &\to \mathbb{R} \\ x &\mapsto \lim_{n\to\infty} f_{\varphi(n)}(x) \end{aligned} \tag{26.111}$$

où E est de mesure nulle. Montrons que f est bien dans $L^p(\Omega)$; pour cela nous complétons la série d'inégalités (26.109) en

$$\left| f_{\varphi(n+q)}(x) - f_{\varphi(n)}(x) \right| \leqslant g(x) - g_{n-1}(x) \leqslant g(x). \tag{26.112}$$

En prenant la limite $q \to \infty$ nous avons l'inégalité

$$|f(x) - f_{\varphi(n)}(x)| \leqslant g(x) \tag{26.113}$$

pour presque tout $x \in \Omega$, c'est-à-dire pour tout $x \in \Omega\backslash E$. Cette inégalité implique deux choses valables pour presque tout x dans Ω :

$$f(x) \in B\big(g(x), f_{\varphi(n)}(x)\big) \tag{26.114a}$$

$$f_{\varphi(n)}(x) \leqslant |f(x)| + |g(x)|. \tag{26.114b}$$

La première inégalité assure que $|f|^p$ est intégrable sur $\Omega\backslash E$ parce que $|f|$ est majorée par $|g| + |f_{\varphi(n)}|$. Elle prouve par conséquent le point (1) parce que $n \mapsto f_{\varphi(n)}$ est une sous-suite convergente presque partout. La seconde montre le point (2).

Attention : à ce point nous avons prouvé que $n \mapsto f_{\varphi(n)}$ est une suite de fonctions qui converge *ponctuellement presque partout* vers une fonction f qui s'avère être dans L^p. Nous n'avons pas montré que cette suite convergeait au sens de L^p vers f. Ce que nous devons montrer est que

$$\|f - f_{\varphi(n)}\|_p \to 0. \tag{26.115}$$

L'inégalité (26.113) nous donne aussi, toujours pour presque tout $x \in \Omega$:

$$\left| f(x) - f_{\varphi(n)}(x) \right|^p \leqslant g(x)^p \tag{26.116}$$

ce qui signifie que la suite[8] $|f - f_{\varphi(n)}|^p$ est dominée par la fonction $|g|^p$ qui est intégrable sur $\Omega\backslash E$ et tout autant sur Ω parce que E est négligeable ; cela prouve au passage le point (2), et le théorème de la convergence dominée de Lebesgue (13.184) nous dit que

$$\lim_{n\to\infty} \int_\Omega \left| f(x) - f_{\varphi(n)}(x) \right|^p dx = \int_\Omega \lim_{n\to\infty} \left| f(x) - f_{\varphi(n)}(x) \right| dx = 0. \tag{26.117}$$

8. À ce point, [54] se contente de majorer $|f_{\varphi(n)}(x)|$ par $|f(x)| + |g(x)$, mais je ne comprends pas comment cette majoration nous permet d'utiliser la convergence dominée de Lebesgue pour montrer (26.115).

Cette dernière suite d'égalités se lit de la façon suivante :

$$\lim_{n\to\infty} \|f - f_{\varphi(n)}\|_p = \big\| \lim_{n\to\infty} |f - f_{\varphi(n)}| \big\|_p = 0. \tag{26.118}$$

Nous en déduisons que la suite $n \mapsto f_{\varphi(n)}$ est convergente vers f au sens de la norme $L^p(\Omega)$. Or la suite de départ (f_n) était de Cauchy (pour la norme L^p) ; donc l'existence d'une sous-suite convergente implique la convergence de la suite entière vers f, ce qu'il fallait démontrer.

□

Le théorème suivant est souvent cité en disant que L^p est un espace de Hilbert si et seulement si $p = 2$. Comme vous le voyez, il faut un peu plus d'hypothèses.

Je précise que je suis le seul à nommer ce théorème par le nom de Weinersmith. Je ne sais pas si il a déjà un nom ; alors pourquoi pas celui-là plutôt qu'un autre ? La raison de ce choix est dans la constante de Weiner, définition 43.1.

Théorème 26.42 (Théorème de Weinersmith[453, 454])**.**
Nous considérons un espace mesuré $(\Omega, \mathcal{A}, \mu)$ ainsi qu'un nombre $p \in [1, \infty]$. Nous supposons

(1) $L^p(\Omega, \mathcal{A}, \mu)$ est un espace de Hilbert,

(2) Il existe des parties $A, B \subset L^p(\Omega, \mathcal{A}, \mu)$ telles que $A \cap B = \varnothing$ et $0 < \mu(A) < \infty$ et $0 < \mu(B) < \infty$.

Alors $p = 2$.

Démonstration. Vu que L^p est un espace de Hilbert (hypothèse), il vérifie l'identité du parallélogramme de la proposition 7.199, c'est-à-dire

$$\|f + g\|_p^2 + \|f - g\|_p^2 = 2\|f\|_p^2 + 2\|g\|_p^2. \tag{26.119}$$

Pour $1 \leqslant p < \infty$ Soient donc A, B comme dans l'hypothèse. Nous considérons les fonctions

$$f = \frac{1}{\mu(A)^{1/p}} \mathbb{1}_A \tag{26.120a}$$

$$g = \frac{1}{\mu(B)^{1/p}} \mathbb{1}_B. \tag{26.120b}$$

En ce qui concerne les normes L^p de f et g, c'est un calcul simple :

$$\|f\|_p^2 = \left(\int_\Omega |f(\omega)|^p d\mu \right)^{2/p} = \left(\int_A \frac{1}{\mu(A)} d\mu \right)^{2/p} = 1. \tag{26.121}$$

De même pour g : $\|f\|_p^2 = \|g\|_p^2 = 1$. Donc

$$2\|f\|_p^2 + 2\|g\|_p^2 = 4 \tag{26.122}$$

En ce qui concerne la somme,

$$\|f + g\|_p^2 = \left(\int_A \frac{d\mu}{\mu(A)} + \int_B \frac{1}{\mu(B)} \right)^{2/p} = 2^{2/p}. \tag{26.123}$$

Pour la différence, la seule subtilité à voir est que

$$\int_\Omega |\mathbb{1}_A - \mathbb{1}_B|^p = \int_A |\mathbb{1}_A - \mathbb{1}_B|^p + \int_B |\mathbb{1}_A - \mathbb{1}_B| = \int_A |\mathbb{1}_A| + \int_B |-\mathbb{1}_B| = \int_A \mathbb{1}_A + \int_B \mathbb{1}_B. \tag{26.124}$$

Ce n'est pas de la magie que le moins se change en plus. Bref, pour la différence nous avons

$$\|f-g\|^2 = \left(\int_\Omega |f(\omega)-g(\omega)|^p d\mu\right)^{2/p} \tag{26.125a}$$
$$= \left(\int_\Omega |\frac{1}{\mu(A)^{1/p}}\mathbb{1}_A(\omega) - \frac{1}{\mu(B)^{1/p}}\mathbb{1}_B(\omega)|^p d\mu\right)^{2/p} \tag{26.125b}$$
$$= \left(\int_A |\frac{\mathbb{1}_A(\omega)}{\mu(A)^{1/p}}|^p + \int_B |-\frac{\mathbb{1}_B(\omega)}{\mu(B)^{1/p}}|^p\right)^{2/p} \tag{26.125c}$$
$$= 2^{2/p}. \tag{26.125d}$$

Donc $\|f+g\|_p^2 + \|f-g\|_p^2 = 2\times 2^{2/p}$.
Vu que L^p est un espace de Hilbert, nous avons finalement

$$4 = 2\times 2^{2/p}. \tag{26.126}$$

Cela est uniquement valable pour $p=2$.

Pour $p=\infty$ Il suffit de prendre $f=\mathbb{1}_A$ et $g=\mathbb{1}_B$. Nous avons alors $\|f\|_{L^\infty}^2 = \|g\|_{L^\infty}^2 = \|f+g\|_{L^\infty}^2 = \|f-g\|_{L^\infty}^2 = 1$.
L'égalité (26.119) devient $2=4$, ce qui est faux.

□

Si $(\Omega,\mathcal{A},\mu)$ est un espace mesuré, est-ce que $L^p(\Omega,\mathcal{A},\mu)$ est assuré de n'être pas de Hilbert ? Non.

Exemple 26.43 ([454]).
Soit la mesure de Dirac sur $\mathbb{R}$, c'est-à-dire

$$\delta(A) = \begin{cases} 1 & \text{si } 0\in A \\ 0 & \text{sinon.} \end{cases} \tag{26.127}$$

Nous allons prouver que pour tout $p\in[0,\infty]$, l'espace $L^p(\mathbb{R},\mathcal{B}\text{or}(\mathbb{R}),\delta)$ est un espace de Hilbert. Pour cela nous introduisons le produit hermitien

$$\langle f,g\rangle = f(0)\overline{g(0)}. \tag{26.128}$$

Nous avons par ailleurs la norme

$$\|f\|^p = \int_{\mathbb{R}} |f(x)|^p d\delta = f(0)^p. \tag{26.129}$$

Donc oui, $\|f\|_p = \sqrt{\langle f,f\rangle}$. △

26.2.8 Théorèmes d'approximation

Proposition 26.44 ([284], thème 3).
Soient $1\leqslant p\leqslant\infty$ et un espace mesuré $(\Omega,\mathcal{A},\mu)$. Alors les fonctions étagées[9] *dans $L^p(\Omega)$ sont denses dans $L^p(\Omega)$.*

Démonstration. Soit $f\in L^1(\Omega,\mathcal{A},\mu)$. Nous supposons dans un premier temps que $f\colon\Omega\to[0,\infty[$. Pour les fonctions à valeurs dans $\mathbb{C}$, nous verrons plus bas.

Notons que la partie $\{x\in\Omega \text{ tel que } f(x)=\infty\}$ est de mesure nulle, donc nous pouvons vraiment choisir un représentant à valeurs dans $[0,\infty[$ et non à valeurs dans $[0,\infty]$ comme le serait un représentant un peu quelconque.

9. Définition 13.100. Pour rappel, une fonction est simple lorsqu'elle prend un nombre fini de valeurs, et elle est étagée lorsqu'elle est en outre mesurable.

Par le théorème 13.105, il existe une suite croissante de fonctions étagées $\phi_n \colon \Omega \to [0,\infty[$ telles que $\phi_n \to f$ ponctuellement. Notons que ce théorème fonctionne parce que les fonctions L^p (en tout cas leurs représentants) sont mesurables parce que c'est dans la définition 26.10. Notre devoir est maintenant de prouver que sous l'hypothèse que f est dans L^p, alors la convergence $\phi_n \to f$ est une convergence dans L^p.

$1 \leqslant p < \infty$ Vu que f et ϕ_n sont à valeurs positives nous avons $|f-\phi_n|^p \leqslant |f|^p$. Mais par hypothèse $|f|^p \in L^1(\Omega)$. Donc la suite $g_n = |f-\phi_n|^p$ est majorée (uniformément en n) par $|f|^p$ qui est dans L^1. Le théorème de la convergence dominée permet de permuter

$$\lim_{n\to\infty} \int g_n = \int \lim_{n\to\infty} g_n = 0. \tag{26.130}$$

Cela revient à dire que

$$\lim_{n\to\infty} \|f-\phi_n\|_p^p = \lim_{n\to\infty} \int |f-\phi_n|^p = \int \lim_{n\to\infty} |f-\phi_n|^p = 0, \tag{26.131}$$

ce qui signifie que $\phi_n \xrightarrow{L^p(\Omega)} f$.

Pour $p=\infty$ Si $f \in L^\infty(\Omega)$, alors nous pouvons prendre un représentant borné. Avec lui, nous avons $\phi_n \xrightarrow{\|\cdot\|_\infty} f$. Avec cela nous avons, pour ϵ donné, un n assez grand pour avoir

$$N_\infty(\phi_n - f) \leqslant \|\phi_n - f\|_\infty < \epsilon. \tag{26.132}$$

Si f est à valeurs dans $\mathbb{C}$ au lieu de $[0,\infty[$, il suffit de faire valoir le travail que nous venons de faire quatre fois, pour les valeurs réelles, imaginaires, positives et négatives. □

26.2.9 Densité des fonctions infiniment dérivables à support compact

Définition 26.45.
*Une fonction est **étagée par rapport à** L^p si elle est de la forme*

$$f = \sum_{k=1}^N c_k \mathbb{1}_{B_k} \tag{26.133}$$

où les B_k sont des mesurables disjoints et $\mathbb{1}_{B_k} \in L^p$ pour tout k.

Lemme 26.46.
Si f est une fonction étagée en même temps qu'être dans L^p, alors elle est étagée par rapport à L^p.

Démonstration. Nous pouvons écrire

$$f = \sum_{k=1}^N c_k \mathbb{1}_{B_k} \tag{26.134}$$

où les B_k sont disjoints. Par hypothèse $\|f\|_p$ existe. Donc chacune des intégrales $\int_\Omega |\mathbb{1}_{B_k}|^p$ doit exister parce que les B_k étant disjoints, nous pouvons inverser la norme et la somme ainsi que la somme et l'intégrale :

$$\int_\Omega |f|^p = \int_\Omega \sum_{k=1}^N |c_k \mathbb{1}_{B_k}(x)|^p dx = \sum_{k=1}^N \int |c_k \mathbb{1}_{B_k}(x)|^p dx = \sum_{k=1}^N |c_k|^p \int_\Omega |\mathbb{1}_{B_k}(x)|^p dx. \tag{26.135}$$

□

Le contraire n'est pas vrai : la fonction étagée sur $\mathbb{R}$ qui vaut n sur $B(n,\frac{1}{4})$ est étagée par rapport à L^p, mais n'est pas dans L^p.

L'ensemble $C_c^\infty(\mathbb{R}^d)$ des fonctions de classe C^∞ et à support compact sur $\mathbb{R}^d$ est souvent également noté $\mathscr{D}(\mathbb{R}^d)$.

Théorème 26.47 ([449]).
Nous avons des densités emboitées. Ici D est un borélien borné de $\mathbb{R}^d$ contenu dans $B(0,r)$ et K est un compact contenant $B(0,r+2)$.

(1) Les fonctions étagées par rapport à L^p sur $\mathbb{R}^d$ sont denses dans $L^p(\mathbb{R}^d)$. A fortiori les fonctions étagées sont denses dans L^p, mais nous n'en aurons pas besoin ici.

(2) Il existe une suite f_n dans $C(K,\mathbb{C})$ telle que

$$f_n \xrightarrow{L^p} \mathbb{1}_D. \tag{26.136}$$

(3) Si A est un borélien tel que $\mathbb{1}_A \in L^p(\mathbb{R}^d)$ [10] et si $\epsilon > 0$, alors il existe une suite de boréliens bornée $(D_n)_{n\in\mathbb{N}}$ tels que

$$\mathbb{1}_{D_n} \xrightarrow{L^p} \mathbb{1}_A. \tag{26.137}$$

(4) Il existe une suite φ_n dans $\mathscr{D}(\mathbb{R}^d) = C_c^\infty(\mathbb{R}^d)$ telle que

$$\varphi_n \xrightarrow{L^p} \mathbb{1}_D. \tag{26.138}$$

(5) L'ensemble $\mathscr{D}(\mathbb{R}^d) = C_c^\infty(\mathbb{R}^d)$ est dense dans $L^p(\mathbb{R}^d)$ pour tout $1 \leqslant p < \infty$.

Démonstration. Nous allons montrer les choses point par point.

(1) Si $f \in L^1(\mathbb{R}^d)$, nous savons par le théorème 13.105 qu'il existe une suite f_n de fonctions étagées convergeant ponctuellement vers f telle que $|f_n| \leqslant |f|$. La proposition 26.21 nous dit qu'alors $f_n \xrightarrow{L^p} f$.

La fonction f_n étant étagée et dans L^p en même temps, elle est automatiquement étagée par rapport à L^p par le lemme 26.46.

(2) C'est le théorème d'approximation 13.206 appliqué au borélien D contenu dans l'espace mesuré K.

(3) En vertu du point (2), il existe $f \in C^0(K,\mathbb{R})$ telle que

$$\|f - \mathbb{1}_D\|_p \leqslant \epsilon. \tag{26.139}$$

Ensuite, par le théorème de Weierstrass, il existe $\varphi \in C^\infty(K,\mathbb{R})$ telle que $\|f-\varphi\|_\infty \leqslant \epsilon$. Nous avons aussi

$$\|\varphi - f\|_p^p = \int_K |\varphi(x) - f(x)|^p dx \leqslant \mu(X)\|\varphi - f\|_\infty^p \leqslant \epsilon^p \mu(K). \tag{26.140}$$

Quitte à prendre un φ correspondant à un ϵ plus petit, nous avons

$$\|\varphi - f\| \leqslant \epsilon. \tag{26.141}$$

En combinant et en passant à $\epsilon/2$ nous avons trouvé une fonction $\varphi \in C^\infty(K,\mathbb{R})$ telle que

$$\|\varphi - \mathbb{1}_D\| \leqslant \epsilon. \tag{26.142}$$

(4) Nous considérons les boréliens fermés $D_n = A \cap B(0,n)$. Alors $\mathbb{1}_{D_n} \in L^p$ et nous avons pour n assez grand :

$$\int_{\mathbb{R}^d} |\mathbb{1}_{D_n}(x) - \mathbb{1}_A(x)|^p dx = \int_{\mathbb{R}^d \setminus B(0,n)} |\mathbb{1}_A(x)|^p < \epsilon, \tag{26.143}$$

c'est-à-dire que $\mathbb{1}_{D_n} \xrightarrow{L^p} \mathbb{1}_A$.

10. Je pense que cette hypothèse manque dans [449]. En tout cas je vois mal comment je pourrais justifier les différentes étapes de la preuve en prenant par exemple $A = \mathbb{R}^d$.

(5) Il suffit de remettre tout ensemble. Si $f \in L^p(\mathbb{R}^d)$, par le point (2) nous commençons par prendre σ étagée par rapport à L^p telle que

$$\|\sigma - f\|_p \leqslant \epsilon. \tag{26.144}$$

Ensuite nous écrivons σ sous la forme

$$\sigma = \sum_{k=1}^{N} c_k \mathbb{1}_{B_k} \tag{26.145}$$

et nous appliquons le point (3) à chacune des $\mathbb{1}_{B_k}$ pour trouver des boréliens bornés D_k tels que

$$\|\mathbb{1}_{D_k} - \mathbb{1}_{B_k}\|_p \leqslant \epsilon. \tag{26.146}$$

Enfin nous appliquons le point (4) pour trouver des fonctions $\varphi_k \in C_c^\infty(\mathbb{R}^d)$ telles que

$$\|\varphi_k - \mathbb{1}_{D_k}\|_p \leqslant \epsilon. \tag{26.147}$$

Il n'est pas compliqué de calculer que

$$\Big\| \sum_{k=1}^{N} c_k \varphi_k - f \Big\|_p \leqslant 2\epsilon \sum_k c_k + \epsilon. \tag{26.148}$$

□

Corolaire 26.48.
Si $1 < p < \infty$ alors la partie[11] $L^2\big([0,1]\big) \cap L^p\big([0,1]\big)$ *est dense dans* $L^p\big([0,1]\big)$.

Démonstration. Nous savons du théorème 26.47(5) que $C_c^\infty\big([0,1]\big)$ est dense dans L^p. Mais nous avons évidemment $C_c^\infty \subset L^2 \cap L^p$, donc $L^2 \cap L^p$ est dense dans L^p. □

Lemme 26.49 ([449, 430])**.**
Soit $1 \leqslant p < \infty$ et $f \in L^p(\Omega)$. Nous notons τ_v l'opérateur de translation par v :

$$\begin{aligned} \tau_v \colon L^p(\Omega) &\to L^p(\Omega) \\ f &\mapsto \Big[x \mapsto f(x-v)\Big]. \end{aligned} \tag{26.149}$$

Pour chaque $f \in L^p(\Omega)$, l'application

$$\begin{aligned} \tau(f) \colon \mathbb{R}^d &\to L^p(\Omega) \\ v &\mapsto \tau_v(f) \end{aligned} \tag{26.150}$$

est continue en $v = 0$, c'est-à-dire

$$\lim_{v \to 0} \|\tau_v(f) - f\|_p = 0. \tag{26.151}$$

Démonstration. Nous commençons par supposer que f est dans $\mathscr{D}(\Omega)$, et nous verrons ensuite comment généraliser.

Si $f \in \mathscr{D}(\Omega)$ Soit une suite $v_i \xrightarrow{\mathbb{R}^d} 0$, et posons $f_i = \tau_{v_i}(f)$; le but est de montrer que $f_i \xrightarrow{L^p} f$. Pour cela, la fonction $f - f_i$ est également à support compact, et qui plus est, si $\operatorname{supp}(f) \subset B(0,r)$, alors $\operatorname{supp}(f - f_i) \subset B(0, r + |v_i|)$, et l'ensemble

$$S = \overline{B\big(0, r + \max_i |v_i|\big)} \tag{26.152}$$

11. Nous parlons bien ici de l'*ensemble* L^2 parce que nous le considérons sans norme ou topologie particulière. La densité dont nous parlons ici est celle pour la topologique de L^p.

est un compact contenant les supports de tous les $f - f_i$. Le maximum existe parce que $v_i \to 0$. Voilà qui « majore » le domaine de $f - f_i$ uniformément en i.

Majorons maintenant $|f - f_i|^p$ de façon uniforme en i. Soit le nombre

$$M = 2 \max_{x \in \mathbb{R}^d} \{f(x)\}. \tag{26.153}$$

La fonction qui vaut M^p sur S et zéro ailleurs est une fonction intégrable qui majore $|f - f_i|^p$. Nous pouvons donc utiliser la convergence dominée de Lebesgue (théorème 13.184) pour écrire

$$\lim_{i \to \infty} \|f - f_i\|_p^p = \lim_{i \to \infty} \int_\Omega |f(x) - f(x - v_i)|^p dx = \int_\Omega \lim_{i \to \infty} |f(x) - f(x - v_i)| dx = 0. \tag{26.154}$$

Pour $f \in L^p(\Omega)$ Soit $\epsilon > 0$, $f \in L^p(\Omega)$ et $\varphi \in \mathscr{D}(\Omega)$ tel que $\|f - \varphi\|_p \leqslant \epsilon$. Cela est possible par la densité de $\mathscr{D}(\Omega)$ dans $L^p(\Omega)$ vue en 26.47(5). Nous choisissons de plus $|v|$ assez petit pour avoir $\|\tau_v(\varphi) - \varphi\|_p < \epsilon$, qui est possible en vertu de ce que nous venons de démontrer à propos des fonctions à support compact. De plus τ_v étant une isométrie de L^p nous avons $\|\tau_v(\varphi) - \tau_v(f)\| = \|\varphi - f\| < \epsilon$. Nous avons tout pour majorer :

$$\|f - \tau_v(f)\| \leqslant \|f - \varphi\| + \|\varphi - \tau_v(\varphi)\| + \|\tau_v(\varphi) - \tau_v(f)\| \leqslant 3\epsilon. \tag{26.155}$$

Nous avons donc bien $\lim_{v \to 0} \|f - \tau_v(f)\| = 0$.

□

26.2.10 Approximation

Lemme 26.50 (Théorème fondamental d'approximation [277]).
Soit Ω un espace mesurable et $f \colon \Omega \to [0, \infty]$ une application mesurable. Alors il existe une suite croissante d'applications étagées $\varphi_n \colon \Omega \to \mathbb{R}^+$ dont la limite est f.

De plus si f est bornée, la convergence est uniforme.

Théorème 26.51 ([144]).
Soit I un intervalle de $\mathbb{R}$. L'espace $\mathscr{D}(I)$ des fonctions continues à support compact sur I est dense dans $L^2(I)$.

Ce théorème sera généralisé à tous les $L^p(\mathbb{R}^d)$ par le théorème 26.47. Cependant L^p n'étant pas un Hilbert, il faudra travailler sans produit scalaire.

Démonstration. Soit $g \in L^2(I)$ une fonction telle que $g \perp f$ pour toute fonction $f \in C_c(I)$. Nous avons donc

$$\langle f, g \rangle = \int_I f \bar{g} = 0. \tag{26.156}$$

En passant éventuellement aux composantes réelles et imaginaires nous pouvons supposer que les fonctions sont toutes réelles. Nous décomposons g en parties positives et négatives : $g = g^+ - g^-$. Notre but est de montrer que $g^+ = g^-$, c'est-à-dire que g est nulle. La proposition 24.14 conclura que $C_c(I)$ est dense dans $L^2(I)$.

Soit un intervalle $[a, b] \subset I$ et une suite croissante de fonctions $f_n \in C_c(I)$ qui converge vers $\mathbb{1}_{[a,b]}$. Par hypothèse pour chaque n nous avons

$$\int_I f_n g^+ = \int_I f_n g^-. \tag{26.157}$$

La suite étant croissante, le théorème de la convergence monotone (théorème 13.160) s'applique et nous avons

$$\lim_{n \to \infty} \int_I f_n g^+ = \int_a^b g^+, \tag{26.158}$$

de telle sorte que nous ayons, pour tout intervalle $[a,b] \subset I$ l'égalité

$$\int_a^b g^+ = \int_a^b g^-. \tag{26.159}$$

De plus ces intégrales sont finies parce que

$$\int_a^b g^+ \leqslant \int_a^b |g| = \int_I |g| \mathbb{1}_{[a,b]} = \langle |g|, \mathbb{1}_{[a,b]} \rangle \leqslant \|g\|_{L^2} \sqrt{b-a} < \infty \tag{26.160}$$

par l'inégalité de Cauchy-Schwarz.

Soit maintenant un ensemble mesurable $A \subset I$. La fonction caractéristique $\mathbb{1}_A$ est mesurable et il existe une suite croissante de fonctions étagées (φ_n) convergente vers a par le lemme 26.50. À multiples près, les fonctions φ_n sont des sommes de fonctions caractéristiques du type $\mathbb{1}_{[a,b]}$, par conséquent, en vertu de (26.159) nous avons

$$\int_I \varphi_n g^+ = \int_I \varphi_n g^-. \tag{26.161}$$

Une fois de plus nous pouvons utiliser le théorème de la convergence monotone et obtenir

$$\int_A g^+ = \int_A g^- \tag{26.162}$$

pour tout ensemble mesurable $A \subset I$. Si nous notons dx la mesure de Lebesgue, les mesures g^+dx et g^-dx sont par conséquent égales et dominées par dx. Par le corolaire 13.199 du théorème de Radon Nikodym, les fonctions g^+ et g^- sont égales. □

26.3 Convolution

Définition 26.52.
Pour toutes fonctions $f, g \colon \mathbb{R}^n \to \mathbb{C}$ et pour tout $x \in \mathbb{C}$ tels que l'intégrale de droite ait un sens[12]*, nous définissons*

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y) g(x-y) dy. \tag{26.163}$$

L'éventuelle fonction $f * g$ ainsi définie est le ***produit de convolution*** *de f et g.*

Le théorème qui permet de dire que le produit de convolution n'est pas tout à fait ridicule est le suivant.

Théorème 26.53 ([298])**.**
Soient $f, g \in L^1(\mathbb{R}^n)$.

(1) Pour presque tout $x \in \mathbb{R}^n$, la fonction

$$y \mapsto g(x-y) f(y) \tag{26.164}$$

est dans $L^1(\mathbb{R}^n)$.

(2) $f * g \in L^1(\mathbb{R}^n)$.

(3) $\|f * g\|_1 \leqslant \|f\|_1 \|g\|_1$.

Proposition 26.54.
L'ensemble $L^1(\mathbb{R}^n)$ devient alors une algèbre de Banach.

Lemme 26.55.
Le produit de convolution est commutatif : $f * g = g * f$.

12. Attention divlgâchi : ce sera le cas pour $f, g \in L^1(\mathbb{R}^n)$ par le théorème 26.53.

Démonstration. Le théorème de Fubini (théorème 13.262) permet d'écrire

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y)g(x-y)dy = \int_{-\infty}^{\infty} dy_1 \dots \int_{-\infty}^{\infty} dy_n f(y)g(x-y). \tag{26.165}$$

En effectuant le changement de variable $z_i = x_i - y_i$ dans chacune des intégrales nous obtenons

$$(f * g)(x) = \int_{\mathbb{R}^n} g(z)f(x-z)dz = (g * f)(x). \tag{26.166}$$

Attention : on pourrait croire qu'un signe apparaît du fait que $z = x - y$ donne $dz = -dy$. Mais en réalité, l'intégrale $\int_{-\infty}^{+\infty}$ devient par le même changement de variables $\int_{+\infty}^{-\infty}$ qui redonne un nouveau signe au moment de remettre dans l'ordre. □

La proposition suivante est une conséquence de l'inégalité de Minkowski sous forme intégrale de la proposition 26.37(3).

Proposition 26.56.
Si $1 \leqslant p \leqslant \infty$ et si $f \in L^p(\mathbb{R}^d)$ et $g \in L^1(\mathbb{R}^d)$ alors

(1) $f * g \in L^p$

(2) $\|f * g\|_p \leqslant \|f\|_p \|g\|_1$.

Proposition 26.57 ([455])**.**
Si $f \in L^1(\mathbb{R})$ et si g est dérivable avec $g' \in L^\infty$, alors $f * g$ est dérivable et $(f * g)' = f * g'$.

Démonstration. La fonction qu'il faut intégrer pour obtenir $f * g$ est $f(t)g(x-t)$, dont la dérivée par rapport à x est $f(t)g'(x-t)$. La norme de cette dernière est majorée (uniformément en x) par $G(t) = |f(t)|\|g'\|_\infty$. La fonction f étant dans $L^1(\mathbb{R})$, la fonction G est intégrable et le théorème de dérivation sous l'intégrale (théorème 16.18) nous dit que $f * g$ est dérivable et

$$(f * g)'(x) = \frac{d}{dx} \int_{\mathbb{R}} f(t)g(x-t)dt = \int_{\mathbb{R}} f(t)g'(x-t)dt = (f * g')(x). \tag{26.167}$$

□

Corolaire 26.58.
Si $f \in L^1(\mathbb{R}^d)$ et si g est de classe C^∞, alors $f * g$ est de classe C^∞.

Démonstration. Il s'agit d'itérer la proposition 26.57. □

Lemme 26.59.
Soit $f \in L^2(I)$ telle que

$$\int_I f\varphi = 0 \tag{26.168}$$

pour toute fonction $\varphi \in C_c^\infty(I)$. Alors $f = 0$ presque partout sur I.

Démonstration. Nous considérons la forme linéaire

$$\begin{aligned} \phi\colon L^2(I) &\to \mathbb{C} \\ g &\mapsto \langle f, g\rangle = \int_I f\bar{g}. \end{aligned} \tag{26.169}$$

Par densité [13] nous pouvons aussi considérer une suite (φ_n) dans $C_c^\infty(I)$ convergeant dans L^2 vers f. Alors nous avons pour tout n :

$$\langle f, \varphi_n\rangle = 0. \tag{26.170}$$

En passant à la limite, $\langle f, f\rangle = 0$, ce qui implique $f = 0$ dans L^2 et donc $f = 0$ presque partout en tant que bonne fonction. □

Ce résultat est encore valable dans les espaces L^p (proposition 26.132), mais il demande le théorème de représentation de Riesz [14].

13. Théorème 26.47(5).
14. Théorème 26.130.

26.3.1 Approximation de l'unité

Définition 26.60 ([1, 449]).
Nous considérons $\Omega = \mathbb{R}^d$ ou $(S^1)^d$. Une ***approximation de l'unité*** *sur Ω autour de $a \in \Omega$ est une suite (φ_n) de fonctions à valeurs réelles dans $L^1(\Omega)$ telle que*

(1) $\sup_k \|\varphi_k\|_1 < \infty$,

(2) pour chaque n nous avons $\int_\Omega \varphi_n = 1$,

(3) si V est un voisinage de a, alors

$$\lim_{k\to\infty} \int_{\Omega\setminus V} |\varphi_k| = 0. \tag{26.171}$$

En pratique, nous allons, sur $\mathbb{R}^d$ toujours considérer des approximations de l'unité autour de 0, même si nous ne le préciserons pas. Vous noterez que dans le cas de S^1, le choix du « point de base » est plus arbitraire.

Ce sont des fonctions dont la masse vient s'accumuler autour de zéro. En effet quel que soit le voisinage $B(0,\alpha)$, si k est assez grand, il n'y a presque plus rien en dehors.

Pour le point ((3)), si Ω est S^1, la mesure que nous considérons est $\frac{dx}{2\pi}$.

Exemple 26.61.
Une façon de construire une approximation de l'unité sur $\mathbb{R}$ est de considérer une fonction $\varphi \in L^1(\Omega)$ telle que $\int \varphi = 1$ puis de poser

$$\varphi_k(x) = k^d \varphi(kx). \tag{26.172}$$

Ici, Ω peut être $\mathbb{R}$ ou S^1. △

Le lemme suivant permet de construire des approximations de l'unité intéressantes. Nous aurons une version pour S^1 dans le lemme 26.94.

Lemme 26.62 ([449]).
Soit φ est une fonction continue et positive à support compact sur $\mathbb{R}^d$ telle que $\varphi(x) > \varphi(0)$ pour tout $x \neq 0$. Si nous posons

$$\varphi_n(x) = \left(\int \varphi(y)^n\right)^{-1} \varphi(x)^n, \tag{26.173}$$

alors la suite (φ_n) est une approximation de l'unité.

Voici un théorème qui donne les propriétés à propos du produit de convolution avec une approximation de l'unité dans $\mathbb{R}^d$. Une version pour S^1 sera le théorème 26.95.

Théorème 26.63 ([449]).
Soit (φ_k) une approximation de l'unité sur $\mathbb{R}^d$.

(1) Si g est mesurable et bornée sur $\mathbb{R}^d$ et si g est continue en x_0 alors

$$(\varphi_k * g)(x_0) \to g(x_0). \tag{26.174}$$

(2) Si $g \in L^p(\mathbb{R}^d)$ $(1 \leqslant p < \infty)$ alors

$$\varphi_k * g \xrightarrow{L^p} g. \tag{26.175}$$

(3) Si g est uniformément continue et bornée, alors

$$\varphi_k * g \xrightarrow{L^\infty} g \tag{26.176}$$

Démonstration. En plusieurs points.

(1) Nous notons $d_k = (\varphi_k * g)(x_0) - g(x_0)$ et nous devons prouver que $d_k \to 0$. Vu que φ_k est d'intégrale 1 sur $\mathbb{R}^d$ nous pouvons écrire

$$d_k = \int_{\mathbb{R}^d} \varphi_k(y) g(x_0 - y) dy - \int_{\mathbb{R}^d} g(x_0)\varphi_k(y) dy, \tag{26.177}$$

et donc

$$|d_k| = \Big| \int_{\mathbb{R}^d} \big(g(x_0 - y) - g(x_0)\big)\varphi_k(y) dy \Big| \leqslant \int_{\mathbb{R}^d} \big|g(x_0 - y) - g(x_0)\big| |\varphi_k(y)| dy. \tag{26.178}$$

Nous notons $M = \sup_k \|\varphi_k\|_1$, et nous considérons $\alpha > 0$ tel que

$$\big|g(x_0 - y) - g(x_0)\big| \leqslant \epsilon \tag{26.179}$$

pour tout $y \in B(0, \alpha)$. Nous nous restreignons maintenant aux k suffisamment grands pour que $\int_{\complement B(0,\alpha)} |\varphi_k(y)| dy \leqslant \epsilon$. Alors en découpant l'intégrale en $B(0, \alpha)$ et son complémentaire dans $\mathbb{R}^d$,

$$|d_k| \leqslant \epsilon M + \int_{\complement B(0,\alpha)} 2\|g\|_\infty |\varphi_k(y)| dy \leqslant \epsilon M + 2\|g\|_\infty \epsilon \leqslant \epsilon C. \tag{26.180}$$

Donc oui, nous avons $|d_k| \to 0$, et donc le premier point du théorème.

(2) Cette fois $g \in L^p(\mathbb{R}^d)$ et nous cherchons à montrer que $\|d_k\|_p \to 0$. Encore qu'ici d_k soit défini à partir d'un représentant dans la classe de g et que d'ailleurs, nous allons travailler avec ce représentant.

D'abord nous développons un peu ce d_k :

$$\|d_k\|_p = \Big[\int_{\mathbb{R}^d} \Big| \int_{\mathbb{R}^d} \big(g(x - y) - g(x)\big)\varphi_k(y) dy \Big|^p dx \Big]^{1/p} \tag{26.181a}$$

$$\leqslant \Big[\int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} |g(x - y) - g(x)| \cdot |\varphi_k(y)| dy \Big)^p dx \Big]^{1/p}. \tag{26.181b}$$

À cette dernière expression nous appliquons l'inégalité de Minkowski (théorème 26.37) sous la forme (26.71) pour la mesure $d\nu(y) = |\varphi_k(y)| dy$ et $f(x, y) = g(x - y) - g(x)$:

$$\|d_k\|_p \leqslant \int_{\mathbb{R}^d} \Big(\int_{\mathbb{R}^d} \big|g(x - y) - g(x)\big|^p dx \Big)^{1/p} |\varphi_k(y)| dy = \int_{\mathbb{R}^d} \|\tau_y g - g\|_p |\varphi_k(y)| dy. \tag{26.182}$$

Par le lemme 26.49 nous pouvons trouver $\alpha > 0$ tel que $\|\tau_y g - g\|_p \leqslant \epsilon$ pour tout $y \in B(0, \alpha)$. Avec cela nous découpons encore le domaine d'intégration :

$$\|d_k\|_p \leqslant \int_{B(0,\alpha)} \underbrace{\|\tau_y g - g\|_p}_{\leqslant \epsilon} |\varphi_k(y)| dy + \int_{\complement B(0,\alpha)} \underbrace{\|\tau_y g - g\|_p}_{\leqslant 2\|g\|_p} |\varphi_k(y)| dy \leqslant \epsilon M + 2\epsilon \|g\|_p. \tag{26.183}$$

(3) Nous posons $d_k(x) = (\varphi_k * g)(x) - g(x)$ et nous voulons prouver que $\|d_k\|_\infty \to 0$, c'est-à-dire que $d_k(x)$ converge vers zéro uniformément en x. Nous posons aussi

$$\tau_y(g) \colon x \mapsto g(x - y). \tag{26.184}$$

En récrivant le produit de convolution, une petite majoration donne

$$|d_k(x)| \leqslant \int_{\mathbb{R}^d} \|\tau_y(g) - g\|_\infty |\varphi_k(y)| dy. \tag{26.185}$$

L'uniforme continuité de g signifie que pour tout ϵ, il existe un α tel que pour tout $y \in B(0, \alpha)$,

$$\|\tau_y(g) - g\|_\infty \leqslant \epsilon. \tag{26.186}$$

Encore une fois nous découpons le domaine d'intégration en $B = B(0, \alpha)$ et son complémentaire :

$$\|d_k\|_\infty \leqslant \int_B \underbrace{\|\tau_y(g) - g\|_\infty}_{\leqslant \epsilon} |\varphi_k(y)| dy + \int_{\complement B} \underbrace{\|\tau_y(g) - g\|_\infty}_{\leqslant 2\|g\|_\infty} |\varphi_k(y)| \tag{26.187a}$$

$$\leqslant \epsilon M + 2\|g\|_\infty \epsilon \tag{26.187b}$$

où la seconde ligne est justifiée par le choix d'un k assez grand pour que $\int_{\complement B} |\varphi_k(y)| dy \leqslant \epsilon$. Nous avons donc bien $\|d_k\|_\infty \to 0$.

□

Exemple 26.64.
Une petite remarque en passant : aussi triste que cela en ait l'air, la convergence uniforme n'implique pas la convergence $L^p(\Omega)$ si Ω n'est pas borné. En effet si $f \in L^p$, la suite donnée par

$$f_n(x) = f(x) + \frac{1}{n} \tag{26.188}$$

converge uniformément vers f, mais

$$\|f_n - f\|_p = \int_\Omega \frac{1}{n} \tag{26.189}$$

n'existe même pas si le domaine Ω n'est pas borné. △

26.3.2 Densité des polynômes trigonométriques

Définition 26.65.
*Le **système trigonométrique** donné par $\{e_n\}_{n\in\mathbb{Z}}$ est*

$$e_n(t) = \frac{1}{\sqrt{2\pi}} e^{int}. \tag{26.190}$$

Une bonne partie de la douleur qu'évoque mot « densité » consiste à montrer que ce système est total dans $L^2(S^1) = L^2([0, 2\pi])$, et donc en est une base hilbertienne.

Définition 26.66.
*Un **polynôme trigonométrique** est une fonction de la forme*

$$P(t) = \sum_{n=-N}^{N} c_n e_n(t). \tag{26.191}$$

Définition 26.67 (Coefficients de Fourier)**.**
Pour toute fonction pour laquelle ça a un sens (que ce soit des fonctions L^2 ou non), nous posons

$$c_n(f) = \langle f, e_n \rangle. \tag{26.192}$$

*Ces nombres sont les **coefficients de Fourier** de f.*

Ces trois définitions n'ont à priori aucun rapport entre elles, et rien en particulier ne devrait vous faire penser à une égalité du type

$$f(x) = \sum_{n=-\infty}^{\infty} c_n(f) e_n(x). \tag{26.193}$$

Nous avons toutefois quelque liens.

Lemme 26.68.
Deux petits résultats simples mais utiles à propos des polynômes trigonométriques.

(1) Si $f \in L^1(S^1)$, alors nous avons la formule

$$f * e_n = c_n(f)e_n. \tag{26.194}$$

(2) Si P est un polynôme trigonométrique et si $f \in L^1(S^1)$ alors $f * P$ est encore un polynôme trigonométrique.

Démonstration. Le premier point est un simple calcul :

$$(f * e_n)(x) = \int_0^{2\pi} f(x-t)e_n(t) \tag{26.195a}$$

En ce qui concerne le second point, nous notons $P = \sum_{k=-N}^{N} P_k e_k$, et par linéarité de la convolution,

$$f * P = \sum_{k=-N}^{N} P_k f * e_k = \sum_{k=-N}^{n} P_k c_k(f) e_k, \tag{26.196}$$

qui est encore un polynôme trigonométrique. □

Exemple 26.69.
Sur S^1 nous construisons alors l'approximation de l'unité basée sur la fonction $1 + \cos(x)$ et le lemme 26.62. Cette fonction est évidemment un polynôme trigonométrique parce que

$$\cos(x) = \frac{e^{ix} + e^{-ix}}{2}. \tag{26.197}$$

Ensuite les puissances le sont aussi à cause de la formule du binôme :

$$\big(1 + \cos(x)\big)^n = \sum_{k=0}^{n} \binom{n}{k} \cos^n(x), \tag{26.198}$$

dans laquelle nous pouvons remettre $\cos(x)$ comme un polynôme trigonométrique et développer à nouveau la puissance avec (encore) la formule du binôme. La chose importante est qu'il existe une approximation de l'unité (φ_n) formée de polynômes trigonométrique.

Ce qui fait la spécificité des polynômes trigonométriques est qu'ils sont à la fois stables par convolution (lemme 26.68) et qu'ils permettent de créer une approximation de l'unité sur $[0, 2\pi]$. Ce sont ces deux choses qui permettent de prouver l'important théorème suivant. △

Théorème 26.70.
Les polynôme trigonométriques sont dense dans $L^p(S^1)$ pour $1 \leqslant p < \infty$.

Démonstration.

$$\varphi_k * f \xrightarrow{L^p} f \tag{26.199}$$

par le théorème 26.63. Nous avons donc convergence L^p d'une suite de polynômes trigonométrique, ce qui prouve que l'espace de polynômes trigonométriques est dense dans $L^p(S^1)$. □

Remarque 26.71.
Deux remarques.

— Il n'est pas possible que les polynômes trigonométriques soient dense dans L^∞ parce qu'une limite uniforme de fonctions continues est continue (c'est le théorème 11.294). Donc les polynômes trigonométriques ne peuvent engendrer que des fonctions continues.

— Nous donnerons au théorème 27.6 une démonstration indépendante de la densité des polynômes trigonométriques dans $L^p(S^1)$.

26.4 Espaces L^2, généralités

L'espace L^2 est l'espace L^p définit en 26.8 avec $p = 2$. Cependant il possède une propriété extraordinaire par rapport aux autres L^p, c'est que la norme $|.|_2$ dérive d'un produit scalaire. Il sera donc un espace de Hilbert.

26.72.
Nous en rappelons la construction. Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré. Nous considérons l'opération

$$\langle f, g\rangle = \int_\Omega f(\omega)\overline{g(\omega)}d\mu(\omega) \tag{26.200}$$

et la norme associée

$$\|f\|_2 = \sqrt{\langle f, f\rangle}. \tag{26.201}$$

Nous considérons l'ensemble

$$\mathcal{L}^2(\Omega, \mu) = \{f\colon \Omega \to \mathbb{C} \text{ tel que } \|f\|_2 < \infty\} \tag{26.202}$$

et la relation d'équivalence $f \sim g$ si et seulement si $f(x) = g(x)$ pour μ-presque tout x.

Et enfin, nous considérons le quotient

$$L^2(\Omega, \mu) = \mathcal{L}^2(\Omega, \mu)/\sim . \tag{26.203}$$

Lemme 26.73.
Soit un espace mesuré[15] $(\Omega, \mathcal{A}, \mu)$.

(1) Pour tout $f, g \in L^2(\Omega, \mathcal{A}, \mu)$, le produit

$$\langle f, g\rangle = \int_\Omega f\bar{g}\, d\mu \tag{26.204}$$

est bien défini et est un nombre complexe[16].

(2) L'opération $(f, g) \mapsto \langle f, g\rangle$ est un produit hermitien[17].

(3) Le couple $\big(L^2(\Omega, \mathcal{A}, \mu), \langle ., .\rangle\big)$ est un espace de Hilbert[18].

Démonstration. Que $L^2(\Omega)$ soit un espace vectoriel est un cas particulier de la proposition 26.17. Voyons cette histoire de produit scalaire.

Pour de vraies fonctions Nous commençons par analyser l'intégrale (26.204) dans le cas où f et g sont des fonctions, c'est-à-dire des représentants d'éléments de L^2.

Dans ce cas, l'inégalité de Hölder (proposition 26.33) avec $p = q = 2$ nous indique que le produit $f\bar{g}$ est un élément de L^1. Par conséquent la formule a un sens.

Passage aux classes Ensuite nous montrons que la formule passe au quotient. Pour cela, nous considérons des fonctions α et β nulles presque partout et nous regardons le produit de $f_1 = f + \alpha$ par $g_1 = g + \beta$:

$$\langle f_1, g_1\rangle = \int fg + \beta f + \alpha g + \alpha\beta. \tag{26.205}$$

Les fonctions βf, αg et $\alpha\beta$ étant nulles presque partout, leur intégrale est nulle et nous avons bien $\langle f_1, g_1\rangle = \langle f, g\rangle$. Nous pouvons donc considérer le produit sur l'ensemble des classes.

15. Est-ce qu'il ne faudrait pas un peu plus d'hypothèses, comme σ-fini par exemple ? Vérifiez et écrivez-moi quand vous avez la réponse.
16. Par opposition au fait que ce serait l'infini.
17. Définition 7.189. Pour rappel, nous considérons des fonctions à valeurs complexes. Si au contraire nous avions considéré seulement des fonctions à valeurs réelles, nous aurions eu un produit scalaire.
18. Définition 24.1.

Produit hermitien Pour vérifier que la formule est un produit hermitien, le seul point non évidement est de prouver que $\langle f, f\rangle = 0$ implique $f = 0$. Cela découle du fait que

$$\langle f, f\rangle = \int_\Omega |f|^2. \tag{26.206}$$

La fonction $x \mapsto |f(x)|^2$ vérifie les hypothèses du lemme 13.179. Par conséquent $|f(x)|^2$ est presque partout nulle.

Espace de Hilbert En ce qui concerne le fait que $L^2(\Omega)$ soit un espace de Hilbert, il s'agit simplement de se remémorer que c'est un espace complet (théorème 26.40) et dont la norme dérive d'un produit scalaire ou hermitien. Nous sommes donc bien dans la définition 24.1.

□

26.74.
Ces espaces seront utilisés pour de nombreuses applications. Nous en aurons besoin pour plusieurs combinaisons d'ensembles Ω et de mesures μ.

— Pour $\mathbb{R}^d$
— Pour S^1
— Pour $[a, b]$
— Pour $[0, 2\pi[$
— Pour $[-T, T[$

Le premier est non compact et il est raisonnable de penser qu'il sera foncièrement différents des autres. À isomorphismes assez triviaux près, les espaces des fonctions sur les trois autres sont identiques. Nous nous attendons donc à ce qu'ils aient les mêmes propriétés. Notons que du point de vue de L^2, étant donné qu'il y a un quotient par les parties de mesures nulles, prendre $]0, 2\pi[$ ou $[0, 2\pi]$ ou n'importe quelle autre possibilité de ce genre revient au même.

Afin de pouvoir utiliser ces espaces de façon optimale, et entre autres y définir les séries de Fourier, nous avons besoin, pour chacun d'entre eux de définir les éléments suivants :

— mesure
— produit de convolution
— le système trigonométrique (que nous allons montrer être une base hilbertienne)
— coefficients de Fourier

Ça fait pas mal de choses à définir. Il n'est pas besoin de définir un produit scalaire parce que le lemme 26.73 nous en donne un générique.

Les définitions qui viennent sont à prendre « tant que les formules ont un sens ». Nous parlons donc de fonctions dans $\mathrm{Fun}(\Omega, \mathbb{C})$, l'ensemble de toutes les fonctions sur Ω à valeurs dans $\mathbb{C}$. Nous verrons plus tard les espaces de fonctions sur lesquels tout a un sens.

26.5 L'espace $L^2(\mathbb{R}^d)$

La mesure est celle de Lebesgue. Le produit de convolution est donné, pour $f, g \in \mathrm{Fun}(\mathbb{R}^d, \mathbb{C})$, par

$$(f * g)(x) = \int_{\mathbb{R}^d} f(y)g(x - y)dy \tag{26.207}$$

Certaines de ses propriétés ont déjà été vues dans le théorème 26.53.

En ce qui concerne le système trigonométrique, pour tout $\xi \in \mathbb{R}^d$ nous définirions bien

$$e_\xi(x) = e^{i\xi \cdot x}, \tag{26.208}$$

genre pour faire que les transformations de Fourier sont des séries continues ... mais bon. Nous n'allons pas tenter le diable plus que ça, et nous ne définissons

— pas de système trigonométrique,
— pas de coefficients de Fourier non plus,
— pas de théorie des séries de Fourier sur $\mathbb{R}^d$.

Quand je disais que la non-compacité de $\mathbb{R}^d$ allait un peu changer les choses par rapport aux autres, je ne rigolais pas.

26.6 L'espace $L^2(S^1)$

L'espace S^1 sera fait avec forces détails, parce qu'il va servir de base pour les espaces $L^2([0,2\pi[)$, $L^2([-T,T[)$ ainsi que pour l'étude des fonctions périodiques sur $\mathbb{R}$.

En tant qu'ensemble,

$$S^1 = \{e^{it}\}_{t\in\mathbb{R}}, \tag{26.209}$$

sans garanties que ce paramétrage soit une bijection.

Il y a essentiellement deux façons de définir une intégrale sur S^1.

(1) Voir S^1 comme une sous-variété de $\mathbb{R}^2$ et utiliser la définition 19.9. Cette façon a cependant deux inconvénients :
   — Elle ne donne pas la tribu des mesurables sur S^1, c'est-à-dire que cette méthode ne donne pas de façon évidente une théorie de la mesure sur S^1.
   — Il faut au moins deux cartes pour paramétrer le cercle. La fainéantise nous prévient que ça va être technique.

(2) Rapporter la structure d'espace mesuré de $[0,2\pi[$ vers S^1, de force via le premier difféomorphisme qui nous passe par la tête, à savoir $t\mapsto e^{it}$.

Nous allons choisir la seconde possibilité, en gardant en tête qu'elle fonctionne de façon très simple un peu par coup de chance, voir la remarque 13.258(5).

26.6.1 Espace mesuré

Plusieurs choses sont déjà faites.

— Les boréliens de S^1 sont décrits dans la proposition 17.52,
— la tribu de Lebesgue de S^1 est décrite dans la proposition 17.54. Non, ce n'est pas la tribu induite de la tribu de Lebesgue de $\mathbb{C}$.

Proposition 26.75 (Espaces de fonctions sur S^1[1])**.**
Soit l'espace mesuré $\big(S^1,\mathcal{L}eb(S^1),\mu\big)$.

(1) La formule

$$\langle f,g\rangle = \int_{S^1} f\bar{g}d\mu \tag{26.210}$$

est un produit hermitien[19] *sur $L^2(S^1,\mathcal{L}eb(S^1),\mu)$.*

(2) L'espace $L^2(S^1)$ est un espace de Hilbert.

(3) L'application

$$\begin{aligned}\phi\colon L^2(S^1)&\to L^2\big([0,2\pi[\big)\\ f&\mapsto \frac{1}{\sqrt{2\pi}}f\circ\varphi.\end{aligned} \tag{26.211}$$

est une bijection isométrique (isomorphisme d'espaces de Hilbert)

$$L^2(S^1,\mathcal{L}eb(S^1),\mu) = L^2\big(]0,2\pi[,\mathcal{L}eb(\mathbb{R}),\lambda\big) \tag{26.212}$$

où nous avons fait un minuscule abus de notations : ici $\mathcal{L}eb(\mathbb{R})$ est en réalité la tribu induite sur $]0,2\pi[$.

19. Définition 7.189.

Démonstration. Le fait que la formule (26.210) donne bien un produit hermitien est le lemme 26.73. Ce même lemme assure que le tout donne un espace de Hilbert.

Il nous reste à prouver le point (3). En ce qui concerne l'isométrie, nous posons [20]

$$\begin{aligned}\phi\colon L^2(S^1) &\to L^2\big([0,2\pi[\big)\\ f &\mapsto \frac{1}{\sqrt{2\pi}} f\circ\varphi.\end{aligned} \tag{26.213}$$

Injection Si $\phi(f)=\phi(g)$, alors pour tout $x\in[0,2\pi[$ nous avons $f\big(\varphi(x)\big)=g\big(\varphi(x)\big)$. Vu que $\varphi\colon[0,2\pi[\to S^1$ est une bijection nous avons alors $f(s)=g(s)$ pour tout $s\in S^1$.

Surjection Si $f\in L^2\big(]0,2\pi]\big)$, nous posons $g\colon S^1\to\mathbb{C}$ par

$$g(s)=\sqrt{2\pi}f\big(\varphi^{-1}(s)\big). \tag{26.214}$$

Nous avons alors bien $\phi(g)(x)=f(x)$.

Isométrie Nous montrons que ϕ préserve le produit scalaire :

$$\begin{aligned}\langle\phi(f),\phi(g)\rangle &= \int_0^{2\pi}\phi(f)(x)\overline{\phi(g)(x)}d\lambda(x) &(26.215a)\\ &= \frac{1}{2\pi}\int_0^{2\pi}(f\circ\varphi)(x)\overline{(g\circ\varphi)(x)}\,d\lambda(x) &(26.215b)\\ &= \frac{1}{2\pi}\int_0^{2\pi}(fg)\circ\varphi\,d\lambda &(26.215c)\end{aligned}$$

Pour la suite nous devons invoquer la proposition 13.257 pour passer d'une intégrale sur $\big([0,2\pi[,\mathcal{L}\mathrm{eb}\big([0,2\pi[\big),\lambda\big)$ à une intégrale sur $\big(S^1,\mathcal{L}\mathrm{eb}(S^1),\mu\big)$. La première condition de cette proposition est que $\mathcal{L}\mathrm{eb}(S^1)=\varphi\big(\mathcal{L}\mathrm{eb}([0,2\pi[)\big)$. Cela est la proposition 17.54(2). La condition sur la mesure dans la proposition 13.257 n'est vraie ici qu'à un facteur 2π près. Nous avons :

$$\int_{[0,2\pi[} f d\lambda = 2\pi\int_{S^1}(f\circ\varphi^{-1})d\mu. \tag{26.216}$$

Nous continuons le calcul (26.215) :

$$\langle\phi(f),\phi(g)\rangle=\frac{1}{2\pi}\int_0^{2\pi}(fg)\circ\varphi\,d\lambda=\int_{S^1}f\bar{g}d\mu=\langle f,g\rangle. \tag{26.217}$$

□

26.6.2 Topologie

Nous considérons sur S^1 la topologie induite de $\mathbb{C}$. Vu que S^1 est fermé et borné dans $\mathbb{C}$, il en est une partie compacte. Par le lemme 7.51, l'espace S^1 muni de sa topologie est un espace topologique compact.

Nous pouvons donc sans crainte affirmer que toute fonction continue $f\colon S^1\to\mathbb{K}$ est bornée et atteint ses bornes.

Proposition-définition 26.76.
Soit la formule

$$d(e^{ix},e^{iy})=\inf_{k\in\mathbb{Z}}|x-y+2k\pi|. \tag{26.218}$$

(1) Elle est bien définie (ne dépend pas des choix de x et y donnant les mêmes points dans S^1)

(2) L'infimum est en réalité un minimum : il est atteint par un certain $k\in\mathbb{Z}$ (qui, lui, dépend des choix).

20. Notez que cette définition passe aux classes. Nous le répéterons pas.

(3) La formule définit une distance[21] *sur* S^1.

Nous considérons sur S^1 *la topologie* τ_d *découlant de cette distance.*

Démonstration. Point par point.

(1) Soient $x', y' \in \mathbb{R}$ tels que $e^{ix'} = e^{ix}$ et $e^{iy'} = e^{iy}$. Alors $x' = x + 2l\pi$ et $y' = y + 2l'\pi$ pour certains entiers $l, l' \in \mathbb{Z}$ (corolaire 17.18). Nous avons alors $|x'-y'+2k\pi| = |x-y+2\pi(k+l-l')|$ et

$$\inf_{k\in\mathbb{Z}} |x' - y' + 2k\pi| = \inf_{k\in\mathbb{Z}} |x - y + 2k\pi|. \tag{26.219}$$

(2) Quels que soient x et y fixés, nous avons

$$\lim_{k\to\pm\infty} |x - y + 2k\pi| = \infty. \tag{26.220}$$

Donc l'infimum est forcément atteint par un $k \in \mathbb{Z}$.

(3) Pour la distance, il y a plusieurs points à prouver.

— Pour tout $z, z' \in S^1$ nous avons $d(z, z') \geqslant 0$ parce que la distance est donnée par une valeur absolue.

— Si $d(z, z') = 0$, alors il existe k tel que $x = y+2k\pi$. Alors $e^{ix} = e^{i(y+2k\pi)} = e^{iy}e^{2ki\pi} = e^{iy}$. C'est-à-dire $z = z'$.

— Pour la symétrie, nous avons

$$|x - y + 2k\pi| = |y - x - 2k\pi| = |y - x + 2k'\pi| \tag{26.221}$$

en posant $k' = -k$. L'infimum étant pris sur $k \in \mathbb{Z}$, nous avons al symétrique $d(e^{ix}, e^{iy}) = d(e^{iy}, e^{ix})$.

— Pour attaquer l'inégalité triangulaire, nous considérons $z_1 = e^{ix_1}$, $z_2 = e^{ix_2}$ et $z_3 = e^{ix_3}$. Nous posons également $k_1, k_2, k_3 \in \mathbb{Z}$ tels que $d(z_1, z_3) = |x_1 - x_3 + 2k_1\pi|$, $d(z_1, z_2) = |x_1 - x_2 + 2k_2\pi|$ et $d(z_2, z_3) = |x_2 - x_3 + 2k_3\pi|$. Nous avons alors

$$\begin{aligned}
d(z_1, z_3) &= |x_1 - x_3 + 2k_1\pi| = |x_1 - x_2 + x_2 - x_3 + 2k_1\pi| && (26.222a)\\
&= \inf_{k\in\mathbb{Z}} |(x_1 - x_2 + 2k_2\pi) + (x_2 - x_3 + 2k_3\pi) + 2k\pi| && (26.222b)\\
&= \inf_{k\in\mathbb{Z}} \Big(|x_1 - x_2 + 2k_1\pi| + |x_2 - x_3 + 2k_3\pi| + 2k\pi\Big) && (26.222c)\\
&= d(z_1, z_3) + d(z_2, z_3) && (26.222d)
\end{aligned}$$

parce que le dernier infimum est réalisé par $k = 0$.

□

Le cercle est bien connu pour être symétrique et en particulier avoir une symétrie sous les rotations. Nous allons voir quelques résultats qui vont dans le sens de dire que la distance définie sur S^1 respecte cette symétrie.

Lemme 26.77.

Plusieurs points à propos de l'invariance de la topologie sous les rotations.

(1) La distance est invariante sous les rotations, c'est-à-dire que si $a, b \in S^1$ *et si* $s \in \mathbb{R}$, *alors*

$$d(e^{is}a, e^{is}b) = d(a, b). \tag{26.223}$$

(2) Les boules sont préservées sous les rotations[22], *c'est-à-dire que*

$$e^{is}B_d(a, r) = B_d(e^{is}a, r). \tag{26.224}$$

21. Définition 7.93.
22. Pas chaque boule séparément, mais l'ensemble des boules

(3) La topologie est invariante sous les rotations : $e^{is}\tau_d = \tau_d$.

Démonstration. Point par point.

(1) Si $a = e^{ix}$ et $b = e^{iy}$, nous avons

$$d(a,b) = \inf_{k\in\mathbb{Z}} |x - y + 2k\pi| = \inf_{k\in\mathbb{Z}} |(x-s) - (y-s) + 2k\pi| = d(e^{is}a, e^{is}b). \tag{26.225}$$

Et de un.

(2) Il faut une inclusion dans chaque sens.

$e^{is}B_d(a,r) \subset B_d(e^{is}a,r)$ Soit $b \in e^{is}B_d(a,r)$. Alors $b = e^{is}b'$ pour un certain $b' \in B_d(a,r)$. Nous avons alors, en utilisant le premier point,

$$d(b, e^{is}) = d(e^{-is}b, a) = d(b', a) < r. \tag{26.226}$$

Donc $b \in B_d(e^{is}a, r)$.

$B_d(e^{is}a,r) \subset e^{is}B_d(a,r)$ Soit $b \in B_d(e^{is}a,r)$. Nous devons prouver que $b \in e^{is}B_d(a,r)$, c'est-à-dire que $b = e^{is}b'$ pour un certain $b' \in B_d(a,r)$ ou encore que $e^{-is}b \in B_d(a,r)$. En utilisant encore le premier point,

$$d(e^{-is}b, a) = d(b, e^{is}a) < r. \tag{26.227}$$

Donc oui, $e^{-is}b \in B(a,r)$.

(3) Soit $A \in \tau_d$. Si $a \in e^{is}A$, alors $a = e^{is}a'$ pour un certain $a' \in A$. Notre but est de prouver que $e^{is}A$ contient un voisinage de a.

Vu que $a' \in A$, il existe $r > 0$ tel que $B_d(a',r) \subset A$. Nous avons alors

$$e^{is}aB_d(a',r) \subset e^{is}A, \tag{26.228}$$

et comme $e^{is}B_d(a',r) = B_d(e^{is}a',r) = B_d(a,r)$ nous avons bien

$$B_d(a,r) \subset e^{is}A. \tag{26.229}$$

□

Nous allons voir maintenant quelques résultats à propos de $B_d(1,r)$ qui a la bonne figure d'être un ouvert qui s'étale symétriquement en partant de 1 (le point le plus à droite du cercle). Par rapport à la figure 26.1, il s'agit ni plus ni moins que de voir qu'une boule de rayon r autour de 1 est bien la partie indiquée (symétrique par rapport à 1 et de longueur d'arc r des deux côtés). De plus, ce voisinage n'est autre que la partie du cercle située à droite de la ligne en pointillés.

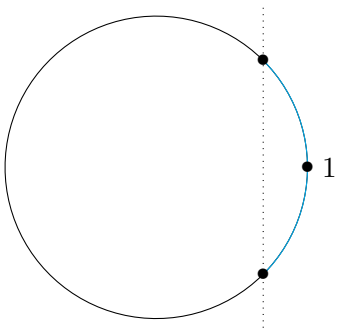


FIGURE 26.1 – Un voisinage de 1 dans S^1.

Ce lemme-ci montre que $B_d(1,r)$ est une partie de S^1 qui s'étale symétriquement autour de 1.

Lemme 26.78.
Soit l'application

$$\begin{aligned}\varphi\colon \mathbb{R} &\to S^1\\ x &\mapsto e^{ix}.\end{aligned} \tag{26.230}$$

Nous avons $B_d(1,r) = \varphi\big(]-r,r[\big)$.

Démonstration. Soit $b \in B_d(1,r)$ de la forme $b = e^{iy}$ avec y choisi de telle sorte que $d(1,b) = |y|$. Vu que $d(1,b) < r$, nous avons $|y| < r$ et donc $b \in \varphi\big(]-r,r[\big)$.

Dans l'autre sens, si $y \in]-r,r[$, alors

$$d(1,e^{iy}) = \inf_{k\in\mathbb{Z}} |y + 2k\pi| \leqslant |y| < r. \tag{26.231}$$

Nous avons utilisé le fait que l'infimum sur $k \in \mathbb{Z}$ est plus petit ou égal à la valeur pour $k = 0$. Les inégalités (26.231) montrent que $e^{iy} \in B_d(1,r)$. □

Le lemme suivant montre que que les boules autour de 1 sont délimitées par la droite en pointillé de la figure 26.1.

Lemme 26.79.
Soit $r \in [0,\pi]$. *Nous avons*

$$B_d(1,r) = S^1 \cap \{x + iy \text{ tel que } x > \cos(r)\}. \tag{26.232}$$

Démonstration. Si $r = \pi$, alors $B(1,r) = S^1\backslash\{-1\}$, alors que $\cos(\pi) = -1$.

Si $r < \pi$, alors nous partons de la formule (17.20) qui dit que $e^{ir} = \cos(e) + i\sin(r)$. D'après le lemme 26.78, un élément de $B_d(1,r)$ est de la forme e^{iy} avec $y \in]-r,r[$. Nous voudrions donc prouver que $\cos(y) > \cos(r)$ dès que $y \in]-r,r[$ et $r < \pi$.

Sur $]-r,0[$, la fonction cos est croissante, donc si $y < 0$ alors

$$\cos(y) > \cos(-r) = \cos(r). \tag{26.233}$$

De la même façon, sur $]0,r[$, la fonction cos est décroissante, de telle sorte que si $y > 0$, alors $\cos(y) > \cos(r)$.

Nous avons prouvé que $B_d(1,r) \subset S^1 \cap \{x + iy \text{ tel que } x > \cos(r)\}$.

Lançons nous dans la preuve de l'inclusion inverse.

Soit $x > \cos(r)$. Si $x + iy \in S^1$, nous avons $x + iy = e^{is} = \cos(s) + i\sin(s)$ pour un certain $s \in [-\pi,\pi[$. Notons que $s = -\pi$ correspondrait au point $-1 \in S^1$, qui est exclu de notre étude parce que nous supposons $r < \pi$. Donc $s \in]-\pi,\pi[$.

Nous avons donc $\cos(r) < x = \cos(s)$. Et voila. □

Proposition 26.80.
La topologie τ_d *sur* S^1 [23] *est la topologique induite depuis* $\mathbb{C}$.

Démonstration. Nous notons τ_i la topologie induite (c'est-à-dire l'ensemble des ouverts) et τ_d la topologie de la distance fraichement définie. Nous allons également noter $B_{\mathbb{C}}(a,r)$ la boule dans $\mathbb{C}$ de centre a et de rayon r, et $B_d(z,r)$ celle dans S^1, de centre $z \in S^1$ et de rayon r pour notre distance d.

$\underline{\tau_i \subset \tau_d}$ Un élément général de τ_i est de la forme $\mathcal{O} \cap S^1$ où $\mathcal{O}$ est un ouvert de $\mathbb{C}$. Soit $a \in \mathcal{O} \cap S^1$ et prouvons qu'il existe un ouvert de τ_d contenant a et contenu dans $\mathcal{O} \cap S^1$; cela prouvera que $\mathcal{O} \cap S^1$ est ouvert de τ_d par le théorème 7.4.

Soient $a = e^{ix}$ et r tel que $B_{\mathbb{C}}(a,r) \subset \mathcal{O}$. Nous allons montrer que $B_d(a,r) \subset B_{\mathbb{C}}(a,r)$. Un élément général de $B_d(a,r)$ est $b = e^{iy}$ tel que

$$d(a,b) = \inf_{k\in\mathbb{Z}} |x - y + 2k\pi| \leqslant r. \tag{26.234}$$

23. Définition 26.76.

Quitte à redéfinir x ou y nous pouvons supposer que l'infimum est atteint en $k = 0$. En utilisant la proposition 20.15 nous majorons :

$$|a - b| = |e^{ix} - e^{iy}| \leqslant |x - y| = d(a, b) \leqslant r. \tag{26.235}$$

Donc nous avons bien $b \in B_{\mathbb{C}}(a, r)$ dès que $b \in B_d(a, r)$.

$\underline{\tau_d \subset \tau_i}$ Ce sens est plus délicat parce que, si nous voulons suivre les mêmes pas que le premier sens, nous devrons nous appuyer sur la continuité de l'application $\ln\colon \mathbb{C}^* \to \mathbb{C}$, laquelle n'est pas vraie en -1 (voir par exemple le lemme 25.54).

Soient $A \in \tau_d$ et $a \in A$. Nous devons prouver l'existence d'un ouvert $\mathcal{O}$ de $\mathbb{C}$ tel que $S^1 \cap \mathcal{O}$ soit inclus dans A et contienne a. Nous allons prouver cela dans le cas $a = 1$ et ensuite propager le résultat en utilisant la symétrie de S^1.

Si $a = 1$ Vu que A est ouvert pour la topologie de la distance d, et vu que $1 \in A$, il existe $r > 0$ tel que $B_d(1, r) \subset A$. Pour ce r le lemme 26.79 donne

$$B_d(1, r) = S^1 \cap \{x + iy \text{ tel que } x > \cos(r)\}. \tag{26.236}$$

Nous montrons que $\mathcal{O} = B_{\mathbb{C}}(1, \delta)$ avec $\delta < 1 - \cos(r)$ fait l'affaire. Si $x + iy \in B_{\mathbb{C}}(1, \delta)$, alors $x > 1 - \delta$ et donc

$$1 - \delta > 1 - (1 - \cos(r)) = \cos(r), \tag{26.237}$$

ce qui prouve que la partie de $\mathcal{O}$ qui est dans S^1 est bien dans $B_d(1, r)$.

Si $a \neq 1$ Soient un ouvert quelconque $A \in \tau_d$ ainsi que $a = e^{ix} \in A$. Nous considérons $r > 0$ tel que $B_d(a, r) \subset A$; nous avons $e^{-ix} B_d(d, r) \subset e^{-ix} A$ et donc, en tenant compte du lemme 26.77(2) :

$$B_d(1, r) \subset e^{-ix} A. \tag{26.238}$$

Par le premier point, il existe un ouvert $\mathcal{O}$ de $\mathbb{C}$ tel que $1 \in \mathcal{O}$ et

$$\mathcal{O} \cap S^1 \subset B_d(1, r) \subset e^{-ix} A. \tag{26.239}$$

Nous avons évidemment que $a \in e^{ix}\mathcal{O}$ et

$$e^{ix}(\mathcal{O} \cap S^1) \subset e^{ix} B_d(1, r) \subset A. \tag{26.240}$$

Donc $e^{ix}\mathcal{O} \cap S^1 \subset A$. Vu que $e^{ix}\mathcal{O}$ est un ouvert de $\mathbb{C}$, l'ensemble $e^{ix}\mathcal{O} \cap S^1$ est un ouvert de τ_i.

□

Lemme 26.81 ([1]).
Deux résultats de limites dans S^1.

(1) Pour tout $a_0 \in S^1$, nous avons

$$\lim_{s \to 1} d(a, as) = 0. \tag{26.241}$$

(2) Si $f\colon S^1 \to \mathbb{C}$ est continue en $a \in S^1$, alors

$$\lim_{s \to 1} f(as) = f(a). \tag{26.242}$$

Démonstration. Point par point.

(1) Soient $a, s \in S^1$. Donnons une formule pour $d(a, as)$. Si $a = e^{ix}$ et $s = e^{iy}$ nous avons $as = e^{i(x+y)}$ et donc

$$d(a, as) = \inf_{k \in \mathbb{Z}} |x - (x + y) + 2k\pi| = \int_{k \in \mathbb{Z}} |y + 2k\pi|. \tag{26.243}$$

Voila pour la formule. Maintenant la preuve de notre point.

Soit $\epsilon > 0$. Si $\delta < \epsilon$ et si $s \in B(1, \delta$, alors il existe $y \in]-\delta, \delta[$ tel que $s = e^{iy}$ par le lemme 26.78. Pour un tel s nous avons

$$d(a, sa) = \inf_{k \in \mathbb{Z}} |y + 2k\pi| \leqslant |y| < \delta < \epsilon. \tag{26.244}$$

Nous avons trouvé $\delta > 0$ tel que $s \in B(1, \delta)$ implique $d(a, as) < \epsilon$. Cela est la limite que nous devions prouver.

(2) Soit $\epsilon > 0$. Soit $r > 0$ tel que si $b \in B(a, r)$, alors $|g(b) - g(a)| < \epsilon$; l'existence d'un tel r est la continuité de g en a. Nous considérons $\delta > 0$ tel que $s \in B(1, \delta)$ implique $sa \in B(a, r)$; l'existence d'un tel δ est le point (1) de ce lemme.

Avec tout cela nous avons $|g(as) - g(a)| < \epsilon$ dès que $s \in B(1, \delta)$. Nous avons donc, comme nous le voulions, la limite $\lim_{s \to 1} g(as) = g(a)$.

□

Lemme 26.82.
Pour $s \in S^1$, nous considérons l'application

$$\begin{aligned} \alpha_s \colon \operatorname{Fun}(S^1) &\to \operatorname{Fun}(S^1) \\ \alpha_s(g)(u) &= g(u\bar{s}) - g(u). \end{aligned} \tag{26.245}$$

Quelques propriétés avec $1 \leqslant p < \infty$:

(1) Si $f \in L^p(S^1)$, alors $\alpha_s(f) \in L^p(S^1)$.

(2) Si f est continue dans $L^p(S^1)$ nous avons la limite

$$\lim_{s \to 1} \alpha_s(f) = 0 \tag{26.246}$$

dans $L^p(S^1)$.

Démonstration. D'abord un calcul de norme :

$$\|\alpha_s(f)\|_p^p = \int_{S^1} |f(u\bar{s}) - f(u)|^p du \leqslant \int_{S^1} |f(u\bar{s})|^p du + \int_{S^1} |f(u)|^p du = 2\|f\|_p^p. \tag{26.247}$$

Donc oui pour que $\alpha_s(f) \in L^p(S^1)$.

La fonction f étant supposée continue sur le compact S^1, elle est majorée. Nous savons qu'en posant $\|f\|_\infty$ nous avons $|\alpha_s(f)| \leqslant 2M$. Donc la fonction constante

$$\begin{aligned} g \colon S^1 &\to \mathbb{C} \\ u &\mapsto 2M \end{aligned} \tag{26.248}$$

est une fonction intégrable sur S^1 qui majore $|\alpha_s(f)|$ uniformément en s. Soit une suite $s_i \to 1$ dans S^1, et posons $f_i = \alpha_{s_i}(f)$. Alors nous avons

$$\|f_i\|_p^p = \int_{S^1} |f_i(u)|^p du \tag{26.249}$$

et aussi $|f_i|^p \leqslant (2M)^p$. Le théorème de la convergence dominée de Lebesgue 13.184 nous permet de permuter limite et intégrale :

$$\lim_{i \to \infty} \|f_i\|_p^p = \int_{S^1} \lim_{i \to \infty} |f_i(u)|^p du. \tag{26.250}$$

Mais

$$\lim_{i \to \infty} f_i(u) = \lim_{i \to \infty} \alpha_{s_i}(f)(u) = \lim_{i \to \infty} \big(f(u\bar{s}_i) - f(u)\big) = 0. \tag{26.251}$$

La dernière limite est due au fait que $\lim_{s \to 1} g(us) = g(u)$ (lemme 26.81(2)). □

26.6.3 Système trigonométrique

Définition 26.83.
*La **famille trigonométrique** sur S^1 est l'ensemble de fonctions $\{e_n\}_{n\in\mathbb{Z}}$ données par*

$$\begin{aligned} e_n\colon S^1 &\to \mathbb{C} \\ z &\mapsto z^n \end{aligned} \tag{26.252}$$

*avec $n \in \mathbb{Z}$. Un **polynôme trigonométrique** est une application $S^1 \to \mathbb{C}$ de la forme*

$$\sum_{k=-n}^{n} a_k e_k \tag{26.253}$$

pour des nombres $a_k \in \mathbb{C}$, peut-être pas tous non-nuls (autrement dit, il n'est pas forcé d'avoir autant de termes négatifs que positifs).

Le but de $z \mapsto z^n$ dans cette définition est d'être lu $t \mapsto e^{int}$ lorsqu'on considère les fonctions sur $[0, 2\pi[$.

Proposition 26.84.
La famille trigonométrique est une famille orthonormale pour le produit scalaire $L^2(S^1, \mathcal{A}, \mu)$.

Démonstration. En utilisant la proposition 26.75(3) nous avons :

$$\langle e_n, e_n\rangle = \int_{S^1} e_n\overline{e_n} = \frac{1}{2\pi}\int_{[0,2\pi[} e_n\big(\varphi(x)\big)\overline{e_n\big(\varphi(x)\big)} \tag{26.254a}$$

$$= \frac{1}{2\pi}\int_{[0,2\pi[} e^{inx}e^{-inx}dx = \frac{1}{2\pi}\int_0^{2\pi} 1dx = 1. \tag{26.254b}$$

Et nous avons également, pour $m \neq n$:

$$\langle e_n, e_m\rangle = \frac{1}{2\pi}\int_{[0,2\pi[} e^{i(n-m)x}dx = \frac{1}{2\pi}\left[\frac{1}{i(n-m)e^{i(n-m)x}}\right]_0^{2\pi} = 0. \tag{26.255}$$

□

Remarque 26.85.
Vous aurez noté que le facteur $\frac{1}{2\pi}$ qui permet d'avoir $\langle e_n, e_n\rangle = 1$ ne provient ni de la définition du produit scalaire ni de celle de la famille trigonométrique, mais bien de la mesure, voir la définition 17.136.

Remarque 26.86.
Notez aussi que nous avons bien $\langle e_n, e_{-n}\rangle = 0$. Il faut donc bien prendre tous les e_n avec $n \in \mathbb{Z}$ et non seulement $n \in \mathbb{N}$.

Proposition 26.87.
Les polynômes trigonométriques forment une partie dense dans $\big(C(S^1, \mathbb{C}), \|.\|_\infty\big)$.

Démonstration. Pour préciser les notations, $C(S^1, \mathbb{C})$ est l'ensemble des fonctions continues de S^1 vers $\mathbb{C}$, et l'espace topologique que nous considérons est cet ensemble sur lequel nous considérons la distance supremum.

Nous utilisons le théorème de Stone-Weierstrass 11.317.

Le système contient une fonction constante non nulle, à savoir e_0.

Il sépare les points grâce à la fonction e_1 qui n'est autre que la fonction identité $z \mapsto z$. De plus l'ensemble des polynômes trigonométriques est stable par conjugaison parce que si

$$P = \sum_{k=-n}^{n} a_k e_k, \tag{26.256}$$

alors $\bar{P} = \sum_{k=-n}^{n} \overline{a_k e_k a} = \sum_{k=-n}^{n} \overline{a_k} e_{-k}$ qui est encore un polynôme trigonométrique. □

Définition 26.88.
Si nous avons une fonction $f\colon S^1 \to \mathbb{C}$, nous définissons ses ***coefficients de Fourier*** *par*

$$c_n(f) = \langle f, e_n \rangle \tag{26.257}$$

pourvu que l'intégrale existe.

26.6.4 Convolution

La convolution sur $\mathbb{R}^n$ est donnée par la définition 26.52. Nous voyons maintenant comment cela s'adapte à S^1.

Définition 26.89.
Si f et g sont des fonctions sur S^1 à valeurs dans $\mathbb{C}$, nous définissons la ***convolution*** *de f et g comme étant la fonction sur S^1 définie par*

$$(f * g)(z) = \int_{S^1} f(s)g(z\bar{s})d\mu(s). \tag{26.258}$$

Cette définition appelle plusieurs remarques.

— Dès que $z, s \in S^1$, nous avons $z\bar{s} \in S^1$, de telle sorte qu'au moins l'intégrande ait un sens.

— Nous ne prétendons pas que l'intégrale (26.258) converge pour toutes les fonctions f et g. Cela est une définition « pour tous les couples f, g pour lesquels l'intégrale fonctionne ».

— Le lemme 26.90 nous dira que $L^1(S^1)$ est stable par convolution : si f et g sont dans L^1, alors $f * g$ y est aussi.

— Dans la formule (26.258), la variable s est vraiment une variable muette. Cette formule aurait également pu être écrite

$$(f * g)(z) = \int_{S^1} \big[s \mapsto f(s)g(z\bar{s})\big]d\mu. \tag{26.259}$$

Lemme 26.90 ([456]).
Si $f, g \in L^1(S^1)$, alors pour presque tout $z \in S^1$, la fonction $s \mapsto f(s)g(z\bar{s})$ est dans $L^1(S^1)$.

Démonstration. Nous considérons la fonction

$$\begin{aligned} \psi\colon S^1 \times S^1 &\to \mathbb{C} \\ (z, s) &\mapsto f(s)g(z\bar{s}). \end{aligned} \tag{26.260}$$

$\underline{\psi \in L^1(S^1 \times S^1)}$ Nous utilisons le corolaire 13.261, et pour cela nous calculons les intégrales en chaine [24] :

$$\int_{S^1} \left[\int_{S^1} |f(s)g(z\bar{s})|dz\right]]ds = \int_{S^1} |f(s)| \underbrace{\left[\int_{S^1} |g(z\bar{s})|dz\right]}_{=A<\infty} ds \tag{26.261a}$$

$$= A \int_{S^1} |f(s)ds| \tag{26.261b}$$

$$< \infty. \tag{26.261c}$$

Le fait que $A < \infty$ provient directement de l'hypothèse $g \in L^1(S^1)$ [25].

Par le corolaire sus-cité nous avons bien $\psi \in L^1(S^1 \times S^1)$.

24. Dans les expressions suivantes, les symboles « ds » et « dz » n'ont pas d'autres valeurs que purement de notation pour indiquer le nom de la variable d'intégration.

25. Avec un changement de variables $z \mapsto z\bar{s}$ que je vous conseille d'être capable de justifier.

Et par Fubini Le théorème de Fubini 13.262(1) nous renseigne que pour presque tout $z \in S^1$, l'application

$$s \mapsto \psi(z, s) \tag{26.262}$$

est dans $L^1(S^1)$. Et la partie 13.262(2) ajoute que l'application

$$z \mapsto \int_{S^1} \psi(s, z) ds \tag{26.263}$$

est également $L^1(S^1)$.

Conclusion L'application donnée en (26.263) est précisément $(f * g)$. Donc $f * g \in L^1(S^1)$.

□

Lemme 26.91 ([1]).
Si $f \in L^1(S^1)$ et si g est continue sur S^1, alors $f * g$ existe et est continue sur S^1.

Démonstration. Vu que S^1 est compact, la continuité de g implique que g est bornée et donc dans $L^1(S^1)$. Le lemme 26.90 dit alors que $f * g$ est bien définie sur S^1.

Soit $z_0 \in S^1$. Nous montrons que $f * g$ est continue en z_0 ; pour cela nous considérons $\epsilon > 0$ et ensuite nous réfléchissons un peu.

Vu que g est continue sur S^1 qui est compact, g y est uniformément continue par le théorème de Heine11.90. Il existe donc un $\delta > 0$ tel que pour tout $z_0 \in S^1$, si $z \in B(z_0, \delta)$, alors $|g(z_0) - g(z)| < \epsilon$.

Soit $s \in S^1$. Si $z \in B(z_0, \delta)$, alors $\bar{s}z \in B(\bar{s}z_0, \delta)$ par le lemme 26.77(2). Dans ce cas nous avons aussi

$$|g(\bar{s}z_0) - g(\bar{s}z)| < \epsilon. \tag{26.264}$$

Un peu de calcul maintenant. D'une part

$$(f * g)(z_0) - (f * g)(z) = \int_{S^1} f(s)\big(g(z_0\bar{s}) - g(z\bar{s})\big) ds, \tag{26.265}$$

et donc

$$|(f * g)(z_0) - (f * g)(z)| \leqslant \int_{S^1} |f(s)| \underbrace{\big|g(z_0\bar{s}) - g(z\bar{s})\big|}_{<\epsilon} ds \tag{26.266a}$$

$$\leqslant \epsilon \int_{S^1} |f| \tag{26.266b}$$

$$= A\epsilon \tag{26.266c}$$

pour une certaine constante A ne dépendant pas de z_0.

Nous avons prouvé que pour tout $\epsilon > 0$, il existe un δ tel que $z \in B(z_0, \alpha)$ implique

$$|(f * g)(z_0) - (f * g)(z)| \leqslant A\epsilon, \tag{26.267}$$

ce qui signifie que $f * g$ est continue en z_0. □

Notez que dans cette démonstration, l'uniforme continuité de g a été utilisée pour effectuer d'un seul coup la majoration pour tout s dans l'intégrale.

Proposition 26.92.
Si $f \in L^1(S^1)$, nous avons

$$f * e_n = c_n(f) e_n. \tag{26.268}$$

Démonstration. Il s'agit d'un bon calcul. En considérant $z = e^{i\theta}$ nous avons

$$(f * e_n)(z) = \int_{S^1} f(s)e_n(z\bar{s})ds \tag{26.269a}$$
$$= \frac{1}{2\pi}\int_{[0,2\pi[} f(e^{ix})e_n(e^{i(\theta-x)})dx \tag{26.269b}$$
$$= e^{in\theta}\frac{1}{2\pi}\int_{[0,2\pi[} f(e^{ix})e^{-inx}dx \tag{26.269c}$$
$$= e_n(e^{i\theta})\frac{1}{2\pi}\int_{[0,2\pi[} f(e^{ix})\overline{e_n(e^{ix})}dx \tag{26.269d}$$
$$= e_n(z)\int_{S^1} f(s)\overline{e_n(s)}ds \tag{26.269e}$$
$$= e_n(z)\langle f, e_n\rangle. \tag{26.269f}$$

Donc $(f * e_n)(z) = \langle f, e_n\rangle e_n(z)$, c'est-à-dire que

$$f * e_n = c_n(f)e_n. \tag{26.270}$$

□

Lemme 26.93.
Si P est un polynôme trigonométrique et si $f \in L^1(S^1)$, alors $f * P$ est également un polynôme trigonométrique.

Démonstration. Soit $P = \sum_{k=-n}^{n} a_k e_k$. Par la linéarité du produit de convolution,

$$f * P = \sum_{k=-n}^{n} a_k f * e_k = \sum_k a_k c_k(f)e_k \tag{26.271}$$

où nous avons également utilisé la proposition 26.92. Nous avons donc un polynôme trigonométrique dont les coefficients sont $a_k c_k(f)$ au lieu de a_k. □

26.6.5 Approximation de l'unité

Lemme 26.94 ([449]).
Soient une fonction continue $f\colon S^1 \to [0,\infty[$ et $a \in S^1$ telle que $f(z) < f(a)$ pour tout $z \in S^1\backslash\{a\}$. Alors la suite de fonctions $f_n\colon S^1 \to \mathbb{R}$ donnée par

$$f_n(z) = \left(\int_{S^1} f^n\right)^{-1} f(z)^n \tag{26.272}$$

est une approximation de l'unité[26] *autour de a.*

Démonstration. En plusieurs points, dont d'abord une série de vérifications pour voir que la formule a un sens.

Strictement positive D'abord, vu que f prend ses valeurs dans $[0,\infty[$ et vu que $f(z) < f(a)$, nous avons $f(a) > 0$ (strict). Peut-être que f s'annule à certains endroits de S^1, mais pas a.

f^n est intégrable sur S^1 La fonction f^n est dans les hypothèses de la proposition 13.195 parce que S^1 est compact, f^n y est continue et la mesure sur S^1 est compatible avec la topologie (voir les hypothèses précises).

L'intégrale n'est pas nulle Vu que $f(a) > 0$, il existe un ouvert A contenant a sur lequel $f > 0$. Nous avons alors

$$\int_K f^n \geqslant \int_A f^n > 0. \tag{26.273}$$

Cela pour dire que l'inverse dans (26.272) ne pose pas de problèmes.

26. Définition 26.60.

Norme Vu que toutes les fonctions tant f que f_n sont positives, les valeurs absolues ne jouent aucun rôle et nous avons

$$\|f_n\|_1 = \int_{S^1} f_n d\mu = \left(\int_{S^1} f^n\right)^{-1} \int_{S^1} |f(z)|^n d\mu = 1. \tag{26.274}$$

Ce calcul donne d'un seul coup les deux conditions

— $\sup_k \|f_k\| = 1$

— $\int_{S^1} f_n = 1$ pour tout n.

Nous passons maintenant au vrai travail. Soit un voisinage V de a dans S^1. Soit une suite croissante (t_k) qui converge vers $f(a)$, c'est-à-dire $0 < t_k < f(a)$. Nous posons

$$A_k = \{x \in S^1 \backslash V \text{ tel que } f(x) \geqslant t_k\}. \tag{26.275}$$

Cet ensemble est contenu dans S^1 et est donc borné (pour la métrique de S^1).

A_k est fermé Attention : ici nous démontrons que A_k est fermé dans S^1, et les complémentaires sont pris dans S^1.

Nous montrons que le complémentaire est ouvert en prenant $y \in A^c$ et en montrant que y admet un voisinage contenu dans A^c (le fameux théorème 7.4 que nous ne nous lasserons jamais de citer). Si $y \in A^c$, il y a deux possibilités (non exclusives) : soit $y \in V$ soit $f(y) < t_k$. Si $y \in V$, alors le voisinage V lui-même est encore dans A^c. Si par contre $f(y) < t_k$, alors par continuité, il existe un voisinage de y sur lequel $f < t_k$.

A_k est compact L'espace S^1 est compact, par exemple grâce au lemme 7.51. La partie A_k est fermée dans le compact S^1, donc elle est compacte par le lemme 7.64.

$A_{k+1} \subset A_k$ Si $x \in A_{k+1}$, alors $f(x) \geqslant t_{k+1} > t_k$. Donc $f(x) > t_k$ et $x \in A_k$.

Intersection vide Si $x \in \bigcap_{k\in\mathbb{N}} A_k$, alors $f(x) \geqslant t_k$ pour tout k. En passant à la limite et en sachant que $\lim_{k\to\infty} t_k = f(a)$, nous avons $f(x) \geqslant f(a)$. Par hypothèse, cela n'est pas. Donc

$$\bigcap_{k\in\mathbb{N}} A_k = \varnothing. \tag{26.276}$$

Nous avons, dans un compact, des fermés emboîtés dont l'intersection est vide. Le corolaire 7.58 nous dit qu'il existe un indice à partir duquel tous les A_k sont vides.

Soit $\delta = t_k$ pour un k tel que A_k est vide. Nous avons

$$\{x \in S^1 \backslash V \text{ tel que } f(x) \geqslant \delta\} = \varnothing, \tag{26.277}$$

c'est-à-dire que sur $S^1 \backslash V$, nous avons $f < \delta$ et donc

$$\int_{S^1\backslash V} f(s)^n < \int_{S^1\backslash V} \delta^n = \mathrm{Vol}(S^1\backslash V)\delta^n \tag{26.278}$$

où $\mathrm{Vol}(S^1\backslash V) = \int_{S^1\backslash V} 1 = \mu'(S^1\backslash V)$ est une constante réelle strictement positive.

Nous avons aussi $\delta < f(a)$ parce que δ est un des t_k (et que cette suite croissante converge vers $f(a)$ sans l'atteindre par hypothèse). Soit δ_1 tel que $\delta < \delta_1 < f(a)$ [27].

Nous posons

$$W = \{x \in S^1 \text{ tel que } f(x) > \delta_1\}. \tag{26.279}$$

Cet ensemble n'est pas vide parce qu'il contient a et est ouvert parce que f est continue. Nous avons

$$\int_{S^1} f(s)^n ds \geqslant \int_W f(s)^n ds \geqslant \delta_1^n \,\mathrm{Vol}(W). \tag{26.280}$$

27. Dans [449], il prend $\delta < \delta_1 < 1$ et je crois qu'il aurait dû écrire $\varphi(0)$ au lieu de 1.

Nous avons donc déjà ces deux inégalités :

$$\int_{S^1\backslash V} f(s)^n <= \mathrm{Vol}(S^1\backslash V)\delta^n \tag{26.281}$$

et

$$\int_{S^1} f(s)^n ds \geqslant \delta_1^n \,\mathrm{Vol}(W). \tag{26.282}$$

En ce qui concerne les fonctions f_n que nous voulions étudier,

$$f_n(z) = \left(\int_{S^1} f(s)^n ds\right)^{-1} f(z)^n \leqslant \big(\mathrm{Vol}(W)\delta_1^n\big)^{-1} f(z)^n, \tag{26.283}$$

et donc

$$\int_{S^1\backslash V} f_n \leqslant \big(\mathrm{Vol}(W)\delta_1^n\big)^{-1}\mathrm{Vol}(S^1\backslash V)\sigma^n = \frac{\mathrm{Vol}(S^1\backslash V)}{\mathrm{Vol}(W)}\left(\frac{\delta}{\delta_1}\right)^n. \tag{26.284}$$

Étant donné que $\delta < \delta_1$, nous avons $(\delta/\delta_1)^n \to 0$. Donc aussi

$$\lim_{n\to\infty}\int_{S^1\backslash V} f_n(z)dz = 0. \tag{26.285}$$

□

Le théorème suivant est une version pour S^1 du théorème 26.63. Le produit de convolution dans S^1 est la définition 26.89.

Théorème 26.95 ([449, 1]).
Soient (φ_k) une approximation de l'unité sur $\Omega = S^1$ ainsi qu'une fonction $g\colon S^1 \to \mathbb{C}$.

(1) Si g est mesurable et bornée sur Ω et si g est continue en a_0 alors

$$(\varphi_k * g)(a_0) \to g(a_0). \tag{26.286}$$

Démonstration. Pour chaque $k \in \mathbb{N}$ nous posons

$$d_k = \varphi_k * g - g. \tag{26.287}$$

Le but de ce théorème est de montrer que $d_k \to 0$ pour diverses notions de convergence.

Preuve du point (1) Soit $a_0 \in S^1$. Par définition de l'approximation de l'unité, $\int_{S^1}\varphi_k = 1$ et donc on peut écrire $g(a_0) = \int_{S^1} g(a_0)\varphi_k(s)ds$. En ce qui concerne $d_k(a_0)$ nous avons alors

$$d_k(a_0) = \int_{S^1}\varphi_k(s)g(a_0\bar{s})ds - \int_{S^1} g(a_0)\varphi_k(s)ds \tag{26.288a}$$

$$= \int_{S^1}\varphi_k(s)\big(g(a_0\bar{s}) - g(a_0)\big). \tag{26.288b}$$

Nous pouvons passer à la norme (et non la valeur absolue parce que d_k prend ses valeurs dans $\mathbb{C}$) :

$$|d_k(a_0)| \leqslant \int_{S^1}|\varphi_k(s)|\big|g(a_0\bar{s}) - g(a_0)\big|ds. \tag{26.289}$$

La définition d'une approximation de l'unité nous permet de considérer $M = \sup_k \|\varphi_k\|_1 < \infty$. Le lemme 26.81(2) nous permet, lui, de considérer $\alpha > 0$ tel que

$$|g(a_0\bar{s}) - g(a_0)| < \epsilon \tag{26.290}$$

dès que $s \in B(1,\alpha)$ [28]. Vu que la suite (φ_k) est une approximation de l'unité, nous avons

$$\int_{S^1\backslash B(1,\alpha)}|\varphi_k| = 0. \tag{26.291}$$

28. Notez que $s \in B(1,\alpha)$ si et seulement si $\bar{s} \in B(1,\alpha)$. Il n'y a donc pas d'incohérence entre l'hypothèse sur s et notre condition sur $g(a_0\bar{s})$

Soit k suffisamment grand pour avoir $\int_{S^1\setminus B(1,\alpha)}|\varphi_k| < \epsilon$. Avec tout cela nous avons les majorations

$$|d_k(a_0)| \leqslant \int_{S^1\setminus B(1,\alpha)} |\varphi_k(s)|\big|g(a_0\bar{s}) - g(a_0)\big| ds \tag{26.292a}$$

$$= \int_{S^1\setminus B(1,\alpha)} |\varphi_k(s)| \underbrace{\big|g(a_0\bar{s}) - g(a_0)\big|}_{\leqslant 2\|g\|_\infty} ds + \int_{B(1,\alpha)} |\varphi_k(s)| \underbrace{\big|g(a_0\bar{s}) - g(a_0)\big|}_{<\epsilon} ds \tag{26.292b}$$

$$\leqslant 2\|g\|_\infty \int_{S^1\setminus B(1,\alpha)} |\varphi_k(s)| ds + \epsilon \int_{S^1} |\varphi_k(s)| ds \tag{26.292c}$$

$$\leqslant \epsilon\big(1 + 2\|g\|_\infty\big). \tag{26.292d}$$

Nous avons donc bien $\lim_{k\to\infty} |d_k(a_0)| = 0$ et donc la continuité de $\varphi_k * g$ en a_0.

□

Voici une version un peu forte sous l'hypothèse de continuité. Vu que S^1 est compact, la continuité est en réalité une hypothèse assez forte : ça implique l'uniforme continuité et l'existence d'un maximum et d'un minimum.

Proposition 26.96.
Soient (φ_k) une approximation de l'unité sur $\Omega = S^1$ ainsi qu'une fonction continue $g\colon S^1 \to \mathbb{C}$.

(1) Si $g \in L^p(\Omega)$ $(0 \leqslant p < \infty)$ et si g est continue, alors[29]

$$\varphi_k * g \xrightarrow{L^p} g. \tag{26.293}$$

(2) Si g est continue sur S^1, alors

$$\varphi_k * g \xrightarrow{L^\infty} g \tag{26.294}$$

Démonstration. En plusieurs points

$\lim_{u\to 1} \|\tau_u(g) - g\| = 0$ Ceci est un petit point intermédiaire. Pour des besoins de notations, nous posons

$$\begin{aligned} \tau_u(g)\colon S^1 &\to \mathbb{C} \\ s &\mapsto g(s\bar{u}) \end{aligned} \tag{26.295}$$

pour $u \in S^1$.

La fonction g est continue sur le compact S^1, et y est donc uniformément continue[30]. Nous allons en déduire que $\lim_{u\to 1} \|\tau_u(g) - g\|_\infty = 0$.

Soit $\epsilon > 0$. L'uniforme continuité de g signifie qu'il existe $\delta > 0$ tel que pour tout $a \in S^1$ si $s \in B(a,\delta)$, alors $\big|g(s) - g(a)\big| < \epsilon$. Si $u \in B(1,\delta)$ nous avons aussi $\bar{u} \in B(1,\delta)$ et donc $s\bar{u} \in B(s,\delta)$; ça c'est le lemme 26.77(2).

Pour tout $a \in S^1$ nous avons la chaine

$$u \in B(1,\delta) \Rightarrow a\bar{u} \in B(a,\delta) \Rightarrow \big|g(a\bar{u}) - g(a)\big| < \epsilon. \tag{26.296}$$

Cela étant valable pour tout a, c'est encore valable en passant au supremum[31] :

$$u \in B(1,\delta) \Rightarrow \sup_{a\in S^1} \big|g(a\bar{u}) - g(a)\big| \leqslant \epsilon \tag{26.297}$$

et donc d'accord pour

$$\lim_{u\to 1} \|\tau_u(g) - g\| = 0. \tag{26.298}$$

29. Vous noterez les $p \in]0,1[$ en bonus par rapport au cas de $\mathbb{R}^n$.
30. Théorème de Heine 11.90. C'est fondamentalement ce fait qui unifie les parties (1) et (2) de cette preuve.
31. Notez l'inégalité qui n'est plus stricte.

$\|d_k\|_\infty \to 0$ Nous prouvons la convergence uniforme sur S^1 de d_k vers zéro. Ensuite nous verrons que la compacité de S^1 permet d'en déduire les points (1) et (2).

En utilisant la notation τ_u, nous pouvons écrire

$$d_k(s) = \int_{S^1} \varphi_k(u) g(s\bar{u}) du - g(s) = \int_{S^1} \varphi_k(u) \big(\underbrace{g(s\bar{u})}_{=\tau_u(g)(s)} - g(s) \big) du, \tag{26.299}$$

et donc

$$|d_k(s)| \leqslant \int_{S^1} |\varphi_k(u)| \|\tau_u(g) - g\|_\infty du. \tag{26.300}$$

Nous posons $M = \sup_k \|\varphi_k\|_1$, et nous considérons δ tel que $\|\tau_u(g) - g\|_\infty < \epsilon$ pour tout $u \in B(1,\delta)$. Ensuite nous subdivisons S^1 en $B(1,\delta)$ et $B(1,\delta)^c$:

$$\|d_k\|_\infty \leqslant \int_{B(1,\delta)} |\varphi_k(u)| \underbrace{\|\tau_u(g) - g\|_\infty}_{\leqslant \epsilon} du + \int_{B(1,\delta)^c} |\varphi_k(u)| \|\tau_u(g) - g\|_\infty du \tag{26.301a}$$

$$\leqslant \epsilon M + 2\|g\|_\infty \int_{B(1,\delta)^c} |\varphi_k(u)| du \tag{26.301b}$$

$$\leqslant \epsilon \big(M + 2\|g\|_\infty \big) \tag{26.301c}$$

parce que pour chaque $s \in S^1$ nous avons $\tau_u(g)(s) - g(s)$ et donc $\|\tau_u(g) - g\|_\infty \leqslant 2\|g\|_\infty$.

Tout cela montre que $d_k \xrightarrow{\|.\|_\infty} 0$.

Convergence L^p, $0 < p < \infty$ Soit $\epsilon > 0$. Nous avons

$$\|d_k\|_p^p \leqslant \int_{S^1} \big| (\varphi_k * g)(s) - g(s) \big|^p ds. \tag{26.302}$$

Il existe un k à partir duquel $\|\varphi_k * g - g\|_\infty < \epsilon$. Pour de tels k nous avons

$$\|d_k\|_p^p < \epsilon^p. \tag{26.303}$$

Ce passage est très possible dans le cas de S^1 parce que $\int_{S^1} 1 = 1$. Dans le cas de $\mathbb{R}^d$, c'est pas du tout bon ; c'est pour cela que nous avons un résultat un peu plus fort dans S^1. La croissance de la fonction puissance (proposition 11.346) nous permet de conclure que $\|d_k\|_p < \epsilon$.

Nous avons donc la convergence L^p pour $0 < p < \infty$.

Convergence L^∞ Non, la convergence L^∞ n'est pas la convergence pour la norme $\|.\|_\infty$. Voir la sous-section 26.2.3. Il n'en reste pas moins que si $\epsilon > 0$ et si k est assez grand pour que $\|f_k - f\|_\infty < \epsilon$, nous aurons

$$N_\infty(f_k - f) \leqslant \|f_k - f\| < \epsilon. \tag{26.304}$$

□

26.6.6 Base hilbertienne (suite des polynômes trigonométriques)

Voici le plan pour la suite :

— Construire un polynôme trigonométrique qui vérifie les hypothèse du lemme 26.94.

— En déduire une approximation de l'unité constituée de polynômes trigonométriques.

— Dire que si $f \in L^2(S^1)$, alors $f * \varphi_k$ est un polynôme trigonométrique dès que φ_k en est un.

— Invoquer le théorème 26.95(1) pour déduire que $\varphi_k * f$ est une suite de polynômes trigonométriques dans $L^2(S^1)$ qui converge $\varphi_k * f \xrightarrow{L^2} f$.

Lemme 26.97.
La fonction $P = e_1 + e_{-1}$ est continue à valeurs réelles sur S^1.

Démonstration. Nous avons $e_1(z) = z$ et $e_{-1}(z) = z^{-1}$, c'est-à-dire que pour $z = e^{ix}$ ($x \in \mathbb{R}$), nous avons $e_{-1}(e^{ix}) = e^{-ix}$, de telle sorte que, en utilisant le lemme 17.10 qui donne e^{ix} en termes des fonctions trigonométriques usuelles :

$$(e_1 + e_{-1})(e^{ix}) = e^{ix} + e^{-ix} = \cos(x) + i\sin(x) + \cos(x) - i\sin(x) = 2\cos(x). \tag{26.305}$$

Nous avons donc la continuité et les valeurs réelles. □

Lemme 26.98.
Il existe un polynôme trigonométrique à valeurs dans $[0, \infty[$ et tel que $f(a) < f(1)$ pour tout $a \neq 1$ dans S^1.

Démonstration. Le lemme 26.97 nous dit déjà que $P = e_1 + e_{-n}$ est continue à valeurs réelles. Or qui est continue sur un compact (ici S^1), atteint donc ses bornes. Il est donc facile de considérer [32] $M > 0$ tel que $Q = M + e_1 + e_{-1}$ est à valeurs dans $[0, \infty[$.

Une forme explicite de Q est que

$$Q(e^{ix}) = M + 2\cos(x). \tag{26.306}$$

Le maximum de $\cos(x)$ est obtenu en $x = 0$ et vaut 1. Le maximum de Q est alors $Q(1) = 2 + M$. Il n'est atteint qu'une seule fois sur S^1 parce que pour avoir $Q(e^{ix}) = 2 + M$, il faut avoir $2\cos(x) = 2$, c'est-à-dire $x = 2k\pi$. Mais $e^{i2k\pi} = 1$.

Donc $Q(a) < M + 2 = Q(1)$ pour tout $a \neq 1$ dans S^1. □

Proposition 26.99.
Les polynômes trigonométriques $\{e_n\}_{n\in\mathbb{Z}}$ forment une base hilbertienne de $L^2(S^1)$.

Démonstration. Le fait que les e_n soient orthonormée est la proposition 26.84. Il reste à prouver que ce soit un système total.

Soit $f \in L^2(S^1)$. Soit un polynôme Q vérifiant le lemme 26.98 ; nous posons

$$\varphi_k(z) = \left(\int_{S^1} Q^n\right)^{-1} Q(z)^n. \tag{26.307}$$

Cela est une approximation de l'unité par la proposition 26.94. Les φ_k sont des polynômes trigonométriques parce que les Q^n le sont et que que $\int_{S^1} Q^n$ est seulement un nombre.

Le lemme 26.93 nous dit alors que pour tout k, la fonction

$$\varphi_k * f \tag{26.308}$$

est un polynôme trigonométrique. □

Nous nous permettons de confirmer la remarque 26.86 comme quoi il faut bien tous les e_n avec $n \in \mathbb{Z}$, parce que le polynôme trigonométrique Q est bien construit à partir de $e_1 + e_{-1}$.

26.6.7 Convolution, bis

Lemme 26.100.
Nous considérons l'application

$$\begin{aligned} \varphi\colon \mathbb{R} &\to S^1 \\ t &\mapsto e^{it}. \end{aligned} \tag{26.309}$$

Soient $t, u \in \mathbb{R}$ tels que $\varphi(t) = \varphi(u)$. Alors pour toutes fonctions pour lesquelles les intégrales convergent,

$$\int_0^{2\pi} (f \circ \varphi)(\theta)(g \circ \varphi)(t - \theta)\frac{d\theta}{2\pi} = \int_0^{2\pi} (f \circ \varphi)(\theta)(g \circ \varphi)(u - \theta)\frac{d\theta}{2\pi}. \tag{26.310}$$

32. Par exemple, M est le maximum de $|P|$.

Démonstration. Si $\varphi(u) = \varphi(t)$, alors $u = t + 2k\pi$ pour un certain $k \in \mathbb{Z}$. Cette condition implique que $\varphi(t-\theta) = \varphi(u-\theta)$, et donc l'égalité

$$\int_0^{2\pi} (f \circ \varphi)(\theta)(g \circ \varphi)(t-\theta)\frac{d\theta}{2\pi} = \int_0^{2\pi} (f \circ \varphi)(\theta)(g \circ \varphi)(u-\theta)\frac{d\theta}{2\pi}. \tag{26.311}$$

□

Définition 26.101 (Convolution sur S^1)**.**
Le lemme 26.100 permet de définir

$$(f * g)\big(\varphi(t)\big) = \int_0^{2\pi} (f \circ \varphi)(\theta)(g \circ \varphi)(t-\theta)\frac{d\theta}{2\pi} \tag{26.312}$$

pour toutes les paires de fonctions $f, g \in \mathrm{Fun}(S^1, \mathbb{C})$ pour lesquelles l'intégrale converge.

26.7 L'espace $L^2\big([a,b]\big)$

L'espace $L^2\big([a,b]\big)$ est l'espace générique sur lequel nous allons construire les espaces L^2 sur $[-T,T]$ et $[0,2\pi]$.

Si f et g sont dans $L^2\big([a,b]\big)$, il n'est pas possible de définir $f * g$ par la formule intégrale usuelle parce que $f(x_0 + t)$ n'existe pas pour tout x_0 et t dans $[a,b]$. Donc soit nous utilisons un truc pas très net comme étendre les fonctions sur $[a,b]$ en fonctions périodiques sur $\mathbb{R}$, soit nous intégrons vraiment seulement sur $[a,b]$.

Nous n'allons suivre aucune de ces deux voies ou plutôt les deux en même temps. Nous allons seulement tout ramener de S^1 que nous venons de travailler.

Définition 26.102.
Sur $[a,b]$ nous considérons la mesure de Lebesgue et le produit

$$\langle f, g\rangle = \frac{1}{b-a}\int_a^b f(x)\bar{g}(x)dx. \tag{26.313}$$

Si vous voulez une raison inavouable pour justifier ce facteur $\frac{1}{b-a}$, remarquez que dx a pour unité le mètre. En mettant la facteur $b-a$ (qui a aussi le mètre comme unité), le tout a les unités de fg, comme il se doit pour le produit scalaire.

Proposition 26.103.
L'application

$$\begin{aligned} \phi\colon L^2\big([a,b]\big) &\to L^2(S^1) \\ \phi(f)(z) &= f\big((s^{-1} \circ \varphi^{-1})(z)\big) \end{aligned} \tag{26.314}$$

où s est donnée par

$$\begin{aligned} s\colon [a,b] &\to [0,2\pi] \\ x &\mapsto 2\pi\frac{x-a}{b-a} \end{aligned} \tag{26.315}$$

et φ est l'application usuelle

$$\begin{aligned} \varphi\colon [0,2\pi[&\to S^1 \\ t &\mapsto e^{it} \end{aligned} \tag{26.316}$$

est une bijection isométrique.

Démonstration. La preuve du fait que ϕ est isométrique suffira pour prouver qu'elle prend bien ses valeurs dans $L^2(S^1)$.

Isométrique C'est un calcul :

$$\begin{aligned}
\|\phi(f)\|^2 &= \langle \phi(f), \phi(f)\rangle && (26.317a)\\
&= \int_{S^1} |\phi(f)|^2 && (26.317b)\\
&= \frac{1}{2\pi}\int_{[0,2\pi[} |\phi(f)\big(\varphi(u)\big)|^2 du && (26.317c)\\
&= \frac{1}{2\pi}\int_0^{2\pi} |f\big((s^{-1}\circ\varphi^{-1}\circ\varphi)(u)\big)|^2 du && (26.317d)\\
&= \frac{1}{2\pi}\int_0^{2\pi} |f\big(s^{-1}(u)\big)|^2 du. && (26.317e)
\end{aligned}$$

Il est temps de faire le changement de variables[33] $y = s^{-1}(u)$, c'est-à-dire

$$y = \frac{b-a}{2\pi}u + a. \tag{26.318}$$

En ce qui concerne la différentielle,

$$dy = \frac{b-a}{2\pi}du \tag{26.319}$$

et pour les bornes, si $u = 0$ alors $y = a$ et si $u = 2\pi$, $y = b$. Donc

$$\begin{aligned}
\|\phi(f)\|^2 &= \frac{1}{2\pi}\int_a^b |f(y)|^2 \frac{2\pi}{b-a} dy && (26.320a)\\
&= \frac{1}{b-a}\int_a^b |f|^2 && (26.320b)\\
&= \|f\|^2. && (26.320c)
\end{aligned}$$

Injectif Soit f telle que $\phi(f) = 0$. Alors pour tout $z \in S^1$ nous avons

$$f\big((s^{-1}\circ\varphi^{-1})(z)\big) = 0. \tag{26.321}$$

Vu que $s^{-1}\circ\varphi^{-1}\colon S^1 \to [a,b[$ est une bijection, pour tout $u \in [a,b[$ nous avons $f(u) = 0$. Donc $f = 0$ dans $L^2\big([a,b]\big)$ parce que du point de vue de L^2, que l'on prenne ou non les bornes, ce n'est pas important.

Surjectif Si $g \in L^2(S^1)$, alors en posant

$$f(u) = g\big((\varphi\circ s)(u)\big) \tag{26.322}$$

nous avons $g = \phi(f)$.

□

Définition 26.104.
En ce qui concerne le produit de convolution, si f et g sont des fonctions sur $[a,b]$ nous définissons

$$f * g = \phi^{-1}\big(\phi(f) * \phi(g)\big) \tag{26.323}$$

tant que les formules ont un sens.

33. Nous le faisons de façon un peu informelle ; soyez capable de bien justifier.

26.8 Sur $[0, 2\pi[$

Le produit de convolution est un peut subtil parce que $f(t-x)$ n'est pas défini à priori pour tout $t, x \in [0, 2\pi[$, vu que f n'est définie que sur $[0, 2\pi[$. Au moins trois solutions s'offrent à nous :

— considérer implicitement la fonction prolongée par périodicité.

— considérer les fonctions sur $\mathbb{R}/2\pi$, et définir un peut toutes les opérations modulo 2π (fastidieux)

— utiliser une bijection ayant les bonnes propriétés avec S^1 sur lequel tout est déjà fait.

Nous sélectionnons la troisième voie. Pour cela nous considérons la fonction (attention, elle n'est pas tout à fait la même que celle plus haut)

$$\begin{aligned} \varphi\colon [0, 2\pi[&\to S^1 \\ t &\mapsto e^{it} \end{aligned} \tag{26.324}$$

qui est une bijection par la proposition 17.48. Pour le produit de convolution,

$$(f * g)(x) = (f \circ \varphi^{-1}) * (g \circ \varphi^{-1})\big(\varphi(x)\big) \tag{26.325}$$

pour toutes les fonctions $f, g\colon [0, 2\pi[\to \mathbb{C}$ pour lesquelles les intégrales en jeu ont un sens.

26.9 Sur $[-T, T[$

Pour rappel, les éléments de L^2 sont des classes de fonctions à valeurs dans $\mathbb{C}$.

Proposition 26.105.
Les fonctions

$$\begin{aligned} e_n\colon [-T, T] &\to \mathbb{C} \\ t &\mapsto \frac{1}{\sqrt{2T}} e^{2\pi i n t/T}. \end{aligned} \tag{26.326}$$

forment une base hilbertienne de $L^2\big([-T, T[\big)$.

Démonstration. Nous utilisons le théorème de Stone-Weierstrass 11.317. Pour cela, nous considérons $X = [-T, T]$ qui est compact et Hausdorff ainsi que A, l'ensemble des polynômes trigonométrique :

$$A = \{ \sum_{k=-n}^{n} a_k e_k \text{ tel que } a_k \in \mathbb{C}\}_{n\in\mathbb{N}}. \tag{26.327}$$

Nous vérifions que ce A satisfait aux hypothèses de Stone-Weierstrass 11.317.

A est une algèbre Il s'agit seulement de vérifier que

$$e_k e_l = \frac{1}{\sqrt{2T}} e_{k+l}. \tag{26.328}$$

A sépare les points Soient $x, y \in [-T, T]$. Si $e_k(x) = e_k(y)$

□

26.9.1 Le cas dans $[0, 2\pi]$

En pratique, nous n'allons pas souvent travailler avec des fonctions sur intervalle symétrique $[-T, T]$, mais le plus souvent nous serons sur $[0, 2\pi]$.

Nous notons ici une conséquence du théorème 26.41 dans le cas de l'espace L^2. La proposition suivante est une petite partie du corolaire 24.49, qui sera d'ailleurs démontré de façon indépendante.

Proposition 26.106.
Si nous avons une suite de réels (a_k) telle que $\sum_{k=0}^{\infty}|a_k|^2<\infty$ alors la suite

$$f_n(x)=\sum_{k=0}^{n}a_ke^{ikx} \tag{26.329}$$

converge dans $L^2\big(]0,2\pi[\big)$.

Démonstration. Quitte à séparer les parties réelles et imaginaires, nous pouvons faire abstraction du fait que nous parlons d'une série de fonctions à valeurs dans $\mathbb{C}$ au lieu de $\mathbb{R}$.

Un simple calcul est :

$$\|f_n-f_m\|^2\leqslant\int_0^{2\pi}\sum_{k=n}^{m}|a_k|^2dx\leqslant 2\pi\sum_{k=n}^{m}|a_k|^2. \tag{26.330}$$

Par hypothèse le membre de droite est $|s_m-s_n|$ où s_k dénote la suite des sommes partielles de la série des $|a_k|^2$. Cette dernière est de Cauchy (parce que convergente dans $\mathbb{R}$) et donc la limite $n\to\infty$ (en gardant $m>n$) est zéro. Donc la suite des f_n est de Cauchy dans L^2 et donc converge dans L^2. □

Adaptons tout cela pour l'espace $L^2\big([0,2\pi]\big)$. Nous posons

$$\langle f,g\rangle=\int_0^{2\pi}f(t)\overline{g(t)}dt \tag{26.331}$$

et

$$e_n(t)=\frac{1}{\sqrt{2\pi}}e^{int}. \tag{26.332}$$

L'importance du système trigonométrique défini en 26.65 est d'être une base de $L^2\big([0,2\pi]\big)$, comme précisé dans le lemme suivant.

Lemme 26.107.
Le système trigonométrique $\{e_n\}_{n\in\mathbb{Z}}$ est une base hilbertienne de $L^2\big([0,2\pi]\big)$.

Démonstration. Pour rappel, une base hilbertienne est la définition 24.26. Nous prouvons d'abord que le système est orthogonal. Nous avons

$$\langle e_n,e_m\rangle=\frac{1}{2\pi}\int_0^{2\pi}e^{i(n-m)t}dt. \tag{26.333}$$

Si $n=m$, alors cela est égal à 1. Sinon c'est une intégrale simple :

$$\langle e_n,e_m\rangle=\left[\frac{i}{n-m}e^{i(n-m)t}\right]_0^{2\pi}=0. \tag{26.334}$$

Cela est pour l'orthogonalité.

Pour que le système soit total, il faut que son espace vectoriel engendré soit dense. Cela est le théorème 26.70. □

Note : le théorème 27.6 donné aussi la densité, mais sera démontré plus tard, indépendamment. Voir aussi les thèmes 13 et 14.

Pour un élément donné $f\in L^2\big([0,2\pi]\big)$, nous définissons

$$S_nf=\sum_{k=-n}^{n}\langle f,e_k\rangle e_k \tag{26.335}$$

et nous avons le théorème suivant, qui récompense les efforts consentis à propos de la densité des polynômes trigonométriques dans L^2.

Théorème 26.108.
Soit $f \in L^2\big([0,2\pi]\big)$. Nous avons égalité[34]

$$f = \sum_{n\in\mathbb{Z}} c_n(f) e_n \tag{26.336}$$

dans L^2.

Nous avons aussi la convergence

$$S_n f \xrightarrow{L^2} f. \tag{26.337}$$

Démonstration. Le système trigonométrique $\{e_n\}_{n\in\mathbb{Z}}$ est total pour l'espace de Hilbert $L^2\big([0,2\pi]\big)$ (sans périodicité particulière). Donc le point (1) du théorème 24.44 nous donne l'égalité demandée.

La convergence (26.337) est une reformulation de l'égalité (26.336). □

26.109.
Obtenir la convergence L^2 ne demande pas d'hypothèses de périodicité : la convergence (26.337) est automatique du fait que le système trigonométrique soit total. Ce n'est cependant pas plus qu'une convergence L^2 et elle ne demande pas $f(0) = f(2\pi)$, même si pour chacun des e_k nous avons $e_k(0) = e_k(2\pi)$.

Si $f(2\pi) \neq f(0)$, alors il existe tout de même une suite (f_n) convergente vers f au sens L^2 telle que $f_n(0) = f_n(2\pi)$. Cela ne contredit en rien le fait que $e_k(0) = e_k(2\pi)$ parce que dans L^2, la valeur d'un point seul n'a pas d'importance.

Si nous voulons une vraie convergence ponctuelle voir uniforme $(S_n f)(x) \to f(x)$, alors il faut ajouter des hypothèses sur la continuité de f, sa périodicité ou le comportement des coefficients c_n. Voir aussi le thème 37.

Exemple 26.110.
Si $f \in L^2\big([0,2\pi]\big)$ est (la classe de) une fonction à valeurs réelles, alors on peut la développer avec nettement moins de termes. D'abord nous savons que $e_{-n} = \overline{e_n}$, et donc

$$\langle f, e_n\rangle = \overline{\langle f, e_{-n}\rangle}, \tag{26.338}$$

ce qui donne

$$f = \sum_{n\in\mathbb{Z}} \langle f, e_n\rangle e_n = \sum_{n\in\mathbb{N}} \langle f, e_n\rangle e_n + \overline{\langle f, e_n\rangle e_n} = \sum_{n\in\mathbb{N}} \Re\big(\langle f, e_n\rangle e_n\big). \tag{26.339}$$

Or

$$\Re\big(\langle f, e_n\rangle e_n\big) = \frac{1}{(2\pi)^{3/2}} \cos(nx) \int_0^{2\pi} f(t)\cos(nt)dt - \frac{1}{(2\pi)^{3/2}} \sin(nx) \int_0^{2\pi} f(t)\sin(nt)dt. \tag{26.340}$$

Considérons la fonction impaire $\tilde{f} \in \text{Ł}^2\big([-2\pi,2\pi]\big)$ créée à partir de f. Elle se développe de même et nous avons la même formule (26.340) à part quelques coefficients et le fait que les intégrales sont entre -2π et 2π. Vu que $\tilde{f}$ est impaire, l'intégrale avec $\cos(nt)$ s'annule et

$$\tilde{f}(x) = \sum_{n\in\mathbb{N}} c_n \sin(nx) \tag{26.341}$$

pour certains coefficients réels c_n. Cette égalité est à considérer dans L^2, c'est-à-dire presque partout et en particulier presque partout sur $[0,2\pi]$.

Donc les fonctions réelles sur $[0,2\pi]$ peuvent être écrites sous la forme d'une série de seulement des sinus.

Note : en choisissant $\tilde{f}$ paire, nous aurions eu une série de cosinus. △

34. Notons que la somme sur $\mathbb{Z}$ dans (26.336) est commutative ; il n'est donc pas besoin d'être plus précis.

26.10 Théorème de la projection normale

26.10.1 Espace uniformément convexe

Définition 26.111 (Espace uniformément convexe[457]).
*Un espace de Banach B est **uniformément convexe** si il existe une fonction $\delta\colon]0,\infty[\to\mathbb{R}^+$ telle que si*

(1) $\|x\|\leqslant\|y\|\leqslant 1$,

(2) $\|x-y\|\geqslant\epsilon$,

alors

$$\|\frac{x+y}{2}\|\leqslant\|y\|-\delta(\epsilon). \tag{26.342}$$

Lemme 26.112 ([1]).
Si B est un espace de Banach uniformément convexe, alors pour tout $k>0$, il existe une fonction $\delta_k\colon]0,\infty[\to\mathbb{R}^2$ telle que si

(1) $\|x\|\leqslant\|y\|\leqslant k$,

(2) $\|x-y\|\geqslant\epsilon$,

alors

$$\|\frac{x+y}{2}\|\leqslant\|y\|-\delta_k(\epsilon). \tag{26.343}$$

Démonstration. Nous posons $x'=x/\|x\|$ et $y'=y/\|y\|$. Nous avons alors

$$\|x'-y'\|=\frac{\|x-y\|}{k}>\frac{\epsilon}{k}. \tag{26.344}$$

L'uniforme convexité de B dit alors que

$$\|\frac{x'+y'}{2}\|>\|y'\|-\delta(\epsilon/k). \tag{26.345}$$

En multipliant cette inégalité par k nous trouvons

$$\|\frac{x+y}{2}\|>\|y\|-k\delta(\epsilon/k). \tag{26.346}$$

Donc en posant $\delta_k(\epsilon)=k\delta(\epsilon/k)$, nous avons le résultat escompté. □

Définition 26.113 (Projection normale[457]).
Soient un espace de Banach B ainsi que $V\subset B$. Soit $a\in B$. La fonction

$$\begin{aligned}f\colon V&\to\mathbb{R}\\ x&\mapsto d(x,a)\end{aligned} \tag{26.347}$$

possède un infimum[35] *m. Si $x\in V$ est tel que $d(x,a)=m$, alors x est une **projection normale** de a sur V.*

Proposition 26.114 ([457]).
Soient un espace de Banach B et un sous-espace vectoriel $V\subset B$. Si une projection normale de $a\in B$ sur V existe, alors elle est unique.

Démonstration. Soient deux projections normales b,b' de a sur V.

Si $m=0$, alors $\|a-b\|=0$ et $\|a-b'\|=0$, ce qui donne $a=b$ et $a=b'$. Donc d'accord pour $b=b'$.

Si $m>0$ alors nous utilisons l'inégalité $\|x+y\|\leqslant\|x\|+\|y\|$ sous la forme

$$\|a-\frac{b+b'}{2}\|=\|\frac{a-b}{2}+\frac{a-b'}{2}\|\leqslant\|\frac{a-b}{2}\|+\|\frac{a-b'}{2}\|=\frac{m}{2}+\frac{m}{2}=m. \tag{26.348}$$

35. Toute fonction à valeurs positives possède un infimum, c'est la proposition 1.157.

Mais $\frac{b+b'}{2} \in V$, donc

$$\|a - \frac{b+b'}{2}\| \geqslant m. \tag{26.349}$$

Nous en déduisons que dans (26.348), toutes les inégalités sont des égalités et en particulier

$$\|\frac{b+b'}{2} - a\| = m. \tag{26.350}$$

Nous avons donc les deux égalités suivantes :

$$2m = \|a-b\| + \|a-b'\| \tag{26.351}$$

et

$$2m = \|b+b'-2a\|. \tag{26.352}$$

Cela donne

$$\|a-b\| + \|a-b'\| = \|(a-b)+(a-b')\|. \tag{26.353}$$

Vu que B est strictement convexe, cela n'est possible que si $a-b = a-b'$, ce qui signifie que $b = b'$. □

Théorème 26.115 ([458, 457]).
Si B est un espace de Banach uniformément convexe, si $V \subset B$ est un sous-espace vectoriel complet et si $a \in B$, alors la projection normale de a sur V existe et est unique.

Démonstration. En deux parties.

Unicité Soient deux projections normales b et b' de a sur V. Nous avons $\|a-b\| = \|a-b'\| = m$. Si $b \neq b'$, il existe $\epsilon > 0$ tel que

$$\|b-b'\| > \epsilon > 0. \tag{26.354}$$

En posant

$$x = \frac{1}{2}\frac{b-a}{m}, \quad y = \frac{1}{2}\frac{b'-a}{m}, \tag{26.355}$$

nous avons $\|x\| = \|y\| = \frac{1}{2} < 1$. L'uniforme convexité de B donne alors

$$\|\frac{x+y}{2}\| \leqslant \|y\| - \delta(\epsilon). \tag{26.356}$$

Mais

$$x+y = \frac{\frac{1}{2}(b+b')-a}{m} \tag{26.357}$$

et

$$\|x+y\| \leqslant 2\|y\| - 2\delta(\epsilon) = 1 - 2\delta(\epsilon) < 1. \tag{26.358}$$

Nous avons donc prouvé que

$$\|\frac{1}{2}(b+b')-a\| < m, \tag{26.359}$$

ce qui est impossible parce que cela dirait que $\frac{b+b'}{2}$ est une « meilleure » projection normale que b et b'.

Existence Soient b_k dans V tels que $\|a-b_k\| \to m$. Nous supposons (quitte à passer à une sous-suite) que

$$\|a-b_{k+1}\| \leqslant \|a-b_k\|. \tag{26.360}$$

La suite (b_k) converge Nous supposons qu'elle ne converge pas. Elle n'est donc pas de Cauchy parce que B est de Banach[36] et donc complet (ce n'est pas la complétude de V qui joue ici, mais bien celle de B). Il existe $\epsilon > 0$ tel que pour tout $N \in \mathbb{N}$ il existe $p, q > N$ tels que

$$\|b_p - b_q\| > \epsilon. \tag{26.361}$$

Nous effectuons quelque choix.

36. Définition 7.279.

(1) nous choisissons $q > p$ de telle sorte que $\|a - b_p\| \leqslant \|a - b_q\|$,

(2) nous choisissons N assez grand pour avoir

$$\|a - b_p\| \leqslant \|a - b_q\| < m. \tag{26.362}$$

Nous avons $\|(a - b_p) - (a - b_q)\| = \|a_q - a_p\| > \epsilon$, ce qui avec l'uniforme convexité donne

$$\frac{\|(a - b_p) + (a - b_q)\|}{2} \leqslant \|a - b_q\| - \delta(\epsilon). \tag{26.363}$$

Donc

$$m \leqslant \|a - \frac{b_p - b_q}{2}\| = \|\frac{(a - b_p) + (a - b_q)}{2}\| \leqslant \|a - b_q\| - \delta(\epsilon) < m - \delta(\epsilon) < m. \tag{26.364}$$

Cela signifie que $m < m$, ce qui est impossible.

Conclusion La suite (b_k) converge dans B. Vu que V est complet, la limite est dans V. Cette limite, que nous nommons b, vérifie

$$\|a - b\| \leqslant \|a - b_k\| \tag{26.365}$$

pour tout k. Mais comme nous avons $m \leqslant \|a - b_k\| \to m$, nous avons $\|a - b\| = m$, c'est-à-dire que b est une projection normale de a sur V.

□

26.10.2 Des inégalités

Avant d'entrer dans le vif du sujet, nous nous fendons d'une petite étude de fonction. Soit

$$\begin{aligned} \phi\colon [0,1] &\to \mathbb{R} \\ x &\mapsto \frac{(1+x)^r}{1+x^r}. \end{aligned} \tag{26.366}$$

Un peu de calcul montre que

$$\frac{\phi'(x)}{\phi(x)} = \frac{r(1 - x^{r-1})}{(1 + x^r)(1 + x)}. \tag{26.367}$$

Lemme 26.116.
Soient $a, b > 0$ et $r > 1$. Nous avons les inégalités

$$a^r + b^r \leqslant (a + b)^r \leqslant 2^{r-1}(a^r + b^r). \tag{26.368}$$

Démonstration. Pour la première inégalité, nous posons $f(x) = a^r + x^r$ et $g(x) = (a + x)^r$. Nous avons $f(0) = g(0) = a^r$, et

$$f'(x) = rx^{r-1} \tag{26.369a}$$

$$g'(x) = r(a + x)^{r-1}. \tag{26.369b}$$

Vu que $r > 1$, la fonction $t \mapsto t^{r-1}$ est croissante par la proposition 11.346.

Nous passons à la seconde inégalité. Le lemme 16.85 nous dit que la fonction $f\colon x \mapsto x^r$ est convexe. Donc

$$f\left(\frac{a}{2} + \frac{b}{2}\right) \leqslant \frac{1}{2}f(a) + \frac{1}{2}f(b). \tag{26.370}$$

De là nous déduisons

$$\frac{(a + b)^r}{2^r} \leqslant \frac{1}{2}(a^r + b^r), \tag{26.371}$$

c'est-à-dire la seconde inégalité. □

Nous allons démontrer les inégalités de Hanner dans le théorème 26.121. Vu que ce sera un peu longuet, nous faisons un lemme.

Lemme 26.117.
Soient $z_1, z_2 \in \mathbb{C}$. Nous avons

$$|z_1 + z_2|^p + |z_1 - z_2|^p \geqslant \big(|z_1| + |z_2|\big)^p + \big||z_1| - |z_2|\big|^p. \tag{26.372}$$

Démonstration. Soient $z_1, z_2 \in \mathbb{C}$. Nous posons

$$d = |z_1 + z_2|^p + |z_1 - z_2|^p. \tag{26.373}$$

Pour $|z_1|$ et $|z_2|$ fixés, nous nous demandons quel est le minimum possible de d.

Si $|z_1| = 0$, alors le minimum est $2|z_2|^p$ et si $|z_2| = 0$ alors il est $2|z_1|^p$. Pour les autres cas, nous posons $|z_1| = a > 0$ ainsi que $b \in \mathbb{R}$ et $\theta \in \mathbb{R}$ tels que[37]

$$z_2 = z_1 a^{-1} b e^{i\theta}. \tag{26.374}$$

Nous avons déjà que $z_1 + z_2 = z_1(1 + a^{-1}be^{i\theta})$ et donc

$$|z_1 + z_2| = a|1 + a^{-1}be^{i\theta}| = |a + be^{i\theta}| \tag{26.375}$$

parce que $a > 0$. De plus,

$$|a + be^{i\theta}|^2 = (a + be^{i\theta})(a + be^{-i\theta}) = a^2 + b^2 + 2ab\cos(\theta) \tag{26.376}$$

parce que $e^{i\theta} + e^{-i\theta} = \cos(\theta)$. Nous posons

$$d(\theta) = |a + be^{i\theta}|^p + |a - be^{i\theta}|^p. \tag{26.377}$$

En développant,

$$d(\theta) = \big(a^2 + b^2 + 2ab\cos(\theta)\big)^{p/2} + \big(a^2 + b^2 - 2ab\cos(\theta)\big)^{p/2}. \tag{26.378}$$

Trouvons le minimum de cette fonction de θ. D'abord sa dérivée :

$$ad'(\theta) = pab\sin(\theta)\big[\big(a^2 + b^2 - 2ab\cos(\theta)\big)^{p/2-1} - \big(a^2 + b^2 + 2ab\cos(\theta)\big)^{p/2-1}\big] \tag{26.379a}$$
$$= pab\sin(\theta)s(\theta). \tag{26.379b}$$

Nous avons $s(\theta) = 0$ pour $\theta = \pi/2$ et $\theta = 3\pi/2$. Il faut surtout remarquer que $1 < p < 2$, ce qui donne $\frac{p}{2} - 1 < 0$. La fonction $x \mapsto x^{p/2-1}$ est donc décroissante. Cela pour dire que

$$s(0) = \big(|a - b|^2\big)^{p/2-1} - \big(|a + b|^2\big)^{p/2-1} > 0. \tag{26.380}$$

De la même façon, $s(\pi) = -s(0) < 0$. Cela permet d'écrire un petit tableau de signe de d', et de conclure que $d(\theta)$ a un minimum en 0 et en π. Calcul fait, nous avons

$$d(0) = d(\pi) = |a + b|^p + |a - b|^p. \tag{26.381}$$

En reliant à (26.373) nous avons l'inégalité

$$|z_1 + z_2|^p + |z_1 - z_2|^p \geqslant (a + b)^p - |a - b|^p. \tag{26.382}$$

Nous rappelons que $a = |z_1|$ et que $z_2 = z_1 a^{-1} b e^{i\theta}$. Notons au passage que $|z_2| = b$, donc que ce que nous dit l'équation (26.382) est que

$$|z_1 + z_2|^p + |z_1 - z_2|^p \geqslant \big(|z_1| + |z_2|\big)^p + \big||z_1| - |z_2|\big|^p. \tag{26.383}$$

□

37. Proposition 17.49

Encore dans la catégorie des lemmes pour les inégalités de Hanner, nous avons celui-ci.

Lemme 26.118 ([1, 459])**.**
La fonction

$$\begin{aligned}\eta\colon]0,\infty[&\to \mathbb{R}\\ a &\mapsto (a^{1/p}+1)^p+|a^{1/p}-1|^p\end{aligned}\tag{26.384}$$

est strictement convexe.

Démonstration. La fonction η est une fonction de classe C^∞ sur $]0,\infty[\backslash\{1\}$. Quelle est sa régularité en $a=1$? Le fait qu'elle y soit dérivable pas clair à cause de la valeur absolue. En tout cas, la fonction $x\mapsto|x-1|$ n'est pas dérivable en $x=1$, mais peut-être que les exposants aident à lisser. Nous y reviendrons.

Afin de suivre les calculs nous introduisons quelques fonctions :

$$so(x)=1+x^{1/p}\tag{26.385a}$$
$$di(x)=1-x^{1/p}\tag{26.385b}$$
$$dj(x)=x^{1/p}-1\tag{26.385c}$$

Pour les dérivées, nous avons

$$so'(x)=\frac{1}{p}x^{1/p-1}\tag{26.386a}$$
$$di'(x)=-so'(x)\tag{26.386b}$$
$$dj'(x)=so'(x).\tag{26.386c}$$

En divise les cas selon $a<1$ ou $a>1$.

Pour $a<1$ Nous avons

$$\eta(a)=so(a)^p+di(a)^p,\tag{26.387}$$

et la première dérivée donne :

$$\eta'(a)=p\,so'(a)\big(so(a)^{p-1}-di(a)^{p-1}\big).\tag{26.388}$$

Pour la seconde dérivée nous trouvons d'abord

$$\begin{aligned}\eta''(a)=&\left(\frac{1-p}{p}\right)a^{\frac{1}{p}-2}\big(so(a)^{p-1}-di(a)^{p-1}\big)\\ &+\frac{p-1}{p}a^{\frac{2}{p}-2}\big(so(a)^{p-2}+di(a)^{p-2}\big).\end{aligned}\tag{26.389}$$

À partir de là, le truc est de substituer les expressions suivantes :

$$so(a)^{p-1}=so(a)^{p-2}so(a)=so(a)^{p-2}+so(a)^{p-2}a^{1/p}\tag{26.390a}$$
$$di(a)^{p-1}=di(a)^{p-2}-x^{1/p}di(a)^{p-2}.\tag{26.390b}$$

Plein de trucs se simplifient et nous obtenons

$$\eta''(a)=\frac{p-1}{p}a^{\frac{1}{p}-2}\big(di(a)^{p-1}-so(a)^{p-2}\big).\tag{26.391}$$

Pour $a>1$ Les calculs sont essentiellement les mêmes, en partant de

$$\eta(a)=so(a)^p+dj(a)^p.\tag{26.392}$$

Les résultats sont :

$$\eta'(a)=p\,so'(a)\big(so(a)^{p-1}+dj(a)^{p-1}\big),\tag{26.393}$$

et

$$\eta''(a)=\frac{p-1}{p}a^{\frac{1}{p}-2}\big(dj(a)^{p-2}-so(a)^{p-2}\big).\tag{26.394}$$

Au final, nous avons pour tout $a \neq 1$:

$$\eta''(a) = \frac{p-1}{p} a^{\frac{1}{p}-2} \big(|1 - a^{1/p}|^{p-2} - (1 + a^{1/p})^{p-2} \big). \tag{26.395}$$

Ce qu'il se passe en $a = 1$ est encore une question ouvert que nous traitons maintenant.

Pour $a = 1$ Les limites des expressions (26.388) et (26.393) en $a = 1$ sont vite calculées et c'est 2^{p-1} dans les deux cas. Donc la dérivée admet une prolongation continue en $a = 1$. Nous allons prouver que la fonction η est en réalité dérivable en $a = 1$ et que la dérivée vaut 2^{p-1}.

Nous nous concentrons sur la partie difficile donnée par $f(x) = |x^{1/p} - 1|^p$. Elle est donnée par

$$f(x) = \begin{cases} di(x)^p & \text{si } x < 1 \\ dj(x)^p & \text{si } x > 1 \\ 0 & \text{si } x = 1. \end{cases} \tag{26.396}$$

Si $f'(1)$ existe, alors elle est égale à la limite

$$f'(1) = \lim_{\epsilon \to 0} \frac{f(1) - f(1-\epsilon)}{\epsilon}. \tag{26.397}$$

Les deux limites à calculer sont :

$$\lim_{\epsilon \to 0^+} \frac{\big((1+\epsilon)^{1/p} - 1\big)^p}{\epsilon} \tag{26.398}$$

et

$$\lim_{\epsilon \to 0^-} \frac{\big(1 - (1+\epsilon)^{1/p}\big)^p}{\epsilon}. \tag{26.399}$$

La première se traite par la règle de l'Hospital [38], et le résultat est zéro. Pour la seconde, il faut juste transformer

$$\lim_{\epsilon \to 0^+} \frac{\big((1+\epsilon)^{1/p} - 1\big)^p}{\epsilon} = \lim_{h \to 0^+} \frac{\big(1 - (1-h)^{1/p}\big)^p}{-h}, \tag{26.400}$$

qui se traite également par la règle de l'Hospital. Le résultat est également zéro.

Donc η est dérivable en $a = 1$ et la dérivée vaut $\eta'(1) = 2^{p-1}$.

En récapitulant, nous avons $\eta'' > 0$ sur $]0, \infty[\setminus\{0\}$, donc η' est croissante sur cette partie (proposition 11.125). Vu que η' est continue sur $]0, \infty[$, elle est même croissante (strictement) sur tout $]0, \infty[$.

La proposition 16.81 conclu que η est strictement convexe sur $]0, \infty[$. □

Toujours dans la catégorie des lemmes pour les inégalités de Hanner, nous avons celui-ci.

Lemme 26.119 ([459]).
Soit $1 < p < 2$. La fonction

$$\begin{aligned} \xi \colon \mathbb{R}^+ \times \mathbb{R}^+ &\to \mathbb{R} \\ (a, b) &\mapsto \big(a^{1/p} + b^{1/p}\big)^p + |a^{1/p} - b^{1/p}|^p \end{aligned} \tag{26.401}$$

est convexe.

Pour rappel, les conventions de données en 1.137 donnent $\mathbb{R}^+ = [0, \infty[$.

Démonstration. La fonction ξ vérifie facilement les conditions suivante :

— $\xi(a, b) = \xi(b, a)$,

— $\xi(0, 0) = 0$,

38. Proposition 11.132

— $\xi(ta, tb) = t\xi(a, b)$ pour tout $t \geqslant 0$.

Nous posons

$$\begin{aligned} \eta\colon \mathbb{R}^+ &\to \mathbb{R} \\ a &\mapsto \xi(a, 1). \end{aligned} \tag{26.402}$$

Le lemme 26.118 dit que η est strictement convexe, et le lemme 16.91 conclu que ξ est convexe. □

Lemme 26.120 ([460]).
Soit $1 < p < 2$. Nous considérons les fonctions

$$\begin{aligned} \alpha\colon [0,1] &\to \mathbb{R} \\ x &\mapsto (1+x)^{p-1} + (1-x)^{p-1} \end{aligned} \tag{26.403}$$

et

$$\begin{aligned} \beta\colon [0,1] &\to \mathbb{R} \\ x &\mapsto x^{1-p}\big((1+x)^{p-1} - (1-x)^{p-1}\big). \end{aligned} \tag{26.404}$$

Soient $A, B \in \mathbb{R}$. Nous avons

$$\alpha(x)|A|^p + \beta(x)|B|^p \leqslant |A+B|^p + |A-B|^p. \tag{26.405}$$

Démonstration. Plusieurs étapes.

$\underline{\beta(x) \leqslant \alpha(x)}$ Nous avons $\alpha(1) = \beta(1) = 2^{p-1}$. Pour les autres valeurs de x, nous allons raisonner avec la dérivée. La valeur de $\alpha'(x)$ est facile à calculer

$$\alpha'(x) = (p-1)(x+1)^{p-2} - (p-1)(1-x)^{p-2}. \tag{26.406}$$

Pour $\beta'(x)$ c'est un peu plus lourd. En substituant $(1+x)^{p-1} = (1+x)^{p-2}(1+x)$ et $(1-x)^{p-1} = (1-x)^{p-2}(1-x)$ nous pouvons regrouper les termes en $(1+x)^{p-2}$ et $(1-x)^{p-2}$. Après un peu de travail,

$$\beta'(x) = \frac{p-1}{x^p}\big((1-x)^{p-2} - (1+x)^{p-2}\big). \tag{26.407}$$

Cela nous permet de calculer $\alpha' - \beta'$:

$$\alpha'(x) - \beta'(x) = (p-1)\big(1 + \frac{1}{x^p}\big)\big((1+x)^{p-2} - (1-x)^{p-2}\big). \tag{26.408}$$

Vu que $1 < p < 2$, le nombre $p-2$ est strictement négatif ; afin de travailler avec des exposants positifs, nous écrivons

$$\alpha'(x) - \beta'(x) = \underbrace{(p-1)}_{>0}\underbrace{\big(1 + \frac{1}{x^p}\big)}_{>0}\underbrace{\left(\frac{1}{(1+x)^{2-p}} - \frac{1}{(1-x)^{2-p}}\right)}_{<0}. \tag{26.409}$$

Nous avons $\alpha'(x) - \beta'(x) < 0$ pour tout $x \in]0,1]$. Du fait qu'en plus nous ayons $\alpha(1) = \beta(1)$, nous déduisons que $\alpha(x) \geqslant \beta(x)$.

Une petite étude de fonction Soit $R \in [0,1]$. Nous considérons la fonction

$$\begin{aligned} F\colon [0,1]\mathbb{R} &\to \mathbb{R} \\ x &\mapsto \alpha(x) + R^p\beta(x). \end{aligned} \tag{26.410}$$

Nous montrons maintenant que cette fonction a un maximum global pour $x = R$. D'abord sa dérivée :

$$F'(x) = \underbrace{(p-1)}_{>0}\underbrace{\big((1-x)^{p-1} - (1+x)^{p-2}\big)}_{<0}\left(1 - \left(\frac{R}{x}\right)^p\right) \tag{26.411}$$

Nous avons

— $F'(x) = 0$ pour $x = R$,
— $F'(x) < 0$ pour $x > R$,
— $F'(x) > 0$ pour $x < R$.

Donc $x = R$ est bien un maximum global.

Pause Nous avons les petits résultats utiles pour commencer à prouver. Petite pause avant de commencer ; pas de panique, ça ne va pas être trop violent.

Pour $0 < B < A$ Nous devons prouver que

$$\alpha(x)A^p + \beta(x)B^p \leqslant (A+B)^p + (A-B)^p. \tag{26.412}$$

En divisant par A^p et en posant $R = B/A$, l'inéquation (26.412) est équivalente à

$$\alpha(x) + \beta(x)R^p \leqslant (1+R)^p + (1-R)^p \tag{26.413}$$

où $R \in]0,1[$ parce que nous avons supposé $0 < B < A$. Nous avons (il y a un petit calcul pour $F(R)$)

$$(1+R)^p + (1-R)^p = F(R) \geqslant F(x) = \alpha(x) + \beta(x)R^p. \tag{26.414}$$

ok.

Pour $0 < A < B$ Lorsque $0 < A < B$ nous avons

$$\begin{aligned}
\alpha(x)|A|^p + \beta(x)|B|^p &= \alpha(x)A^p + \beta(x)^p && (26.415a)\\
&\leqslant \alpha(x)B^p + \beta(x)A^p && (26.415b)\\
&\leqslant (B+A)^p + (B-A)^p && (26.415c)\\
&= |A+B|^p + |A-B|^p. && (26.415d)
\end{aligned}$$

Justification :

— Pour (26.415b), c'est parce que $\alpha(x) > \beta(x)$; alors en mettant le plus grand de A et B devant le α au lieu du β, nous majorons.
— Pour (26.415c), c'est l'inégalité dans le cas $0 < B < A$, mais en inversant les noms de A et B.

Pour $0 < A = B$ Toutes les expressions sont continues par rapport à B (fixons x et A). Nous avons prouvé pour $B < A$ et pour $B > A$. Par continuité, l'inégalité est encore valide pour $A = B$.

Pour $A < 0$, $B > 0$ En posant $A' = -A$ nous avons $A' > 0$ et nous pouvons écrire

$$|A+B|^p + |A-B|^p = |-A'+B|^p + |-A'-B|^p = |B-A'|^p + |B+A'|^p \geqslant \alpha(x)|A'|^p + \beta(x)|B|^p. \tag{26.416}$$

Nous avons utilisé, avec A' et B le cas déjà prouvé $A', B > 0$.

Pour $A > 0$, $B < 0$ Celui-là, je vous le laisse.

Pour $A < 0$, $B < 0$ Posez $A' = -A$ et $B' = -B$ et hop.

□

Théorème 26.121 (Inégalités de Hanner[459, 460]).
Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$. Soit $1 < p < 2$ and $f, g \in L^p(\Omega, \mathcal{A}, \mu)$; nous avons

$$\big(\|f\|_p + \|g\|_p\big)^p + \Big|\|f\|_p - \|g\|_p\Big|^p \leqslant \|f+g\|_p^p + \|f-g\|_p^p \tag{26.417}$$

Il y a égalité si et seulement si $f(t)$ et $g(t)$ sont colinéaires pour presque tout t.

Démonstration. Nous supposons que $\|f\|_p \geqslant \|g\|_p$ pour fixer les idées. De toutes façons, la symétrie des formules nous fait passer de ce cas à l'autre sans difficulté.

Soit $x \in [0,1]$. Nous écrivons l'inégalité du lemme 26.120 pour $A = |f(\omega)|$ et $B = |g(\omega)|$:

$$\alpha(x)|f(\omega)|^p + \beta(x)|g(\omega)|^p \leqslant \big|f(\omega) + g(\omega)\big|^p + \big|f(\omega) - g(\omega)\big|^p. \tag{26.418}$$

Nous intégrons cela par rapport à ω sur Ω :

$$\alpha(x)\|f\|_p^p + \beta(x)\|g\|_p^p \leqslant \|f+g\|_p^p + \|f-g\|_p^p. \tag{26.419}$$

Et là vient l'idée qu'on se demande ce qui est passé par l'esprit du mec qui a tout combiné : nous évaluons cela pour $x = \frac{\|g\|_p}{\|f\|_p}$, ce qui est permis parce que nous avons supposé $\|f\|_p \geqslant \|g\|_p$. Faites le calcul, collectez les termes identiques, vous obtiendrez

$$\big(\|f\|_p + \|g\|_p\big)^p + \big(\|f\|_p - \|g\|_p\big)^p \leqslant \|f+g\|_p^p + \|f-g\|_p^p. \tag{26.420}$$

Et vu que $\|f\|_p \geqslant \|g\|_p$, nous pouvons gratuitement faire

$$\|f\|_p - \|g\|_p = \big|\|f\|_p - \|g\|_p\big|. \tag{26.421}$$

Fini pour Hanner. □

26.10.3 Inégalités de Clarkson

Lemme 26.122 ([461]).
Si $p \geqslant 2$ et si $a, b \in \mathbb{C}$, alors

$$\left|\frac{a+b}{2}\right|^p + \left|\frac{a-b}{2}\right|^p \leqslant \frac{1}{2}\big(|a|^p + |b|^p\big). \tag{26.422}$$

Démonstration. Nous prouvons l'inégalité en prenant montent petit à petit en généralité.

Avec $x > 0$ Soit $x \geqslant 0$. Nous montrons dans cette partie l'inégalité

$$x^p + 1 \leqslant (x+1)^{p/2}. \tag{26.423}$$

Pour cela nous considérons la fonction

$$\begin{aligned} f\colon [0,\infty[&\to \mathbb{R} \\ t &\mapsto (t^2+1)^{p/2} - t^p - 1. \end{aligned} \tag{26.424}$$

Nous avons $f(0) = 0$, mais aussi, en utilisant les règle de dérivation [39] nous trouvons vite

$$f'(t) = p(t^2+1)^{p/2-1}t - pt^{p-1}. \tag{26.425}$$

Vu que $(t^2+1)^{p/2-1} \geqslant t^{p-2}$, le signe de $f'(t)$ est toujours strictement positif pour $t > 0$. La proposition 11.125 fait que f est strictement croissante et que $f(t) > 0$ pour tout $t > 0$.

Avec $x, y \geqslant 0$ Soient $x, y \geqslant 0$ dans $\mathbb{R}$. Nous prouvons dans cette partie que

$$(x^2+y^2)^{p/2} \geqslant x^p + y^p. \tag{26.426}$$

Il s'agit d'appliquer l'inégalité (26.423) à x/y :

$$\left(\left(\frac{x}{y}\right)^2 + 1\right)^{p/2} \geqslant \left(\frac{x}{y}\right)^p + 1. \tag{26.427}$$

En multipliant par y^p et en simplifiant un peu, nous trouvons le résultat (26.426).

39. Par exemple celle de la proposition 13.245.

Avec $a, b \in \mathbb{C}$ Nous appliquons l'inégalité (26.426) à $x = |\frac{a+b}{2}|$ et $y = |\frac{a-b}{2}|$. Cela donne :

$$\left|\frac{a+b}{2}\right|^p + \left|\frac{a-b}{2}\right|^p \leqslant \left(\left|\frac{a+b}{2}\right|^2 + \left|\frac{a-b}{2}\right|^2\right)^{p/2} \tag{26.428a}$$

$$= \left(\frac{2|a|^2 + 2|b|^2}{4}\right)^{p/2} \tag{26.428b}$$

$$= \frac{1}{2}|a|^p + \frac{1}{2}|b|^p. \tag{26.428c}$$

□

Lemme 26.123 ([462]).
Soient $a, b \in \mathbb{C}$ ainsi que $1 < p < 2$. Nous notons q l'exposant conjugué de p. Nous avons l'inégalité

$$|a+b|^q + |a-b|^q \leqslant 2\big(|a|^p + |b|^p\big)^{q-1}. \tag{26.429}$$

Démonstration. Soient $a, b \in \mathbb{C}$; nous nommons $|a|$ le module du nombre complexe a. Nous avons :

$$|a+b|^q + |a-b|^q = \left\|\begin{pmatrix}|a+b|\\|a-b|\end{pmatrix}\right\|_q^q \tag{26.430a}$$

$$\leqslant \left(2^{\frac{1}{q}-\frac{1}{2}}\left\|\begin{pmatrix}|a+b|\\|a-b|\end{pmatrix}\right\|_2\right)^q \tag{26.430b}$$

$$= 2^{1-q/2}\left\|\begin{pmatrix}|a+b|\\|a-b|\end{pmatrix}\right\|_2^q \tag{26.430c}$$

$$= 2^{1-q/2}\big(|a|^2 + |b|^2\big)^{q/2} \tag{26.430d}$$

$$= 2\left\|\begin{pmatrix}|a|\\|b|\end{pmatrix}\right\|_2^q \tag{26.430e}$$

$$\leqslant 2\left\|\begin{pmatrix}|a|\\|b|\end{pmatrix}\right\|_p^q \tag{26.430f}$$

$$= 2\big(|a|^p + |b|^p\big)^{q/p} \tag{26.430g}$$

$$= 2\big(|a|^p + |b|^p\big)^{q-1}. \tag{26.430h}$$

Quelques justifications.

— Pour (26.430b). C'est Hölder de la proposition 16.106.

— Pour (26.430d). C'est le calcul suivant, basé sur le fait que $|z|^2 = z\bar{z}$: Un petit calcul pour la norme 2 :

$$\left\|\begin{pmatrix}|a+b|\\|a-b|\end{pmatrix}\right\|_2^2 = |a+b| + |a-b|^2 \tag{26.431a}$$

$$= |a|^2 + a\bar{b} + \bar{a}b + |b|^2 \tag{26.431b}$$

$$+ |a|^2 - a\bar{b} - \bar{a}b + |b|^2 \tag{26.431c}$$

$$= 2|a|^2 + 2|b|^2. \tag{26.431d}$$

— Pour (26.430f). Du fait que $p < 2 < q$, la proposition 16.106 donne $\|x\|_p \geqslant \|x\|_2$.

— Pour (26.430h). Simplement multiplier par q l'équation $\frac{1}{p} + \frac{1}{q} = 1$ fournit $\frac{q}{p} = q - 1$.

□

Proposition 26.124 (Inégalité de Clarkson[458]).
Soient $f, g \in L^p(\Omega, \mathcal{A}, \mu)$.

(1) Si $p \geqslant 2$, alors

$$\|\frac{f+g}{2}\|_p^p + \|\frac{f-g}{2}\|_p^p \leqslant \frac{1}{2}\Big(\|f\|_p^p + \|g\|_p^p\Big). \tag{26.432}$$

(2) Si $1 < p < 2$ et si q est l'exposant conjugué de p, alors

$$\|f+g\|_p^q + \|f-g\|_p^q \leqslant 2\Big(\|f\|_p^p + \|g\|_p^p\Big)^{q-1}, \tag{26.433}$$

ou

$$\|\frac{f+g}{2}\|_p^q + \|\frac{f-g}{2}\|_p^q \leqslant 2^{1-q}\big(\|f\|_p^p + \|g\|_p^p\big)^{q-1}. \tag{26.434}$$

Démonstration. En deux parties.

Pour $p \geqslant 2$ Soient $f, g \in L^p(\Omega, \mathcal{A}, \mu)$; ce sont des fonctions à valeurs dans $\mathbb{C}$. Pour chaque $\omega \in \Omega$ nous considérons les nombres complexes $f(\omega)$ et $g(\omega)$; nous pouvons écrire l'inégalité du lemme 26.122 :

$$\Big|\frac{f(\omega)+g(\omega)}{2}\Big|^p + \Big|\frac{f(\omega)-g(\omega)}{2}\Big|^p \leqslant \frac{1}{2}\big(|f(\omega)|^p + |g(\omega)|^p\big). \tag{26.435}$$

Nous avons les substitutions évidentes $f(\omega)+g(\omega) = (f+g)(\omega)$ et $f(\omega)-g(\omega) = (f-g)(\omega)$. En intégrant alors (26.435) sur Ω nous trouvons l'inégalité demandée.

Pour $1 < p < 2$ Il s'agit de faire la même chose, en utilisant l'inégalité de Clarkson du lemme 26.123.

Pour obtenir (26.434), il s'agit simplement de multiplier et diviser le member de gauche de (26.433) par 2^q.

□

26.10.4 Uniforme convexité des espaces de Lebesgue

Proposition 26.125 ([463]).
Si $1 < p < \infty$, l'espace $L^p(\Omega, \mathcal{A}, \mu)$ est uniformément convexe[40].

Démonstration. En deux parties.

$1 < p \leqslant 2$ Nous montrons que la fonction $\delta(\epsilon) = 2^{-q}\epsilon^q$ fonctionne.

Soient $f, g \in L^p$ telles que $\|f\|_p \leqslant \|g\|_p \leqslant 1$ et $\|f-g\|_p \geqslant \epsilon$. Nous commençons par écrire l'inégalité de Clarkson (26.433) :

$$\|\frac{f+g}{2}\|_p^q + \|\frac{f-g}{2}\|_p^q \leqslant 2^{1-q}\big(\|f\|_p^p + \|g\|_p^p\big)^{q-1}. \tag{26.436}$$

Par hypothèse, $\|f\|_p$ et $\|g\|_p$ sont plus petites que 1. Vu que $p > 1$, nous avons

$$\|f\|_p^p + \|g\|_p^p \leqslant 1 + 1 = 2. \tag{26.437}$$

En remplaçant dans le membre de droite de (26.436) nous avons

$$\|\frac{f+g}{2}\|_p^q + \|\frac{f-g}{2}\|_p^q \leqslant 2^{1-q}2^{q-1} = 1, \tag{26.438}$$

et donc

$$\|\frac{f+g}{2}\|_p^q \leqslant 1 - \|\frac{f-g}{2}\|_p^q. \tag{26.439}$$

Par ailleurs nous avons supposé $\|f-g\|_p \geqslant 1$. Donc aussi [41]

$$\|\frac{f-g}{2}\|_p^q \geqslant 2^{-q}\epsilon^q. \tag{26.440}$$

40. Définition 26.111.
41. Ici j'ai un coefficient un peu différent que celui de [463]. Écrivez-moi pour confirmer ou infirmer mes calculs.

Et par un autre ailleurs,

$$\|\frac{f+g}{2}\|_p = \frac{1}{2}\|f+g\|_p \leqslant \frac{1}{2}\big(\|f\|_p + \|g\|_p\big) \leqslant 1. \tag{26.441}$$

Vu que nous avons $q \geqslant 2$, cela donne aussi

$$\|\frac{f+g}{2}\|_p \leqslant \|\frac{f+g}{2}\|^q. \tag{26.442}$$

Avec les inégalités (26.440) et 26.442 nous finissons l'inégalité (26.439) :

$$\|\frac{f+g}{2}\|_p \leqslant \|\frac{f+g}{2}\|_p^q \leqslant 1 - 2^{-q}\epsilon^q \leqslant \|g\|_p - \delta(\epsilon). \tag{26.443}$$

Okay, c'est bon.

$2 \leqslant p < \infty$ Il s'agit de faire la même chose en partant de Clarkson (26.432). Le résultat est que la fonction $\delta(\epsilon) = (\sigma/2)^p$, ça fonctionne.

□

26.10.5 Théorème de la projection normale

Proposition 26.126.
Si $1 < p < \infty$, et si V est un sous-espace vectoriel fermé de $L^p(\Omega, \mathcal{A}, \mu)$, alors la projection normale[42] *de $a \in L^p$ sur V existe et est unique.*

Démonstration. La proposition 26.125 nous indique que l'espace $L^p(\Omega, \mathcal{A}, \mu)$ est uniformément convexe. Or le théorème 26.115 nous indique que les espaces uniformément convexes vérifient la présente proposition. □

Nous pouvons donner une preuve directe, sans passer par l'uniforme convexité, dans les cas $p \geqslant 2$.

Théorème 26.127 (Théorème de la projection normale[464])**.**
Nous considérons $p \geqslant 2$. Soit un sous-espace vectoriel fermé $W \subset L^p(\Omega, \mathcal{A}, \mu)$ et $u_0 \in L^p$. Nous notons

$$d(u_0, W) = \inf_{w \in W} d(u_0, W). \tag{26.444}$$

Alors il existe $w_0 \in W$ tel que $\|u_0 - w_0\| = d(u_0, W)$.

Démonstration. Nous allons séparer trois cas : $p = 2$ et $p > 2$.

$p = 2$ Pour $p = 2$, nous savons que L^2 est un espace de Hilbert[43], et nous avons déjà le théorème de la projection 24.5.

$p > 2$ Pour chaque $x \in \Omega$ nous avons $f(x), g(x) \in \mathbb{C}$ et donc l'identité du parallélogramme[44] :

$$\big|f(x) - g(x)\big|^2 + \big|f(x) + g(x)\big| = 2|f(x)| + 2|g(x)|^2. \tag{26.445}$$

Vu que $p > 2$, la fonction $s\colon x \mapsto x^{p/2}$ est convexe (lemme 16.85). Calcul :

$$\begin{aligned}
|f(x) - g(x)|^p + |f(x) + g(x)|^p &= \big(|f(x) - g(x)|^2\big)^{p/2} + \big(|f(x) + g(x)|^2\big)^{p/2} && \text{(26.446a)}\\
&= s\big(|\dots|^2\big) + s\big(|\dots|^2\big) && \text{(26.446b)}\\
&\leqslant \big(|f(x) - g(x)|^2 + |f(x) + g(x)|^2\big)^{p/2} && \text{(26.446c)}\\
&= \big(2|f(x)|^2 + 2|g(x)|^2\big)^{p/2} && \text{(26.446d)}\\
&= 2^{p/2}\big(|f(x)|^2 + |g(x)|^2\big)^{p/2} && \text{(26.446e)}\\
&\leqslant 2^{p/2}2^{p/2-1}\big(|f(x)|^p + |g(x)|^p\big) && \text{(26.446f)}\\
&= 2^{p-1}\big(|f(x)|^p + |g(x)|^p\big) && \text{(26.446g)}
\end{aligned}$$

42. Définition 26.113.
43. Lemme 26.73.
44. Proposition 7.192 en remarquant que $(z_1, z_2) \mapsto z_1\bar{z}_2$ est un produit scalaire hermitien sur $\mathbb{C}$.

Justifications :

— Pour (26.446c) : la convexité de s.

— Pour (26.446d) : la relation (26.445).

— Pour (26.446f) : la seconde inégalité du lemme 26.446f.

Nous isolons $|f(x)-g(x)|^p$:

$$|f(x)-g(x)|^p \leqslant 2^{p-1}\big(|f(x)|^p+|g(x)|^p\big)-|f(x)+g(x)|^p \tag{26.447a}$$

$$= 2^p\left(\frac{|f(x)|^p+|g(x)|^p}{2}-\left|\frac{|f(x)|+|g(x)|}{2}\right|^p\right) \tag{26.447b}$$

Cette inégalité étant valable pour tout x, nous pouvons intégrer sur Ω et découper l'intégrale en petits morceaux :

$$\|f-g\|_p^p \leqslant 2^p\left(\frac{\|f\|_p^p+\|g\|_p^p}{2}-\|\frac{f+g}{2}\|_p^p\right). \tag{26.448}$$

Voila une bonne chose de prouvée. Nous pouvons maintenant passer au vif du sujet.

Soit une suite w_j dans W telle que $\|u_0-w_j\|\to d(u_0,W)$. Trois choses à savoir sur cette suite :

(1) Une telle suite existe parce que $d(u_0,W)$ est défini comme un infimum.

(2) Rien ne garanti qu'elle converge.

(3) Même si elle convergeait, rien ne garantirait que la limite soit encore dans W.

Le troisième point est facile à régler : vu que W est fermé par hypothèse, une suite convergente contenue dans W a sa limite dans W. Nous allons régler la convergence de w_j en prouvant qu'elle est de Cauchy.

Remarquons que W est vectoriel, donc $(w_j+w_k)/2$ est dans W pour tout j et k ; donc

$$\|\frac{w_j+w_k}{2}-u_0\| \geqslant d(u_0,W). \tag{26.449}$$

En tenant compte de cela, nous écrivons l'inégalité (26.448) avec $f=w_j-u_0$ et $g=w_k-u_0$:

$$\|f-g\|_p^p=\|w_j-w_k\|_p^p \leqslant 2^p\left(\frac{\|w_j-u_0\|^p+\|w_k-u_0\|^p}{2}-d(u_0,W)\right). \tag{26.450}$$

Soit $\epsilon>0$ et $0<\epsilon_1,\epsilon_2<\epsilon$ tels que $\epsilon_1+\epsilon_2<\epsilon$. Il existe un N tel que si $j,k>N$ alors $\|w_j-u_0\|^p\leqslant d(u_0,W)^p+\epsilon_1$ et $\|w_k-u_0\|^p\leqslant d(u_0,W)^p+\epsilon_2$. Pour de telles valeurs de j et k, nous avons

$$\|w_j-w_k\|_p \leqslant 2\left(\frac{\epsilon_1+\epsilon_2}{2}\right)<2\epsilon^{1/p}. \tag{26.451}$$

Donc la suite (w_j) est de Cauchy.

L'espace L^p étant complet par le théorème 26.40, nous en déduisons que (w_j) converge dans L^p. Mais comme W est fermé, nous avons $w_j\xrightarrow{L^p}w\in W$.

En termes de normes, nous avons

$$\|w-u_0\|=\lim_j\|w_j-u_0\|=d(W,u_0). \tag{26.452}$$

□

26.11 Dualité et théorème de représentation de Riesz

Dans la suite E' est le dual topologique, c'est-à-dire l'espace des formes linéaires et continues sur E.

Voici déjà un bel énoncé. Pour des espaces mesurés $(\Omega, \mathcal{A}, \mu)$ plus généraux, voir l'arme totale en el théorème 26.130.

Proposition 26.128 ([202], thème 34)**.**
Soit $1 < p < 2$ et q tel que $\frac{1}{p} + \frac{1}{q} = 1$. L'application

$$\begin{aligned} \Phi\colon L^q\big([0,1]\big) &\to L^p\big([0,1]\big)' \\ \Phi_g(f) &= \int_{[0,1]} f\bar{g}. \end{aligned} \tag{26.453}$$

est une isométrie linéaire surjective.

Démonstration. Pour la simplicité des notations nous allons noter L^2 pour $L^2\big([0,1]\big)$, et pareillement pour L^p.

Φ_g est un élément de $(L^p)'$ Si $f \in L^p$ et $g \in L^q$ nous devons prouver que $\Phi_q(f)$ est bien définie. Pour cela nous utilisons l'inégalité de Hölder[45] qui dit que $fg \in L^1$; par conséquent la fonction $f\bar{g}$ est également dans L^1 et nous avons

$$|\Phi_g(f)| \leqslant \int_{[0,1]} |f\bar{g}| = \|fg\|_1 \leqslant \|f\|_p \|g\|_q. \tag{26.454}$$

En ce qui concerne la norme de l'application Φ_g nous avons tout de suite

$$\|\Phi_g\| = \sup_{\|f\|_p=1} \big|\Phi_g(f)\big| \leqslant \|g\|_q. \tag{26.455}$$

Cela signifie que l'application Φ_g est bornée et donc continue par la proposition 10.18. Nous avons donc bien $\Phi_g \in (L^p)'$.

Isométrie Afin de prouver que $\|\Phi_g\| = \|g\|_q$ nous allons trouver une fonction $f \in L^p$ telle que $\frac{|\Phi_g(f)|}{\|f\|_p} = \|g\|_q$. De cette façon nous aurons prouvé que $|\Phi_g| \geqslant \|g\|_q$, ce qui conclurait que $|\Phi_g| = \|g\|_q$.

Nous posons $f = g|g|^{q-2}$, de telle sorte que $|f| = |g|^{q-1}$ et

$$\|f\|_p = \left(\int |g|^{p(q-1)}\right)^{1/p} = \left(\int |g|^q\right)^{1/p} = \|g\|_q^{q/p} \tag{26.456}$$

où nous avons utilisé le fait que $p(q-1) = q$. La fonction f est donc bien dans L^p. D'autre part,

$$\Phi_g(f) = \int f\bar{g} = \int g|g|^{q-2}\bar{g} = \int |g|^q = \|g\|_q^q. \tag{26.457}$$

Donc

$$\frac{|\Phi_g(f)|}{\|f\|_p} = \|g\|_q^{q-\frac{q}{p}} = \|g\|_q \tag{26.458}$$

où nous avons encore utilisé le fait que $q - \frac{q}{p} = \frac{q(p-1)}{p} = 1$.

Surjectif Soit $\ell \in (L^p)'$; c'est une application $\ell\colon L^p \to \mathbb{C}$ dont nous pouvons prendre la restriction à L^2 parce que la proposition 26.38 nous indique que $L^2 \subset L^p$. Nous nommons $\phi\colon L^2 \to \mathbb{C}$ cette restriction.

45. Proposition 26.33.

$\phi \in (L^2)'$ Nous devons montrer que ϕ est continue pour la norme sur L^2. Pour cela nous montrons que sa norme opérateur (subordonnée à la norme de L^2 et non de L^p) est finie :

$$\sup_{f\in L^2} \frac{|\phi(f)|}{\|f\|_2} \leqslant \sup_{f\in L^2} \frac{|\ell(f)|}{\|f\|_p} < \infty. \tag{26.459}$$

Nous avons utilisé l'inégalité de norme $\|f\|_p \leqslant \|f\|_2$ de la proposition 26.38(2).

Utilisation du dual de L^2 Étant donné que L^2 est un espace de Hilbert (lemme 26.73) et que $\phi \in (L^2)'$, le théorème 24.17 nous donne un élément $g \in L^2$ tel que $\phi(f) = \Phi_g(f)$ pour tout $f \in L^2$.

Nous devons prouver que $g \in L^q$ et que pour tout $f \in L^p$ nous avons $\ell(f) = \Phi_g(f)$.

$g \in L^q$ Nous posons $f_n = g|g|^{q-2}\mathbb{1}_{|g|<n}$. Nous avons d'une part

$$\Phi_g(f_n) = \int_0^1 f_n \bar{g} = \int_{|g|<n} |g|^q. \tag{26.460}$$

Et d'autre part comme $f_n \in L^2$ nous avons aussi $\phi(f_n) = \Phi_g(f_n)$ et donc

$$0 \leqslant \Phi(f_n) = \phi(f_n) \leqslant \|\ell\|\|f_n\|_p \tag{26.461a}$$

$$= \|\ell\| \left(\int_{|g|<n} |g|^{(q-1)p} \right)^{1/p} \tag{26.461b}$$

$$= \|\ell\| \left(\int_{|g|<n} |g|^q \right)^{1/p}. \tag{26.461c}$$

où nous avons à nouveau tenu compte du fait que $p(q-1) = q$. En combinant avec (26.460) nous trouvons

$$\int_{|g|<n} |g|^q \leqslant \|\ell\| \left(\int_{|g|<n} |g|^q \right)^{1/p}, \tag{26.462}$$

et donc

$$\left(\int_{|g|<n} |g|^q \right)^{1-\frac{1}{p}} \leqslant \|\ell\|, \tag{26.463}$$

c'est-à-dire

$$\left(\int_{|g|<n} |g|^q \right)^{1/q} \leqslant \|\ell\|. \tag{26.464}$$

Si ce n'était pas encore fait nous nous fixons un représentant de la classe g (qui est dans L^2), et nous nommons également g ce représentant. Nous posons alors

$$g_n = |g|^q \mathbb{1}_{|g|<n} \tag{26.465}$$

qui est une suite croissante de fonctions convergeant ponctuellement vers $|g|^q$. Le théorème de Beppo-Levi 13.160 nous permet alors d'écrire

$$\lim_{n\to\infty} \int_{|q|<n} |g|^q = \int_0^1 |g|^q. \tag{26.466}$$

Mais comme pour chaque n nous avons $\int_{|g|<n} |q|^q \leqslant \|\ell\|^q$, nous conservons l'inégalité à la limite et

$$\int_0^1 |g|^q \leqslant \|\ell\|^q. \tag{26.467}$$

Cela prouve que $g \in L^p$.

$\ell(f) = \Phi_g(f)$ Soit $f \in L^p$. En vertu de la densité de L^2 dans L^p prouvée dans le corolaire 26.48 nous pouvons considérer une suite (f_n) dans L^2 telle que $f_n \xrightarrow{L^p} f$. Pour tout n nous avons

$$\ell(f_n) = \Phi_g(f_n). \tag{26.468}$$

Mais ℓ et Φ_g étant continues sur L^p nous pouvons prendre la limite et obtenir

$$\ell(f) = \Phi_g(f). \tag{26.469}$$

□

Lemme 26.129 ([6]).
Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré fini. Soit $g \in L^1(\Omega)$ et S fermé dans $\mathbb{C}$. Si pour tout $E \in \mathcal{A}$ nous avons

$$\frac{1}{\mu(E)} \int_E g d\mu \in S, \tag{26.470}$$

alors $g(x) \in S$ pour presque tout $x \in \Omega$.

Démonstration. Soit $D = \overline{B(a,r)}$ un disque fermé dans le complémentaire de S (ce dernier étant fermé, le complémentaire est ouvert). Posons $E = g^{-1}(D)$. Prouvons que $\mu(E) = 0$ parce que cela prouverait que $g(x) \in D$ pour seulement un ensemble de mesure nulle. Mais S^c pouvant être écrit comme une union dénombrable de disques fermés [46], nous aurions $g(x) \in S^c$ presque nulle part.

Vu que $\frac{1}{\mu(E)} \int_E a = a$ nous avons

$$\Big|\frac{1}{\mu(E)} g d\mu - a\Big| = \Big|\frac{1}{\mu(E)} \int_E (g-a)\Big| \leqslant \frac{1}{\mu(E)} \int_E |g-a| \leqslant \frac{1}{\mu(E)} \mu(E) r = r. \tag{26.471a}$$

Donc

$$\frac{1}{\mu(E)} \int_E g d\mu \in D, \tag{26.472}$$

ce qui est une contradiction avec le fait que $D \subset S^c$. □

Dans toute la partie d'analyse fonctionnelle, sauf mention du contraire, nous considérons dans L^p des fonctions à valeurs complexes, et donc les éléments du dual sont des applications linéaires continues à valeurs dans $\mathbb{C}$.

Théorème 26.130 (Théorème de représentation de Riesz, thème 34, [6, 296, 465, 284]).
Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$. Soit q tel que $\frac{1}{p} + \frac{1}{q} = 1$ avec la convention que $q = \infty$ si $p = 1$. Alors l'application

$$\begin{aligned} \Phi \colon L^q &\to (L^p)' \\ \Phi_g(f) &= \int_\Omega f \bar{g} d\mu \end{aligned} \tag{26.473}$$

est une bijection isométrique dans les cas suivants :

(1) si $1 < p < \infty$ et $(\Omega, \mathcal{A}, \mu)$ est un espace mesuré quelconque,

(2) si $p = 1$ et $(\Omega, \mathcal{A}, \mu)$ est σ-fini.

Démonstration. Par petits bouts.

Φ est injective Nous commençons par prouver que Φ est injectif. Soient $g, g' \in L^q$ tels que $\Phi_g = \Phi_{g'}$. Alors pour tout $f \in L^p$ nous avons

$$\int_\Omega f(g - g') d\mu = 0. \tag{26.474}$$

46. Tout ouvert peut être écrit comme union dénombrable d'éléments d'une base de topologie par la proposition 7.60 et $\mathbb{C}$ a une base dénombrable de topologie par la proposition 7.97.

Soient des parties A_i de mesures finies telles que $\Omega = \bigcup_{i=1}^{\infty} A_i$. Étant donné que $\mu(A_i)$ est fini, nous avons $\mathbb{1}_{A_i} \in L^p(\Omega)$ et donc

$$\int_{A_i} (g - g')d\mu = \int_{\Omega} \mathbb{1}_{A_i}(x)(g - g')(x)d\mu(x) = 0. \tag{26.475}$$

La proposition 26.22 nous dit alors que $g - g' = 0$ dans $L^q(A_i)$. Pour chaque i, la partie $N_i = \{x \in A_i \text{ tel que } (g - g')(x) = 0\}$ est de mesure nulle.

Vu que Ω est l'union de tous les A_i, la partie de Ω sur laquelle $g - g'$ est non nulle est l'union des N_i et donc de mesure nulle parce que une réunion dénombrable de parties de mesure nulle est de mesure nulle. Donc $g - g' = 0$ presque partout dans Ω, ce qui signifie $g - g' = 0$ dans $L^q(\Omega)$.

La suite La partie difficile est de montrer que Φ est surjective.

Soit $\phi \in L^p(\Omega)'$. Si $\phi = 0$, c'est bien dans l'image de Φ ; nous supposons donc que non. Nous allons commencer par prouver qu'il existe une (classe de) fonction $g \in L^1(\Omega)$ telle que $\Phi_g(f) = \phi(f)$ pour tout $f \in L^\infty(\Omega, \mu)$; nous montrerons ensuite que $g \in L^q$ et que le tout est une isométrie.

Une mesure complexe Si $E \in \mathcal{A}$ nous notons $\nu(E) = \phi(\mathbb{1}_E)$. Nous prouvons maintenant que ν est une mesure complexe[47] sur $(\Omega, \mathcal{A})$. La seule condition pas facile est la condition de dénombrable additive. Il est déjà facile de voir que A et B sont disjoints, $\nu(A \cup B) = \nu(A) + \nu(B)$. Soient ensuite des ensembles A_n deux à deux disjoints et posons $E_k = \bigcup_{i \leqslant k} A_i$ pour avoir $\bigcup_k A_k = \bigcup_k E_k$ avec l'avantage que les E_k soient emboîtés. Cela donne

$$\|\mathbb{1}_E - \mathbb{1}_{E_k}\|_p = \mu(E \backslash E_k)^{1/p}, \tag{26.476}$$

mais vu que $1 \leqslant p < \infty$, avoir $x_k \to 0$ implique d'avoir $x_k^{1/p} \to 0$. Prouvons que $\mu(E \backslash E_k) \to 0$. En vertu du lemme 13.22 nous avons pour chaque k :

$$\mu(E \backslash E_k) = \mu(E) - \mu(E_k), \tag{26.477}$$

et vu que $E_k \to E$ est une suite croissante, le lemme 13.23(1), sachant que μ est une mesure « normale », donne

$$\lim_{n \to \infty} \mu(E_k) = \mu\big(\bigcup_k E_k\big). \tag{26.478}$$

Donc effectivement $\mu(E_k) \to \mu(E)$ et donc oui : $\mu(E \backslash E_k) \to 0$. Jusqu'à présent nous avons

$$\lim_{k \to \infty} \|\mathbb{1}_E - \mathbb{1}_{E_k}\|_p = 0, \tag{26.479}$$

c'est-à-dire $\mathbb{1}_{E_k} \xrightarrow{L^p} \mathbb{1}_E$. La continuité de ϕ sur L^p donne alors

$$\lim_{k \to \infty} \nu(E_k) = \lim_{k \to \infty} \phi(\mathbb{1}_{E_k}) = \phi(\lim_{k \to \infty} \mathbb{1}_{E_k}) = \phi(\mathbb{1}_E) = \nu(E). \tag{26.480}$$

Par additivité finie de ν nous avons

$$\nu(E_k) = \sum_{i \leqslant k} \nu(A_i) \tag{26.481}$$

et en passant à la limite, $\sum_{i=1}^{\infty} \nu(A_i) = \nu(\bigcup_i A_i)$. L'application ν est donc une mesure complexe.

Mesure absolument continue En prime, si $\mu(E) = 0$ alors $\nu(E) = 0$ parce que

$$\mu(E) = 0 \Rightarrow \|\mathbb{1}_E\|_p = 0 \Rightarrow \mathbb{1}_E = 0 \text{ (dans } L^p) \Rightarrow \phi(\mathbb{1}_E) = 0 \tag{26.482}$$

47. Définition 13.201.

Utilisation de Radon-Nikodym Nous sommes donc dans un cas où $\nu \ll \mu$ et nous utilisons le théorème de Radon-Nikodym 13.202 sous la forme de la remarque 13.203 : il existe une fonction intégrable $g\colon \Omega \to \mathbb{C}$ [48] telle que pour tout $A \in \mathcal{A}$,

$$\nu(A) = \int_A \bar{g} d\mu. \tag{26.483}$$

C'est-à-dire que

$$\phi(\mathbb{1}_A) = \int_A \bar{g} d\mu = \int_\Omega \bar{g} \mathbb{1}_A d\mu. \tag{26.484}$$

Nous avons donc exprimé ϕ comme une intégrale pour les fonctions caractéristiques d'ensembles.

Pour les fonctions étagées Par linéarité si f est mesurable et étagée nous avons aussi

$$\phi(f) = \int f\bar{g} d\mu = \Phi_g(f). \tag{26.485}$$

Pour $f \in L^\infty(\Omega)$ Une fonction $f \in L^\infty$ est une fonction presque partout bornée. Nous supposons que f est presque partout bornée par M. Par ailleurs cette f est limite uniforme de fonctions étagées : $\|f_k - f\|_\infty \to 0$ en posant $f_k = f\mathbb{1}_{|f|\leqslant k}$. Pour chaque k nous avons l'égalité

$$\Phi_g(f_k) = \phi(f_k). \tag{26.486}$$

Par ailleurs la fonction $f_k\bar{g}$ est majorée par la fonction intégrable $M\bar{g}$ et le théorème de la convergence dominée 13.184 nous donne

$$\lim_{k\to\infty} \Phi_g(f_i) = \lim_{k\to\infty} \int f_k\bar{g} = \int f\bar{g} = \Phi_g(f). \tag{26.487}$$

Et la continuité de ϕ sur L^p couplée à la convergence $f_k \xrightarrow{L^p} f$ donne $\lim_{k\to\infty} \phi(f_k) = (f)$. Bref prendre la limite dans (26.486) donne

$$\Phi_g(f) = \phi(f) \tag{26.488}$$

pour tout $f \in L^\infty(\Omega)$.

La suite ... Voici les prochaines étapes.

— Nous avons $\int f\bar{g} = \phi(f)$ tant que $f \in L^\infty$. Nous allons étendre cette formule à $f \in L^p$ par densité. Cela terminera de prouver que notre application est une bijection.

— Ensuite nous allons prouver que $\|\phi\| = \|\Phi_g\|$, c'est-à-dire que la bijection est une isométrie.

De L^∞ à L^p Soit $f \in L^p$. Si nous avions une suite (f_n) dans L^∞ telle que $f_n \xrightarrow{L^p} f$ alors $\lim \phi(f_n) = \phi(f)$ par continuité de ϕ. La difficulté est de trouver une telle suite de façon à pouvoir permuter l'intégrale et la limite :

$$\lim_{n\to\infty} \int_\Omega f_n\bar{g} = \int_\Omega \lim_{n\to\infty} f_n\bar{g} = \int_\Omega f\bar{g} = \Phi_g(f). \tag{26.489}$$

Nous allons donc maintenant nous atteler à la tâche de trouver $f_n \in L^\infty$ avec $f_n \xrightarrow{L^p} f$ et telle que (26.489) soit valide.

Nous allons d'abord supposer que $f \in L^p$ est positive à valeurs réelles. Nous avons alors par le théorème 13.105 qu'il existe une suite croissante de fonction étagées (et donc L^∞) telles que $f_n \to f$ ponctuellement. De plus étant donné que $|f_n| \leqslant |f|$, la proposition 26.21 nous dit que $f_n \xrightarrow{L^p} f$. Pour chaque n nous avons

$$\int_\Omega f_n\bar{g} = \phi(f_n). \tag{26.490}$$

48. On peut écrire, pour utiliser de la notation compacte que $g \in L^1(\Omega, \mathbb{C})$.

Soit g^+ la partie réelle positive de $\bar{g}$. Alors nous avons la limite croissante ponctuelle $f_n g^+ \to fg^+$ et le théorème de la convergence monotone 13.160 nous permet d'écrire

$$\lim_{n\to\infty} \int f_n g^+ = \int f g^+. \tag{26.491}$$

Faisant cela pour les trois autres parties de $\bar{g}$ nous avons prouvé que si $f \in L^p$ est réelle et positive,

$$\int f\bar{g} = \phi(f), \tag{26.492}$$

c'est-à-dire que $\Phi_g(f) = \phi(f)$.
Refaisant le tout pour les trois autres parties de f nous montrons que

$$\Phi_g(f) = \phi(f) \tag{26.493}$$

pour tout $f \in L^p(\Omega)$. Nous avons donc égalité de ϕ et Φ_g dans $(L^p)'$ et donc bijection entre $(L^p)'$ et L^q.

Isométrie : mise en place Nous devons prouver que cette bijection est isométrique. Soit $\phi \in (L^p)'$ et $g \in L^q$ telle que $\Phi_g = \phi$. Il faut prouver que

$$\|g\|_q = \|\phi\|_{(L^p)'}. \tag{26.494}$$

$\|\phi\| \leqslant \|g\|_q$ Nous savons que $\phi(f) = \int f\bar{g}$, et nous allons écrire la définition de la norme dans $(L^p)'$:

$$\|\phi\|_{(L^p)'} = \sup_{\|f\|_p=1} \big|\phi(f)\big| \tag{26.495a}$$

$$= \sup \big|\int f\bar{g}\big| \tag{26.495b}$$

$$\leqslant \sup \underbrace{\int |f\bar{g}|}_{=\|f\bar{g}\|_1}. \tag{26.495c}$$

Il s'agit maintenant d'utiliser l'inégalité de Hölder 26.33 :

$$\|\phi\| \leqslant \sup_{\|f\|_p=1} \|f\|_p \|\bar{g}\|_q = \|g\|_q. \tag{26.496}$$

L'inégalité dans l'autre sens sera démontrée en séparant les cas $p = 1$ et $1 < p < \infty$.

Si $p = 1$, une formule Si E est un ensemble mesurable de mesure finie, alors

$$\big|\int_E g d\mu\big| = \big|\phi(\mathbb{1}_E)\big|. \tag{26.497}$$

Mais le fait que $\mu(E) < \infty$ donne que $\mathbb{1}_E \in L^1(\Omega)$. Donc $\mathbb{1}_E \in L^\infty \cap L^1$; nous pouvons alors écrire $\phi(\mathbb{1}_E) = \int_\Omega \mathbb{1}_E \bar{g} d\mu$ et donc

$$\big|\int_\Omega \mathbb{1}_E \bar{g} d\mu\big| = \big|\int_E g d\mu\big| = \big|\phi(\mathbb{1}_E)\big| \leqslant \|\phi\|_{(L^1)'} \|\mathbb{1}_E\|_1 = \|\phi\|\mu(E). \tag{26.498}$$

Nous écrivons cela dans l'autre sens :

$$\|\phi\| \geqslant \frac{1}{\mu(E)} \big|\int_\Omega \mathbb{1}_E \bar{g} d\mu\big| = \big|\frac{1}{\mu(E)} \int_E \bar{g} d\mu\big|. \tag{26.499}$$

Si nous prenons $S = \{t \in \mathbb{C} \text{ tel que } |t| \leqslant \|\phi\|\}$, c'est un fermé vérifiant que

$$\frac{1}{\mu(E)} \int_E \bar{g} d\mu \in S. \tag{26.500}$$

Voila une petite formule qui va nous aider à utiliser le lemme 26.129. Nous ne pouvons cependant pas l'utiliser immédiatement parce que l'appartenance (26.500) n'est vraie que pour les parties de mesure finie.

Si $p = 1$, conclusion[1] Pour utiliser le lemme 26.129, nous utilisons l'hypothèse que Ω est σ-fini. Soient des mesurables A_i de mesure fine tels que $\bigcup_{i\in\mathbb{N}} A_i = \Omega$.

Pour chaque i nous considérons la restriction $g_i\colon A_i \to \mathbb{C}$ de g à A_i. Par le point précédent, elle vérifie

$$\frac{1}{\mu(A_i)}\int_{A_i} \bar{g}_i d\mu = \frac{1}{\mu(A_i)}\int_{A_i} \bar{g} d\mu \in S. \tag{26.501}$$

En appliquant le lemme 26.129 à l'espace restreint $(A_i, \mathcal{A}_i, \mu_i)$, nous concluons $\bar{g}_i \in S$ presque partout, ce qui signifie que $\|g_i\|_\infty \in S$. Nous en concluons que

$$\|g_i\|_\infty \leqslant \|\phi\| \tag{26.502}$$

où, dans ce contexte, $\|g_i\|_\infty$ signifie $\sup_{x\in A_i} |g_i(x)|$.

Nous avons alors

$$\|g\|_\infty = \sup_{x\in\Omega} |g(x)| = \sup_{i\in\mathbb{N}} \|g_i\|_\infty \leqslant \|\phi\|. \tag{26.503}$$

Une petite justification pour cela ? Prenons une suite x_k telle que $|g(x_k)| \to \|g\|_\infty$. Vu que les A_i recouvrent Ω, existe un naturel $i(k)$ tel que $x_k \in A_{i(k)}$. Nous avons alors

$$|g(x_k)| \leqslant \|g_{i(k)}\|_\infty \leqslant \|\phi\|. \tag{26.504}$$

Cela pour conclure que $g \in L^\infty$.

Notons que cet argument ne tient pas avec $p > 1$ parce que l'équation (26.498) terminerait sur $\|\phi\|\mu(E)^{1/p}$. Du coup l'ensemble S à prendre serait $S = \{t \in \mathbb{C} \text{ tel que } |t| \leqslant \|\phi\|\mu(E)^{1/p-1}\}$ et nous sommes en dehors des hypothèses du lemme parce qu'il n'y a pas d'ensemble *indépendant* de E dans lequel l'intégrale $\frac{1}{\mu(E)}\int_E \bar{g} d\mu$ prend ses valeurs.

$1 < p < \infty$ La fonction

$$\alpha(x) = \begin{cases} \frac{g(x)}{|g(x)|} & \text{si } g(x) \neq 0 \\ 1 & \text{si } g(x) = 0 \end{cases} \tag{26.505}$$

a la propriété de faire $\alpha g = |g|$ en même temps que $|\alpha(x)| = 1$ pour tout x. Nous définissons

$$E_n = \{x \text{ tel que } |g(x)| \leqslant n\} \tag{26.506}$$

et

$$f_n = \mathbb{1}_{E_n} |g^{q-1}|\alpha. \tag{26.507}$$

Ce qui est bien avec ces fonctions c'est que [49]

$$|f_n|^p = |g^{p(q-1)}|\alpha|^p = |g|^q \tag{26.508}$$

sur E_n. Dans E_n nous avons $|f_n| = |g^{q-1}| \leqslant n^{q-1}$ et dans E_n nous avons $f_n = 0$. Au final, $f_n \in L^\infty$. Par ce que nous avons vu plus haut, nous avons alors

$$\phi(f_n) = \Phi_g(f_n). \tag{26.509}$$

Par ailleurs,

$$f_n\bar{g} = \mathbb{1}_{E_n} |g^{q-1}| \frac{g}{|g|}\bar{g}, \tag{26.510}$$

49. C'est ici que nous utilisons le lien entre p et q. En l'occurrence, de $1/p + 1/q = 1$ nous déduisons $q(p-1) = p$.

donc [50]

$$\left|\int_{E_n} |g|^q d\mu\right| = |\int_\Omega f_n \bar{g} d\mu| \tag{26.511a}$$
$$= |\phi(f_n)| \tag{26.511b}$$
$$\leqslant \|\phi\| \|f_n\|_p \tag{26.511c}$$
$$= \|\phi\| \left(\int_{E_n} |f_n|^p\right)^{1/p} \tag{26.511d}$$
$$= \|\phi\| \left(\int_{E_n} |g|^q\right)^{1/p}. \tag{26.511e}$$

Nous avons de ce fait une inégalité de la forme $A \leqslant \|\phi\| A^{1/p}$ et donc aussi $A^{1/p} \leqslant \|\phi\|^{1/p} A^{1/p^2}$, et donc $A \leqslant \|\phi\| \|\phi\|^{1/p} A^{1/p^2}$. Continuant ainsi à injecter l'inégalité dans elle-même, pour tout $k \in \mathbb{N}$ nous avons :

$$\left|\int_{E_n} |g|^q d\mu\right| \leqslant \|\phi\|^{1+\frac{1}{p}+\cdots+\frac{1}{p^k}} \left(\int_{E_n} |g|^q d\mu\right)^{1/p^k}. \tag{26.512}$$

Nous pouvons passer à la limite $k \to \infty$. Sachant que $p > 1$ nous savons $A^{1/k} \to 1$ et

$$1 + \frac{1}{p} + \cdots + \frac{1}{p^k} \to \frac{p}{p-1} = q. \tag{26.513}$$

Nous avons alors

$$\int_{E_n} |g|^q d\mu \leqslant \|\phi\|^q. \tag{26.514}$$

L'intégrale s'écrit tout aussi bien sous la forme $\int_\Omega |g|^q \mathbb{1}_{E_n}$. La fonction dans l'intégrale est une suite croissante de fonctions mesurables à valeurs dans $[0, \infty]$. Nous pouvons alors permuter l'intégrale et la limite $n \to \infty$ en utilisant la convergence monotone (théorème 13.160) qui donne alors $\int_\Omega |g|^q \leqslant \|\phi\|^q$ ou encore

$$\|g\|_q \leqslant \|\phi\|. \tag{26.515}$$

Ceci achève de prouver que l'application $\phi \mapsto \Phi_g$ est une isométrie, et donc le théorème.

□

Théorème 26.131.
Soit un espace mesuré $(\Omega, \mathcal{A}, \mu)$.

(1) Si $1 < p < \infty$, alors $L^p(\Omega, \mathcal{A}, \mu)$ est réflexif[51].

(2) Si $(\Omega, \mathcal{A}, \mu)$ est σ-finie, alors

(2a) $(L^1)' = L^\infty$

(2b) $L^1 \subset (L^\infty)'$.

Démonstration. En plusieurs parties, en notant toujours p et q les exposants conjugués, c'est-à-dire $\frac{1}{p} + \frac{1}{q} = 1$.

Pour (1) Le théorème 26.130(1) nous indique que

$$(L^p)' = L^q \tag{26.516}$$

au sens d'une bijection isométrique. Vu que $1 < p < \infty$, nous avons aussi $1 < q < \infty$ et donc $(L^q)' = L^p$. En prenant le dual des deux côtés de (26.516),

$$(L^p)'' = (L^q)' = L^p, \tag{26.517}$$

et nous avons prouvé que L^p est réflexif.

50. Dans [6], cette équation arrive sans modules, ce qui me laisse entendre que $\phi(f_n)$ est réel et positif pour pouvoir écrire que $\phi(f_n) \leqslant \|\phi\| \|f_n\|_p$, mais je ne comprends pas pourquoi.

51. Définition 10.163.

Pour (2a) Il s'agit du théorème 26.130(2).

Pour (2b)[1] Il nous reste à couvrir le cas de $(L^\infty)'$. Pour $g \in L^1$ nous prouvons que $\Phi_g \in (L^\infty)'$.

$\Phi_g(f)$ est bien définie Nous prouvons d'abord que si $f \in L^\infty$, alors l'intégrale $\int_\Omega f\bar{g}$ est bien définie. Par définition du supremum essentiel[52], il existe $M > 0$ tel que $|f(x)| < M$ pour tout x hors d'une partie A de mesure nulle. Nous avons alors

$$\int_\Omega |f\bar{g}| = \int_{\Omega\setminus A} |f\bar{g}| \leqslant M \int_{\Omega\setminus A} |f| = M \int_\Omega |f| < \infty. \tag{26.518}$$

Φ_g est continue Soit une suite $f_k \xrightarrow{L^\infty} f$ ainsi que $g \in L^1$. Pour chaque k, il existe une partie de mesure nulle A_k et un nombre $M_k = \|f_k\|_{L^\infty}$ tel que $|f_k(x)| < \|f_k\|_{L^\infty}$ pour tout x hors de A_k. Nous avons alors

$$|\Phi_g(f_k)| \leqslant \int_{\Omega\setminus A_k} |f_k\bar{g}| d\mu \leqslant \|f_k\|_{L^\infty} \int_{\Omega\setminus A_k} |g| \leqslant \|f_k\|_{L^\infty} \|g\|_1. \tag{26.519}$$

Vu que par hypothèse $f_k \to 0$ dans L^∞, nous avons $\|f_k\|_{L^\infty \to 0}$, et donc aussi

$$|\Phi_g(f_k)| \to 0. \tag{26.520}$$

□

Proposition 26.132.
Soit $f \in L^p(\Omega)$ telle que

$$\int_\Omega f\varphi = 0 \tag{26.521}$$

pour tout $\varphi \in C_c^\infty(\Omega)$. Alors $f = 0$ presque partout.

Démonstration. Nous considérons la forme linéaire $\Phi_f \in (L^q)'$ donnée par

$$\begin{aligned} \Phi_f \colon L^p &\to \mathbb{C} \\ u &\mapsto \int_\Omega fu \end{aligned} \tag{26.522}$$

Par hypothèse cette forme est nulle sur la partie dense $C_c^\infty(\Omega)$. Si (φ_n) est une suite dans $C_c^\infty(\Omega)$ convergente vers u dans L^p, nous avons pour tout n que

$$0 = \Phi_f(\varphi_n) \tag{26.523}$$

En passant à la limite, nous voyons que Φ_f est la forme nulle. Elle est donc égale à Φ_0. La partie « unicité » du théorème de représentation de Riesz 26.130 nous indique alors que $f = 0$ dans L^p et donc $f = 0$ presque partout. □

Proposition 26.133.
Si $f \in L^1_{loc}(I)$ est telle que

$$\int_I f\varphi' = 0 \tag{26.524}$$

pour tout $\varphi \in C_c^\infty(I)$, alors il existe une constante C telle que $f = C$ presque partout.

Démonstration. Soit $\psi \in C_c^\infty(I)$ une fonction d'intégrale 1 sur I. Si $w \in C_c^\infty(I)$ alors nous considérons la fonction

$$h = w - \psi \int_I w, \tag{26.525}$$

52. Voir les définitions 26.23 et 26.24.

qui est dans $C_c^\infty(I)$ et dont l'intégrale sur I est nulle. Par la proposition 16.2, la fonction h admet une primitive dans $C_c^\infty(I)$; et nous notons φ cette primitive. L'hypothèse appliquée à φ donne

$$0 = \int_I f\varphi' = \int_I f\left(w - \psi\int_I w\right) = \int_I fw - \underbrace{\left(\int_I f(x)\psi(x)dx\right)}_{C}\left(\int_I w(y)dy\right) = \int_I w(f-C). \tag{26.526}$$

L'annulation de la dernière intégrale implique par la proposition 26.132 que $f - C = 0$ dans L^2, c'est-à-dire $f = C$ presque partout. □

Dans [466], il est dit que « la preuve [du lemme suivant], un peu fastidieuse mais en rien ingénieuse, est laissée en exercice ». La preuve est donc de moi ; elle est un tout petit peu ingénieuse mais en rien fastidieuse. J'espère ne pas m'être trompé et me demande bien ce que l'auteur avait en tête. Ma preuve s'appuie sur la proposition 16.99 dont la preuve ne me paraît pas non plus « fastidieuse mais en rien ingénieuse ».

Lemme 26.134 ([466, 1]).
Soient $r > 0$. Il existe $\delta > 0$ tel que pour tout $s, t \in \mathbb{C}$ vérifiant $|s| \leqslant 1$, $|t| \leqslant 1$ et $|s - t| \geqslant r$ nous ayons

$$\left|\frac{s+t}{2}\right|^p \leqslant (1-\delta)\frac{|s|^p + |t|^p}{2}. \tag{26.527}$$

Démonstration. Soit $r > 0$. La partie de $\mathbb{C}^2$ donnée par

$$D = \{(s,t) \in \mathbb{C}^2 \text{ tel que } |s| \leqslant 1, |t| \leqslant 1, |s-t| \geqslant r\} \tag{26.528}$$

est compacte. En effet elles est bornée (par la sphère de rayon $\sqrt{2}$) et fermée comme intersection de fermée[53]. Nous considérons la fonction $\Delta\colon D \to \mathbb{R}$ donnée par

$$\left|\frac{s+t}{2}\right|^p = \Delta(s,t)\frac{|s|^p + |t|^p}{2}. \tag{26.529}$$

Si vous voulez une expression explicite,

$$\Delta(s,t) = \frac{2^{p-1}|s+t|^p}{|s|^p + |t|^p}. \tag{26.530}$$

Cela est bien défini et continu sur D parce que le complémentaire D^c (qui est ouvert) contient $(0,0)$ et donc aussi un voisinage de $(0,0)$.

La proposition 16.99 nous dit que la fonction $z \mapsto |z|^p$ est strictement convexe. En prenant la définition 16.92 de la stricte convexité avec $\theta = \frac{1}{2}$, nous trouvons que

$$\Delta(s,t) < 1 \tag{26.531}$$

pour tout $(s,t) \in D$. Vu que par ailleurs Δ est une fonction continue sur le compact D, elle atteint un minimum dans D. Soit Δ_0 ce minimum quivérifie forcément $\Delta_0 < 1$.

En posant $1 - \delta = \Delta_0$ nous avons le résutat. □

26.12 Théorèmes de Hahn-Banach

Théorème 26.135 (Hahn-Banach[467, 153]).
Soit E, un espace vectoriel réel et une application $p\colon E \to \mathbb{R}$ satisfaisant

(1) $p(\lambda x) = \lambda p(x)$ pour tout $x \in E$ et pour tout $\lambda > 0$,

(2) $p(x+y) \leqslant p(x) + p(y)$ pour tout $x, y \in E$.

53. Lemme 7.3 suivit du théorème de Borel-Lebesgue 7.127.

Soit de plus $G \subset E$ un sous-espace vectoriel muni d'une application $g\colon G \to \mathbb{R}$ vérifiant $g(x) \leqslant p(x)$ pour tout $x \in G$. Alors il existe $f \in \mathcal{L}(E, \mathbb{R})$ telle que $f(x) = g(x)$ pour tout $x \in G$ et $f(x) \leqslant p(x)$ pour tout $x \in E$.

Démonstration. Si h une application linéaire définie sur un sous-espace de E, nous notons D_h ledit sous-espace.

Un ensemble inductif Nous considérons P, l'ensemble des fonctions linéaires suivant

$$P = \left\{ h\colon D_h \to \mathbb{R} \text{ tel que } \begin{cases} G \subset D_h \\ h(x) = g(x) & \forall x \in G \\ h(x) \leqslant p(x) & \forall x \in D_h \end{cases} \right\} \tag{26.532}$$

Cet ensemble est non vide parce que g est dedans. Nous le munissons de la relation d'ordre $h_1 \leqslant h_2$ si et seulement si $D_{h_1} \subset D_{h_2}$ et h_2 prolonge h_1. Nous montrons à présent que P est un ensemble inductif. Soit un sous-ensemble totalement ordonné $Q \subset P$; nous définissons une fonction h de la façon suivante. D'abord $D_h = \sup_{l \in Q} D_l$ et ensuite

$$\begin{aligned} h\colon D_h &\to \mathbb{R} \\ x &\mapsto l(x) \quad \text{si } x \in D_l \end{aligned} \tag{26.533}$$

Cela est bien définit parce que si $x \in D_l \cap D_{l'}$ alors, vu que Q est totalement ordonné (i.e. $l \leqslant l'$ ou $l' \leqslant l$), on a obligatoirement $D_l \subset D_{l'}$ et l' qui prolonge l (ou le contraire). Donc h est un majorant de Q dans P parce que $h \geqslant l$ pour tout $l \in Q$. Cela montre que P est inductif (définition 1.14). Le lemme de Zorn 1.15 nous dit alors que P possède un maximum f qui va être la réponse à notre théorème.

Le support de f La fonction f est dans P ; donc $f(x) \leqslant p(x)$ pour tout $x \in D_h$ et $f(x) = g(x)$ pour tout $x \in G$. Pour terminer nous devons montrer que $D_f = E$. Supposons donc que $D_f \neq E$ et prenons $x_0 \notin D_f$. Nous allons contredire la maximalité de f en considérant la fonction h donnée par $D_h = D_f + \mathbb{R}x_0$ et

$$h(x + tx_0) = f(x) + t\alpha \tag{26.534}$$

où α est une constante que nous allons fixer plus tard.

Nous commençons par prouver que f est dans P. Nous devons prouver que

$$h(x + tx_0) = f(x) + t\alpha \leqslant p(x + tx_0) \tag{26.535}$$

Pour cela nous allons commencer par fixer α pour avoir les relations suivantes :

$$\begin{cases} f(x) + \alpha \leqslant p(x + x_0) & \text{(26.536a)} \\ f(x) - \alpha \leqslant p(x - x_0) & \text{(26.536b)} \end{cases}$$

pour tout $x \in D_f$. Ces relations sont équivalentes à demander α tel que

$$\begin{cases} \alpha \leqslant p(x + x_0) - f(x) & \text{(26.537a)} \\ \alpha \geqslant f(x) - p(x - x_0) & \text{(26.537b)} \end{cases}$$

Nous nous demandons donc s'il existe un α qui satisfasse

$$\sup_{y \in D_f} \big(f(y) - p(y - x_0)\big) \leqslant \alpha \leqslant \inf_{z \in D_f} \big(p(z + x_0) - f(z)\big). \tag{26.538}$$

Ou encore nous devons prouver que pour tout $y, z \in D_f$,

$$p(z + x_0) - f(x) \geqslant f(y) - p(y - x_0) \geqslant 0. \tag{26.539}$$

Par les propriétés de p et de f,

$$p(z+x_0)+p(y-x_0)-f(z)-f(y)\geqslant p(z+y)-f(z+y)\geqslant 0. \tag{26.540}$$

La dernière inégalité est le fait que $f\in P$. Un choix de α donnant les inéquations (26.536) est donc possible.

À partir des inéquations (26.536) nous obtenons la relation (26.535) de la façon suivante. Si $t>0$ nous multiplions l'équation (26.536a) par t :

$$tf(x)+t\alpha\leqslant tp(x+x_0). \tag{26.541}$$

Et nous écrivons cette relation avec x/t au lieu de x en tenant compte de la linéarité de f :

$$f(x)+t\alpha\leqslant tp\big(\frac{x}{t}+x_0\big)=p(x+tx_0). \tag{26.542}$$

Avec $t<0$, c'est similaire, en faisant attention au sens des inégalités.

Nous avons donc construit $h\colon D_h\to\mathbb{R}$ avec $h\in P$, $D_f\subset D_h$ et $h(x)=f(x)$ pour tout $x\in D_f$. Cela pour dire que $h>f$, ce qui contredit la maximalité de f. Le domaine de f est donc E tout entier.

La fonction f est donc une fonction qui remplit les conditions.

□

Définition 26.136.
Un espace topologique est ***localement convexe*** *si tout point possède un système fondamental de voisinages formé de convexes.*

Définition 26.137 (Hyperplan qui sépare)**.**
Soit E un espace vectoriel topologique ainsi que A, B des sous-ensembles de E. Nous disons que l'hyperplan d'équation $f=\alpha$ ***sépare au sens large*** *les parties A et B si $f(x)\leqslant\alpha$ pour tout $x\in A$ et $f(x)\geqslant\alpha$ pour tout $x\in B$.*

La séparation est ***au sens strict*** *s'il existe $\epsilon>0$ tel que*

$$f(x)\leqslant\alpha-\epsilon \qquad \text{pour tout } x\in A \tag{26.543a}$$
$$f(x)\geqslant\alpha+\epsilon \qquad \text{pour tout } x\in B. \tag{26.543b}$$

Théorème 26.138 (Hahn-Banach, première forme géométrique[153])**.**
Soit E un espace vectoriel topologique et A, B deux convexes non vides disjoints de E. Si A est ouvert, il existe un hyperplan fermé qui sépare A et B au sens large.

Théorème 26.139 (Hahn-Banach, seconde forme géométrique)**.**
Soient un espace vectoriel topologique localement convexe[54] *ainsi que des convexes non vides disjoints A et B tels que A soit compact et B soit fermé. Alors il existe un hyperplan fermé qui sépare strictement A et B.*

Démonstration. Vu que B est fermé, A est dans l'ouvert $E\backslash B$. Donc si $a\in A$, il existe un voisinage ouvert convexe de a inclus dans A. Soit U_a un voisinage ouvert et convexe de 0 tel que $(a+U_a)\cap B=\varnothing$.

Vu que la fonction $(x,y)\mapsto x+y$ est continue, nous pouvons trouver un ouvert convexe V_a tel que $V_a+V_a\subset U_a$. L'ensemble $a+V_a$ est alors un voisinage ouvert de a et bien entendu $\bigcup_a(a+V_a)$ recouvre A qui est compact. Nous en extrayons un sous-recouvrement fini, c'est-à-dire que nous considérons $a_1,\dots,a_n\in A$ tels que

$$A\subset\bigcup_{i=1}^{n}(a_i+V_{a_i}). \tag{26.544}$$

54. Définition 26.136.

Nous posons alors

$$V = \bigcap_{i=1}^{n} V_{a_i}. \tag{26.545}$$

Cet ensemble est non vide parce et il contient un voisinage de zéro parce que c'est une intersection finie de voisinages de zéro. Soit $x \in A + V$. Il existe i tel que

$$x \in a_i + U_{a_i} + V \subset a_i + V_{a_i} + V_{a_i} \subset a_i + U_{a_i} \subset E \backslash B. \tag{26.546}$$

Donc $(A + V) \cap B = \emptyset$. L'ensemble $A + V$ est alors un ouvert convexe disjoint de B. Par la première forme géométrique du théorème de Hahn-Banach 26.138 nous avons un hyperplan qui sépare $A + V$ de B au sens large : il existe $f \in E' \backslash \{0\}$ tel que $f(a) + f(v) \leqslant f(b)$ pour tout $a \in A$, $v \in V$ et $b \in B$.

Il suffit donc de trouver un $v \in V$ tel que $f(v) \neq 0$ pour avoir la séparation au sens strict. Cela est facile : V étant un voisinage de zéro et f étant linéaire, si elle était nulle sur V, elle serait nulle sur E. □

26.13 Théorème de Tietze

Définition 26.140.
*Si E et F sont des espaces normés, une application $f \colon E \to F$ est **presque surjective** s'il existe $\alpha \in]0, 1[$ et $C > 0$ tels que pour tout $y \in \overline{B_F(0, 1)}$, il existe $x \in \overline{B_E(0, C)}$ tel que $\|y - f(x)\| \leqslant \alpha$.*

Lemme 26.141 ([54]).
Soient E et F des espaces de Banach et $f \in L(E, F)$[55]*. Si f est presque surjective, alors*

(1) f est surjective

(2) pour tout $y \in \overline{B_F(0, 1)}$, il existe $x \in \overline{B_E(0, \frac{C}{1-\alpha})}$ tel que $y = f(x)$.

Le point (2) est une précision du point (1) : il dit quelle est la taille de la boule de E nécessaire à obtenir la boule unité dans F.

Démonstration. Soit $y \in \overline{B_F(0, 1)}$. Nous allons construire $x \in B\big(0, \frac{C}{1-\alpha}\big)$ qui donne $f(x) = y$. Ce x sera la limite d'une série que nous allons construire par récurrence. Pour $n = 1$ nous utilisons la presque surjectivité pour considérer $x_1 \in \overline{B_E(0, C)}$ tel que $\|y - f(x_1)\| \leqslant \alpha$. Ensuite nous considérons la récurrence

$$x_n \in \overline{B_E(0, C)} \tag{26.547}$$

tel que

$$\Big\|y - \sum_{i=1}^{n} \alpha^{i-1} f(x_i)\Big\| \leqslant \alpha^n \tag{26.548}$$

Pour montrer que cela existe nous supposons que la série est déjà construire jusqu'à $n > 1$:

$$\frac{1}{\alpha^n}\Big(y - \sum_{i=1}^{n} \alpha^{i-1} f(x_i)\Big) \in \overline{B_F(0, 1)} \tag{26.549}$$

À partir de là, par presque surjectivité il existe un $x_{n+1} \in \overline{B_E(0, C)}$ tel que

$$\Big\|\frac{y - \sum_{i=1}^{n} \alpha^{i-1} f(x_i)}{\alpha^n} - f(x_{n+1})\Big\| \leqslant \alpha. \tag{26.550}$$

En multipliant par α^n, le terme $\alpha^n f(x_{n+1})$ s'intègre bien dans la somme :

$$\Big\|y = \sum_{i=1}^{n+1} \alpha^{i-1} f(x_i)\Big\| \leqslant \alpha^{n+1}. \tag{26.551}$$

55. L'ensemble des applications linéaires continues

Nous nous intéressons à une éventuelle limite à la somme des $\alpha^{n-1}x_n$. D'abord nous avons la majoration $\|\alpha^{n-1}x_n\| \leqslant \alpha^{n-1}C$, et vu que par la définition de la presque surjectivité $0 < \alpha < 1$, la série

$$\sum_{n=1}^{\infty} \alpha^{n-1}x_n \tag{26.552}$$

converge absolument [56] parce que la suite des normes est une suite géométrique de raison α. Vu que E est de Banach, la convergence absolue implique la convergence simple (la suite des sommes partielles est de Cauchy et Banach est complet). Nous posons

$$x = \sum_{n=1}^{\infty} \alpha^{n-1}x_n \in E, \tag{26.553}$$

et en termes de normes, ça vérifie

$$\|x\| \leqslant \sum_{n=1}^{\infty} \alpha^{n-1}\|x_n\| \leqslant C \sum_{n=1}^{\infty} \alpha^{n-1} = \frac{C}{1-\alpha}. \tag{26.554}$$

Donc c'est bon pour avoir $x \in B\big(0, \frac{C}{1-\alpha}\big)$. Nous devons encore vérifier que $y = f(x)$. Pour cela nous remarquons que

$$\|y - f\Big(\sum_{n=1}^{N} \alpha^{n-1}x_n\Big)\| \leqslant \alpha^N. \tag{26.555}$$

Nous pouvons prendre la limite $N \to \infty$ et permuter f avec la limite (par continuité de f). Vu que $0 < \alpha < 1$ nous avons

$$\|y - f(x)\| = 0. \tag{26.556}$$

□

Théorème 26.142 (Tietze[54, 311]).
Soit un espace métrique (X, d) et un fermé $Y \subset X$. Soit $g_0 \in C^0(Y, \mathbb{R})$. Alors g_0 admet un prolongement continu sur X.

Démonstration. Soit l'opération de restriction

$$\begin{aligned} T\colon (C_b^0(X,\mathbb{R}), \|.\|_\infty) &\to (C_b^0(Y,\mathbb{R}), \|.\|_\infty) \\ f &\mapsto f|_Y. \end{aligned} \tag{26.557}$$

L'application T est évidemment linéaire. Elle est de plus borné pour la norme opérateur usuelle donnée par la proposition 10.3 parce que $\|T(f)\| \leqslant \|f\| < \infty$. L'application T est alors continue par la proposition 10.18.

Presque surjection Soit $g \in C_b^0(Y, \mathbb{R})$ avec $\|g\|_\infty \leqslant 1$. Nous posons

$$Y^+ = \{x \in Y \text{ tel que } \frac{1}{3} \leqslant g(x) \leqslant 1\} \tag{26.558a}$$

$$Y^- = \{x \in Y \text{ tel que } -1 \leqslant g(x) \leqslant -\frac{1}{3}\}. \tag{26.558b}$$

Nous considérons alors

$$\begin{aligned} f\colon X &\to \mathbb{R} \\ x &\mapsto \frac{1}{3}\frac{d(x,Y^-) - d(x,Y^+)}{d(x,Y^-) + d(x,Y^+)} \end{aligned} \tag{26.559}$$

Vu qu'en valeur absolue le dénominateur est plus grand que le numérateur nous avons $\|f\|_\infty \leqslant \frac{1}{3}$. Notons que

— Si $x \in Y^+$ alors $f(x) = \frac{1}{3}$ et $g(x) \in [\frac{1}{3}, 1]$;

56. Définition 10.51.

— Si $x \in Y^-$ alors $f(x) = -\frac{1}{3}$ et $g(x) \in [-1, -\frac{1}{3}]$;

— Si x n'est ni dans Y^+ ni dans Y^- alors nous avons[57] $g(x) \in [-\frac{1}{3}, \frac{1}{3}]$ et donc $\big|f(x) - g(x)\big| \leqslant \big|f(x)\big| + \big|g(x)\big| \leqslant \frac{2}{3}$.

Dans les deux cas nous avons $\big|f(x) - g(x)\big| \in [0, \frac{2}{3}]$ pour tout $x \in X$. Cela prouve que

$$\|T(f) - g\|_{Y,\infty} \leqslant \frac{2}{3}. \tag{26.560}$$

En résumé nous avons pris g dans la boule $\overline{B(0,1)}$ de $\big(C_b^0(Y,\mathbb{R}), \|.\|_\infty\big)$ et nous avons construit une fonction f dans la boule $\overline{B(0,\frac{1}{3})}$ de $\big(C_b^0(X,\mathbb{R}), \|.\|_\infty\big)$ telle que $\|T(f) - g\|_\infty \leqslant \frac{2}{3}$. L'application T est donc une presque surjection avec $\alpha = \frac{1}{3}$ et $C = \frac{2}{3}$.

Prolongement dans les boules unité fermées La proposition 11.296 nous assure que les espaces $C_b^0(X,\mathbb{R})$ et $C_b^0(Y,\mathbb{R})$ sont de Banach (complets), et le lemme 26.141 nous dit alors que T est surjective et que pour tout $g \in \overline{B(0,1)}$, il existe

$$f \in \overline{B\left(0, \frac{1/3}{1-\frac{2}{3}}\right)} = \overline{B(0,1)}. \tag{26.561}$$

telle que $g = T(f)$.

Prolongement pour les boules ouvertes Jusqu'à présent nous avons montré qu'une fonction $g \in \overline{B(0,1)}$ admet une prolongement continu dans $\overline{B(0,1)}$. Nous allons montrer que si g est dans la boule ouverte $B(0,1)$ de $\big(C_b^0(Y,\mathbb{R}), \|.\|_\infty\big)$ alors g admet un prolongement dans la boule ouverte $B(0,1)$ de $\big(C_b^0(X,\mathbb{R}), \|.\|_\infty\big)$.

Soit $g \in B_{C_b^0(Y)}(0,1)$ et son prolongement $h \in \overline{B_{C_b^0(X)}(0,1)}$. Si $\|h\|_\infty < 1$ alors le résultat est vrai. Sinon nous considérons l'ensemble

$$Z = \{x \in X \text{ tel que } |h(x)| = 1\}. \tag{26.562}$$

Nous avons $Y \cap Z = \emptyset$ parce que nous avons $h = g$ sur Y et nous avons choisi $\|g\|_\infty < \infty$. Par ailleurs Y est fermé par hypothèse et Z est fermé parce que h est continue ; par conséquent $Y \cap Z$ est fermé, donc[58]

$$\bar{Y} \cap \bar{Z} = Y \cap Z = \emptyset. \tag{26.563}$$

Nous posons

$$\begin{aligned} u\colon X &\to \mathbb{R}^+ \\ x &\mapsto \frac{d(x,Z)}{d(x,Y) + d(x,Z)} \end{aligned} \tag{26.564}$$

Le dénominateur n'est pas nul parce qu'il faudrait $d(x,Y) = d(x,Z) = 0$, ce qui demanderait $x \in \bar{Y} \cap \bar{Z}$, ce qui n'est pas possible. Nous posons $f = uh$. Si $x \in Y$ alors $u(x) = 1$, donc f est encore un prolongement de g. De plus f est encore continue, et donc encore un bon candidat. Enfin si x est hors de Y alors $d(x,Y) > 0$ (strictement parce que Y est fermé) et donc $0 < u(x) < 1$, ce qui donne $|f(x)| < |h(x)| \leqslant 1$. Donc $\|f\|_\infty < 1$.

Nous avons donc trouvé qu'une fonction dans la boule ouverte $B_{C_b^0(Y)}(0,1)$ se prolonge en une fonction dans la boule ouverte $B_{C_b^0(X)}(0,1)$.

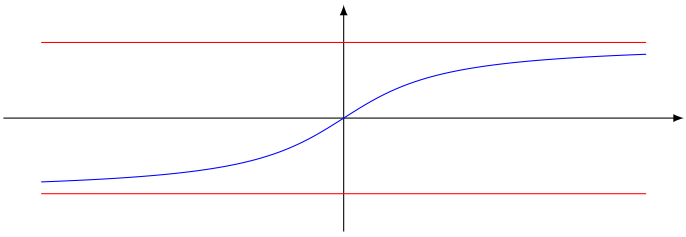
Le cas non borné Soit enfin $g_0 \in C^0(Y,\mathbb{R})$. Nous allons nous ramener au cas de la boule unité ouverte en utilisant un homéomorphisme $\phi\colon \mathbb{R} \to]-1,1[$. L'application $g = \phi \circ g_0$ est dans la boule unité ouvert de $C^0(Y,\mathbb{R})$ et donc admet un prolongement f dans la boule unité ouverte de $C^0(X)$. L'application $f_0 = \phi^{-1} \circ f$ est un prolongement continu de g_0.

□

57. Nous rappelons que $\|g\| = 1$, donc $g(x)$ est forcément ente -1 et 1.

58. Si vous avez l'intention de dire que $\overline{Y \cap Z} = \bar{Y} \cap \bar{Z} = Y \cap Z = \emptyset$, allez d'abord voir l'exemple 7.246. Ici c'est correct parce que Y et Z sont fermés.

Un homéomorphisme $\phi\colon \mathbb{R} \to]-1,1[$ est donné par exemple par la fonction $\phi(t) = \frac{2}{\pi}\arctan(t)$ dont le graphique est donné ci-dessous :



26.14 Espace de Schwartz

Pour un multiindice $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$, nous notons

$$\partial^\alpha \varphi = \partial_{x_1}^{\alpha_1} \dots \partial_{x_d}^{\alpha_d} \varphi \tag{26.565}$$

pour peu que la fonction φ soit $|\alpha| = \alpha_1 + \dots + \alpha_d$ fois dérivable.

Définition 26.143.
*Soit $\Omega \subset \mathbb{R}^d$. L'**espace de Schwartz** $\mathscr{S}(\Omega)$ est le sous-ensemble de $C^\infty(\Omega)$ des fonctions dont toutes les dérivées décroissent plus vite que tout polynôme :*

$$\mathscr{S}(\Omega) = \big\{\varphi \in C^\infty(\Omega) \text{ tel que } \forall \alpha, \beta \in \mathbb{N}^d, p_{\alpha,\beta}(\varphi) < \infty\big\} \tag{26.566}$$

où nous avons considéré

$$p_{\alpha,\beta}(\varphi) = \sup_{x\in\Omega} |x^\beta (\partial^\alpha \varphi)(x)| = \|x^\beta \partial^\alpha \varphi\|_\infty. \tag{26.567}$$

Pour simplifier les notations (surtout du côté de Fourier), nous allons parfois écrire $M_i\varphi$ pour la fonction $x \mapsto x_i\varphi(x)$.

Exemple 26.144.
La fonction e^{-x^2} est une fonction à décroissance rapide sur $\mathbb{R}$. △

Définition 26.145.
*Une fonction $f\colon \mathbb{R}^d \to \mathbb{C}$ est dite à **décroissance rapide** si elle décroît plus vite que n'importe quel polynôme. Plus précisément, si pour tout polynôme Q, il existe un $r > 0$ tel que $|f(x)| < \frac{1}{|Q(x)|}$ pour tout $\|x\| \geqslant r$.*

Proposition 26.146.
Une fonction Schwartz est à décroissance rapide.

Démonstration. Nous commençons par considérer un polynôme P donné par

$$P(x) = \sum_k c_k x^{\beta_k} \tag{26.568}$$

où les β_k sont des multiindices, les c_k sont des constantes et la somme est finie. Nous avons la majoration

$$\sup_{x\in\mathbb{R}^d} |\varphi(x)P(x)| \leqslant \sum_k \sup_x \big|c_k\varphi(x)x^{\beta_k}\big| \leqslant \sum_k |c_k| p_{0,\beta_k}(\varphi) < \infty. \tag{26.569}$$

Nous allons noter M_P la constante $\sum_k |c_k| p_{0,\beta_k}(\varphi)$, de sorte que pour tout $x \in \mathbb{R}^d$ nous ayons $|\varphi(x)P(x)| \leqslant M_P$ et donc

$$|\varphi(x)| \leqslant \frac{M_P}{|P(x)|} = \frac{1}{|\frac{1}{M_P}P(x)|}. \tag{26.570}$$

Notons que cette inégalité est a fortiori correcte pour les x sur lesquels P s'annule.

Soit maintenant un polynôme Q. Nous considérons le polynôme $P(x) = \|x\|Q(x)$. Étant de plus haut degré, pour toute constante C il existe un rayon r_C tel que $|P(x)| \geqslant C|Q(x)|$ pour tout $|x| \geqslant r_C$. En particulier pour $|x| \geqslant r_{M_P}$ nous avons

$$|P(x)| \geqslant M_P|Q(x)| \tag{26.571}$$

et donc, pour ces x,

$$|\varphi(x)| \leqslant \frac{1}{|\frac{1}{M_P}P(x)|} \leqslant \frac{1}{|Q(x)|}. \tag{26.572}$$

La première inégalité est valable pour tout x, et la seconde pour $\|x\| \geqslant r_{M_P}$. □

Corolaire 26.147 ([1]).
Soit φ une fonction Schwartz sur $\mathbb{R}^m \times \mathbb{R}^n$. Alors la fonction

$$y \mapsto \sup_{x\in\mathbb{R}^n} |\varphi(x,y)| \tag{26.573}$$

est intégrable.

Démonstration. Soit un polynôme Q en la variable y. Par la proposition 26.146, il existe $r > 0$ tel que

$$|\varphi(x,t)| < \frac{1}{Q(y)} \tag{26.574}$$

pour tout $\|(x,y)\| > r$. A fortiori l'inégalité tient pour tout $|y| > r$. Donc

$$\int_{\mathbb{R}^m} \sup_{x\in\mathbb{R}^n} |\varphi(x,y)|dy = \int_{\|y\|\leqslant r} \sup_x |\varphi(x,y)|dy + \int_{\|y\|>r} \sup_x |\varphi(x,y)|dy. \tag{26.575}$$

La première intégrale est bornée par $\mathrm{Vol}\big(B(0,r)\big)\|f\|_\infty$ tandis que la seconde est bornée par l'intégrale de $\frac{1}{Q(y)}$. En prenant Q de degré suffisamment élevé en toutes les composantes de y nous avons intégrabilité. □

26.14.1 Topologie

Lemme-définition 26.148.
Les $p_{\alpha,\beta}$ donnés par l'équation (26.567) ci-dessus sont des semi-normes[59]*. La topologie considérée sur $\mathscr{S}(\mathbb{R}^d)$ est celle des semi-normes $p_{\alpha,\beta}$.*

26.149.
Nous avons un enchainement de résultats qui nous aident à prouver la continuité d'une application $T\colon \mathscr{S}(\mathbb{R}^d) \to X$.

(1) La topologie de $\mathscr{S}(\mathbb{R}^d)$ est donnée par une famille dénombrable de semi-normes. Donc la proposition 7.331 nous dit que $\mathscr{S}(\mathbb{R}^d)$ est métrisable.

(2) La proposition 7.302 nous dit alors que si X est métrique, toute application séquentiellement continue $T\colon \mathscr{S}(\mathbb{R}^d) \to X$ est continue.

(3) Donc si X est métrique, il suffit de prouver que pour $f_n \xrightarrow{\mathscr{S}(\mathbb{R}^d)} 0$ nous avons $T(f_n) \xrightarrow{X} 0$ où $f_n\colon \mathscr{S}(\mathbb{R}^d) \to X$. Dans les cas usuels, T sera une distribution et $X = \mathbb{C}$.

(4) En vertu de la proposition 7.329, la convergence $f_n \xrightarrow{\mathscr{S}(\mathbb{R}^d)} 0$ signifie que pour tout choix de multiindice α et β, $p_{\alpha,\beta}(f_n) \to 0$, c'est-à-dire

$$\|x^\beta \partial^\alpha f_n\|_\infty \to 0. \tag{26.576}$$

(5) Et enfin, la technique pour montrer que $T\colon \mathscr{S}(\mathbb{R}^d) \to \mathbb{C}$ est continue est de montrer que sous l'hypothèse d'avoir (26.576) pour tout choix de α et β, nous avons $T(f_n) \to 0$ dans $\mathbb{C}$.

59. Définition 7.326.

Lemme 26.150 ([4]).
La topologie sur $\mathscr{S}(\mathbb{R}^d)$ est donnée aussi par les semi-normes

$$q_{n,m} = \max_{|\alpha|\leqslant n} \sup_{x\in\mathbb{R}^d} \big(1+\|x\|\big)^m \big|\partial^\alpha \varphi(x)\big|. \tag{26.577}$$

Autrement dit, une suite $\varphi_n \overset{\mathscr{S}(\mathbb{R}^d)}{\to} 0$ si et seulement si $q_{n,m(\varphi)} \to 0$ pour tout n et m.

Le fait que les $q_{n,m}(\varphi)$ restent bornés est la proposition 26.146. Cependant ce lemme est plus précis parce qu'en disant seulement que φ est majoré par des polynôme, nous ne disons pas que les polynômes correspondants aux φ_n tendent vers zéro si $\varphi_n \overset{\mathscr{S}}{\to} 0$. Et d'ailleurs on ne sait pas très bien ce que signifierait $P_n \to 0$ pour une suite de polynômes.

Proposition 26.151.
Pour $p \in [1,\infty]$, l'espace $\mathscr{S}(\mathbb{R}^d)$ s'injecte continument dans $L^p(\mathbb{R}^d)$.

Démonstration. L'injection dont nous parlons est l'identité ou plus précisément l'identité suivie de la prise de classe. Il faut vérifier que cela est correct et continu, c'est-à-dire d'abord qu'une fonction à décroissance rapide est bien dans L^p et ensuite que si $f_n \overset{\mathscr{S}}{\to} 0$, alors $f_n \overset{L^p}{\to} 0$.

Commençons par $p=\infty$. Alors $\|f_n\|_\infty = p_{0,0}(f_n) \to 0$ parce que si $f_n \overset{\mathscr{S}}{\to} 0$, alors en particulier $p_{0,0}(f_n) \to 0$.

Au tour de $p<\infty$ maintenant. Nous savons qu'en dimension d, la fonction

$$x \mapsto \frac{1}{(1+\|x\|)^s} \tag{26.578}$$

est intégrable dès que $s>d$. Pour toute valeur de m nous avons

$$\|\varphi\|_p^p = \int_{\mathbb{R}^d} |\varphi(x)|^p dx = \int_{\mathbb{R}^d} \frac{\big|(1+\|x\|)^m\varphi(x)\big|^p}{\big(1+\|x\|\big)^{mp}} \leqslant \int_{\mathbb{R}^d} \frac{q_{0,m}(\varphi)^p}{\big(1+\|x\|\big)^{mp}}. \tag{26.579}$$

En choisissant m de telle sorte que $mp>d$, nous avons convergence de l'intégrale et donc $\|\varphi\|_p<\infty$. Nous retenons que

$$\|\varphi\|_p^p \leqslant C q_{0,m}(\varphi)^p \tag{26.580}$$

pour une certaine constance C et un bon choix de m.

Ceci prouve que $\mathscr{S}(\mathbb{R}^d) \subset L^p(\mathbb{R}^d)$. Nous devons encore vérifier que l'inclusion est continue. Si $\varphi_n \overset{\mathscr{S}}{\to} 0$, alors en particulier nous avons $q_{0,m}(\varphi_n) \to 0$ par le lemme 26.150. Par conséquent la majoration (26.580) nous dit que $\|\varphi_n\|_p \to 0$ également.

□

En résumé, si $\varphi_n \overset{\mathscr{S}(\mathbb{R}^d)}{\to} \varphi$ alors $\varphi_n \overset{L^p}{\to} \varphi$.

Théorème 26.152 ([298]).
Soit μ une mesure sur les boréliens de $\mathbb{R}^n$ finie sur les compacts. Alors $\mathscr{D}(\mathbb{R}^n)$ est dense dans $L^1(\mathbb{R}^n, \mathcal{B}or(\mathbb{R}^n), \mu)$.

Proposition 26.153 ([468]).
La partie $\mathscr{D}(\mathbb{R}^d)$ est dense dans $\mathscr{S}(\mathbb{R}^d)$.

Démonstration. Soit $f \in \mathscr{S}(\mathbb{R}^d)$, et ϕ, une fonction de $\mathscr{D}(\mathbb{R}^d)$ telle que $\phi(x)=1$ pour $|x|\leqslant 1$ (l'existence de telles fonctions est discutée en 14.13.1). Soit aussi $\phi_k(x)=\phi(x/k)$. Nous posons

$$f_k(x) = \phi_k(x) f(x), \tag{26.581}$$

et nous allons prouver que pour tout multiindices α et γ,

$$p_{\alpha,\gamma}(f_k - f) = \|x^\gamma \partial^\alpha (f_k - f)\|_\infty \to 0. \tag{26.582}$$

Pour cela nous allons noter $\beta \leqslant \alpha$ lorsque β est un multiindice contenu dans α. En utilisant la dérivée du produit nous avons

$$(\partial^\alpha f_k)(x) = \sum_{\beta\leqslant\alpha} (\partial^{\alpha-\beta}\phi_k)(x)\partial^\beta f(x) \tag{26.583a}$$

$$= \sum_{\beta\leqslant\alpha} k^{-|\alpha-\beta|}(\partial^{\alpha-\beta}\phi)(x/k)(\partial^\beta f)(x) \tag{26.583b}$$

$$= \sum_{\beta<\alpha} k^{-|\alpha-\beta|}(\partial^{\alpha-\beta}\phi)(x/k)(\partial^\beta f)(x) + \phi(x/k)(\partial^\alpha f)(x). \tag{26.583c}$$

Nous devons donc étudier et majorer

$$\begin{aligned}\sup_{x\in\mathbb{R}^d} |x^\gamma \partial^\alpha(f_k - f)| \leqslant \sup \Big|x^\gamma \sum_{\beta<\alpha} k^{-|\alpha-\beta|}(\partial^{\alpha-\beta}\phi)(x/k)(\partial^\beta f)(x)\Big| \\ + \sup \big|x^\gamma\big(\phi(x/k) - 1\big)(\partial^\alpha f)(x)\big|\end{aligned} \tag{26.584}$$

En ce qui concerne le second terme, soit $\epsilon > 0$, vu que f est Schwartz, il existe R tel que

$$|x^\gamma(\partial^\alpha f)(x)| < \epsilon \tag{26.585}$$

dès que $\|x\| > R$. En prenant $k > R$,

$$|x^\gamma(\partial^\alpha f)(x)| \begin{cases} = 0 & \text{si } \|x\| < R \\ \leqslant \epsilon & \text{si } \|x\| > R. \end{cases} \tag{26.586}$$

En ce qui concerne le premier terme,

$$\sup_{x\in\mathbb{R}^d} \Big|x^\gamma \sum_{\beta<\alpha} k^{-|\alpha-\beta|}(\partial^{\alpha-\beta}\phi)(x/k)(\partial^\beta f)(x)\Big| \tag{26.587a}$$

$$\leqslant \frac{1}{k}\sup_x \Big|\sum_{\beta<\alpha}(\partial^{\alpha-\beta}\phi)(x/k)(x^\gamma\partial^\beta f)(x)\Big| \tag{26.587b}$$

$$= \frac{1}{k}\sup_x \Big|\sum_{\beta<\alpha}(\partial^{\alpha-\beta}\phi)(x/k)p_{\beta,\gamma}(f)\Big| \tag{26.587c}$$

La somme ne contient qu'un nombre fini de β différents, donc nous pouvons considérer un nombre K qui majore tous les $p_{\beta,\gamma}(f)$ en même temps. La partie avec ϕ peut être majorée par $\|\partial^{\alpha-\beta}\phi\|_\infty$ (qui est fini) dont nous pouvons prendre le maximum sur $\beta < \alpha$. Toute l'expression dans la somme est donc majorée par un nombre qui ne dépend ni de x ni de β. Vu que la somme est finie, elle est majorée par ce nombre multiplié par le nombre de termes dans la somme et au final

$$\sup_{x\in\mathbb{R}^d} \Big|x^\gamma \sum_{\beta<\alpha} k^{-|\alpha-\beta|}(\partial^{\alpha-\beta}\phi)(x/k)(\partial^\beta f)(x)\Big| \leqslant \frac{K'}{k}. \tag{26.588}$$

La limite $k \to \infty$ ne fait alors plus de doutes. □

Remarque 26.154.

Vu la topologie de $\mathscr{S}(\mathbb{R}^d)$ (définition 26.148), la convergence $f_k \xrightarrow{\mathscr{S}(\mathbb{R}^d)} f$ peut être exprimée par le fait que pour tout k, l,

$$t^k f_n^{(l)} \xrightarrow{unif} t^k f^{(l)}. \tag{26.589}$$

C'est-à-dire convergence uniforme de toutes les dérivées multipliées par n'importe quel polynôme.

26.14.2 Produit de convolution

Proposition 26.155 (Stabilité de Schwartz par convolution[60] [455]).
Si $\varphi \in L^1(\mathbb{R}^d)$ et $\psi \in \mathscr{S}(\mathbb{R}^d)$, alors $\varphi * \psi \in \mathscr{S}(\mathbb{R}^d)$.

Démonstration. Nous devons prouver que

$$p_{\alpha,\beta}(\varphi * \psi) = \sup_{x\in\mathbb{R}^d} |x^\beta(\partial^\alpha(\varphi * \psi))(x)| \tag{26.590}$$

est borné pour tout multiindices α et β. En appliquant $|\alpha|$ fois la proposition 26.57, nous mettons toutes les dérivées sur ψ : $\partial^\alpha(\varphi * \psi) = (\varphi * \partial^\alpha\psi)$. Cela étant fait, nous majorons

$$\big|x^\beta(\varphi * \partial^\alpha\psi)(x)\big| \leqslant |x^\beta| \int_{\mathbb{R}^d} |\varphi(y)| \underbrace{\big|(\partial^\alpha\psi)(x-y)\big|}_{\leqslant\|\partial^\alpha\psi\|_\infty} dy\big| \tag{26.591a}$$

$$\leqslant |x^\beta| \|\partial^\alpha\psi\|_\infty \int_{\mathbb{R}^d} |\varphi(y)| dy \tag{26.591b}$$

$$\leqslant p_{\alpha,\beta}(\psi)\|\varphi\|_{L^1}. \tag{26.591c}$$

Par conséquent, $p_{\alpha,\beta}(\varphi * \psi) \leqslant \|\varphi\|_{L^1} p_{\alpha,\beta}(\psi) < \infty$. □

26.15 Théorème de Montel

Théorème 26.156 (Montel[54]).
Soient Ω un ouvert de $\mathbb{C}$ et $\mathcal{F}$ une famille de fonctions holomorphes sur Ω, uniformément bornée sur tout compact de Ω. Alors de toute suite dans $\mathcal{F}$ nous pouvons extraire une sous-suite convergeant uniformément sur tout compact de Ω.

Démonstration. **Un ensemble équicontinu** Nous commençons par prendre une suite de compacts dans Ω comme dans le lemme 7.312, et une suite δ_n de réels strictement positifs tels que

$$B(z, 2\delta_n) \subset K_{n+1} \tag{26.592}$$

pour tout $z \in K_n$. Soient $x, y \in K_n$ tels que $|x - y| < \delta_n$; nous notons $\partial B(x, 2\delta_n)$ le cercle de rayon $2\delta_n$ autour de x, parcouru dans le sens positif. La formule de Cauchy 25.36 nous donne

$$f(x) - f(y) = \frac{1}{2\pi i}\int_{\partial B}\left(\frac{f(\xi)}{\xi - x} - \frac{f(\xi)}{\xi - y}\right) d\xi = \frac{x-y}{2\pi i}\int_{\partial B}\frac{f(\xi)}{(\xi - x)(\xi - y)} d\xi \tag{26.593}$$

Nous majorons ça par

$$\big|f(x) - f(y)\big| \leqslant \frac{|x-y|}{2\pi}\int_{\partial B}\frac{|f(\xi)|}{2\delta_n^2} d\xi \leqslant \frac{|x-y|}{\delta_n} M_n. \tag{26.594}$$

Justifications :

— $|\xi - x| = 2\delta_n$ et $|\xi - y| \geqslant \delta_n$ parce que ξ est au mieux sur le rayon passant par x et y.

— $|f(\xi)| \leqslant M_n$ où M_n est la borne uniforme de $\mathcal{F}$ sur le compact K_n.

— Nous avons aussi fini par calculer l'intégrale dans laquelle il ne restait plus rien, ça a donné la circonférence du cercle de rayon $2\delta_n$.

Jusqu'à présent nous avons prouvé que l'ensemble

$$\mathcal{F}_n = \{f|_{K_n} \text{ tel que } f \in \mathcal{F}\} \tag{26.595}$$

est équicontinu. Il est aussi équiborné par hypothèse.

60. Définition 26.52.

Application du théorème d'Ascoli L'ensemble $\mathcal{F}_n$ vérifie les hypothèses du théorème d'Ascoli 26.5. Donc l'ensemble $\mathcal{F}_n$ est relativement compact dans $C(K_n, \mathbb{C})$ pour la norme uniforme. Autrement dit l'ensemble $\bar{\mathcal{F}}$ est compact et si nous avons une suite de fonctions dans $\mathcal{F}_n$, il existe une sous-suite convergeant dans $\bar{\mathcal{F}}_n$, c'est-à-dire uniformément. Autrement dit il existe une fonction strictement croissante $\varphi\colon \mathbb{N} \to \mathbb{N}$ telle que la suite $k \mapsto f_{\varphi(k)}$ converge uniformément sur K_n. La limite n'est cependant pas spécialement dans $\mathcal{F}_n$.

L'argument diagonal La suite $k \mapsto f_{\varphi_1 \circ \ldots \varphi_k(k)}$ converge uniformément sur tous les K_n. Si K est un compact de Ω, alors les petites propriétés sympas du lemme 7.312 nous disent que $K \subset \mathrm{Int}(K_m)$ pour un certain m. Ladite suite convergeant uniformément sur K_m, elle converge uniformément sur K et nous avons montré la convergence uniforme sur tout compact de Ω. □

Corolaire 26.157 ([54]).
Soient Ω un ouvert connexe borné de $\mathbb{C}$ et $a \in \Omega$. Soit f holomorphe sur Ω telle que $f(a) = a$ et $|f'(a)| < 1$.

Alors de (f^n) on peut extraire une sous-suite convergeant uniformément sur tout compact de Ω vers la fonction constante a.

Démonstration. Nous considérons un voisinage de a inclus dans Ω ; sachant que $|f(a)| < 1$, nous trouvons un voisinage encore plus petit de a sur lequel $|f'(z)| < 1$. Soit donc r tel que $\overline{B(a,r)} \subset \Omega$ et tel que $|f'(z)| < 1$ sur $\overline{B(a,r)}$. Étant donné que $f'(z)$ est continue sur le compact $\overline{B(a,r)}$, nous en prenons le maximum λ (qui est strictement inférieur à 1) et nous avons au final

$$|f'(z)| \leqslant \lambda < 1 \tag{26.596}$$

pour tout $z \in \overline{B(a,r)}$. Le théorème des accroissements finis 11.257 nous dit que

$$\big|f(z) - a\big| \leqslant \lambda |z - a| \tag{26.597}$$

pour tout $z \in \overline{B(a,r)}$. C'est ici que nous utilisons l'hypothèse de convexité de Ω. Nous montrons alors par récurrence que

$$\big|f^n(z) - a\big| \leqslant \lambda^n |z - a| \leqslant \lambda^n r \leqslant r. \tag{26.598}$$

L'ensemble $A = \{f^n \text{ tel que } n \geqslant 1\}$ est donc uniformément borné sur $\overline{B(a,r)}$ par $a + r$. Autre manière de le dire : pour tout $z \in \overline{B(a,r)}$ nous avons

$$f^n(z) \in \overline{B(a,r)}. \tag{26.599}$$

La suite (f^n) est donc uniformément bornée sur tout compact de $B(a,r)$. Le théorème de Montel 26.156 nous indique que l'on peut extraire une sous-suite convergente uniformément sur tout compact. Au vu de (26.598) cette convergence ne peut avoir lieu que vers une fonction g qui vaut la constante a sur $B(a,r)$.

D'autre par la fonction g est holomorphe en tant que limite uniforme de fonctions holomorphes, théorème 25.65. Or une fonction holomorphe constante sur un ouvert est constante sur tout son domaine d'holomorphie (principe d'extension analytique, théorème 16.128). □

26.16 Espaces de Bergman

Source : [311].

Soit Ω un borné dans $\mathbb{C}$ et D le disque unité ouvert de $\mathbb{C}$.

Définition 26.158.
*L'**espace de Bergman** sur Ω, noté $A^2(\Omega)$ est l'espace des fonctions holomorphes sur Ω qui sont en même temps dans $L^2(\Omega)$.*

Nous mettons sur $A^2(\Omega)$ le produit scalaire usuel hérité de L^2 :

$$\langle f, g\rangle = \int_\Omega f(z)\overline{g(z)}dz. \tag{26.600}$$

Lemme 26.159.
Soient un compact $K \subset \Omega$ et une fonction $f \in A^2(\Omega)$. Alors

$$\max_{z\in K} |f(z)| \leqslant \frac{1}{\sqrt{\pi}}\frac{1}{d(K,\partial\Omega)}\|f\|_2. \tag{26.601}$$

Démonstration. Soient $a \in \Omega$ et $r > 0$ tels que $B(a,r) \subset \Omega$. Nous considérons aussi $\rho \leqslant r$. La formule de Cauchy (25.36) nous donne

$$f(a) = \frac{1}{2\pi i}\int_{B(a,\rho)} \frac{f(\xi)}{\xi - a} f\xi = \frac{1}{2\pi}\int_0^{2\pi} f(a + \rho e^{i\theta})d\theta \tag{26.602}$$

où nous avons utilisé le chemin $\gamma(\theta) = a + \rho e^{i\theta}$, $\gamma'(\theta) = i\rho e^{i\theta}$ et $\rho = |\xi - a|$. Maintenant une astuce est d'écrire

$$\frac{r^2}{2}f(a) = \int_0^r f(a)\rho d\rho, \tag{26.603}$$

et d'y substituer la valeur de $f(a)$ que nous venons de calculer :

$$\begin{aligned}
\frac{r^2}{2}f(a) &= \int_0^r \frac{1}{2\pi}\int_0^{2\pi} f(a + \rho e^{i\theta})d\theta\rho d\rho && \text{(26.604a)}\\
&= \frac{1}{2\pi}\int_{B(a,r)} f(z)dz && \text{passage aux polaires} \quad \text{(26.604b)}\\
&= \frac{1}{2\pi}\langle 1, f\rangle_B && \text{produit scalaire sur } B(a,r) \quad \text{(26.604c)}\\
&\leqslant \frac{1}{2\pi}\sqrt{\langle 1,1\rangle_B\langle f,f\rangle_B} && \text{(26.604d)}
\end{aligned}$$

Nous avons donc

$$r^2 f(a) \leqslant \frac{1}{\pi}\sqrt{\langle 1,1\rangle_B\langle f,f\rangle_B}, \tag{26.605}$$

et donc

$$\pi r^2 f(a) \leqslant \sqrt{\pi r^2}\|f\|_2, \tag{26.606}$$

parce que $\langle f, f\rangle_B \leqslant \|f\|_2^2$. En effet le produit scalaire $\|.\|_2$ est donné par une intégrale sur Ω alors que $B(a,r) \subset \Omega$ et que la fonction qu'on y intègre est positive (c'est $|f(z)|^2$). En simplifiant,

$$f(a) \leqslant \frac{1}{\sqrt{\pi}r}\|f\|_2. \tag{26.607}$$

Mais r a été choisi pour avoir $B(a,r) \subset \Omega$, donc $r \leqslant d(a,\partial\Omega)$ et

$$|f(a)| \leqslant \frac{1}{d(a,\partial\Omega)\sqrt{\pi}}\|f\|_2. \tag{26.608}$$

Maintenant si nous prenons $a \in K$, nous avons encore la minoration $d(a,\partial K) \leqslant d(a,\partial\Omega)$ et donc

$$|f(a)| \leqslant \frac{1}{d(a,\partial K)\sqrt{\pi}}\|f\|_2. \tag{26.609}$$

□

Théorème 26.160.
Soit Ω un ouvert de $\mathbb{C}$.

(1) L'espace $A^2(\Omega)$ est un espace de Hilbert.

(2) Si D est la boule unité dans $\mathbb{C}$, une base hilbertienne de $A^2(D)$ est donnée par les fonctions

$$e_n(z) = \sqrt{\frac{n+1}{\pi}} z^n \tag{26.610}$$

pour $n \geqslant 0$.

Démonstration. Nous commençons par montrer que $A^2(\Omega)$ est complet. Pour cela nous considérons une suite de Cauchy (f_n) dans $A^2(\Omega)$ et un compact $K \subset \Omega$. Nous savons par le lemme 26.159 que

$$\max_{z\in K} \big|f_n(z) - f_m(z)\big| \leqslant \frac{1}{\sqrt{\pi} d(K, \partial\Omega)} \|f_n - f_m\|_2. \tag{26.611}$$

Donc f_n converge uniformément sur K. Par le théorème de Weierstrass 25.65, la fonction f est holomorphe. Il existe donc une fonction holomorphe f qui est limite uniforme sur tout compact de Ω de la suite (f_n).

Mais $L^2(\Omega)$ étant complet, la suite (f_n) a une limite $g \in L^2(\Omega)$. Ce que nous voudrions faire est prouver que $f = g$. Notons que tel quel, ce n'est pas vrai parce que f est une vraie fonction alors que g est une classe. Ce que nous enseigne la proposition 26.19 est qu'il existe une sous-suite (qu'on note (g_n)) qui converge vers g presque partout. Dans cette dernière phrase, g_n et g sont de vraies fonctions, des représentants des classes dans L^2.

Nous déduisons que $f = g$ presque partout (ici f et g sont les fonctions) parce que la sous-suite converge uniformément vers f en même temps que presque partout vers g. Donc $f = g$ dans $L^2(\Omega)$ (ici f et g sont les classes). Donc $f \in L^2(\Omega)$ et l'espace $A^2(\Omega)$ est de Hilbert.

Il nous faut encore prouver que $(e_n)_{n\geqslant 0}$ est une base orthonormale. En ce qui concerne les produits scalaires,

$$\begin{aligned}
\langle e_m, e_n\rangle &= \sqrt{\frac{(m+1)(n+1)}{\pi}} \int_D z^n \overline{z^m} dz && (26.612a)\\
&= \sqrt{\frac{(m+1)(n+1)}{\pi^2}} \int_0^1 \rho\, d\rho \int_0^{2\pi} d\theta \rho^{m+n} e^{i\theta(n-m)} && (26.612b)\\
&= \sqrt{\frac{(m+1)(n+1)}{\pi^2}} \frac{1}{m+n+2} \underbrace{\int_0^{2\pi} e^{i\theta(n-m)} d\theta}_{2\pi\delta_{mn}} && (26.612c)\\
&= \sqrt{\frac{(n+1)^2}{\pi^2}} \frac{1}{2n+2} 2\pi\delta_{nm} && (26.612d)\\
&= \delta_{nm}. && (26.612e)
\end{aligned}$$

Donc les fonctions données sont bien orthonormales. Nous devons montrer qu'elles sont denses dans $A^2(D)$. Soit $f \in A^2(D)$ et $c_n(f) = \langle f, e_n\rangle$; nous allons montrer que

$$\|f\|_2^2 = \sum_{n=0}^{\infty} |\langle f, e_n\rangle|^2, \tag{26.613}$$

parce que le point (5) du théorème 24.44 nous indique que ce sera suffisant pour avoir une base hilbertienne.

Étant donné que f est holomorphe sur D, le théorème 25.13 nous développe f en série entière :

$$f(z) = \sum_{k=0}^{\infty} a_k z^k. \tag{26.614}$$

En permutant la somme avec le produit scalaire,

$$c_n(f) = \int_D f(z)\bar{e}_n(z) = \sqrt{\frac{n+1}{\pi}} \int_D f(z)\bar{z}^n dz. \tag{26.615}$$

Afin de profiter de la convergence uniforme de la série (26.614) à l'intérieur de D, nous allons exprimer l'intégrale sur D comme une intégrale sur $|z| < r$ en faisant tendre r vers 1 (par le bas). Pour ce faire nous considérons les fonctions

$$g_k(z) = \begin{cases} f(z)\bar{z}^n & \text{si } |z| < 1 - 1/k \\ 0 & \text{sinon.} \end{cases} \tag{26.616}$$

Ces fonctions sont intégrables sur D et dominées par $f(z)\bar{z}^n$ qui est intégrable sans dépendre de k. Mais nous avons évidemment $g_k(z) \to f(z)\bar{z}^n$. Le théorème de la convergence dominée permet alors de permuter l'intégrale et la limite $k \to \infty$. Cela nous permet d'écrire

$$c_n(f) = \sqrt{\frac{n+1}{\pi}} \lim_{r\to 1^-} \int_{|z|<r} \bar{z}^n f(z) dz = \sqrt{\frac{n+1}{\pi}} \lim_{r\to 1^-} \int_{|z|<r} \sum_{k=0}^{\infty} a_k z^k \bar{z}^n. \tag{26.617}$$

Par la convergence uniforme de la série entière *à l'intérieur* du disque D nous pouvons permuter l'intégrale et la somme (proposition 14.30) :

$$c_n(f) = \sqrt{\frac{n+1}{\pi}} \lim_{r\to 1^-} \sum_{k=0}^{\infty} a_k \int_{|z|<r} z^k \bar{z}^n dz. \tag{26.618}$$

L'intégrale proprement dite est vite calculée et vaut

$$\int_{|z|<1} \bar{z}^n z^k dz = \frac{\pi r^{2n+2}}{n+1} \delta_{kn}. \tag{26.619}$$

Nous pouvons donc continuer le calcul de $c_n(f)$ en effectuant la somme sur k qui se réduit à changer k en n puis en effectuant la limite :

$$c_n(f) = \sqrt{\frac{n+1}{\pi}} \lim_{r\to 1^-} \sum_k a_k \frac{\pi r^{2n+2}}{n+1} \delta_{kn} = \sqrt{\frac{\pi}{n+1}} a_n. \tag{26.620}$$

Nous effectuons le même genre de calculs pour évaluer $\|f\|_2^2$:

$$\|f\|_2^2 = \int_D |f(z)|^2 dz \tag{26.621a}$$

$$= \lim_{r\to 1^-} \int_{|z|<r} f(z) \sum_{k=0}^{\infty} \bar{a}_k \bar{z}_k dz \tag{26.621b}$$

$$= \lim_{r\to 1^-} \sum_{k=0}^{\infty} \bar{a}_k \int_{|z|<r} f(z) \bar{z}^k dz \qquad \text{permuter } \sum \text{ et } \int \tag{26.621c}$$

$$= \lim_{r\to 1^-} \sum_{k=0}^{\infty} \bar{a}_k a_k \frac{\pi r^{2k+2}}{k+1} \qquad \text{intégrale déjà faite.} \tag{26.621d}$$

Mais nous savons déjà que $c_n(f) = \sqrt{\pi/(n+1)}$, donc ce qui est dans la somme est $\pi \bar{a}_k a_k/(n+1) = |c_k(f)|^2$. Nous avons donc

$$\|f\|_2^2 = \lim_{r\to 1^-} \sum_{k=0}^{\infty} |c_k(f)|^2 r^{2k+2}. \tag{26.622}$$

La fonction (de r) constante $|c_k(f)|^2$ domine $|c_k(f) r^{2k+2}|$ tout en ayant une somme (sur k) qui converge ; en effet la proposition 24.24 nous indique que $\sum_j |c_k(f)|^2 \leqslant \|f\|_2^2$. Le théorème de la convergence dominée nous permet d'inverser la limite et la somme pour obtenir le résultat attendu :

$$\|f\|_2^2 = \sum_{k=0}^{\infty} |c_k(f)|^2. \tag{26.623}$$

□

Chapitre 27

Séries de Fourier

27.1 Densité des polynômes trigonométriques

27.1.1 Convergence pour les fonctions continues (via Weierstrass)

Le résultat fondamental qui nous permet d'utiliser les polynômes trigonométriques comme base pour les fonctions *continues* périodiques est le suivant. Notons que pour les fonctions non continues, il y a encore du travail.

Lemme 27.1.
Si $f\colon \mathbb{R} \to \mathbb{C}$ est une fonction continue 2π-périodique et si $\epsilon > 0$, alors il existe un polynôme trigonométrique P tel que $\|f - P\|_\infty \leqslant \epsilon$.

Démonstration. Nous allons utiliser le théorème de Stone-Weierstrass 11.317. Soit le compact Hausdorff

$$S^1 = \{z \in \mathbb{C} \text{ tel que } |z| = 1\}, \tag{27.1}$$

et $C(S^1, \mathbb{C})$ l'algèbre des fonctions continues de S^1 vers $\mathbb{C}$. Il suffit de vérifier que les polynômes trigonométriques vérifient les hypothèse du théorème de Stone-Weierstrass. Un polynôme trigonométrique est un polynôme en z et $\bar{z}$ défini sur S^1.

(1) Le polynôme constant est dans l'algèbre, ok.
(2) Pour la séparation des points, le polynôme trigonométrique $x \mapsto e^{ix}$.
(3) Si P est un polynôme en z et $\bar{z}$, alors $\bar{P}$ l'est encore.

Donc si $\epsilon > 0$ et $\tilde{f} \in C(S^1, \mathbb{C})$ sont donnés, il existe un polynôme trigonométrique P tel que

$$\sum_t |\tilde{f}(e^{it}) - P(t)| < \epsilon. \tag{27.2}$$

Soit $f\colon \mathbb{R} \to \mathbb{C}$ une fonction continue 2π-périodique. Nous considérons $\tilde{f} \in C(S^1, \mathbb{C})$ donnée par $\tilde{f}(e^{it}) = f(t)$. Alors $\sup_t |f(t) - P(t)| \leqslant \epsilon$. □

27.1.2 Convergence pour les fonctions continues (via Fejér)

Si nous ne voulons pas passer par le gros théorème de Stone-Weierstrass pour prouver la densité des polynômes trigonométrique dans $\big(C^0_{2\pi}, \|.\|_\infty\big)$, nous pouvons passer par le gros théorème de Fejér. C'est ce que nous faisons maintenant.

Le **noyau de Dirichlet** est la fonction

$$D_n(t) = \sum_{k=-n}^{n} e^{int}. \tag{27.3}$$

Le **noyau de Fejér** est la moyenne de Cesaro des noyaux de Dirichlet :

$$F_n(t) = \frac{1}{n} \sum_{k=0}^{n-1} D_k(t). \tag{27.4}$$

Lemme 27.2.
Le noyau de Dirichlet s'exprime sous la forme

$$D_n(t) = \sum_{k=-n}^{n} e^{-ikt} = \frac{\sin\left(\frac{2n+1}{2}t\right)}{\sin(t/2)} \tag{27.5}$$

Note : ce noyau n'est pas positif.

Démonstration. Nous commençons par mettre en facteur le premier terme :

$$D_n(t) = \sum_{k=-n}^{n} e^{int} = e^{-int} \sum_{k=0}^{2n} e^{ikt}. \tag{27.6}$$

En utilisant la formule de la somme géométrique,

$$\begin{aligned}
D_n(t) &= e^{-int} \frac{1-(e^{it})^{2n+1}}{1-e^{it}} && (27.7a)\\
&= e^{-int} \frac{1-e^{(2n+1)it}}{1-e^{it}} && (27.7b)\\
&= e^{-int} \frac{e^{(2n+1)it/2}}{e^{i\frac{t}{2}}} \frac{e^{-(2n+1)it/2} - e^{(2n+1)it/2}}{e^{-it/2} - e^{it/2}} && (27.7c)\\
&= \frac{(-2i)\sin\left(\frac{2n+1}{2}t\right)}{(-2i)\sin\left(\frac{t}{2}\right)}. && (27.7d)
\end{aligned}$$

□

Théorème 27.3 (Théorème de Dirichlet)**.**
Soit f une fonction 2π-périodique et C^1 par morceaux. Pour tout $x \in \mathbb{R}$ nous posons

$$s_n(x) = \sum_{k=-n}^{n} c_k(f) e^{ikx}. \tag{27.8}$$

Alors nous avons

$$\lim_{n\to\infty} s_n(x) = \frac{f(x^+) + f(x^-)}{2}. \tag{27.9}$$

Lemme 27.4.
Le noyau de Fejér s'exprime sous la forme

$$F_n(t) = \frac{1}{n}\left(\frac{\sin\frac{nt}{2}}{\sin\frac{t}{2}}\right)^2. \tag{27.10}$$

Note : ce noyau est positif. C'est important parce qu'on s'en sert dans la preuve du théorème de Fejér.

Démonstration. L'astuce est de noter $\sin(x) = \Im(e^{ix})$ et de repartir du résultat à propos du noyau

de Dirichlet. En utilisant encore la formule de la série géométrique partielle [1],

$$F_n(t) = \frac{1}{n\sin(t/2)}\Im\sum_{k=0}^{n-1} e^{(2k+1)it/2} \tag{27.11a}$$

$$= \frac{1}{n\sin(t/2)}\Im e^{\frac{it}{2}}\sum_{k=0}^{n-1} \tag{27.11b}$$

$$= \frac{1}{n\sin(t/2)}\Im e^{\frac{it}{2}}\left(\frac{1-e^{nit}}{1-e^{it}}\right) \tag{27.11c}$$

$$= \frac{1}{n\sin(t/2)}\Im e^{it/2}\frac{e^{\frac{nit}{2}}\left(e^{-\frac{int}{2}}-e^{\frac{nit}{2}}\right)}{e^{\frac{it}{2}}\left(e^{-it/2}-e^{it/2}\right)} \tag{27.11d}$$

$$= \frac{1}{n\sin(t/2)}\underbrace{\Im e^{\frac{nit}{2}}}_{\sin(nt/2)}\frac{\sin\left(\frac{nt}{2}\right)}{\sin(\frac{t}{2})} \tag{27.11e}$$

$$= \frac{1}{n}\left(\frac{\sin\frac{nt}{2}}{\sin\frac{t}{2}}\right)^2. \tag{27.11f}$$

□

Théorème 27.5 (Fejèr)**.**
Soit $f\colon \mathbb{R}\to\mathbb{C}$ une fonction continue et 2π-périodique. Pour tout $k\in\mathbb{Z}$ nous notons

$$\begin{aligned} e_k\colon \mathbb{R}&\to\mathbb{C}\\ x&\mapsto e^{ikx}. \end{aligned} \tag{27.12}$$

Pour chaque $n\in\mathbb{N}$ nous posons

$$D_n = \sum_{k=-n}^{n} e_k \qquad \tilde{S}_n(f) = \sum_{k=-n}^{n} c_k(f)e_k \tag{27.13a}$$

$$F_n = \frac{D_0+\cdots+D_{n-1}}{n} \qquad \tilde{F}_n = \sigma_n(f) = \frac{1}{n}\sum_{k=0}^{n-1} S_k(f). \tag{27.13b}$$

Alors

(1) $\frac{1}{2\pi}\int_{-\pi}^{\pi} F_n(t)dt = 1$.
(2) Pour tout $\alpha\in\,]0,\pi[$, F_n converge uniformément vers 0 sur $[-\pi,\pi]\backslash[-\alpha,\alpha]$.
(3) La suite $\tilde{F}_n$ converge uniformément sur $\mathbb{R}$ vers f.
(4) Le système trigonométrique $\{e_k\}_{k\in\mathbb{Z}}$ est total pour l'espace $\left(C^0(S^1),\|.\|_\infty\right)$ des fonctions continues 2π-périodiques.

Démonstration. Un calcul usuel montre que

$$\int_{-\pi}^{\pi} e_l(t)dt = \begin{cases} 0 & \text{si } l\neq 0\\ 2\pi & \text{si } l=0 \end{cases} \tag{27.14}$$

Nous avons alors

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} F_n(t)dt = \frac{1}{2\pi}\frac{1}{n}\sum_{k=0}^{n-1}\sum_{l=-k}^{k}\underbrace{\int_{-\pi}^{\pi} e_l(t)dt}_{2\pi\delta_l} = \frac{1}{n}\sum_{k=0}^{n-1} 1 = 1. \tag{27.15}$$

Cela prouve déjà le premier point.

1. Voir l'exemple 10.67.

Pour le second point, en partant de l'expression (27.10) et en considérant $x \in [-\pi, \pi,]\backslash[-\alpha, \alpha]$ (ce qui nous évite l'annulation du dénominateur),

$$|F_n(x)| \leqslant \frac{1}{(n+1)\sin^2(\alpha/2)}, \tag{27.16}$$

et donc $F_n \to 0$ uniformément sur l'ensemble considéré.

Nous passons maintenant à cette histoire de convergence uniforme de la moyenne de Cesaro vers f. Pour tout $n \in \mathbb{N}$ nous avons

$$\tilde{D}_n(x) = \frac{1}{2\pi} \sum_{k=-n}^{n} \left(\int_{-\pi}^{\pi} f(t) e^{-ikt} dt \right) e^{ikx} \tag{27.17a}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \sum_{k=-n}^{n} e_k(x-t) \tag{27.17b}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) D_k(x-t). \tag{27.17c}$$

Par conséquent, en effectuant le changement de variable $u = x - t$ et la périodicité,

$$\tilde{F}_n(x) = \int_{-\pi}^{\pi} f(t) F_n(x-t) dt \tag{27.18a}$$

$$= -\int_{x+\pi}^{x-\pi} f(x-u) F_n(u) du \tag{27.18b}$$

$$= \int_{-\pi}^{\pi} f(x-u) F_n(u) du. \tag{27.18c}$$

Nous prouvons à présent l'uniforme continuité. Soit $\epsilon > 0$; étant donné que f est continue et 2π-périodique, elle est uniformément continue et nous considérons $\delta > 0$ tel que $|x - y| < \delta$ implique $\big|f(x) - f(y)\big| < \epsilon$. Soit M un majorant de $|f|$ sur $\mathbb{R}$. L'équation (27.18) nous donne

$$\big|f(x) - \tilde{F}_n(x)\big| = \Big|\frac{1}{2\pi} \int_{-\pi}^{\pi} \big(f(x-t) - f(x)\big) F_n(t) dt\Big| \tag{27.19a}$$

$$\leqslant \frac{1}{2\pi} \int_{\delta \leqslant |t| \leqslant \pi} |2M F_n(t)| dt + \frac{1}{2\pi} \int_{-\delta}^{\delta} \epsilon |F_n(t)| dt \tag{27.19b}$$

$$\leqslant \frac{2M}{2\pi} \int_{\delta \leqslant |t| \leqslant \pi} F_n(t) dt + \epsilon' \tag{27.19c}$$

Pour obtenir (27.19a) nous avons pu rentrer $f(x)$ dans l'intégrale en utilisant le premier point. Pour obtenir (27.19c) nous avons d'abord utilisé la positivité de F_n (lemme 27.4) pour enlever les valeurs absolues, et nous avons ensuite utilisé le fait que son intégrale valait 2π.

Étant donné que $F_n \to 0$ uniformément sur $[-\pi, \pi,]\backslash[-\alpha, \alpha]$, il existe un N tel que

$$\int_{\delta \leqslant |t| \leqslant \pi} F_n(t) dt \leqslant \epsilon \tag{27.20}$$

dès que $n > N$. Le résultat découle.

Pour le point (4), il suffit de remarquer que chacun des $\tilde{F}_n$ est une combinaison finie d'éléments du système trigonométrique. □

27.1.3 Densité dans L^p

Nous venons de voir (de deux façons différentes) que les polynômes trigonométriques étaient dense dans $\big(C^0_{2\pi}(\mathbb{R}), \|.\|_\infty\big)$. Nous avons aussi déjà vu par le théorème 26.70 que ces polynômes trigonométriques étaient denses dans $L^p(S^1)$. Nous présentons à présent une autre façon de prouver cette dernière densité.

Théorème 27.6.
Les polynômes trigonométriques sont denses dans $L^p(S^1)$ pour $1 \leqslant p < \infty$.

Démonstration. Par les théorèmes 27.1 ou 27.5 (au choix), nous savons que les polynômes trigonométriques sont denses dans $\big(C^0_{2\pi}(S^1), \|.\|_\infty\big)$. Vu que S^1 est compact, la densité est également au sens L^p. En effet si $\|f_n - f\|_\infty \leqslant \epsilon$, alors

$$\|f_n - f\|_\infty = \int_0^{2\pi} |f_n - f|^p \leqslant \int_0^{2\pi} \epsilon^p = 2\pi\epsilon^p. \tag{27.21}$$

Donc les polynômes trigonométriques sont denses dans $\big(C^0_{2\pi}(S^1), \|.\|_p\big)$. Mais nous savons par (un a fortiori sur) le théorème 26.47 que les fonctions continues sont denses dans $L^p(S^1)$.

Par densité de la densité, les polynômes trigonométriques sont denses dans $L^p(S^1)$. □

27.1.4 Suite équirépartie, critère de Weyl

Définition 27.7.
Soit u une suite dans $[0,1]$. Pour $0 \leqslant a \leqslant b \leqslant 1$ nous posons

$$X_n(a,b) = \operatorname{Card}\big\{k \in \{1,\dots,n\} \text{ tel que } u_k \in [a,b]\big\}. \tag{27.22}$$

*Nous disons que la suite u est **équirépartie** si pour tout $0 \leqslant a < b < 1$, on a*

$$\lim_{n\to\infty} \frac{X_n(a,b)}{n} = b - a. \tag{27.23}$$

Voir aussi la remarque 35.140 sur les nombres normaux.

Proposition 27.8 (Critère de Weyl[311, 54])**.**
Soit (x_n) une suite dans $[0,1[$. Les conditions suivantes sont équivalentes.

(1) La suite (x_n) est équirépartie.

(2) Pour toute fonction continue à valeurs réelles sur $[0,1]$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n f(x_k) = \int_0^1 f(x)dx. \tag{27.24}$$

(3) Pour tout $p \in \mathbb{N}^*$ nous avons

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n e^{2i\pi p x_k} = 0. \tag{27.25}$$

Démonstration. On pose

$$S_n(f) = \frac{1}{n}\sum_{k=1}^n f(x_k). \tag{27.26}$$

Une espèce de lemme Supposons connaitre un ensemble de fonctions A dense dans $C^0([0,1])$ pour toutes les fonctions duquel nous avons la limite (27.24). Alors la limite a lieu pour toute fonction de $C^0([0,1])$. En effet, soit $f \in C^0([0,1])$ et $g \in A$ tel que $\|f - g\|_\infty < \epsilon$. Alors

$$\left\|\frac{1}{n}\sum_{k=1}^n f(x_k) - \int_0^1 f(t)dt\right\| \leqslant \left\|\frac{1}{n}\sum_{k=1}^n \big(f(x_k) - g(x_k)\big)\right\| \tag{27.27a}$$

$$+ \left\|\frac{1}{n}\sum_{k=1}^n g(x_k) - \int_0^1 g(t)dt\right\| \tag{27.27b}$$

$$+ \left\|\int_0^1 g(t)dt - \int_0^1 f(t)dt\right\|. \tag{27.27c}$$

Le premier terme se majore par ϵ. Le troisième est la même majoration : $\int_0^1 \big(f(t) - g(t)\big)dt \leqslant \|f - g\|_\infty = \epsilon$. Par hypothèse sur l'espace A, le second terme se majore par ϵ lorsque n est grand.

(1)⇒(2) Nous supposons que la suite est équirépartie et nous commençons par montrer le résultat pour les fonctions en escalier. Soit donc la fonction en escalier $\eta(x) = c_j$ sur $a_{j-1} < x < a_j$. Sur le point a_j lui-même, la fonction η vaut soit c_j soit c_{j+1}. Nous avons

$$\frac{1}{n}\sum_{k=1}^{n}\eta(x_k) = \frac{1}{n}\left[\sum_{j=1}^{m} c_j X_n(a_j, a_{j+1}) - \sum_{j=1}^{m} c_j X_n(a_j, a_j) + \sum_{j=1}^{m}\eta(a_j)X_n(a_j, a_j)\right]. \tag{27.28}$$

À la limite $n \to \infty$, les deux derniers termes tombent [2] et il reste

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\eta(x_k) = \sum_{j=1}^{m} c_j(a_{j-1} - a_j). \tag{27.29}$$

Or par construction, pour une fonction en escalier,

$$\sum_{j=1}^{m} c_j(a_{j-1} - a_j) = \int_0^1 \eta. \tag{27.30}$$

Étant donné que les fonctions en escalier sont denses dans les fonctions continues, l'espèce de lemme plus haut conclut.

(2)⇒(1) Nous prouvons maintenant le sens inverse. C'est-à-dire que pour toute fonction continue sur $[0, 1]$, nous avons

$$\int_0^1 f(x)dx = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f(x_k). \tag{27.31}$$

Nous devons en déduire que (x_n) est équirépartie. Pour ce faire, soit $x \in [0, 1[$ et $\epsilon > 0$ tel que $x + \epsilon < 1$. Nous considérons $\varphi = \mathbb{1}_{[x,1[}$ et

$$\varphi_{\epsilon(t)} = \begin{cases} 0 & \text{si } t \in [0, x[\\ \frac{t-x}{\epsilon} & \text{si } t \in [x, x+\epsilon[\\ 1 & \text{si } t \geqslant x + \epsilon. \end{cases} \tag{27.32}$$

Cela est une fonction continue, donc

$$\lim_{n\to\infty} S_n\big(\varphi_\epsilon(t)\big) = \int_0^1 \varphi_\epsilon(t)dt = \int_x^{x+\epsilon}\frac{t-x}{\epsilon}dt + \int_{x+\epsilon}^1 1dt = 1 - x - \frac{\epsilon}{2}. \tag{27.33}$$

Mais $\varphi_\epsilon \leqslant \varphi$, donc $S_n(\varphi_\epsilon) \leqslant S_n(\varphi)$ et donc

$$\liminf_{n\to\infty} S_n(\varphi) \geqslant 1 - x. \tag{27.34}$$

Notons que nous ne savons pas si la *vraie* limite de gauche existe ; c'est pourquoi nous prenons la limite inférieure, qui existe toujours.

Nous définissons aussi

$$\psi_\epsilon(t) = \begin{cases} 0 & \text{si } t \in [0, x-\epsilon[\\ \frac{t-x+\epsilon}{\epsilon} & \text{si } t \in [x-\epsilon, x[\\ 1 & \text{si } t > x. \end{cases} \tag{27.35}$$

C'est encore une fonction continue et nous trouvons [3]

$$\int_0^1 \psi_\epsilon(t)dt = 1 - x + \frac{\epsilon}{2}. \tag{27.36}$$

2. J'en profite pour mentionner que mon équation (27.28) n'est pas la même que celle de [311] dans laquelle il me semble voir une faute ; quoi qu'il en soit, les termes litigieux tombent.

3. Je recommande chaudement de dessiner les fonctions φ_ϵ et ψ_ϵ pour avoir une idée de la situation.

Vu que $\psi_\epsilon \geqslant \varphi$, nous avons $S_n(\psi_\epsilon) \geqslant S_n(\varphi)$ et donc

$$\limsup_n S_n(\varphi) \leqslant 1 - x. \tag{27.37}$$

Nous avons déjà obtenu que

$$1 - x \leqslant \liminf S_n(\varphi) \leqslant \limsup S_n(\varphi) \leqslant 1 - x, \tag{27.38}$$

donc la limite existe et vaut

$$\lim_{n\to\infty} S_n(\varphi) = 1 - x. \tag{27.39}$$

Cela est pour la fonction caractéristique $\varphi = \mathbb{1}_{[x,1[}$. Si nous prenons une fonction caractéristique $\mathbb{1}_{[a,b]}$, nous avons la même chose parce que $\mathbb{1}_{[a,b[}$ est une combinaisons linéaire de fonctions du type $\mathbb{1}_{[x,1[}$.
Nous avons donc

$$\lim_{n\to\infty} S_n\big(\mathbb{1}_{[a,b]}\big) = b - a, \tag{27.40}$$

alors que le membre de gauche n'est autre que

$$S_n\big(\mathbb{1}_{[a,b]}\big) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{[a,b]}(x_k) = \frac{1}{n}N(n,a,b). \tag{27.41}$$

(2)⇒(3) Vu que $e^{2i\pi p x_k} = \cos(2\pi p x_k) + \sin(2\pi i x_k)$ est une fonction périodique, c'est immédiat.

(3)⇒(2) Par linéarité, le point (2) montre que si f est un polynôme trigonométrique, alors

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} f(x_k) = \int_0^1 f(t)dt. \tag{27.42}$$

Densité des polynômes trigonométriques Il nous reste à prouver que les polynômes trigonométriques sont denses dans les fonctions continues sur $[0,1]$. Soit une fonction continue sur $[0,1]$ avec $f(0) = f(1)$. Alors le théorème de Stone-Weierstrass dans sa version trigonométrique (lemme 27.1) nous donne la densité.

Si $f(1) \neq f(0)$ c'est pas très grave : on peut trouver une fonction g vérifiant $g(0) = g(1)$ et $\|f - g\|_\infty \leqslant \epsilon$. Ensuite un polynôme trigonométrique approxime très bien g. .

□

27.2 Fonctions de Dirichlet

Définition 27.9.
*Une fonction $f\colon \mathbb{R} \to \mathbb{C}$ est une **fonction de Dirichlet** si*

(1) elle est 2π-périodique,

(2) elle est continue par morceaux,

(3) pour tout $x \in \mathbb{R}$ nous avons

$$f(x) = \frac{f(x^+) + f(x^-)}{2}. \tag{27.43}$$

Nous notons $\mathcal{D}$ l'ensemble des fonctions de Dirichlet.

Lemme 27.10 ([469]).
L'ensemble $C^0(S^1)$ est dense dans $\big(\mathcal{D}, \|.\|_2\big)$.

Démonstration. Nous commençons par supposer que $f \in \mathcal{D}$ n'ait qu'un seul point de discontinuité, x_0. Alors nous considérons la fonction f_n qui est égale à f sur $S^1 \backslash B(x_0, \frac{1}{n})$ et qui sur $B(x_n, \frac{1}{n})$ est le segment de droite joignant $f(x_0 - \frac{1}{n})$ et $f(x_0 + \frac{1}{n})$. Cela est une fonction continue, et de plus nous avons

$$|f_n(x)| \leqslant \|f\|_\infty \tag{27.44}$$

pour tout x. En effet si x est en dehors de $B(x_0, \frac{1}{n})$ c'est évident, et si $x \in B(x_0, \frac{1}{n})$, alors $|f_n(x)|$ est majoré soit par $f(x_0 - \frac{1}{n})$ soit par $f(x_0 + \frac{1}{n})$ suivant que le raccord affin soit croissant ou décroissant. Avec ça nous avons

$$\|f_n - f\|_2^2 = \int_{x_0-1/n}^{x_0+1/n} |f(x) - f_n(x)|^2 dx \leqslant \int_{x_0-1/n}^{x_0+1/n} 4\|f\|_\infty = \frac{8\|f\|_\infty}{n}. \tag{27.45}$$

Et nous voyons que $\|f_n - f\|_2 \to 0$.

Si f contient plusieurs points de continuité, on fait le même coup autour de chaque point, en prenant n assez grand pour que si x_0 est un point de discontinuité, $B(x_0, \frac{1}{n})$ n'en contienne pas d'autres. □

Notons que la densité de $C^0(S^1)$ dans $\big(\mathcal{D}, \|.\|_\infty\big)$ est impossible parce qu'une limite uniforme de fonctions continues est continue.

Théorème 27.11.
Le système trigonométrique $\{e_n\}_{n\in\mathbb{Z}}$ est total dans $\big(\mathcal{D}, \|.\|_2\big)$.

Démonstration. Soit $f \in \mathcal{D}$. Si elle est continue, le théorème de Fejèr 27.5 nous donne convergence uniforme sur S^1 d'une suite de polynômes trigonométriques vers f. Cette convergence est également une convergence L^2 parce que S^1 est compact.

Prenons donc $f \in \mathcal{D}$ non continue et $\epsilon > 0$[4]. Par le lemme 27.10, il existe une fonction $g \in C^0(S^1)$ telle que

$$\|g - f\|_2 \leqslant \epsilon. \tag{27.46}$$

Le théorème de Fejèr donne aussi un polynôme trigonométrique P tel que $\|P - g\|_2 < \epsilon$; nous avons alors

$$\|P - f\|_2 \leqslant \|P - g\|_2 + \|g - f\|_2 \leqslant 2\epsilon. \tag{27.47}$$

□

Notons que cette histoire de fonctions de Dirichlet n'a pas attaquée le vrai fond du problème de la densité des polynômes trigonométriques dans $L^2(S^1)$ parce que nous restons avec une hypothèse de continuité, alors que les représentants des éléments de $L^2(S^1)$ n'ont strictement aucune régularité à priori.

27.3 Coefficients et série de Fourier

Définition 27.12.
*La **série de Fourier** associée à f est*

$$f(x) \sim \sum_{n=-\infty}^{\infty} c_n(f) e^{2\pi i \frac{n}{T} x}. \tag{27.48}$$

Cette expression est pour l'instant purement formelle. Cela ne présume ni de la convergence de la série, ni, au cas où elle serait convergente, que la limite soit f.

Pour la suite nous allons considérer des fonctions périodiques de période 2π, et les coefficients de Fourier de f (quand ils existent) sont alors

$$c_n(f) = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int} dt \tag{27.49}$$

4. Par exemple $\epsilon = 0.4$, mais ce n'est qu'un exemple hein. Si vous en voulez un autre, prenez p, un nombre premier puis calculez $\epsilon = 1/p$.

Proposition 27.13 ([470]).
Soit f une fonction continue et périodique telle que sa série de Fourier converge uniformément. Alors la convergence est vers f.

Démonstration. Notons d'abord que f étant continue sur $[0,2\pi]$, elle y est bornée et L^2. Par conséquent Parseval nous enseigne que

$$\|S_N(f)-f\|_{L^2}\to 0. \tag{27.50}$$

Cela signifie que

$$\lim_{N\to\infty}\frac{1}{2\pi}\int_0^{2\pi}|f(t)-S_N(t)|^2dt=0. \tag{27.51}$$

L'hypothèse de convergence uniforme nous dit que la fonction $|f(t)-S_N(t)|^2$ converge uniformément vers la fonction $|f(t)-S(t)|^2$ où nous avons écrit S la limite de S_N. En permutant la limite et l'intégrale,

$$\frac{1}{2\pi}\int_0^{2\pi}|f(t)-S(t)|^2dt=0, \tag{27.52}$$

ce qui signifie que la fonction $t\mapsto|f(t)-S(t)|^2$ est la fonction nulle. Nous en déduisons que $f=S$. □

Proposition 27.14.
Soit f une fonction 2π-périodique. Si $\sum_{n\in\mathbb{Z}}|c_n(f)|<\infty$, alors pour tout $x\in\mathbb{R}$ nous avons

$$f(x)=\sum_{n\in\mathbb{Z}}c_n(f)e^{inx}. \tag{27.53}$$

De plus, la suite (S_nf) converge uniformément vers f.

Démonstration. Nous posons

$$g(x)=\sum_{n\in\mathbb{Z}}c_n(f)e^{inx}. \tag{27.54}$$

Étant donné les hypothèses, la série de droite converge absolument, la fonction g est continue sur $\mathbb{R}$. Nous avons

$$\big|g(x)-(S_nf)(x)\big|\leqslant\sum_{|k|>n}|c_k(f)|, \tag{27.55}$$

mais le terme de droite tend vers zéro lorsque $n\to\infty$ parce que c'est le reste d'une série convergente. Cela signifie que S_nf converge uniformément vers g.

Par ailleurs nous savons que dans L^2 nous avons la convergence $S_nf\to f$ (parce que f est continue sur le compact $[0,2\pi]$ et donc y est bornée et L^2), ce qui signifie que $g=f$ presque partout au sens L^2. Ces deux fonctions étant continues, elles sont égales partout. □

Théorème 27.15.
Soit f, une fonction C^1 et 2π-périodique. Nous notons $(c_n)_{n\in\mathbb{Z}}$ la suite de ses coefficients de Fourier. Alors $(c_n)\in\ell^1(\mathbb{Z})$ et pour tout $x\in\mathbb{R}$ nous avons

$$f(x)=\sum_{n\in\mathbb{Z}}c_n(f)e^{inx}. \tag{27.56}$$

Démonstration. Soit $n\in\mathbb{Z}$. Nous posons $g(t)=f(t)e^{-int}$. Nous avons

$$0=g(2\pi)-g(0)=\int_0^{2\pi}g'(t)dt=\int_0^{2\pi}\big[f'(t)e^{-int}-inf(t)e^{-int}\big]. \tag{27.57}$$

Du coup, $c_n(f')=inc_n(f)$. La fonction f' étant bornée (parce que continue sur $[0,2\pi]$), elle est de carré intégrable sur $[0,2\pi]$ et par les inégalités de Parseval (théorème 24.44) nous avons

$$\sum_{n\in\mathbb{Z}}|c_n(f')|^2<\infty. \tag{27.58}$$

Par conséquent $(c_n(f')) \in \ell^2(\mathbb{Z})$ et a fortiori $(c_n(f'))_{n\in\mathbb{N}} \in \ell^2(\mathbb{N})$. L'inégalité de Cauchy-Schwarz nous indique alors

$$\sum_{n\in\mathbb{N}} |c_n(f)| = \sum_{n\in\mathbb{N}} \frac{1}{n}|c_n(f')| \leqslant \left(\sum_n \frac{1}{n^2}\right)^{1/2} \left(\sum_n |c_n(f')|^2\right)^{1/2} < \infty. \tag{27.59}$$

Nous procédons de même pour $n < 0$. Cela prouve que

$$\sum_{n\in\mathbb{Z}} |c_n(f)| < \infty. \tag{27.60}$$

□

Corolaire 27.16.
Soient f, g deux fonctions continues et 2π-périodiques. Si $c_n(f) = c_n(g)$ alors $f = g$.

Démonstration. Dans le cas de fonctions continues, le théorème de Fejér nous enseigne que si nous posons

$$S_n(x) = \sum_{k=-n}^{n} c_k(f)e^{ikx} \tag{27.61}$$

alors nous avons la convergence

$$\frac{1}{N+1}\sum_{n=0}^{N} S_n(f)(x) \to f(x). \tag{27.62}$$

C'est-à-dire qu'une fonction continue est déterminée par ses coefficients de Fourier. □

Exemple 27.17.
Considérons la fonction

$$f(x) = 1 - \frac{x^2}{\pi^2} \tag{27.63}$$

sur $[-\pi, \pi]$. Nous la développons en série trigonométrique, et étant paire il n'y a pas de sinus. Un calcul montre que

$$a_0 = \frac{4}{3} \tag{27.64}$$

et

$$a_n = (-1)^{n+1}\frac{4}{n^2\pi^2}, \tag{27.65}$$

de telle sorte que

$$f(x) = \frac{2}{3} - \frac{4}{\pi^2}\sum_{n=1}^{\infty}(-1)^n\frac{\cos(nx)}{n^2}. \tag{27.66}$$

Nous avons $f(\pi) = 0$, mais vu le développement,

$$f(\pi) = \frac{2}{3} - \frac{4}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}, \tag{27.67}$$

donc

$$\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{\pi^2}{6}. \tag{27.68}$$

△

27.3.1 Le contre-exemple que nous attendions tous

Nous montrons maintenant que la continuité et la périodicité ne sont pas suffisantes pour avoir convergence de la série de Fourier.

Proposition 27.18 ([54]).
Soit $C^0_{2\pi}$ l'ensemble des fonctions continues muni de la norme uniforme. Nous définissons

$$S_n(f)(x) = \sum_{k=-n}^{n} c_k(f)e^{ikx}. \tag{27.69}$$

Alors il existe $f \in C^0_{2\pi}$ tel que la suite $n \mapsto S_n(f)(0)$ soit divergente. En particulier f n'est pas la somme de sa série de Fourier.

Démonstration. Nous considérons la forme linéaire

$$\begin{aligned} l_n \colon C^0_{2\pi} &\to \mathbb{C} \\ f &\mapsto S_n(f)(0) = \sum_{k=-n}^{n} c_k(f). \end{aligned} \tag{27.70}$$

La forme est continue Nous montrons d'abord que $\|l_n\|$ est continue en montrant que $\|l_n\| < \infty$ et en utilisant la proposition 10.18. Pour cela nous calculons un peu :

$$l_n(f) = \sum_{k=-n}^{n} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\sum_{k=-n}^{n} e^{-ikt}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)D_n(t)dt \tag{27.71}$$

où $D_n(t)$ est le noyaux de Dirichlet dont nous savons une formule par le lemme 27.2. Nous avons donc

$$|l_n(f)| \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi} |D_n(t)|\|f\|_\infty dt. \tag{27.72}$$

En prenant $\|f\|_\infty = 1$ nous avons la borne suivante pour la norme de l_n :

$$\|l_n\| \leqslant \frac{1}{2n}\int_{-\pi}^{\pi} |D_n(t)|dt < \infty. \tag{27.73}$$

Notons que la convergence de l'intégrale vient de la continuité de la fonction

$$t \mapsto \frac{\sin\left(\frac{2n+1}{2}t\right)}{\sin\left(\frac{t}{2}\right)} \tag{27.74}$$

qui, elle même, se prouve avec une règle de l'Hospital :

$$\lim_{t\to 0}\frac{\sin(at)}{\sin(t)} = \lim_{t\to 0}\frac{a\cos(at)}{\cos(t)} = a. \tag{27.75}$$

Donc $D_n(t)$ a une limite bien définie pour $t \to 0$ et est alors une fonction continue sur le compact $[-\pi, \pi]$.

La norme de l_n (début) Nous avons prouvé que $\|l_n\| \leqslant \frac{1}{2\pi}\int_{-\pi}^{\pi}|D_n(t)|dt$. Nous allons à présent prouver que cela est effectivement la norme de l_n. Pour $\epsilon > 0$ nous considérons la fonction

$$\begin{aligned} f_\epsilon \colon \mathbb{R} &\to \mathbb{C} \\ x &\mapsto \frac{D_n(x)}{|D_n(x)| + \epsilon}. \end{aligned} \tag{27.76}$$

C'est une fonction continue et 2π-périodique satisfaisant $\|f_\epsilon\| \leqslant 1$ parce que le dénominateur est toujours plus grand que le numérateur. Nous nous proposons de calculer

$$l_n(f_\epsilon) = \sum_{k=-n}^{n} \frac{1}{2\pi}\int_{-\pi}^{\pi} f_\epsilon(t)e^{-ikt}dt. \tag{27.77}$$

Vu que $f_\epsilon(t)e^{-ikt}$ vaut en norme $|f_\epsilon(t)|$ qui est une fonction intégrable (ne dépendant pas de k) sur $[-\pi,\pi]$, le théorème de la convergence dominée 13.184 nous permet de permuter la somme et l'intégrale :

$$l_n(f_\epsilon) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{D_n(t)}{|D_n(t)|+\epsilon}\underbrace{\sum_{k=-n}^{n}e^{-ikt}}_{=D_n(t)}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\left|D_n(t)\right|^2}{|D_n(t)|+\epsilon}dt. \tag{27.78}$$

Nous avons donc

$$\lim_{\epsilon\to 0} l_n(f_\epsilon) = \frac{1}{2\pi}\int_{-\pi}^{\pi}|D_n(t)|dt. \tag{27.79}$$

Mais vue l'inégalité (27.73) nous avons

$$\|l_n\| = \frac{1}{2\pi}\int_{-\pi}^{\pi}|D_n(t)|dt. \tag{27.80}$$

Notre tâche est maintenant de donner une valeur à cette intégrale.

Norme de l_n tend vers ∞ D'abord nous écrivons

$$\|l_n\| = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\left|\sin\left(\frac{2n+1}{2}t\right)\right|}{\left|\sin(t/2)\right|}dt, \tag{27.81}$$

ensuite nous nous souvenons que $|\sin(x)| \leqslant |x|$ pour tout x, ce qui nous permet de changer le dénominateur :

$$\|l_n\| \geqslant \frac{2}{\pi}\int_0^{\pi}\frac{\left|\sin\left(\frac{2n+1}{2}t\right)\right|}{|t|}dt \tag{27.82}$$

Nous y effectuons le changement de variable $u = \frac{2n+1}{2}t$ qui donne

$$\|l_n\| \geqslant \frac{2}{\pi}\int_0^{(n+\frac{1}{2})\pi}\frac{\left|\sin(u)\right|}{|u|}. \tag{27.83}$$

Nous y reconnaissons l'intégrale (16.483) du sinus cardinal que nous savons diverger. Cela donne

$$\lim_{n\to\infty}\|l_n\| = \infty. \tag{27.84}$$

La conclusion L'espace $\left(C^0_{2\pi},\|.\|_\infty\right)$ est complet[5], donc le théorème de Banach-Steinhaus 10.84 s'applique. Par rapport aux notations de l'énoncé de Banch-Steinhaus, nous posons

$$E = \left(C^0_{2\pi},\|.\|_\infty\right) \tag{27.85a}$$
$$F = \mathbb{R} \tag{27.85b}$$
$$H = \{l_n\}_{n\in\mathbb{N}}. \tag{27.85c}$$

Vu que la suite $(\|l_n\|)$ n'est pas bornée, il existe $f\in C^0_{2\pi}$ tel que

$$\sup_n\|l_n(f)\| = \infty. \tag{27.86}$$

Pour cette fonction nous avons

$$\sup_{n\geqslant 0}S_n(f)(0) = \infty, \tag{27.87}$$

et donc la série de Fourier de f ne converge pas en zéro.

□

5. Parce qu'une limite uniforme de fonctions continues est continue, théorème 11.294.

27.3.2 Inégalité isopérimétrique

Le théorème suivant dit que parmi les courbes C^1, le cercle a la plus grande surface possible à périmètre donné.

Théorème 27.19 (Inégalité isopérimétrique[54])**.**
Soit $f\colon S^1 \to \mathbb{C}$ une courbe de Jordan[6] *de classe C^1. Nous notons L sa longueur et S l'aire contenue de la surface délimitée*[7] *par f. Alors*

*(1) Nous avons l'**inégalité isopérimétrique** : $L^2 \geqslant 4\pi S$.*

(2) Nous avons l'égalité $L^2 = 4\pi S$ si et seulement si la courbe donnée par f est un cercle.

Démonstration. Nous commençons par considérer un chemin dont la longueur est 2π et nous en considérons son paramétrage normal. Nous allons exprimer l'aire S en utilisant le théorème de Green, et plus particulièrement la formule de surface (19.245).

Si $f(s) = x(s) + iy(s)$, nous devons intégrer $y'x - x'y$, qui n'est rien d'autre que la partie imaginaire de $f'(s)\overline{f(s)}$. Donc

$$S = \frac{1}{2}\operatorname{Im}\int_0^{2\pi} f'(s)\overline{f(s)}ds \tag{27.88}$$

Nous considérons les coefficients de Fourier de f donnés par la formule (27.49) :

$$c_n(f) = \frac{1}{2\pi}f(s)e^{-ins}. \tag{27.89}$$

Ceux de f' (qui est aussi continue sur le compact S^1 et donc tout autant L^2) sont donnés par

$$c_n(f') = inc_n(f). \tag{27.90}$$

D'autre part en vertu du théorème 20.10, la longueur de γ s'exprime en termes de l'intégrale de la norme de sa dérivée :

$$2\pi = L = \int_0^{2\pi} |f'(s)|ds = \int_0^{2\pi} |f'(s)|^2 ds \tag{27.91}$$

parce que nous avons choisi un paramétrage normal qui vérifie automatiquement $|f'(s)| = 1$ pour tout s. L'identité de Parseval sous sa forme (24.109) appliquée à f' nous enseigne que

$$L = 2\pi = \int_0^{2\pi} |f'(s)|^2 ds = 2\pi\langle f', f'\rangle = 2\pi\sum_{n=-\infty}^{\infty}|c_n(f')|^2 = 2\pi\sum_n n^2|c_n(f)|^2. \tag{27.92}$$

Par ailleurs le système trigonométrique étant une base hilbertienne, et les fonctions f et f' étant dans $L^2\big([0,2\pi]\big)$ (parce que continues sur un compact), elles sont égales à leurs séries de Fourier (au sens L^2), c'est-à-dire que nous avons l'égalité (26.336). Nous avons alors

$$\langle f', f\rangle_{L^2} = \langle\sum_{n\in\mathbb{Z}} c_n(f')e_n, \sum_{m\in\mathbb{Z}} c_m(f)e_m\rangle \tag{27.93a}$$

$$= \sum_m\sum_n c_n(f')\overline{c_m(f)}\underbrace{\langle e_n, e_m\rangle}_{\delta_{m,n}} \tag{27.93b}$$

$$= \sum_{n\in\mathbb{Z}} c_n(f')\overline{c_n(f)} \tag{27.93c}$$

$$= \sum_n in|c_n(f)|^2 \tag{27.93d}$$

6. Définition 20.108
7. C'est la partie connexe bornée de $\mathbb{C}\backslash\gamma$ dont l'existence est donnée par le théorème de Jordan 20.110.

où nous avons utilisé la continuité du produit scalaire pour sortir les sommes. Avec cela nous pouvons exprimer l'aire (27.88) en termes de coefficients de Fourier :

$$S = \frac{1}{2}\operatorname{Im} 2\pi\langle f', f\rangle = \pi \sum_{n\in\mathbb{Z}} n|c_n(f)|^2. \tag{27.94}$$

En utilisant les expressions (27.92) et (27.94) pour L et S, et en écrivant $L = 2\pi L$, nous avons

$$L^2 - 4\pi S = 4\pi^2 \sum_{n\in\mathbb{Z}} (n^2 - n)|c_n(f)|^2 \geqslant 0. \tag{27.95}$$

Cela prouve l'inégalité demandée dans le cas où $L = 2\pi$.

Si γ n'est pas de longueur 2π mais L, alors nous considérons le chemin $\sigma(t) = \frac{2\pi\gamma(t)}{L}$. Sa longueur est 2π et son aire, au vu de la formule de Green (27.88), son aire est $4\pi^2\frac{S}{L^2}$. L'inégalité isopérimétrique appliquée au chemin σ donne alors $L^2 \geqslant 4\pi S$.

Le cas d'égalité s'obtient uniquement si $c_n = 0$ pour tout n différent de 0 ou 1. Dans ce cas nous avons

$$f(s) = c_0(f) + c_1(f)e^{is}, \tag{27.96}$$

qui est un cercle de centre $c_0(f)$ et de rayon $|c_1(f)|$. □

27.3.3 À propos des coefficients

Nous considérons l'application

$$\begin{aligned} c\colon \big(L^1_{2\pi}, \|.\|_1\big) &\to \big(C_0, \|.\|_\infty\big) \\ f &\mapsto (c_n(f))_{n\in\mathbb{Z}} \end{aligned} \tag{27.97}$$

qui à une fonction 2π-périodique fait correspondre la suite (bornée) de ses coefficients de Fourier. Nous rappelons la définition

$$c_n(f) = \frac{1}{2\pi}\int_0^{2\pi} f(t)e^{-int}. \tag{27.98}$$

Nous allons montrer que cette application est linéaire, continue, injective et non surjective. Pour la continuité, par la linéarité il suffit de la montrer en 0. Nous devons donc montrer que si nous avons une suite de fonctions f_k qui tend vers 0 au sens L^1, alors $c(f_k) \to 0$ au sens de la norme $\|.\|_\infty$ sur l'ensemble des suites.

Si nous posons $r_k = \int_0^{2\pi} |f_k(t)|dt$, alors $r_k = \|f_k\|_1$ et nous avons $r_k \to 0$. Mais par définition

$$|c_n(f_k)| \leqslant r_k, \tag{27.99}$$

et donc $\|c(f_k)\|_\infty \leqslant r_k$. L'application c est donc continue. L'injectivité est donnée par le corolaire 27.16.

Si nous supposons que l'application c est continue, alors le théorème d'isomorphisme de Banach (26.1) nous dit que cela devrait être un homéomorphisme, c'est-à-dire que c^{-1} serait également continue. Nous allons montrer qu'il n'en est rien.

Nous considérons la suite de suite

$$(c_n)_k = \begin{cases} 1 & \text{si } k < n \\ 0 & \text{sinon.} \end{cases} \tag{27.100}$$

Ici $(c_n)_k$ est le terme numéro k de la suite n. Par injectivité de l'application qui à une fonction fait correspondre la suite de ses coefficients de Fourier, la seule fonction qui possède ces coefficients est

$$f_n(t) = \sum_{k\in\mathbb{N}} c_{n,k}e^{ikt}. \tag{27.101}$$

Étant donné que $\|f_n\|_1 = n$, la suite $(\|f_n\|_1)$ n'est pas bornée alors que a suite de suites (27.100) est bornée dans l'ensemble des suites parce que $\|c_n\|_\infty = 1$.

Chapitre 28

Transformation de Fourier

Définition 28.1.
Soit une fonction f sur $\mathbb{R}^d$, dont nous ne précisons pas la régularité. Sa ***transformée de fourier*** *est la fonction $\hat{f}$ définie par*

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-i\xi \cdot x} dx \tag{28.1}$$

si elle existe.

28.2.
Ce qui est bien avec cette définition est que si la formule (28.1) ne définit pas $\hat{f}$ (parce que l'intégrale n'existe pas, par exemple), nous nous réservons le droit de définir tout de même $\hat{f}$ par d'autres biais. Ce sera d'ailleurs l'objet du théorème 28.32 qui définira $\hat{f}$ pour tout $f \in L^2(\mathbb{R}^d)$ alors que la formule (28.1) ne fonctionne pas sur toutes ces fonctions.

Une bonne partie de ce qui va suivre aura pour objet de déterminer des espaces de fonctions sur lesquels la transformée est bien définie, et sur lesquels elle a de bonnes propriétés.

Nous allons par ailleurs utiliser indifféremment les notations $\mathcal{F}(f)$ ou $\hat{f}$ pour la transformée de Fourier de f. La notation $\mathcal{F}$ est pratique pour les transformées de looooooongues expressions ainsi que pour parler de l'application « transformée de Fourier » d'un espace de fonction vers un autre.

28.3.
Nous verrons dans le théorème 28.32 que la transformée de Fourier n'est pas une isométrie de L^2. Pour avoir une isométrie, il aurait fallu choisir des coefficients moins simples.

Proposition 28.4.
La transformée de Fourier est $\mathbb{C}$-linéaire au sens suivant. Soient des fonctions $f, g \colon \mathbb{R}^d \to \mathbb{C}$ telles que $\mathcal{F}(f)$ et $\mathcal{F}(g)$ existent. Alors

(1) La transformée de Fourier de $f + g$ existe et vaut $\mathcal{F}(f + g) = \mathcal{F}(f) + \mathcal{F}(g)$

(2) Pour tout $\lambda \in \mathbb{C}$, la transformée de Fourier de λf existe et vaut $\mathcal{F}(\lambda f) = \lambda \mathcal{F}(f)$.

Démonstration. C'est la proposition 13.173 qui nous permet de séparer l'intégrale dans le calcul suivant :

$$\mathcal{F}(f + g)(\xi) = \int_{\mathbb{R}^d} (f + g)(x) e^{-i\xi \cdot x} dx \tag{28.2a}$$

$$= \int_{\mathbb{R}^d} f(x) e^{-i\xi \cdot x} dx + \int_{\mathbb{R}^d} g(x) e^{-i\xi \cdot x} dx \tag{28.2b}$$

$$= \mathcal{F}(f)(\xi) + \mathcal{F}(g)(\xi). \tag{28.2c}$$

En ce qui concerne $\mathcal{F}(\lambda f)$, c'est la même chose, en utilisant la linéarité de l'intégrale. □

28.1 Transformée de Fourier sur $L^1(\mathbb{R}^d)$

Nous rappelons que les espaces L^p sont des ensembles de classes de fonctions, définition 26.8. La transformée de Fourier, comme presque tout ce qui a trait aux intégrales, passe aux classes.

Lemme 28.5.
Soit une fonction $f\colon \mathbb{R}^d \to \mathbb{C}$ telle que $\mathcal{F}(f)$ existe. Alors pour toute fonction $g \in [f]$ la transformée $\mathcal{F}(g)$ existe et $\hat{g} = \hat{f}$.

Démonstration. Par définition des classes, il existe une fonction $s\colon \mathbb{R}^d \to \mathbb{C}$ presque partout nulle telle que $g = f + s$. Soit $\xi \in \mathbb{R}^d$ fixé. La fonction $x \mapsto s(x)e^{-i\xi x}$ est presque partout nulle et donc intégrable d'intégrale null. La proposition 13.173 nous permet alors d'affirmer que $f + s$ est intégrable et que

$$\mathcal{F}(f+s)(\xi) = \int_{\mathbb{R}^d} (f+s)(x)e^{-i\xi x}dx = \int_{\mathbb{R}^d} f(x)e^{-i\xi x}dx + \int_{\mathbb{R}^d} s(x)e^{-i\xi x}dx = \mathcal{F}(f)(x). \tag{28.3}$$

□

À partir de maintenant, lorsque nous parlons de transformée de Fourier d'une fonction dans L^p, nous parlons indifféremment d'une vraie fonction ou d'une classe.

Lemme 28.6.
Si $f \in L^1(\mathbb{R}^d)$, alors $\hat{f}$ existe.

Démonstration. Par définition, si $f \in L^1(\mathbb{R}^d)$, alors l'intégrale $\int_{\mathbb{R}^d} |f|$ existe et est finie. Alors la fonction qui arrive dans la transformée de Fourier en ξ, la fonction $s\colon x \to f(x)e^{-i\xi x}$ est également dans $L^1(\mathbb{R}^d)$ parce que $|s| = |f|$. □

28.7.
Si $f \in L^1(\mathbb{R})$, la transformée de Fourier $\hat{f}$ existe, mais il n'est pas garanti qu'elle soit elle-même dans L^1, même si il est vrai que $\hat{f}$ est continue (proposition 28.11) et bornée (proposition 28.14). Nous allons immédiatement montrer un exemple de fonction L^1 dont la transformée de Fourier n'est pas dans L^1.

Lemme 28.8.
La transformée de Fourier de

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \begin{cases} 1 & \text{si } x \in [-\frac{1}{2}, \frac{1}{2}] \\ 0 & \text{sinon} \end{cases} \end{aligned} \tag{28.4}$$

est

$$\hat{f}(\xi) = -\frac{\sin(\xi/2)}{\xi/2}. \tag{28.5}$$

La fonction $\hat{f}$ n'est pas dans $L^1(\mathbb{R})$.

Démonstration. Pour $\xi \in \mathbb{R}$ fixé, la fonction $x \mapsto f(x)e^{i\xi x}$ est continue à support compact, donc

nous n'avons aucun problèmes avec l'intégrale

$$\begin{aligned}
\hat{f}(\xi) &= \int_{\mathbb{R}} f(x)e^{-i\xi x}dx && (28.6a)\\
&= \int_{[-1/2,1/2]} e^{-i\xi x} && (28.6b)\\
&= \Big[-\frac{1}{i\xi}e^{-i\xi x}\Big]_{x=-1/2}^{x=1/2} && (28.6c)\\
&= -\frac{1}{i\xi}\big(e^{-i\xi/2}-e^{-i\xi/2}\big) && (28.6d)\\
&= -\frac{2i\sin(\xi/2)}{i\xi} && (28.6e)\\
&= \frac{\sin(\xi/2)}{\xi/2}. && (28.6f)
\end{aligned}$$

Le fait que cette fonction ne soit pas dans $L^1(\mathbb{R})$ est le lemme 19.206. □

Lemme 28.9.
Si $f \in L^1(\mathbb{R}^d)$ et si $g(x) = f(\lambda x)$ alors

$$\hat{g}(\xi) = \lambda^{-d}\hat{f}(\xi/\lambda). \tag{28.7}$$

Démonstration. Il s'agit de faire le changement de variable $y = \lambda x$ dans l'intégrale

$$\hat{g}(\xi) = \int_{\mathbb{R}^d} f(\lambda x)e^{-i\xi x}dx. \tag{28.8}$$

Dans le changement de variables, vient le coefficient $dx = \lambda^{-d}dy$. □

Proposition 28.10.
La transformée de Fourier est un morphisme vis-à-vis de la convolution sur $L^1(\mathbb{R}^n)$:

$$\widehat{f*g} = \hat{f}\hat{g}. \tag{28.9}$$

Démonstration. Nous devons étudier l'intégrale

$$\widehat{f*g}(\xi) = \int_{\mathbb{R}}\Big[\int_{\mathbb{R}} f(y)g(t-y)\Big]e^{-it\xi}dt. \tag{28.10}$$

Ici nous avons choisi des représentants f et g dans les classes de L^1. Montrons que f est borélienne. D'abord $f(x) = f_+(x) - f_-(x)$ où f_+ et f_- sont des fonctions positives. Afin d'alléger les notations nous supposons un instant que f est positive et nous posons

$$f_n(x) = \sum_{k=1}^{2^n}\frac{k}{n}\mathbb{1}_{f(x)\in[\frac{k}{n},\frac{k+1}{n}[}. \tag{28.11}$$

Le fait que f soit dans L^1 implique que chacune des fonctions f_n est borélienne[1] et donc que f l'est aussi en tant que limite ponctuelle de fonctions boréliennes[2].

Nous allons appliquer le théorème de Fubini 13.261 à la fonction

$$\phi(x,y) = f(x)g(y)e^{-i\xi(x+y)} \tag{28.12}$$

1. Ceci demanderait plus de justification. Dites moi si vous savez comment justifier que les f_n soient boréliennes.
2. Le fait que f soit borélienne est une conséquence du théorème 26.152.

qui est borélienne en tant que produit et composé de fonctions boréliennes. Nous avons

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x)e^{-i\xi x}||g(y)e^{-i\xi y}|dy \right) dx = \int_{\mathbb{R}} \left(|f(x)| \int_{\mathbb{R}} |g(y)|dy \right) dx \tag{28.13a}$$

$$= \int_{\mathbb{R}} |f(x)|\|g\|_1 \tag{28.13b}$$

$$= \|f\|_1\|g\|_1 < \infty. \tag{28.13c}$$

Le théorème est donc applicable. D'abord nous avons :

$$\hat{f}(\xi)\hat{g}(\xi) = \left(\int_{\mathbb{R}} f(x)e^{-i\xi x}dx \right) \left(\int_{\mathbb{R}} g(y)e^{-i\xi y}dy \right) \tag{28.14a}$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x)g(y)e^{-i\xi(x+y)}dy \right) dx \tag{28.14b}$$

$$= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x)g(t-x)e^{-i\xi t} \right) dx. \tag{28.14c}$$

Jusqu'ici nous n'avons pas utilisé Fubini. Nous avons seulement introduit le nombre $\int_{\mathbb{R}} g(y)e^{-i\xi y}dy$ dans l'intégrale par rapport à x et effectué le changement de variables $y \mapsto t = x + y$. Maintenant nous appliquons le théorème de Fubini pour inverser l'ordre des intégrales :

$$\hat{f}(\xi)\hat{g}(\xi) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x)g(t-x)e^{-it\xi}dx \right) dy \tag{28.15a}$$

$$= \int_{\mathbb{R}} e^{-it\xi} \left(\int_{\mathbb{R}} f(x)g(t-x)dx \right) dt \tag{28.15b}$$

$$= \int_{\mathbb{R}} e^{-it\xi}(f * g)(t)dt \tag{28.15c}$$

$$= \widehat{f * g}(\xi). \tag{28.15d}$$

□

Proposition 28.11.
Soit une fonction $f \in L^1(\mathbb{R}^d)$. Alors sa transformée de Fourier est continue.

Démonstration. Nous considérons une fonction f définie sur $\mathbb{R}^d$ et à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$. Sa transformée de Fourier est donnée par

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} e^{-i\xi x}f(x)dx. \tag{28.16}$$

Pour montrer que cette fonction $\hat{f}$ est continue en ξ_0 nous considérons une suite $(\xi_n) \to \xi_0$ et nous voulons montrer que $\hat{f}(\xi_n) \to \hat{f}(\xi_0)$. Pour cela nous considérons les fonctions

$$g_n(x) = e^{-i\xi_n x}f(x) \tag{28.17}$$

qui convergent simplement vers $g(x) = e^{-i\xi x}f(x)$. Étant donné que

$$|g_n(x)| < |f(x)|, \tag{28.18}$$

le théorème de la convergence dominée donne alors

$$\lim_{n\to\infty} \int g_n(x) = \int \lim_{n\to\infty} g_n(x), \tag{28.19}$$

c'est-à-dire $\lim_{n\to\infty} \hat{f}(\xi_n) = \hat{f}(\xi)$. La fonction $\hat{f}$ est donc continue. □

Lemme 28.12.
Pour tout $f \in L^1(\mathbb{R}^n)$ nous avons $\|\hat{f}\|_\infty \leqslant \|f\|_1$.

Démonstration. Cela est une simple vérification :

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-ix\xi}dx, \tag{28.20}$$

nous avons, pour tout ξ,

$$|\hat{f}(\xi)| \leqslant \int_{\mathbb{R}} |f(x)|dx, \tag{28.21}$$

ce qui signifie exactement $\|\hat{f}\|_\infty \leqslant \|f\|_1$. □

Lemme 28.13 (Lemme de Riemann-Lebesgue[471])**.**
Si f est une fonction $L^1(\mathbb{R})$ alors $\lim_{\xi\to\pm\infty} \hat{f}(\xi) = 0$.

Démonstration. Nous commençons par prouver le résultat dans le cas d'une fonction g en escalier, et plus précisément par une fonction caractéristique d'un compact $K = [a, b]$. Au niveau de la transformée de Fourier nous avons

$$\hat{\mathbb{1}}_K(\xi) = \int_a^b e^{-i\xi x}dx = -\frac{1}{i\xi}(e^{-ib\xi} - e^{-ia\xi}). \tag{28.22}$$

Par conséquent

$$|\hat{\mathbb{1}}_K(\xi)| \leqslant \frac{2}{|\xi|}. \tag{28.23}$$

Plus généralement si $g = \sum_{i=1}^N c_i \mathbb{1}_{K_i}$, alors

$$|\hat{g}(\xi)| \leqslant \frac{2}{|\xi|}\sum_{i=1}^N |c_i|, \tag{28.24}$$

et donc nous avons effectivement $\lim_{\xi\to\pm\infty} |\hat{g}(\xi)| = 0$.

Nous passons maintenant au cas général $f \in L^1(\mathbb{R})$. Étant donné que les fonctions L^1 en escalier sont denses dans L^1, nous considérons une fonction $g \in L^1(\mathbb{R})$ en escalier telle que $\|f - g\|_1 < \epsilon$. Nous avons donc

$$\|\hat{f} - \hat{g}\|_\infty \leqslant \|f - g\|_1 < \epsilon. \tag{28.25}$$

Donc

$$\|\hat{f}(\xi)\| \leqslant \|\hat{f}(\xi) - \hat{g}(\xi)\|_|\hat{g}(\xi)|. \tag{28.26}$$

Le premier terme est plus petit que ϵ. Il nous reste à voir que

$$\lim_{\xi\to\infty} |\hat{g}(\xi)| = 0, \tag{28.27}$$

mais cela est le résultat de la première partie de la preuve. □

Corolaire 28.14.
La transformée de Fourier d'une fonction $L^1(\mathbb{R})$ est bornée.

Démonstration. Par le corolaire 28.11, la transformée de Fourier d'une fonction L^1 est continue. Le lemme de Riemann-Lebesgue 28.13 impliquant qu'elle tend vers zéro en $\pm\infty$, elle doit être bornée. □

28.1.1 Formule sommatoire de Poisson

Proposition 28.15 (Formule sommatoire de Poisson)**.**
Soit $f\colon \mathbb{R} \to \mathbb{C}$ une fonction continue et $L^1(\mathbb{R})$. Nous supposons que

(1) il existe $M > 0$ et $\alpha > 1$ tels que

$$|f(x)| \leqslant \frac{M}{(1 + |x|)^\alpha}, \tag{28.28}$$

(2) $\sum_{n=-\infty}^{\infty} |\hat{f}(2\pi n)| < \infty$.

Alors nous avons

$$\sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(2\pi n). \tag{28.29}$$

Démonstration. **Convergence normale** Nous commençons par montrer qu'il y a convergence normale sur tout compact séparément des séries sur les $n \geqslant 0$ et sur les $n < 0$.

Soit K un compact de $\mathbb{R}$ contenu dans $[-A, A]$ et $n \in \mathbb{Z}$ tel que $|n| \geqslant 2A$. Pour $x \in K$ nous avons

$$|x+n| \geqslant |n| - |x| \geqslant |n| - A \geqslant \frac{|n|}{2}. \tag{28.30}$$

Du coup nous avons un $\alpha > 1$ tel que

$$|f(x+n)| \leqslant \frac{M}{\big(1+|x+n|\big)^{\alpha}} \leqslant \frac{M}{\Big(1+\frac{|n|}{2}\Big)^{\alpha}}. \tag{28.31}$$

Lorsque n est grand, cela a le comportement de $M/|n|^{\alpha}$ et donc la série

$$\sum_{n=0}^{\infty} f(x+n) \tag{28.32}$$

est une série convergent normalement. Les deux séries (usuelles)

$$a_- = \sum_{n\leqslant 0} f(x+n) \tag{28.33a}$$

$$a_- = \sum_{n>0} f(x+n) \tag{28.33b}$$

convergent normalement.

Convergence commutative Au sens de la définition 10.92 nous avons

$$\sum_{n\in\mathbb{Z}} f(x+n) = a_+ + a_-. \tag{28.34}$$

En effet si nous prenons $J_0' \subset \mathbb{N}$ fini tel que $|\sum_{\mathbb{N}\setminus J_0} f(x+n) - a_+| \leqslant \epsilon$ et $J_1' \in -\mathbb{N}$ tel que $|\sum_{n\in-\mathbb{N}\setminus J_1'} f(x+n)| - a_- < \epsilon$, et si nous posons $J_0 = J_0' \cup J_1'$ alors si K est un ensemble fini de $\mathbb{Z}$ contenant J_0 nous avons

$$|\sum_{n\in K} f(n+x) - (a_+ + a_-)| \leqslant |\sum_{n\in K^+} f(n+x) - a_+| + |\sum_{n\in K^-} f(n+x) - a_-| \leqslant 2\epsilon \tag{28.35}$$

où K^+ sont les éléments positifs de K et K^- sont les *strictement* négatifs. Maintenant que la famille $\{f(n+x)\}_{n\in\mathbb{Z}}$ est une famille sommable, nous savons qu'elle est commutativement sommable et que la proposition 10.99 nous permet de sommer dans l'ordre que l'on veut. Nous pouvons donc écrire sans ambigüité l'expression $\sum_{n\in\mathbb{Z}} f(x+n)$ ou $\sum_{n=-\infty}^{\infty} f(x+n)$.

re-convergence normale Nous posons donc sans complexes la série

$$F(x) = \sum_{n\in\mathbb{Z}} f(x+n) \tag{28.36}$$

qui converge tant commutativement que normalement. Notons que nous pouvons maintenant dire que la série sur $\mathbb{Z}$ converge normalement ; pas seulement les deux séries séparément.

Continuité, périodicité Étant donné que chacune des fonctions $f(x+n)$ est continue, la convergence normale nous assure que F est continue.

De plus F est périodique de période 1 parce que

$$F(x+1) = \sum_{n=-\infty}^{\infty} f(x+1+n) = \sum_{p=-\infty}^{\infty} f(x+p) = F(x) \tag{28.37}$$

où nous avons posé $p = 1+n$.

Notons que nous n'avons pas spécialement prouvé que F n'était pas périodique avec des périodes plus petites que 1. Mais cela n'a pas d'importance ici.

Coefficients de Fourier En vertu de la définition (26.192) et de la périodicité de F,

$$\begin{aligned}
c_n(F) &= \int_{-1/2}^{1/2} F(t)e^{-2\pi int}dt && (28.38a)\\
&= \int_0^1 F(t)e^{-2\pi int}dt && (28.38b)\\
&= \int_0^1 \sum_{n\in\mathbb{Z}} f(t+n)e^{-2i\pi nt}dt && (28.38c)\\
&= \sum_{n\in\mathbb{Z}} \int_n^{n+1} f(u)e^{-2\pi i(u-n)t}du && (28.38d)\\
&= \int_{-\infty}^{\infty} f(u)e^{-2\pi inu}du && (28.38e)\\
&= \hat{f}(2\pi n). && (28.38f)
\end{aligned}$$

où nous avons effectué le changement de variables $u = t+n$, et permuté l'intégrale et la somme en vertu du fait que la somme converge normalement.

Conclusion Étant donné l'hypothèse $\sum_{n\in\mathbb{Z}} |\hat{f}(n)| < \infty$ la proposition 27.14 nous dit que

$$F(x) = \sum_{n\in\mathbb{Z}} c_n(F)e^{2\pi inx}, \tag{28.39}$$

c'est-à-dire que

$$\sum_{n-\infty}^{\infty} f(x+n) = \sum_{n=-\infty}^{\infty} \hat{f}(2\pi n)e^{2\pi inx}. \tag{28.40}$$

En écrivant cette égalité en $x = 0$ nous trouvons le résultat :

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \hat{f}(2\pi n). \tag{28.41}$$

□

Exemple 28.16.
La formule sommatoire de Poisson peut être utilisée pour calculer des sommes dans l'espace de Fourier plutôt que dans l'espace direct. Nous allons montrer dans cet exemple l'égalité

$$\sum_{n=-\infty}^{\infty} e^{-\alpha n^2} = \sum_{n=-\infty}^{\infty} \sqrt{\frac{\pi}{\alpha}} e^{-\pi^2 n^2/\alpha}. \tag{28.42}$$

Si α est grand, alors la somme de gauche est plus rapide, tandis que si α est petit, c'est le contraire.

Nous appliquons la formule sommatoire de Poisson à la fonction

$$f(x) = e^{-\alpha x^2}. \tag{28.43}$$

Nous avons

$$\hat{f}(x) = \int_{\mathbb{R}} e^{-\alpha t^2 - ixt} dt \tag{28.44a}$$

$$= e^{-x^2/4\alpha} \int_{\mathbb{R}} e^{-(\sqrt{\alpha}t + \frac{ix}{2\sqrt{\alpha}})^2} \tag{28.44b}$$

$$= e^{-x^2/4\alpha} \frac{1}{\sqrt{\alpha}} \int_{\mathbb{R}+\frac{ix}{2\sqrt{\alpha}}} e^{-u^2} du. \tag{28.44c}$$

Pour traiter cette intégrale nous utilisons la proposition 25.6 en considérant le chemin rectangulaire fermé qui joint les points $-R$, R, $R+ai$, $-R+ai$ et $f(z) = e^{-z^2}$. Calculons l'intégrale sur les deux côtés verticaux. Nous posons

$$\gamma_R(t) = R + tia \tag{28.45}$$

avec $t\colon 0 \to 1$. Nous avons

$$\int_{\gamma_R} f = \int_0^1 f\big(\gamma_R(t)\big)\|\gamma_R'(t)\| dt \tag{28.46a}$$

$$= ae^{-R^2} \int_0^1 e^{-2tRia + at^2} dt, \tag{28.46b}$$

donc en module nous avons

$$|\int_{\gamma_R} f| \leqslant ae^{-R^2} \int_0^1 e^{at^2} dt \leqslant Me^{-R^2}, \tag{28.47}$$

où M est une constante ne dépendant pas de R. Lorsque $R \to \infty$, la contribution des chemins verticaux s'annule et nous trouvons que

$$\int_{\mathbb{R}+ai} e^{-u^2} du = \int_{\mathbb{R}} e^{-u^2} du, \tag{28.48}$$

que nous pouvons utiliser pour continuer le calcul (28.44). Nous avons

$$\hat{f}(x) = \frac{e^{-x^2/4\alpha}}{\sqrt{\alpha}} \int_R e^{-u^2} du = \sqrt{\frac{\pi}{\alpha}} e^{-x^2/4\alpha} \tag{28.49}$$

où nous avons utilisé la formule (13.834). Par conséquent ce qui rentre dans la formule sommatoire de Poisson est

$$\hat{f}(2\pi n) = \sqrt{\frac{\pi}{\alpha}} e^{-\pi^2 n^2/\alpha}. \tag{28.50}$$

△

28.2 Suite régularisante

Définition 28.17.
*Une **suite régularisante** est une suite (ρ_n) dans $L^1(\mathbb{R}^d)$ telle que*

(1) pour tout n, $\rho_n \geqslant 0$ et $\int_{\mathbb{R}^d} \rho_n = 1$;

(2) pour tout $\alpha > 0$,

$$\lim_{n\to\infty} \int_{|t|>\alpha} \rho_n = 0. \tag{28.51}$$

Une telle suite est régularisante parce que souvent $\rho_n \in \mathscr{D}(\mathbb{R}^d)$, ce qui donne $f * \rho_n \in C^\infty$ par le corolaire 26.58.

Proposition 28.18 ([472, 473])**.**
Soit une suite régularisante $\rho_n \in L^1(\mathbb{R}^d)$. Alors :

(1) Si f est continue à support compact, nous avons la convergence uniforme sur $\mathbb{R}^d$:

$$f * \rho_n \xrightarrow{unif} f. \tag{28.52}$$

(2) Si $g \in L^p$ $(1 \leqslant p < \infty)$ alors

$$g * \rho_n \xrightarrow{L^p} g. \tag{28.53}$$

Démonstration. Si f est continue à support compact, elle est uniformément continue[3], et elle est bornée. Soit $\epsilon > 0$ et $\alpha > 0$ tel que pour tout x, y tels que $\|x-y\| < \alpha$ nous ayons $|f(x)-f(y)| < \epsilon$. Nous prenons de plus n suffisamment grand pour avoir $\int_{B(0,\alpha)^c} \rho_n < \epsilon$. Nous avons alors

$$|f(x) - (f * \rho_n)(x)| = |\int_{\mathbb{R}^d} \big(f(x) - f(y)\big)\rho_n(x-y)dy| \tag{28.54a}$$

$$\leqslant \int_{B(x,\alpha)} \underbrace{|f(x)-f(y)|}_{\leqslant\epsilon} \rho_n(x-y)dy + \int_{B(x,\alpha)^c} \underbrace{|f(x)-f(y)|}_{\leqslant 2\|f\|_\infty} \rho_n(x-y)dy \tag{28.54b}$$

$$\leqslant \epsilon(1 + 2\|f\|_\infty). \tag{28.54c}$$

Nous avons prouvé que pour tout $\epsilon > 0$, il existe N tel que $n > N$ implique $\big|f(x) - (f * \rho_n)(x)\big| \leqslant \epsilon$. Cela prouve l'uniforme convergence sur $\mathbb{R}^d$ de $f * \rho_n$ vers f.

Pour le point (2) nous considérons $g \in L^1(\mathbb{R}^d)$ et $\phi \in \mathscr{D}(\mathbb{R}^d)$. Nous avons la majoration

$$\|g * \rho_n - g\|_p \leqslant \|g * \rho_n - \phi * \rho_n\|_p + \|\phi * \rho_n - \phi\|_p + \|\phi - g\|_p \tag{28.55}$$

En ce qui concerne le premier terme ;

$$\|(g - \phi) * \rho_n\|_p \leqslant \|g - \phi\|_p \tag{28.56}$$

par la proposition 26.56. Donc

$$\|g * \rho_n - g\|_p \leqslant 2\|g - \phi\|_p + \|\phi * \rho_n - \phi\|_p. \tag{28.57}$$

Par la densité de $\mathscr{D}$ dans L^p (théorème 26.47(5)) nous pouvons considérer une suite $\phi_i \xrightarrow{L^p} g$ dans $\mathscr{D}(\mathbb{R}^d)$. Pour tout i nous avons

$$\|g * \rho_n - g\|_p \leqslant 2\|g - \phi_i\|_p + \|\phi_i * \rho_n - \phi\|_p. \tag{28.58}$$

Nous effectuons la limite sur $n \to \infty$:

$$\lim_{n\to\infty} \|g * \rho_n - g\|_p \leqslant 2\|g - \phi_i\| + \underbrace{\lim_{n\to\infty} \|\phi_i * \rho_n - \phi_i\|_p}_{=0} \tag{28.59}$$

parce que le point (1) s'applique à ϕ_i. Nous effectuons ensuite la limite sur $i \to \infty$ dans

$$\lim_{n\to\infty} \|g * \rho_n - g\| \leqslant 2\|g - \phi_i\| \to 0. \tag{28.60}$$

□

28.3 Transformée de Fourier dans l'espace de Schwartz

La définition de la transformée de Fourier de $\varphi \in \mathscr{S}(\mathbb{R}^d)$ est

$$\hat{\varphi}(\xi) = \int_{\mathbb{R}^n} \varphi(x) e^{-ix \cdot \xi} dx. \tag{28.61}$$

Si α est un multiindice de taille m, nous notons

$$(M_\alpha f)(x) = x_{\alpha_1} \dots x_{\alpha_m} f(x). \tag{28.62}$$

3. Théorème de Heine 11.88.

Lemme 28.19 (Lemme de transfert)**.**
Si $\varphi \in \mathscr{S}(\mathbb{R}^d)$ et si α est un multiindice, alors

$$\partial^\alpha \hat{\varphi} = (-i)^{|\alpha|} \widehat{M_\alpha \varphi}. \tag{28.63}$$

et

$$\widehat{\partial^\alpha \varphi}(\xi) = (-i)^{|\alpha|} \xi^\alpha \hat{\varphi}(\xi). \tag{28.64}$$

Démonstration. Nous considérons la fonction $h(x,\xi) = \varphi(x)e^{-ix\cdot\xi}$ dont la dérivée par rapport à ξ_i est donnée par $-i(M_i\varphi)(x)e^{x\cdot\xi}$. Cette fonction est majorée en norme par

$$G(x) = M_i\varphi(x), \tag{28.65}$$

qui est encore une fonction à décroissance rapide et donc parfaitement intégrable sur $\mathbb{R}^d$. Le théorème 16.18 nous dit donc que la dérivée de $\hat{\varphi}$ par rapport à ξ_i existe et vaut

$$\frac{\partial \hat{\varphi}}{\partial \xi_i}(\xi) = -i \int_{\mathbb{R}^n} x_i \varphi(x) e^{-i\xi\cdot x} = -i\widehat{M_i\varphi}(\xi). \tag{28.66}$$

En appliquant ce résultat en chaine, nous trouvons la première formule annoncée.

Nous passons à la seconde formule annoncée. Étant donné que $\varphi \in \mathscr{S}$, ses dérivées le sont aussi et par conséquent, il n'y a pas de problèmes pour écrire

$$\widehat{\partial_{x_k}\varphi}(\xi) = \int_{\mathbb{R}^d} \frac{\partial \varphi}{\partial x_k}(x) e^{-ix\cdot\xi} dx. \tag{28.67}$$

Étant donné que

$$\frac{\partial}{\partial x_k}\left(\varphi(x)e^{-ix\cdot\xi}\right) = \frac{\partial \varphi}{\partial x_k}(x)e^{-ix\cdot\xi} - i\xi_k\varphi(x)e^{-ix\cdot\xi}, \tag{28.68}$$

notre tâche sera de prouver que

$$\int_{\mathbb{R}^d} \frac{\partial}{\partial x_k}\left(\varphi(x)e^{-ix\cdot\xi}\right) dx = 0. \tag{28.69}$$

Autrement dit, nous voulons montrer que le terme au bord d'une intégration par partie s'annule. D'abord le fait que φ soit à décroissance rapide nous assure que l'intégrale (28.69) converge. Pour chaque ξ, la fonction

$$f(x,\xi) = \frac{\partial}{\partial x_k}\left(\varphi(x)e^{-ix\cdot\xi}\right) \tag{28.70}$$

est intégrable par rapport à x. De plus, f est dans $\mathscr{S}(\mathbb{R})$ pour chacune de ses variables (les autres étant fixées). Le théorème de Fubini 13.262 nous permet alors de décomposer l'intégrale en

$$\int_{\mathbb{R}^d} f(x,\xi)dx = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(x_1,\dots,x_d)dx_1\dots dx_d. \tag{28.71}$$

De plus nous pouvons intégrer dans l'ordre de notre choix et nous choisissons évidemment d'intégrer d'abord par rapport à x_k. Étudions donc l'intégrale

$$\int_{\mathbb{R}} \frac{\partial}{\partial x}\left(\varphi(x)e^{-ix\xi}\right) dx = \lim_{A\to\infty} \int_{-A}^{A} \frac{\partial}{\partial x}\left(\varphi(x)e^{-ix\xi}\right) dx \tag{28.72}$$

dans laquelle nous avons un peu allégé les notations. Une primitive de ce qui est intégré est toute trouvée : c'est $\varphi(x)e^{-ix\xi}$, et nous pouvons utiliser le théorème fondamental du calcul intégral pour écrire que

$$\int_{-A}^{A} \left(\varphi(x)e^{-ix\xi}\right)' dx = \left[\varphi(x)e^{-ix\xi}\right]_{x=-A}^{x=A}. \tag{28.73}$$

Vu que φ est dans $\mathscr{S}$, la limite $A \to \infty$ donne zéro.

En substituant maintenant (28.68) dans (28.67) et en tenant compte du terme que nous venons de montrer s'annuler, nous avons

$$\widehat{\partial_k \varphi}(\xi) = -i\xi_k \int_{\mathbb{R}^d} \varphi(x) e^{-ix \cdot \xi} = -i\xi_k \hat{\varphi}(\xi). \tag{28.74}$$

En recommençant la procédure $|\alpha|$ fois nous trouvons la seconde formule annoncée. □

Proposition 28.20 ([298]).
À propos de transformée de Fourier dans l'espace de Schwartz.

(1) L'espace de Schwartz est stable par transformée de Fourier, c'est-à-dire que $\mathcal{F}\big(\mathscr{S}(\mathbb{R}^2)\big) \subset \mathscr{S}(\mathbb{R}^d)$[4].

(2) L'application $\mathcal{F}\colon \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ est continue.

Démonstration. La linéarité découle de celle de l'intégrale. La difficulté est de prouver que pour $\varphi \in \mathscr{S}(\mathbb{R}^d)$ nous avons bien que $\hat{\varphi} \in \mathscr{S}(\mathbb{R}^d)$ et que cette association est continue[5].

Première inclusion : $\mathcal{F}\big(\mathscr{S}(\mathbb{R}^d)\big) \subset \mathscr{S}(\mathbb{R}^d)$ Nous devons prouver que pour tout multiindices α et β, nous avons $p_{\alpha,\beta}(\hat{\varphi}) < \infty$. Nous avons

$$\xi^\beta \partial^\alpha \hat{\varphi}(\xi) = \xi^\beta (-i)^{|\alpha|} \widehat{M_\alpha \varphi}(\xi) = (-i)^{|\alpha|+|\beta|} \widehat{\partial^\beta M_\alpha \varphi}(\xi). \tag{28.75}$$

Ensuite nous nous souvenons que $\|\hat{f}\|_\infty \leqslant \|f\|_1$ parce que

$$|\hat{f}(\xi)| \leqslant \int_{\mathbb{R}^d} \big| f(x) e^{-ix \cdot \xi} \big| = \int_{\mathbb{R}^d} |f(x)| dx = \|f\|_1. \tag{28.76}$$

Donc

$$p_{\alpha,\beta}(\hat{\varphi}) = \|\widehat{\partial^\beta M_\alpha \varphi}\|_\infty \leqslant \|\partial^\beta M_\alpha \varphi\|_1. \tag{28.77}$$

Du fait que φ soit dans $\mathscr{S}$, la dernière expression est finie. Cela prouve déjà que

$$\mathcal{F}\big(\mathscr{S}(\mathbb{R}^d)\big) \subset \mathscr{S}(\mathbb{R}^d). \tag{28.78}$$

Continuité Nous supposons avoir une suite $\varphi_n \xrightarrow{\mathscr{S}} \varphi$, et nous devons prouver que $\hat{\varphi}_n \xrightarrow{\mathscr{S}} \hat{\varphi}$. Pour alléger les notations, nous posons $f_n = \varphi_n - \varphi$. Nous avons

$$\begin{aligned}
\|\hat{f}\|_{\alpha,\beta} &= \|\xi^\beta \partial^\alpha \hat{f}\|_\infty && (28.79a)\\
&= \|\widehat{\partial^\beta M_\alpha f}\|_\infty \text{ lemme 28.19.} && (28.79b)\\
&\leqslant \|\partial^\beta M_\alpha f\|_1 && (28.79c)
\end{aligned}$$

La convergence $f_n \xrightarrow{\mathscr{S}} 0$ nous dit ente autres que $\partial^\beta M_\alpha f_n \xrightarrow{\mathscr{S}} 0$; en particulier la proposition 26.151 nous dit que $\partial^\beta M_\alpha f_n \xrightarrow{L^1} 0$, ce qui signifie, par les majorations (28.79) que

$$\|\hat{f}_n\|_{\alpha,\beta} \leqslant \|\partial^\beta M_\alpha f_n\|_1 \to 0, \tag{28.80}$$

ce qui prouve la continuité de transformée de Fourier dans $\mathscr{S}(\mathbb{R}^d)$.

Bijection Une preuve peut être trouvée dans [474].

□

Proposition 28.21 ([1]).
Soit $\varphi \in \mathscr{S}(\mathbb{R}^n \times \mathbb{R}^m)$ et la transformée de Fourier partielle

$$\tilde{\varphi}(x,k) = \int_{\mathbb{R}^m} e^{-iky} \varphi(x,y) dy. \tag{28.81}$$

Alors $\tilde{\varphi} \in \mathscr{S}(\mathbb{R}^n \times \mathbb{R}^m)$.

4. Nous verrons, avec la formule d'inversion dans la proposition 28.26 que c'est même une bijection.
5. Pour rappel, en dimension infinie, il n'est pas garanti qu'une application linéaire soit continue.

Démonstration. Il s'agit de reprendre les étapes de la partie correspondante de la preuve de la proposition 28.20. Soient des multiindices α, α', β et β' où α et β se réfèrent à la variable x tandis que α' et β' se réfèrent à la variable k.

Vu que la multiplication par $k^{\beta'}$ commute avec ∂^α nous avons

$$x^\beta k^{\beta'} \partial^\alpha \partial^{\alpha'} \tilde{\varphi}(x,k) = x^\beta k^{\beta'} \partial^\alpha (-i)^{|\alpha'|} \widetilde{M_{\alpha'}\varphi}(x,k) = (-i)^{|\alpha'|+|\beta'|} x^\beta \partial^\alpha \partial^{\beta'} \widetilde{M_{\alpha'}\varphi}(x,k). \tag{28.82}$$

D'autre part nous avons $\partial^\alpha \tilde{\varphi} = \widetilde{\partial^\alpha \varphi}$ parce que la fonction $\partial_x \varphi$ étant Schwartz, la fonction

$$G(y) = \sup_{x \in \mathbb{R}^n} |(\partial_x \varphi)(x,y)| \tag{28.83}$$

est dans $L^1(\mathbb{R}^m)$ par le corolaire 26.147. Par conséquent le théorème 16.18 permet de permuter la dérivée et l'intégrale dans

$$\frac{\partial}{\partial x} \tilde{\varphi}(x,k) = \frac{\partial}{\partial x} \int_{\mathbb{R}^m} e^{-iky} \varphi(x,y) dy. \tag{28.84}$$

Dans le même ordre d'esprit des difficultés de permutation de limites nous avons $M_\beta \tilde{\varphi} = \widetilde{M_\beta \varphi}$.

D'autre part nous avons encore $\|\tilde{\varphi}\|_\alpha < \infty$ parce que

$$|\tilde{\varphi}(x,k)| \leqslant \int_{\mathbb{R}^m} |\varphi(x,y)| dy \leqslant \sup_x \int_{\mathbb{R}^m} |\varphi(x,y)| dy \leqslant \int_{\mathbb{R}^m} |\sup_x \varphi(x,y)| dy < \infty \tag{28.85}$$

parce que φ est Schwartz et le corolaire 26.147 donne l'intégrabilité.

Donc nous avons

$$p_{(\alpha\alpha'),(\beta\beta')(\tilde{\varphi})} = \|\partial^{\beta'} \widetilde{M_{\alpha'} M_\beta} \partial^\alpha \varphi\|_\infty < \infty. \tag{28.86}$$

Cela prouve que $\tilde{\varphi}$ est Schwartz. □

28.3.1 Quelques transformées de Fourier

Lemme 28.22 (Transformée de Fourier de la Gausienne [475])**.**
La transformée de Fourier de

$$\begin{aligned} g_\epsilon \colon \mathbb{R}^d &\to \mathbb{R} \\ x &\mapsto e^{-\epsilon \|x\|^2} \end{aligned} \tag{28.87}$$

est donnée par

$$\hat{g}_\epsilon(\xi) = \left(\frac{\pi}{\epsilon}\right)^{d/2} e^{-\|\xi\|^2/4\epsilon} \tag{28.88}$$

Démonstration. Nous commençons par la fonction $g(x) = e^{-\|x\|^2/2}$ et nous prouvons que sa transformée de Fourier est $\hat{g}(\xi) = (2\pi)^{d/2} g(\xi)$.

Réduction à la dimension 1 La fonction g est dans l'espace de Schwartz. Par le théorème de Fubini,

$$\hat{g}(\xi) = \int_{\mathbb{R}^d} \prod_{k=1}^d e^{-x_k^2} e^{-i\xi_k x_k} dx = \prod_{k=1}^d \int_{\mathbb{R}} e^{-t^2/2} e^{-\xi_k x} dt = \prod_{k=1}^d \hat{f}(\xi_k) \tag{28.89}$$

où f est la fonction d'une variable

$$f(x) = e^{-x^2/2}. \tag{28.90}$$

Notons que $f \in \mathscr{D}(\mathbb{R})$.

Une équation différentielle Voyons l'équation différentielle satisfaite par la transformée de Fourier $\hat{f}$ de la fonction (28.90). Grâce au lemme 28.19 nous trouvons l'équation différentielle [6]

$$\xi \hat{f}(\xi) + (\hat{f})'(\xi) = 0. \tag{28.91}$$

6. Une façon alternative, plus directe de déduire cette équation différentielle sera donnée dans l'exemple 28.25.

C'est le moment d'utiliser le théorème de Cauchy-Lipschitz (16.40), appliqué à la fonction $f(t,y) = -ty$ qui est Lipschitz et continue au au problème

$$\begin{cases} y' + ty = 0 & (28.92a) \\ y(0) = y_0 & (28.92b) \end{cases}$$

possède une unique solution maximale, en l'occurrence $y(x) = y_0 e^{-x^2/2}$. En ce qui concerne la condition initiale nous avons

$$\hat{f}(0) = \int_{\mathbb{R}} e^{-x^2/2} dx = \sqrt{2\pi}. \tag{28.93}$$

par l'exemple 13.265. Donc

$$\hat{f}(\xi) = \sqrt{2\pi} e^{-\xi^2/2}. \tag{28.94}$$

En reformant le produit (28.89) nous concluons.

Nous passons maintenant à la fonction g_ϵ. Nous pouvons écrire g_ϵ sous la forme

$$g_\epsilon(x) = g(\sqrt{2\epsilon}x). \tag{28.95}$$

Utilisant successivement la transformée de Fourier de g que nous venons de calculer et 28.9 (facteur d'échelle) nous trouvons

$$\begin{aligned} \hat{g}(\xi) &= (2\pi)^{d/2} g(\xi) && (28.96a) \\ \hat{g}_\epsilon(\xi) &= (2\epsilon)^{-d/2} \hat{g}\big(\xi/\sqrt{2\epsilon}\big) && (28.96b) \\ &= \left(\frac{\pi}{\epsilon}\right)^{d/2} e^{-|\xi|^2/4\epsilon}.. && (28.96c) \end{aligned}$$

Nous voyons que $\hat{g}_\epsilon \in \mathscr{S}(\mathbb{R}^d)$ (c'était gagné d'avance par la proposition 28.20). □

Lemme 28.23.
Si $g_\epsilon(x) = e^{-\epsilon\|x\|^2}$ alors la suite

$$\rho_n = \frac{1}{(2\pi)^d} \hat{g}_{1/n} \tag{28.97}$$

est une suite régularisante (définition 28.17).

Démonstration. Nous savons déjà la transformée de Fourier de g_ϵ par le lemme 28.22. Nous montrons que la suite ρ_n est régularisante. Nous avons $\hat{g}_\epsilon \in L^1(\mathbb{R}^d)$ et $\hat{g}_\epsilon \geqslant 0$ ainsi que $\lim_{\epsilon\to 0} \int_{B(0,\alpha)} \hat{g}_\epsilon = 0$ pour tout α. Il y a seulement un couac avec la norme. Nous calculons $\int_{\mathbb{R}^d} \hat{g}_\epsilon(\xi) d\xi$ avec la forme (28.96c). En utilisant sauvagement Fubini[7] pour séparer les intégrales et en effectuant le changement de variable $u = t/(2\sqrt{\epsilon})$ nous calculons :

$$\begin{aligned} \int_{\mathbb{R}^d} e^{-|\xi|^2/4\epsilon} d\xi &= \prod_{k=1}^d \int_{\mathbb{R}} e^{-t^2/4\epsilon} dt && (28.98a) \\ &= 2\sqrt{\epsilon} \prod_{k=1}^d \int_{\mathbb{R}} e^{-u^2} du && (28.98b) \\ &= \prod_{k=1}^d 2\sqrt{\epsilon}\sqrt{\pi} && (28.98c) \\ &= 2^d (\pi\epsilon)^{d/2}. && (28.98d) \end{aligned}$$

Nous avons utilisé l'exemple 13.265 pour le calcul de l'intégrale gaussienne. Avec tout cela nous avons

$$\int_{\mathbb{R}^d} \hat{g}_\epsilon = (2\pi)^d. \tag{28.99}$$

Donc $\frac{1}{(2\pi)^d} \hat{g}_{1/n}$ est une suite régularisante. □

7. Le pauvre !

Le corolaire suivant regroupe les résultats à propos des suites régularisantes, leur utilité et leur existence.

Corolaire 28.24.
Si la suite régularisante ρ_n est dans $L^1(\mathbb{R}^d) \cap C^\infty(\mathbb{R}^d)$ alors pour $f \in L^p(\mathbb{R}^d)$ en posant $f_n = \rho_n * f$ nous avons

(1) $f_n \in C^\infty(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$

(2) $f_n \xrightarrow{L^p} f$

De plus, de telles suites existent.

Démonstration. Le fait que f_n soit de classe C^∞ est le corolaire 26.58, et la convergence est la proposition 28.18(2).

De telles suites existent, par exemple celle donnée par le lemme 28.23. □

Exemple 28.25 ([54]).
Soit la fonction $g_\epsilon(x) = e^{-\epsilon x^2}$. Sa transformée de Fourier sera déduite dans le lemma 28.22 en utilisant le lemme de transfert 28.19. Nous nous proposons ici de déduire de façon directe l'équation différentielle vérifiée par la transformée de Fourier de g_ϵ.

Nous posons

$$I(k) = \int_{\mathbb{R}} e^{-ikx} e^{-\epsilon x^2} dx. \tag{28.100}$$

et nous considérons la fonction

$$f(k, x) = e^{-ikx} e^{-\epsilon x^2}. \tag{28.101}$$

Elle est de classe C^1 par rapport à k, et intégrable en x pour chaque k. De plus sa dérivée

$$(\partial_k f)(k, x) = -ixe^{-ikx} e^{-\epsilon x^2} \tag{28.102}$$

vérifie $|\partial_k f| \leqslant xe^{-\epsilon x^2}$. La dérivée est donc majorée (uniformément en k) par une fonction intégrable. Le théorème 16.18 permet de permuter la dérivée et l'intégrale :

$$\begin{aligned}
I'(k) &= \int_{\mathbb{R}} -ixe^{-ikx} e^{-\epsilon x^2} dx && \text{(28.103a)}\\
&= i \int_{\mathbb{R}} e^{-ikx} \frac{1}{2\epsilon} \frac{d}{dx} \left(e^{-\epsilon x^2} \right) dx && \text{(28.103b)}\\
&= \frac{-i}{2\epsilon} \int_{\mathbb{R}} \frac{d}{dx} \left(e^{-ikx} \right) e^{-\epsilon x^2} dx && \text{par partie} \quad \text{(28.103c)}\\
&= \frac{-k}{2\epsilon} \int_{\mathbb{R}} e^{-ikx} e^{-\epsilon x^2} dx && \text{(28.103d)}\\
&= \frac{-k}{2\epsilon} I(k). && \text{(28.103e)}
\end{aligned}$$

D'où l'équation différentielle $I'(k) = -\frac{k}{2\epsilon} I(k)$. △

28.3.2 Formule d'inversion

Proposition 28.26 (Formule d'inversion[298, 54]).
À propos d'inversion de la transformée de Fourier.

(1) Si $f \in \mathscr{S}(\mathbb{R})$, alors nous avons la formule d'inversion

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{ikx} \hat{f}(k) dk. \tag{28.104}$$

(2) Nous avons, pour $f \in \mathscr{S}(\mathbb{R})$,

$$f(x) = \frac{1}{2\pi} \mathcal{F}(\hat{f})(-x). \tag{28.105}$$

(3) L'application

$$\mathcal{F}\colon \mathscr{S}(\mathbb{R}) \to \mathscr{S}(\mathbb{R}) \tag{28.106}$$

est une bijection continue.

(4) Cette formule peut d'écrire de plusieurs autres façons : pour tout $f \in \mathscr{S}(\mathbb{R}^d)$ nous avons

$$f(x) = \frac{1}{2\pi}\mathcal{F}(\hat{f})(-x). \tag{28.107a}$$

$$\mathcal{F}^{-1}(f)(\xi) = -\frac{1}{2\pi}\mathcal{F}(f)(-\xi). \tag{28.107b}$$

Démonstration. Le fait que la transformée de Fourier sur $\mathscr{S}(\mathbb{R}^d)$ prenne ses valeurs dans $\mathscr{S}(\mathbb{R}^d)$ est déjà prouvé dans 28.20. Nous commençons maintenant la preuve de (1).

Soit $f \in \mathscr{S}(\mathbb{R}^d)$. Pour $\epsilon > 0$ nous posons

$$f_\epsilon(k) = e^{-\epsilon k^2} e^{ikx} \hat{f}(k). \tag{28.108}$$

Nous allons calculer

$$\lim_{\epsilon \to 0} \int_{\mathbb{R}} e^{-\epsilon k^2} e^{ikx} \hat{f}(k) dk \tag{28.109}$$

de deux façons ; d'abord avec la convergence dominée, et ensuite avec Fubini.

Premier calcul : convergence dominée D'abord en utilisant directement le théorème de la convergence dominée 13.184. La fonction $\hat{f}$ est dans $\mathscr{S}(\mathbb{R})$ (théorème 28.20) et par conséquent $f_\epsilon \in L^1(\mathbb{R})$ parce que le facteur $e^{-\epsilon k^2}$ ne va certainement pas empêcher de converger. De plus $|f_\epsilon| \leqslant |\hat{f}|$ et $\hat{f} \in L^1$. Le théorème est de la convergence dominée est applicable et

$$\lim_{\epsilon \to 0} \int_{\mathbb{R}} e^{-\epsilon k^2} e^{ikx} \hat{f}(k) dk = \int_{\mathbb{R}} e^{ikx} \hat{f}(k) dk. \tag{28.110}$$

Deuxième calcul : Fubini Pour le deuxième calcul nous allons faire appel à Fubini[8] pour la fonction

$$\begin{aligned} u\colon \mathbb{R} \times \mathbb{R} &\to \mathbb{R} \\ (k, y) &\mapsto e^{ik(x-y)} e^{-\epsilon k^2} f(y). \end{aligned} \tag{28.111}$$

D'abord nous nous assurons que $u \in L^1(\mathbb{R} \times \mathbb{R})$ par le corolaire 13.261, et ensuite nous utilisons le théorème de Fubini 13.262 pour manipuler les intégrales (et en particulier les permuter).

$u \in L^1(\mathbb{R} \times \mathbb{R})$ Dans un premier temps nous avons :

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |e^{ik(x-y)} e^{-\epsilon k^2} f(y)| dy\, dk \leqslant \int_{\mathbb{R}} e^{-\epsilon k^2} \Big[\int_{\mathbb{R}} |f(y)| dy \Big] dk < \infty \tag{28.112}$$

parce que f étant dans $\mathscr{S}(\mathbb{R})$, l'intégrale intérieure se réduit à un nombre. Nous savons maintenant que $u \in L^1(\mathbb{R} \times \mathbb{R})$ grâce au corolaire 13.261.

8. Parce qu'il est toujours plus simple de refiler le boulot aux autres que de le faire soi-même... pauvre Fubini !

Calcul Nous pouvons alors calculer un peu ...

$$\begin{aligned}
\int_{\mathbb{R}} e^{ikx}e^{-\epsilon k^2}\hat{f}(k)dk &= \int_{\mathbb{R}}\int_{\mathbb{R}} e^{ikx}e^{-\epsilon k^2}e^{-iky}f(y)dy\,dk && (28.113a)\\
&= \int_{\mathbb{R}}\Big[\int_{\mathbb{R}} e^{ik(x-y)}e^{-\epsilon k^2}f(y)dk\Big]dy && (28.113b)\\
&= \int_{\mathbb{R}} f(y)\Big[\int_{\mathbb{R}} e^{ik(x-y)}e^{-\epsilon k^2}dk\Big]dy && (28.113c)\\
&= \int_{\mathbb{R}} f(y)\hat{g}_\epsilon(y-x)dy && (28.113d)\\
&= \sqrt{\frac{\pi}{\epsilon}}\int_{\mathbb{R}} f(y)e^{-(y-x)^2/4\epsilon}dy && (28.113e)\\
&= 2\sqrt{\epsilon}\sqrt{\frac{\pi}{\epsilon}}\int_{\mathbb{R}} f(x+2\sqrt{\epsilon}t)e^{-t^2}dt && (28.113f)\\
&= 2\sqrt{\pi}\int_{\mathbb{R}} f(x+2\sqrt{\epsilon}t)e^{-t^2}dt && (28.113g)\\
& && (28.113h)
\end{aligned}$$

Justifications :

— Pour (28.113b), c'est Fubini 13.262.

— Pour (28.113d), nous avons reconnu dans le crochet la transformée de Fourier de la fonction $g_\epsilon\colon x\mapsto e^{-\epsilon x^2}$.

— Pour (28.113e), nous utilisons la transformée de Fourier de g_ϵ donnée dans le lemme 28.22.

— Pour (28.113f), nous avons effectué le changement de variables $t=(y-x)/(2\sqrt{\epsilon})$ qui donne $dt=dy/2\sqrt{\epsilon}$.

Second passage à la limite Nous avons obtenu l'égalité

$$\int_{\mathbb{R}} e^{ikx}e^{-\epsilon k^2}\hat{f}(k)dk = 2\sqrt{\pi}\int_{\mathbb{R}} f(x+2\sqrt{\epsilon}t)e^{-t^2}dt, \tag{28.114}$$

et nous voudrions passer à la limite $\epsilon\to 0$. Le membre de gauche est déjà fait en (28.110). Pour le membre de droite, la fonction f étant Schwartz (en particulier bornée), nous pouvons effectuer la majoration

$$f(x+2\sqrt{\epsilon}t)e^{-t^2}\leqslant \|f\|_\infty e^{-t^2}, \tag{28.115}$$

qui est une fonction intégrable de t. Nous avons donc le droit de permuter la limite $\epsilon\to 0$ et l'intégrale dans le calcul suivant :

$$\lim_{\epsilon\to 0}\int_{\mathbb{R}} f(x+2\sqrt{\epsilon}t)e^{-t^2} = \int_{\mathbb{R}} f(x)e^{-t^2}dt. \tag{28.116}$$

Fin Nous avons maintenant les limites des deux membres de (28.114). Récrivons :

$$\lim_{\epsilon\to 0}\int_{\mathbb{R}} e^{ikx}e^{-\epsilon k^2}\hat{f}(k)dk = \lim_{\epsilon\to 0} 2\sqrt{\pi}\int_{\mathbb{R}} f(x+2\sqrt{\epsilon}t)e^{-t^2}dt, \tag{28.117}$$

À gauche nous avons déjà la limite depuis (28.110), et à droite nous obtenons

$$\lim_{\epsilon\to 0} 2\sqrt{\pi}\int_{\mathbb{R}} f(x+2\sqrt{\epsilon}t)e^{-t^2}dt = 2\sqrt{\pi}\int_{\mathbb{R}} f(x)e^{-t^2}dt = 2\sqrt{\pi}f(x)\sqrt{\pi} = 2\pi f(x) \tag{28.118}$$

où nous avons utilisé l'intégrale gaussienne faite dans l'exemple 13.265.
En remettant tout ensemble,

$$2\pi f(x) = \lim_{\epsilon\to 0}\int_{\mathbb{R}} e^{-\epsilon k^2}e^{ikx}\hat{f}(k)dk = \int_{\mathbb{R}} e^{ikx}\hat{f}(k)dk, \tag{28.119}$$

ce qu'il fallait prouver.

Le plus gros est fait ; nous avons prouvé

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{ikx}\hat{f}(k)dk. \tag{28.120}$$

Pour (2), c'est simplement une reformulation de cela.

Nous prouvons maintenant (3).

La transformée de Fourier est injective Vu qu'elle est linéaire, il suffit de démontrer que si $\mathcal{F}(f) = 0$, alors $f = 0$. Si $\hat{f} = 0$, alors la formule (28.120) donne immédiatement $f = 0$.

La transformée de Fourier est surjective Soit $f \in \mathscr{S}(\mathbb{R})$. Nous devons trouver $g \in \mathscr{S}(\mathbb{R})$ tel que $\hat{g} = f$. La formule $\mathcal{F}(\hat{f})(-x) = 2\pi f(x)$ nous incite à essayer

$$g(\xi) = -\frac{1}{2\pi}\hat{f}(-\xi). \tag{28.121}$$

Calculons $\mathcal{F}(g)$:

$$\begin{aligned}
\mathcal{F}(g)(x) &= -\int_{\mathbb{R}} e^{-i\xi x}g(\xi)d\xi && (28.122a)\\
&= -\int_{\mathbb{R}} e^{-i\xi x}\frac{1}{2\pi}\hat{f}(-\xi)d\xi && (28.122b)\\
&= \frac{1}{2\pi}\int_{\mathbb{R}} e^{itx}\hat{f}(t)dt && (28.122c)\\
&= \frac{1}{2\pi}\mathcal{F}(\hat{f})(-x) && (28.122d)\\
&= f(x). && (28.122e)
\end{aligned}$$

Pour la dernière ligne, nous avons utilisé $\mathcal{F}(\hat{f})(-x) = 2\pi f(x)$.

□

Corolaire 28.27.
Nous avons la formule

$$\int_{\mathbb{R}}\int_{\mathbb{R}} e^{-ikx}f(x)dx\,dk = 2\pi f(0). \tag{28.123}$$

Démonstration. Poser $x = 0$ dans l'équation (28.104). □

28.28.
Les physiciens qui n'ont que rarement peur écrivent souvent la formule (28.123) sous la forme

$$\int_{\mathbb{R}} e^{-ikx}dk = \delta(x) \tag{28.124}$$

où δ serait la fonction de Dirac qui vaut zéro partout sauf en $x = 0$ où elle vaudrait l'infini, mais pas n'importe quel infini ; juste celui qu'il faut pour que sont intégrale valle 1.

Lemme 28.29.
Si $\phi \in \mathscr{S}(\mathbb{R}\times\mathbb{R}^n)$, alors

$$\partial_t\hat{\phi} = \widehat{\partial_t\phi} \tag{28.125}$$

où le chapeau dénote la transformée de Fourier par rapport à la variable dans $\mathbb{R}^n$ et non par rapport à celle dans $\mathbb{R}$. Le t par contre est la variable dans $\mathbb{R}$.

Démonstration. Par définition de la transformée de Fourier nous avons

$$(\partial_t\hat{\phi})(t,\xi) = \frac{\partial}{\partial t}\int_{\mathbb{R}^n}\phi(t,x)e^{-ix\xi}dx. \tag{28.126}$$

Notre but est de permuter l'intégrale et la dérivée en utilisant le théorème 16.18. Il nous faut une fonction $G\colon \mathbb{R}^n \to \mathbb{R}$ qui soit intégrable sur $\mathbb{R}^n$ et telle que

$$\left|\frac{\partial\phi}{\partial t}\phi(t,x)\right| \leqslant G(x) \tag{28.127}$$

pour tout $t \in B(t_0,\delta)$. Étant donné que la fonction $\partial_t\phi$ est tout autant Schwartz que ϕ elle-même nous pouvons alléger les notations et chercher une fonction G qui convient pour toute fonction $\varphi \in \mathscr{S}(\mathbb{R}\times\mathbb{R}^n)$. Soit la fonction

$$G(x) = \sup_{t\in B(t_0,\delta)} |\varphi(t,x)|. \tag{28.128}$$

Pour tout multiindice α nous avons alors

$$\sup_{x\in\mathbb{R}^n} \left|x^\alpha G(x)\right| \leqslant \sup_{(t,x)\in\mathbb{R}\times\mathbb{R}^n} \left|x^\alpha\varphi(t,x)\right| \leqslant M_\alpha \in \mathbb{R}. \tag{28.129}$$

Grâce à la proposition 26.146, cela signifie que φ décroît plus vite que n'importe quel polynôme ; G est donc intégrable sur $\mathbb{R}^n$. □

28.4 Transformée de Fourier sur $L^2(\mathbb{R}^d)$

La théorie des transformées de Fourier est intéressante sur $L^2(\mathbb{R}^d)$ parce qu'elle y donne une isométrie. Nous allons la donner avec des fonctions à valeurs dans $\mathbb{C}$.

Remarque 28.30.
Une remarque qui vaut ce qu'elle vaut, mais si u est une classe de fonction pour la relation $u \sim v$ si et seulement si $u(x) = v(x)$ pour presque tout v alors l'intégrale

$$\hat{u}(\xi) = \int_{\mathbb{R}^d} u(x)e^{ix\xi}dx \tag{28.130}$$

ne dépend pas du choix du représentant. Nous pouvons donc parfaitement parler de transformée de Fourier d'une classe de fonctions.

28.4.1 Le problème

Nous avons défini en général la transformée de Fourier d'une fonction $f\colon \mathbb{R}\to\mathbb{C}$ par la formule

$$\hat{f}(\xi) = \int_{\mathbb{R}} e^{i\xi x} f(x)dx \tag{28.131}$$

tant que cette intégrale existe.

Il se fait que cette intégrale n'existe pas toujours pour des fonctions dans $L^2(\mathbb{R})$. Donc nous devons faire mieux pour définir la transformée de Fourier sur L^2.

Exemple 28.31 ([1]).
Prenons la fonction

$$f(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{1}{x} & \text{si } x \geqslant 1. \end{cases} \tag{28.132}$$

Vu que l'intégrale $\int_1^\infty \frac{1}{x^2}dx$ existe et est finie (proposition 13.251(2)), la fonction f est dans $L^2(\mathbb{R})$.

Cependant l'intégrale (28.131) n'existe pas. Pour nous convaincre de cela, nous pouvons simplement nous souvenir de la définition d'une intégrale à valeurs dans un espace vectoriel (définition 13.177). Nous fixons $\xi \in \mathbb{R}$ et nous posons $g(x) = f(x)e^{i\xi x}$.

Bien évidemment, $|f(g)| = \frac{1}{x}$ sur $]1,\infty[$. Donc $\int_{\mathbb{R}} |g| = \infty$, et la fonction g n'est pas intégrable. Fin de l'histoire.

Nous pouvons toujours essayer de comprendre mieux. Vu que $\int_{\mathbb{R}} |g| = \infty$, la proposition 13.178 nous dit qu'au moins une des intégrales parmi

$$\int f_{re}^+, \int f_{im}^+, \int f_{re}^-, \int f_{im}^- \tag{28.133}$$

est égale à $+\infty$.

Note qu'en travaillant un peu, on se convainc qu'en réalité, elles divergent toutes les quatre. △

28.4.2 Extension de $L^1 \cap L^2$ vers L^2

Théorème 28.32 (Extention de la transformée de Fourier vers $L^2(\mathbb{R}^d)$[475])**.**
Soit $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Alors

(1) Nous avons $\mathcal{F}(f) \in L^2$ et $\|\hat{f}\|_{L^2} = (2\pi)^d \|f\|_{L^2}$.

(2) L'application $\mathcal{F}\colon L^1 \cap L^2 \to L^2$ peut être étendue en une application $\mathcal{F}\colon L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ vérifiant

$$\|\hat{f}\|_{L^2} = (2\pi)^d \|f\|_{L^2} \tag{28.134}$$

pour tout $f \in L^2(\mathbb{R}^d)$.

Démonstration. Le fait que $f \in L^1$ implique $\|\mathcal{F}(f)\|_\infty \leqslant \|f\|_1$ (c'est le lemme 28.12). En particulier, $|\mathcal{F}(f)(\xi)|^2$ est majoré et l'intégrale

$$\clubsuit = \int_{\mathbb{R}^d} |\hat{f}|^2 e^{-\epsilon\xi^2} d\xi \tag{28.135}$$

existe et est finie.

Découper l'intégrale Dans un premier temps nous développons les intégrales. Dans les égalités suivantes, $x\xi$ est le produit scalaire $x \cdot \xi$ dans $\mathbb{R}^d$.

$$\clubsuit = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \overline{f(x)} e^{ix\xi} dx \right) \left(\int_{\mathbb{R}^d} f(y) e^{-y\xi} \right) e^{-\epsilon|\xi|^2} d\xi \tag{28.136a}$$

$$= \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d \times \mathbb{R}^d} \overline{f(x)} f(y) e^{i\xi(x-y)} dx dy \right] e^{-\epsilon|\xi|^2} d\xi. \tag{28.136b}$$

Nous avons utilisé le théorème de Fubini pour regrouper les intégrales [9]. Vu que $f \in L^1(\mathbb{R}^d)$, la fonction $(x, y, \xi) \mapsto f(x)f(y)e^{-\epsilon|\xi|^2}$ est dans $L^1(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ et le théorème de Fubini 13.262 avec $\Omega_1 = \mathbb{R}^d \times \mathbb{R}^d$ et $\Omega_2 = \mathbb{R}^d$ nous permet de permuter les intégrales pour avoir

$$\clubsuit = \int_{\mathbb{R}^d \times \mathbb{R}^d} \overline{f(x)} f(y) \left[\int_{\mathbb{R}^d} e^{i\xi(x-y)} e^{-\epsilon|\xi|^2} d\xi \right] dx dy. \tag{28.137}$$

Discuter de cette gaussienne En posant

$$g(x) = e^{-|x|^2/2} \tag{28.138a}$$

$$g_\epsilon(x) = g(\sqrt{2\epsilon}x) = e^{-\epsilon|x|^2} \tag{28.138b}$$

nous avons $g_\epsilon \in \mathscr{S}(\mathbb{R}^d)$ et le lemme 28.23 nous autorise à écrire

$$\hat{g}(\xi) = (2\pi)^{d/2} g(\xi) \tag{28.139a}$$

$$\hat{g}_\epsilon(\xi) = \left(\frac{\pi}{\epsilon}\right)^{d/2} e^{-|\xi|^2/4\epsilon} \tag{28.139b}$$

Nous voyons que $\hat{g}_\epsilon \in \mathscr{S}(\mathbb{R}^d)$ (c'était gagné d'avance par la proposition 28.20) et que $\hat{g}_\epsilon$ est une fonction paire (encore une fois, c'était gagné d'avance parce que la transformée de Fourier d'une fonction paire est paire).

9. Dans la suite nous allons encore utiliser Fubini quelques fois pour regrouper et dégrouper des intégrales.

Tout cela pour dire que l'intégrale entre crochet dans (28.137) est $\hat{g}_\epsilon(y-x) = \hat{g}_\epsilon(x-y)$, et donc

$$\clubsuit = \int_{\mathbb{R}^d\times\mathbb{R}^d} \overline{f(x)} f(y) \hat{g}_\epsilon(x-y) dx dy. \tag{28.140}$$

Encore une fois le théorème de Fubini permet de séparer les intégrales et de calculer l'intégrale sur y en premier. Vu que $f \in L^1$ et que $\hat{g}_\epsilon \in \mathscr{S}(\mathbb{R}^d)$, le produit de convolution $f * \hat{g}_\epsilon$ est un élément de $\mathscr{S}(\mathbb{R}^d)$ par la proposition 26.155. Nous avons donc

$$\clubsuit = \int_{\mathbb{R}^d} \overline{f(x)} (f * \hat{g}_\epsilon)(x) dx. \tag{28.141}$$

Là, nous reconnaissons un produit scalaire dans $L^2(\mathbb{R}^d)$, et donc

$$\int_{\mathbb{R}^d} |\hat{f}|^2 e^{-\epsilon\xi^2} d\xi = \langle f, f * \hat{g}_\epsilon \rangle_{L^2(\mathbb{R}^d)}. \tag{28.142}$$

Notons que tout a un sens : $f \in L^2(\mathbb{R}^d)$ et $f * \hat{g}_\epsilon \in \mathscr{S}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$.

Suite régularisante Nous prenons la suite régularisante du lemme 28.23 donnée par

$$\rho_n = \frac{1}{(2\pi)^d} \hat{g}_{1/n}. \tag{28.143}$$

Première conclusion Nous reprenons (28.142)

$$\int_{\mathbb{R}^d} |\hat{f}|^2 e^{-|\xi^2|/n} d\xi = \langle f, f * \hat{g}_{1/n} \rangle_{L^2(\mathbb{R}^d)} = (2\pi)^d \langle f, f * \rho_n \rangle. \tag{28.144}$$

En prenant la limite $n \to \infty$ nous trouvons

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} |\hat{f}|^2 e^{-\epsilon\xi^2} d\xi = (2\pi)^d \|f\|^2. \tag{28.145}$$

Pour effectuer la limite du membre de gauche nous devons remarquer qu'en posant

$$g_n(\xi) = |\hat{f}(\xi)| e^{-|\xi|^2/n}, \tag{28.146}$$

nous avons une suite décroissante de fonction (c'est-à-dire que à ξ fixé, c'est décroissant en n). Par ailleurs ces fonctions sont toujours à valeurs dans $[0,\infty]$ et nous pouvons utiliser le théorème de la convergence monotone 13.160 pour permuter la limite et l'intégrale. Au final :

$$\|\hat{f}\|_{L^2} = (2\pi)^d \|f\|_{L^2}. \tag{28.147}$$

En ce qui concerne l'extension, soit $f \in L^2(\mathbb{R}^d)$ et une suite (f_n) dans $L^1 \cap L^2$ telle que $f_n \xrightarrow{L^2} f$.

Existence d'une telle suite Si $f \in L^2(\mathbb{R}^d)$, alors nous pouvons poser

$$f_n(x) = f(x) e^{-|x|^2/n^2}. \tag{28.148}$$

Par l'inégalité de Hölder (26.59) nous avons $f_n \in L^1(\mathbb{R}^d)$; de plus $f_n \in L^2(\mathbb{R}^d)$ parce que pour tout x nous avons $|f_n(x)| \leqslant |f(x)|$. Montrons que $f_n \xrightarrow{L^2} f$. Nous avons

$$\|f_n - f\|_{L^2}^2 = \int_{\mathbb{R}^d} |f(x)(1 - e^{-|x^2|/n^2})|^2 dx. \tag{28.149}$$

Nous voulons prendre la limite $n \to \infty$. Pour ce faire à à droite nous remarquons que $e^{-|x|^2/n^2}$ est majoré par 1 ; ce qui se trouve dans l'intégrale est donc majoré (uniformément en n) par $|f(x)|^2$, qui est une fonction L^1 parce que f est L^2. Le théorème de la convergence dominée 13.184 nous permet alors de permuter la limite et l'intégrale, ce qui donne

$$\lim_{n\to\infty} \|f_n - f\|_2^2 = \int_{\mathbb{R}^d} \lim_{n\to\infty} |f(x)(1 - e^{-|x|^2/n^2})|^2 dx = 0. \tag{28.150}$$

Définition de $\mathcal{F}\colon L^2 \to L^2$ La suite (f_n) est une suite convergence dans L^2, et elle est donc de Cauchy. De plus pour chaque n, m nous avons

$$\|\hat{f}_n - \hat{f}_m\| = (2\pi)^d \|f_n - f_m\|. \tag{28.151}$$

Nous voyons donc que la suite $(\hat{f}_n)$ est également de Cauchy, dans l'espace $L^2(\mathbb{R}^d)$ qui est complet (lemme 26.73). Nous posons

$$\hat{f} = \lim_{n\to\infty} \hat{f}_n. \tag{28.152}$$

Indépendance aux choix Nous devons montrer que la définition de $\hat{f}$ ne dépend pas de la suite approximant f dans $L^1 \cap L^2$. Soient dans deux suites $f_n \xrightarrow{L^2} f$ et $g_n \xrightarrow{L^2} f$ telles que $\hat{f}_n \xrightarrow{L^2} F$ et $\hat{g}_n \xrightarrow{L^2} G$. Alors

$$\|\hat{f}_n - \hat{g}_n\| = (2\pi)^d \|f_n - g_n\| \leqslant (2\pi)^d \|f_n - f\| + (2\pi)^d \|g_n - f\| \to 0. \tag{28.153}$$

Par conséquent $(\hat{f}_n - \hat{g}_n)_n$ est une suite qui converge vers zéro. Par unicité de la limite, $F = G$.

□

Remarque 28.33.
Une autre suite possible, à la place de (28.148), est

$$f_n(x) = f(x)\mathbb{1}_{|x|<n}. \tag{28.154}$$

C'est-à-dire la fonction f limitée à une boule de rayon n autour de 0.

Chapitre 29

Distributions

Nous donnons ici une partie de la théorie sur les distributions. L'utilisation des distributions dans le cadre des équations différentielles est mise dans le chapitre sur les équations différentielles, section 31.11.

Proposition 29.1 ([476]).
Soient un ouvert Ω de $\mathbb{R}$ et une fonction intégrable $f\colon \big(\Omega, \mathcal{B}or(\Omega), \lambda\big) \to \mathbb{C}$ telle que

$$\int_\Omega f(t)\varphi(t)dt = 0 \tag{29.1}$$

pour toute fonction $\varphi \in \mathscr{D}(\Omega)$. Alors $f = 0$ presque partout sur Ω.

Démonstration. Nous commençons par prouver que f est nulle sur tout compact de Ω. Soit un compact K de Ω. Le lemme d'Urysohn 14.123 nous donne une fonction θ à support compact qui vaut 1 sur K.

Nous considérons une suite régularisante (ϕ_k) de fonctions toujours strictement positives (par exemple celle du lemme 28.23). Vu que $f\theta$ est à support compact, elle est dans $L^p(\Omega)$ et le corolaire 28.24 s'applique :

$$\phi_k * (\theta f) \to \theta f. \tag{29.2}$$

Mais, x et k étant fixés, nous avons

$$\big(\phi_k * (\theta f)\big)(x) = \int_{\mathbb{R}} \phi_k(x-t)\theta(t)f(t). \tag{29.3}$$

La fonction

$$t \mapsto \phi_k(x-t)\theta(t) \tag{29.4}$$

étant à support compact, l'hypothèse à propos de f fait que l'intégrale (29.3) est nulle :

$$\phi_k * (\theta f) = 0 \tag{29.5}$$

pour tout k. En prenant la limite $k \to \infty$,

$$\theta f = 0. \tag{29.6}$$

Vu que $\theta(x) = 1$ pour tout $x \in K$, nous avons $f(x) = 0$ pour tout $x \in K$.

Nous avons démontré que f était nulle sur tout compact de Ω.

Nous considérons maintenant une suite exhaustive (K_n) de compacts (lemme 7.312). La fonction f est nulle sur chaque K_n, et comme $\Omega = \bigcup_{n=0}^{\infty} K_n$, la fonction f est nulle sur Ω. □

29.1 Dérivée faible

29.1.1 Dérivée partielle au sens faible

Lemme-définition 29.2.
Soit $f \in L^p(I)$ où I est l'intervalle ouvert $]a,b[$. Il existe au maximum[1] *une fonction g telle que*

$$\int_I f\varphi' = -\int_I g\varphi \tag{29.7}$$

pour tout $\varphi \in \mathscr{D}(I)$. Lorsqu'une telle fonction existe, nous la nommons ***dérivée faible*** *de f.*

Démonstration. Soient $g, h \in L^2$ tels que

$$\int_I u\varphi' = -\int_I g\varphi = -\int_I h\varphi \tag{29.8}$$

pour tout $\varphi \in C_c^\infty(I)$. Nous avons alors

$$\int_I (g-h)\varphi = 0. \tag{29.9}$$

Cela implique que $g - h = 0$ presque partout par la proposition 26.132[2]. □

Exemple 29.3 (Dérivée faible de $\mathbb{1}_{\mathbb{Q}}$)**.**
Vu que $\mathbb{Q}$ est de mesure nulle dans $\mathbb{R}$, nous avons

$$\int_{\mathbb{R}} \mathbb{1}_{\mathbb{Q}}\varphi' = 0 \tag{29.10}$$

pour tout $\varphi \in \mathscr{D}(\mathbb{R})$. Pour $g = 0$ nous avons aussi $\int_{\mathbb{R}} g\varphi = 0$. Donc $g = 0$ est la dérivée faible de $\mathbb{1}_{\mathbb{Q}}$.

Cela n'est pas étonnant du fait qu'en théorie de l'intégration, les parties de mesure nulle ne comptent pas. De ce point de vue, $\mathbb{1}_{\mathbb{Q}} = 0$. D'ailleurs cette égalité est vraie dans L^p (les classes et tout ça). △

Exemple 29.4 (La fonction de Heaveside n'a pas de dérivée faible)**.**
Nous montrons que la fonction

$$H(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x \geqslant 0 \end{cases} \tag{29.11}$$

n'a pas dérivée faible. Nous nommons g une hypothétique fonction vérifiant les conditions pour être la dérivée faible de H.

Soit une fonction $\varphi \in \mathscr{D}(\mathbb{R})$, dont le support est contenu dans $]0,\infty[$. Sur le support de φ, et donc aussi de φ', nous avons $H(x) = 1$ et donc

$$\int_{\mathbb{R}} \varphi' = -\int_{\mathbb{R}} g\varphi. \tag{29.12}$$

Vu que φ est à support compact, $\int_{\mathbb{R}} \varphi' = 0$. En effet, si le support de φ est contenu dans $[-M, M]$, alors en utilisant le théorème fondamental de l'analyse 13.236, nous trouvons $\int_{\mathbb{R}} \varphi = \int_{-M}^{M} \varphi' = \varphi(M) - \varphi(-M) = 0 - 0 = 0$.

Donc g doit satisfaire

$$\int_{\mathbb{R}} g(x)\varphi(x) = 0 \tag{29.13}$$

pour tout $\varphi \in \mathscr{D}(x > 0)$. La proposition 29.1 nous dit que $g = 0$ presque partout sur $]0,\infty[$.

1. En réalité, c'est une classe au sens de l'égalité presque partout.
2. Ou alors par le lemme 26.59 qui est moins général mais tout aussi bien pour ici.

Le même raisonnement dit que $g = 0$ presque partout sur les négatifs. Que g soit maintenant nulle ou non en $x = 0$ ne change pas le fait que $g = 0$ presque partout sur $\mathbb{R}$.

Par conséquent, $\int_{\mathbb{R}} g\varphi = 0$ pour toute $\varphi \in \mathscr{D}(\mathbb{R})$. Hélas, nous avons d'autre part

$$\int_{\mathbb{R}} H(x)\varphi'(x)dx = \int_0^{\infty} \varphi'(x)dx, \tag{29.14}$$

qui n'est pas forcément nul. Notons que pour avoir un exemple de φ qui donne $\int_{\mathbb{R}} H\varphi \neq 0$, il faut chercher des fonctions dont le support contient des négatifs et des positifs.

La fonction H n'a donc pas de dérivée faible. Notons cependant que cela ne présume en rien la possibilité d'accepter une dérivée au sens des distributions. △

29.5.
Nous verrons dans la proposition 29.24 que la dérivée de Heaveside au sens des distributions est le delta de Dirac.

Exemple 29.6 (Dérivée faible de la valeur absolue)**.**
L'exemple de base de fonction continue qui n'est pas dérivable est la valeur absolue $f(x) = |x|$ prise en $x = 0$. Nous allons montrer ici que la fonction

$$H(x) = \begin{cases} 1 & \text{si } x < 0 \\ -1 & \text{si } x > 0 \end{cases} \tag{29.15}$$

est la dérivée faible de f.

Commençons par noter que H peut valoir la valeur qu'on veut en zéro ; de toutes façons la dérivée faible n'est définie qu'à partie de mesure nulle près.

Soit $\varphi \in \mathscr{D}(\mathbb{R})$ et $M > 0$ tel que le support de φ soit contenu dans $[-M, M]$. Nous avons d'une part

$$\int_{\mathbb{R}} |x|\varphi'(x)dx = -\int_{-M}^{0} x\varphi'(x)dx + \int_0^M x\varphi'(x)dx \tag{29.16a}$$

$$= -[\varphi x]_{-M}^{0} + \int_{-M}^{0} \varphi + [x\varphi]_0^M - \int_0^M \varphi \tag{29.16b}$$

où nous avons utilisé l'intégration par partie de la proposition 19.130 en posant

$$u = x \qquad v' = \varphi' \tag{29.17a}$$

$$u' = 1 \qquad v = \varphi. \tag{29.17b}$$

Tout cela pour dire que

$$\int_{\mathbb{R}} |x|\varphi'(x)dx = \int_{-M}^{0} \varphi - \int_0^M \varphi. \tag{29.18}$$

D'autre part, l'égalité

$$\int_{\mathbb{R}} H(x)\varphi(x)dx = \int_{-M}^{0} \varphi - \int_0^M \varphi \tag{29.19}$$

est immédiate.

Nous en déduisons que H est bien la dérivée faible de $x \mapsto |x|$. △

Remarque 29.7.
La dérivée faible ne doit pas être confondue avec la dérivée au sens des distributions qui sera définie en 29.22. Nous avons donc trois notions distinctes de dérivation pour une fonction :

— la dérivée usuelle,

— la dérivée au sens des distributions,

— la dérivée faible.

La dérivée faible d'une fonction reste une fonction, tandis que la dérivée distributionnelle d'une fonction est une distribution.

Je vous mets en garde contre l'idée que l'existence de l'une impliquerait trop facilement l'existence d'une autre[3].

29.1.2 Dérivée faible partielle

La notion de dérivée partielle faible est la même que l'autre. Histoire de nous mettre dans le bain, nous écrivons la définition avec les notation du produit scalaire au lieu de l'intégrale.

Définition 29.8.
*Si $i = 1, \dots, n$, la **dérivée faible** de v dans la direction e_i est l'application*[4] *notée $\partial_i v$ définie par*

$$\langle \partial_i v, \phi \rangle = -\langle v, \partial_i \phi \rangle \tag{29.20}$$

pour tout $\phi \in C_c^\infty(\Omega)$.

Lemme 29.9.
Si $v \in L^2$ admet une dérivée faible, alors cette dernière est unique.

Démonstration. Supposons f, g telles que $\langle g, \phi \rangle$ et $\langle f, \phi \rangle$ soient tous deux égaux à $-\langle v, \partial_i \phi \rangle$. En particulier pour tout $\phi \in \mathscr{D}(\Omega)$ nous avons $\langle (f - g), \phi \rangle = 0$.

Cela donne $f - g = 0$ par la proposition 26.132. □

Soit Ω un ouvert de $\mathbb{R}^d$. Le but de notre histoire est de définir une distribution comme étant un élément de l'espace dual (topologique, voir définition 10.32) de l'espace $\mathscr{D}(\Omega)$ des fonctions C^∞ à support compact dans Ω. Pour ce faire nous devons voir un peu de topologie sur différents espaces de fonctions. Notons que l'espace $\mathscr{D}(\Omega)$ n'est pas réduit à la fonction nulle comme en témoigne l'exemple donné par l'équation (14.465).

Pour chaque K compact dans Ω et multiindice $\alpha \in \mathbb{N}^d$ nous considérons sur $C^\infty(\Omega)$ la semi-norme suivante :

$$p_{K,m}(f) = \sum_{|\mu| \leqslant m} \| \partial^\mu f \|_{K,\infty}. \tag{29.21}$$

En particulier,

$$p_{K,0}(f) = \sup_{x \in K} |f(x)| = \| f \|_{\infty,K}. \tag{29.22}$$

29.2 Topologie et convergence sur des espaces de fonctions

Définition 29.10.
Les topologies que nous allons considérer sont :

(1) Sur $C^\infty(\Omega)$, la topologie des semi-normes $p_{K,m}$ (avec K et m comme paramètres).

(2) Sur $\mathscr{D}(K)$, la topologie des semi-normes $p_{K,m}$ (avec seulement m comme paramètre).

(3) Sur $\mathscr{D}(\Omega)$, la topologie induite de $C^\infty(\Omega)$.

Cela n'est pas très explicite, mais heureusement nous n'aurons souvent pas besoin de plus que de la notion de convergence dans $\mathscr{D}'(\Omega)$. Rappelons que la topologie d'un espace donne la notion de convergence par la définition 7.10.

Lemme 29.11 (Convergence dans $\mathscr{D}(K)$)**.**
Si α est un multiindice et si $\varphi_n \xrightarrow{\mathscr{D}(K)} \varphi$, alors nous avons

$$\partial^\alpha \varphi_n \xrightarrow{unif} \partial^\alpha \varphi. \tag{29.23}$$

3. Wikipédia cite l'exemple de la fonction de Cantor qui est dérivable presque partout au sens usuel, mais qui n'est pas faiblement dérivable. Écrivez-moi si vous connaissez des théorèmes qui lient les trois notions de dérivée.

4. En fait c'est une classe au sens de l'égalité presque partout.

Démonstration. Quitte à considérer la suite $\varphi_n - \varphi$ nous pouvons supposer $\varphi_n \xrightarrow{\mathscr{D}(K)} 0$. Nous avons

$$\|\partial^\alpha \varphi_n\| \leqslant \sum_{\mu \leqslant \alpha} \|\partial^\mu \varphi_n\|_{K,\infty}. \tag{29.24}$$

Vu que le membre de droite tend vers zéro, nous avons

$$\lim_{n\to\infty} \|\partial^\alpha \varphi_n\|_{K,\infty} \to 0, \tag{29.25}$$

ce qui revient à dire que $\partial^\alpha \varphi_n$ converge uniformément sur K vers $\partial^\alpha \varphi$. □

Lemme 29.12.
Si une fonction $f\colon \mathscr{D}(\Omega) \to \mathbb{R}$ est continue sur chacun des $\mathscr{D}(K)$ pour tout K compact dans Ω alors est continue sur $\mathscr{D}(\Omega)$.

Démonstration. Soit I ouvert dans $\mathbb{R}$; nous devons trouver un ouvert $\mathcal{O}$ dans $C^\infty(\Omega)$ tel que $f^{-1}(I) = \mathscr{D}(\Omega) \cap \mathcal{O}$. Vu que f est continue sur chacun des $\mathscr{D}(K)$ avec K compact dans Ω, pour tout tel compact nous avons un ouvert $\mathcal{O}_K$ dans $\mathscr{D}(K)$ tel que $f^{-1}(I) \cap \mathscr{D}(K) = \mathcal{O}_K$. En tant qu'union d'ouverts [5], l'ensemble

$$\mathcal{O} = \bigcup_{K \text{ compact de } \Omega} \mathcal{O}_K \tag{29.26}$$

est ouvert dans $C^\infty(\Omega)$. Si $\phi \inf^{-1}(I)$, nous avons $\phi \in \mathscr{D}(K)$ pour un certain K compact de Ω, donc $f^{-1}(I) \subset \mathcal{O}$. A forciori nous avons $f^{-1}(I) \subset \mathcal{O} \cap \mathscr{D}(\Omega)$.

Dans l'autre sens, si $\phi \in \mathcal{O}$, alors ϕ est dans un des $\mathcal{O}_K$ et donc dans $f^{-1}(I)$. Nous avons donc bien $f^{-1}(I) = \mathscr{D}(\Omega) \cap \mathcal{O}$. □

Théorème 29.13 (Convergence dans $\mathscr{D}(\Omega)$[153])**.**
Soit $(\varphi_n)_{n\in\mathbb{N}}$ une suite dans $\mathscr{D}(\Omega)$ et $\varphi \in \mathscr{D}(\Omega)$. Nous avons $\varphi_n \xrightarrow{\mathscr{D}(\Omega)} \varphi$ si et seulement s'il existe K compact dans Ω tel que $\varphi_n \in \mathscr{D}(K)$ pour tout n et $\varphi_n \xrightarrow{\mathscr{D}(K)} \varphi$.

Démonstration. Supposons que $\varphi_n \xrightarrow{\mathscr{D}(\Omega)} \varphi$ et qu'il n'existe pas de compacts contenant tous les supports des φ_n. Alors pour tout compact de Ω il existe un n tel que le support de φ_n ne soit pas dans K. Nous considérons une suite de compacts (K_i) tels que $\mathrm{Int}(K_n) \subset K_{n+1}$ et $\Omega = \bigcup_n K_n$. Une telle suite existe par le lemme 7.312. Ensuite nous construisons des sous-suites de la façon suivante. D'abord $L_1 = K_1$ et $n_1 \in \mathbb{N}$ est choisi de telle sorte que φ_{n_1} ait un support non contenu dans L_1. Ensuite L_i est un compact de la suite (K_n) choisi plus loin que L_{i-1} et tel que $\varphi_{n_{i-1}} \in \mathscr{D}(L_i)$. Le nombre n_i est alors choisitplus grand que n_{i-1} de telle sorte que $\varphi_{n_i} \notin \mathscr{D}(L_i)$. Ce faisant, en posant $\phi_i = \varphi_{n_i}$ nous avons

$$\phi_i \in \mathscr{D}(L_{i+1}) \backslash \mathscr{D}(L_i) \tag{29.27}$$

et $\mathrm{Int}(L_n) \subset L_{n+1}$ et $\Omega = \bigcup_n L_n$. Étant donné que (ϕ_i) et une sous-suite de (φ_i) nous avons encore $\phi_i \xrightarrow{\mathscr{D}(\Omega)} \varphi$.

Soit $i \in \mathbb{N}$. Nous allons utiliser le résultat 26.139, aka la seconde forme géométrique du théorème de Hahn-Banach pour séparer les parties $\{\phi_i\}$ (compact) et $\mathscr{D}(L_i)$ (fermé) dans $\mathscr{D}(\Omega)$. Nous avons $f_i \in \mathscr{D}'(\Omega)$ telle que

$$\begin{cases} f_i(\phi_i) > \alpha & \text{(29.28a)} \\ f\big(\mathscr{D}(L_i)\big) < \alpha. & \text{(29.28b)} \end{cases}$$

Nous redéfinissons immédiatement f_i de façon à avoir

$$\begin{cases} f_i(\phi_i) = 0 & \text{(29.29a)} \\ f\big(\mathscr{D}(L_i)\big) < 0. & \text{(29.29b)} \end{cases}$$

5. Voir définition 7.1.

Nous introduisons la fonction définie sur $\mathscr{D}(\Omega)$ par

$$p(\phi) = \sum_{i=1}^{\infty} i \frac{f_i(\phi)}{|f_i(\phi_i)|}. \tag{29.30}$$

Si $\phi \in L_k$, alors $f_k(\phi) = 0$ et même $f_l(\phi) = 0$ pour tout $l \geqslant k$. Donc pour chaque k, la somme définissant p est finie sur $\mathscr{D}(L_k)$. Nous en déduisons que p est continue sur chacun des $\mathscr{D}(L_k)$ et donc sur $\mathscr{D}(\Omega)$ par le lemme 29.12.

L'image de la suite convergente $\phi_k \xrightarrow{\mathscr{D}(\Omega)} \varphi$ par p doit être bornée parce que p est continue. Mais dans la somme (29.30), tous les termes sont positifs et en particulier le terme $i = k$ vaut k, donc $p(\phi_k) \geqslant k$, ce qui contredit le fait que l'image de la suite soit bornée. Nous en déduisons donc l'existence d'un compact K tel que $\varphi_n \in \mathscr{D}(K)$ pour tout n.

Nous devons encore prouver que $\varphi_n \xrightarrow{\mathscr{D}(K)} \varphi$ pour ce choix de K. Vu que $\varphi_n \xrightarrow{\mathscr{D}(\Omega)} \varphi$, le lemme 7.32 nous dit que nous avons aussi $\varphi_n \xrightarrow{C^\infty(\Omega)} \varphi$, ce qui signifie que pour tout K et m nous avons

$$p_{K,m}(\varphi_n - \varphi) \to 0. \tag{29.31}$$

En particulier pour le K fixé plus haut nous avons $p_m(\varphi_n - \varphi) \to 0$, c'est-à-dire que $\varphi_n \xrightarrow{\mathscr{D}(K)} \varphi$. □

Proposition 29.14.
Si K est compact dans Ω, l'espace $\mathscr{D}(K)$ est métrique et complet.

Démonstration. Nous allons d'abord montrer que $\mathscr{D}(K)$ est complet. Ensuite nous allons montrer que sa topologie peut être donnée par une distance.

Complet Nous considérons une suite de Cauchy (φ_n) dans $\mathscr{D}(K)$ au sens de la définition 7.274. Soient $\epsilon > 0$ et $i \in \mathbb{N}$; si k et l sont assez grands nous avons

$$\varphi_k - \varphi_l \in B_i(0, \epsilon). \tag{29.32}$$

En particulier pour $i = 0$ nous avons l'inégalité

$$\|\varphi_k - \varphi_l\|_\infty \leqslant \epsilon, \tag{29.33}$$

La suite (φ_n) est donc de Cauchy dans $\big(C(K), \|.\|_\infty\big)$ et y converge donc par complétude, proposition 11.296. Il existe donc une fonction $\varphi \in C(K)$ telle que

$$\varphi_n \xrightarrow{unif} \varphi. \tag{29.34}$$

Notre jeu à présent est de prouver que $\varphi \in \mathscr{D}(K)$, c'est-à-dire qu'elle est de classe C^∞.

Soit un multiindice $\alpha = \mu_1, \dots, \mu_n, i$. Si k et l sont assez grands nous avons

$$\|\partial^\alpha(\varphi_k - \varphi_l)\|_\infty \leqslant \epsilon, \tag{29.35}$$

c'est-à-dire que

$$\|\partial_i(\partial^\mu \varphi_k) - \partial_i(\partial^\mu \varphi_l)\|_\infty \leqslant \epsilon. \tag{29.36}$$

Si nous notons $\psi_k = \partial^\mu \varphi_k$ cela signifie que $(\partial_i \psi_n)$ est une suite de Cauchy dans $\big(C(K), \|.\|_\infty\big)$. Elle y converge donc et il existe une fonction $g_i \in C(K)$ telle que

$$\partial_i \psi_n \xrightarrow{unif} g_i. \tag{29.37}$$

Dans ce cas le théorème 11.312 nous indique que ψ est de classe C^1, c'est-à-dire que $\varphi \in C^{n+1}(K)$.

Métrique La proposition 7.332 nous dit que la topologie donnée par l'écart

$$d(\varphi_1,\varphi_2)=\sup_{k\geqslant 1}\min\{\frac{1}{k},p_{k-1}(\varphi_1-\varphi_2)\} \tag{29.38}$$

est la même que celle de $\mathscr{D}(K)$. Il reste à montrer que cette formule est bien une distance au sens de la définition 7.93.

(1) Nous avons bien $d(\varphi_1,\varphi_2)\geqslant 0$ parce que tous les éléments du supremum et du minimum sont positifs.

(2) Si $d(\varphi_1,\varphi_2)=0$ alors pour tout k nous devons avoir $p_{k-1}(\varphi_1-\varphi_2)=0$; en particulier pour $k=1$ cela donne $\varphi_1=\varphi_2$.

(3) Nous avons

$$p_k(\varphi_1-\varphi_2)=p_k\big(-(\varphi_2-\varphi_1)\big)=p_l(\varphi_2-\varphi_1) \tag{29.39}$$

en utilisant la propriété (2) de la définition 7.326 de semi-norme.

(4) Nous avons

$$p_k(\varphi_1-\varphi_2)=p_k(\varphi_1-\varphi_3+\varphi_3-\varphi_2)\leqslant p_k(\varphi_1-\varphi_3)+p_k(\varphi_3-\varphi_2) \tag{29.40}$$

en utilisant la propriété (3) de la définition 7.326.

□

Notons que la proposition 7.332 nous dit que $\mathscr{D}(K)$ est complet tout autant pour la topologie des semi-normes que pour celle de la distance que nous venons de décrire. Ces deux topologies sont les mêmes. Étant métrique et complet, l'espace $\mathscr{D}(\Omega)$ et donc de Baire par le théorème 7.339. Ce qui est bien avec ces deux topologies identiques c'est qu'on peut utiliser la propriété de Baire même en ne parlant que des semi-normes.

29.3 Distributions

Si Ω est un ouvert de $\mathbb{R}^d$, alors l'ensemble $\mathscr{D}(\Omega)$ est contenu dans $C^\infty(\Omega)$. Nous allons commencer par définir une topologie sur $C^\infty(\Omega)$ et ensuite donner à $\mathscr{D}(\Omega)$ la topologie induite[6].

Définition 29.15 (Distribution)**.**
*Une **distribution** sur un ouvert Ω de $\mathbb{R}^d$ est une forme linéaire continue sur $\mathscr{D}(\Omega)=C_c^\infty(\Omega)$. C'est donc un élément de $\mathscr{D}'(\Omega)$.*

Le théorème suivant donne quelques façons de vérifier qu'une forme linéaire soit continue. En particulier il nous dit que pour prouver qu'une forme linéaire est une distribution il suffi de prouver la continuité séquentielle.

Théorème 29.16 ([153, 477])**.**
Soit T une forme linéaire sur $\mathscr{D}(\Omega)$. Nous avons équivalence entre les points suivants.

(1) T est continue.

(2) Pour tout compact $K\subset\Omega$ il existe $m\in\mathbb{N}$ et $C\geqslant 0$ tel que pour tout $\varphi\in\mathscr{D}(K)$ nous ayons

$$\big|T(\varphi)\big|\leqslant Cp_{m,K}(\varphi) \tag{29.41}$$

où $p_{m,K}$ est la semi-norme donnée en (29.21).

(3) T est séquentiellement continue sur $\mathscr{D}(\Omega)$.

(4) T est séquentiellement continue en 0.

(5) Pour tout compact $K\subset\Omega$, la restriction de T à $\mathscr{D}(K)$ est continue.

6. Définition 7.15.

Un certain nombre d'ouvrages prennent le point (2) comme la définition d'une distribution.

Définition 29.17 (Topologie sur $\mathscr{D}'(\Omega)$)**.**
*Nous munissons l'espace $\mathscr{D}'(\Omega)$ de la **topologie $*$-faible**, c'est-à-dire celle de la famille de semi-normes*

$$\begin{aligned} p_\varphi\colon \mathscr{D}'(\Omega) &\to \mathbb{R} \\ T &\mapsto \big|T(\varphi)\big| \end{aligned} \tag{29.42}$$

avec $\varphi \in \mathscr{D}(\Omega)$.

Oui, c'est bien une famille de semi-normes indicée par l'ensemble $\mathscr{D}(\Omega)$. Il n'y en a donc à priori pas du tout une quantité dénombrable.

Proposition 29.18 (Convergence au sens des distributions)**.**
Nous avons $T_n \xrightarrow{\mathscr{D}'(\Omega)} T$ si et seulement si $T_n(\varphi) \to T(\varphi)$ pour tout $\varphi \in \mathscr{D}(\Omega)$.

Démonstration. La convergence $T_n \xrightarrow{\mathscr{D}'(\Omega)} T$ signifie que l'on ait $p_\varphi(T_n - T) \to 0$ pour tout $\varphi \in \mathscr{D}(\Omega)$, ce qui en retour signifie que

$$\big|(T_n - T)(\varphi)\big| \to 0. \tag{29.43}$$

□

Cette proposition suppose que l'on ait une distribution T qui vérifie $T_n(\varphi) \to T(\varphi)$ et conclut qu'on a une convergence dans les distributions. Le théorème suivant est plus fort : il va seulement supposer que $T_n(\varphi)$ converge dans $\mathbb{C}$ et va conclure que $T\colon \varphi \mapsto \lim_{n\to\infty} T_n(\varphi)$ est une distribution.

Théorème 29.19 ([478])**.**
Soit (T_n) une suite dans $\mathscr{D}'(\Omega)$ et nous supposons que pour tout $\varphi \in \mathscr{D}(\Omega)$ la suite $\big(T_n(\varphi)\big)$ converge dans $\mathbb{C}$. Alors il existe $T \in \mathscr{D}'(\Omega)$ telle que $T_n \xrightarrow{\mathscr{D}'(\Omega)} T$.

Proposition 29.20.
L'application

$$\begin{aligned} i\colon L^2(\Omega) &\to \mathscr{D}'(\Omega) \\ f &\mapsto T_f \end{aligned} \tag{29.44}$$

est une injection continue.

Démonstration. Le fait que ce soit une injection est le fait que si $T_f = T_g$ alors pour tout $\phi \in \mathscr{D}(\Omega)$ nous avons $\langle f - g, \phi\rangle = 0$, et cela implique que $f - g$ est nulle presque partout en tant que fonction et est simplement nulle en tant que classe de fonction dans L^2.

En ce qui concerne la continuité, il suffit de la prouver en zéro (par linéarité). Soit donc $f_n \xrightarrow{L^2} 0$ et montrons que $T_{f_n} \xrightarrow{\mathscr{D}'(\Omega)} T_0$. Pour prouver cela, la proposition 29.18 nous indique qu'il est suffisant de tester $T_n(\phi) \to 0$ pour tout $\phi \in \mathscr{D}(\Omega)$.

Notons que si $\phi \in \mathscr{D}$ a fortiori $\phi \in L^2$. Nous avons

$$T_{f_n}(\phi) = \int_\Omega f_n\phi \leqslant \|f_n\phi\|_{L^1} \leqslant \|f_n\|_2\|\phi\|_2 \to 0 \tag{29.45}$$

où nous avons utilisé l'inégalité de Hölder de la proposition 26.33. □

Cette proposition permet de donner un sens à des phrases du type « Soit une distribution T. Si $T \in L^2$, alors … ». Cela signifie qu'il existe $u \in L^2$ tel que $T = T_u$. Notons que dans ce cas, la distribution est définie sur L^2 et non seulement sur $\mathscr{D}$.

29.3.1 Multiplication d'une distribution par une fonction

Définition 29.21.
Si $T \in \mathscr{D}'(\Omega)$ et si $f \in C^\infty(\Omega)$ nous définissons la distribution fT par

$$(fT)(\varphi) = T(f\varphi). \tag{29.46}$$

Souvent écrit sous la forme plus compacte $\langle fu, \phi\rangle = \langle u, f\phi\rangle$.

Cela a un sens parce que si $\varphi \in \mathscr{D}(\Omega)$ alors $f\varphi$ est aussi dans $\mathscr{D}(\Omega)$.

Cette définition est motivée par ce que l'on ferait pour une distribution à densité. Si T est une distribution de densité notée également T, nous avons $T(\phi) = \int T(x)\phi(x)$ et donc

$$(fT)(\phi) = \int (fT)(x)\phi(x) = \int T(x)f(x)\phi(x) = \int T(x)(f\phi)(x) = T(f\phi). \tag{29.47}$$

En ce qui concerne les distributions tempérées, nous pouvons définir le produit avec une fonction $f \in \mathscr{S}(\Omega)$ par la même formule : si $f, \varphi \in \mathscr{S}(\Omega)$ alors le produit $f\varphi$ est encore Schwartz. Notons toutefois que nous ne pouvons pas définir fT dans $\mathscr{S}'(\Omega)$ si f est seulement dans $C^\infty(\Omega)$.

29.3.2 Dérivée de distribution

Proposition-définition 29.22.
Soit T une distribution sur Ω et $\alpha \in \mathbb{N}^d$. Alors la formule

$$(\partial^\alpha T)(\varphi) = (-1)^{|\alpha|}T(\partial^\alpha\varphi) \tag{29.48}$$

définit une distribution $\partial^\alpha T$.

*Cette distribution $\partial^\alpha T$ sera la **dérivée distributionnelle** de T. Notons que le même résultat est encore valide pour des distributions tempérées, et la démonstration est la même.*

Démonstration. La forme linéaire $\partial^\alpha T$ sera continue si elle est séquentiellement continue par le théorème 29.16. Nous considérons donc une suite $\varphi_n \xrightarrow{\mathscr{D}(\Omega)} \varphi$ et nous vérifions que

$$\lim_{n\to\infty}(\partial^\alpha T)(\varphi_n) = (\partial^\alpha T)(\varphi). \tag{29.49}$$

D'abord T étant une distribution (et donc continue) nous pouvons la permuter avec la limite :

$$\lim_{n\to\infty}(\partial^\alpha T)(\varphi_n) = \lim_{n\to\infty}(-1)^{|\alpha|}T(\partial^\alpha\varphi_n) = (-1)^{|\alpha|}T\big(\lim_{n\to\infty}\partial^\alpha\varphi_n\big). \tag{29.50}$$

Notons qu'à gauche la limite est une limite dans $\mathbb{R}$ tandis qu'à droite c'est une limite dans $\mathscr{D}(\Omega)$. Ensuite le lemme 29.11 nous dit que l'hypothèse $\varphi_n \xrightarrow{\mathscr{D}(\Omega)} \varphi$ signifie en particulier que nous avons un compact $K \subset \Omega$ contenant tous les supports des φ_n et que $\partial^\alpha\varphi_n$ converge uniformément (sur K et donc sur Ω) vers $\partial^\alpha\varphi$. Donc

$$\lim_{n\to\infty}(\partial^\alpha T)(\varphi_n) = (-1)^{|\alpha|}T\big(\lim_{n\to\infty}\partial^\alpha\varphi_n\big) = (-1)^{|\alpha|}T\big(\partial^\alpha\varphi\big) = (\partial^\alpha T)(\varphi), \tag{29.51}$$

ce qui est la relation demandée. □

Le lemme suivant montre une compatibilité entre la dérivée des distributions, la dérivée faible et l'injection de L^2 dans l'espace des distributions.

Lemme 29.23.
Soit Ω un ouvert bornée de $\mathbb{R}^n$ et $f \in L^2(\Omega)$. Alors nous avons

$$\partial_i(T_f) = T_{\partial_i f} \tag{29.52}$$

où la dérivée à droite est la dérivée faible définie en 29.8.

Démonstration. En utilisant la définition de la dérivation de distribution, pour tout $\phi \in \mathscr{D}$ nous avons

$$\partial_i(T_f)\phi = -T_f(\partial_i\phi) = -\langle f, \partial_i\phi\rangle = \langle \partial_i f, \phi\rangle = T_{\partial_i f}(\phi). \tag{29.53}$$

Nous avons utilisé la définition (29.20) de la dérivée faible. □

Nous avons déjà vu dans l'exemple 29.4 que la fonction de Heaveside n'a pas de dérivée faible. Nous allons à présent voir que cette fonction a une dérivée au sens des distributions. Intuitivement, la fonction de Heaveside a une dérivée nulle partout sauf en $x = 0$ où sa dérivée serait infinie ; nous nous attendons à un delta de Dirac.

Proposition 29.24.
La dérivée de la fonction de Heaveside

$$\begin{aligned} H\colon \mathbb{R} &\to \mathbb{R}^+ \\ x &\mapsto \begin{cases} 0 & si\ x \leqslant 0 \\ 1 & si\ x > 0 \end{cases} \end{aligned} \tag{29.54}$$

est le delta de Dirac.

Démonstration. Par définition, la dérivée de H au sens des distributions est la distribution H' qui fait

$$H'(\varphi) = -\langle H, \varphi'\rangle \tag{29.55}$$

pour tout élément $\varphi \in \mathscr{D}(\mathbb{R})$. Un petit calcul :

$$\begin{aligned} -\langle H, \varphi'\rangle &= -\int_{\mathbb{R}} H(t)\varphi'(t)dt && (29.56a) \\ &= -\int_0^\infty \varphi'(t)dt && (29.56b) \\ &= -\lim_{x\to\infty}\int_0^x \varphi'(t)dt && (29.56c) \\ &= -\lim_{x\to\infty}\big(\varphi(x) - \varphi(0)\big) && (29.56d) \\ &= \varphi(0). && (29.56e) \end{aligned}$$

Justifications :

— Pour (29.56b), c'est que $H(t) = 0$ pour $t \in]-\infty, 0[$.

— Pour (29.56c), c'est le lemme 13.229.

— Pour (29.56d), c'est le corolaire 13.250.

□

29.3.3 Ordre et support d'une distribution

Définition 29.25 (support d'une distribution[4])**.**
*Soit T une distribution. Le **support** de T est le complémentaire de l'union des ouverts $\mathcal{O}$ tels que $T(\varphi) = 0$ pour tout φ à support dans $\mathcal{O}$.*

Définition 29.26.
*Si T est une distribution sur Ω, nous disons que T est d'**ordre** inférieur ou égal à $p \in \mathbb{N}$ si pour tout compact K de Ω, il existe $C_K \in \mathbb{R}$ tel que pour tout $\varphi \in \mathscr{D}(K)$,*

$$\big|\langle T, \varphi\rangle\big| \leqslant C_K \max_{|\alpha|\leqslant p} \|\partial^\alpha\varphi\|_\infty. \tag{29.57}$$

Ici α est un multiindice.

La distribution T est d'ordre p si elle est d'ordre inférieur ou égal à p mais pas à $p-1$.

Pour la proposition suivante, on peut se remémorer la définition 29.10 de la topologie sur $C^\infty(\Omega)$.

Proposition 29.27 ([479]).
Restriction entre C^∞ et $\mathscr{D}$.

(1) Si $T \in C^\infty(\Omega)'$, alors la restriction de T à $\mathscr{D}(\Omega)$ est une distribution à support[7] *compact.*

(2) Si T est une distribution à support compact alors elle se prolonge de façon unique en une forme linéaire continue sur $C^\infty(\Omega)$.

Proposition 29.28 ([153]).
Une distribution à support compact est d'ordre fini.

Lemme 29.29 ([202]).
Soit $u \in \mathscr{D}'(\mathbb{R})$ et $\phi \in \mathscr{D}(\mathbb{R})$ tels que $\operatorname{supp}(u) \cap \operatorname{supp}(\phi) = \varnothing$. Alors $\langle u, \phi \rangle = 0$.

Démonstration. Soit $x \notin \operatorname{supp}(u)$. Alors il existe un voisinage V_x de x tel que $\langle u, \psi \rangle = 0$ pour tout $\psi \in \mathscr{D}(V_x)$. En particulier, si $x \in \operatorname{supp}(\phi)$, alors x n'est pas dans le support de u et les ensembles $\{V_x \text{ tel que } x \in \operatorname{supp}(\phi)\}$ recouvrent $\operatorname{supp}(\phi)$. Cependant ϕ est à support compact et nous pouvons extraire un sous-recouvrement fini de $\operatorname{supp}(\phi)$: il existe $x_1, \dots, x_p$ tels que

$$\operatorname{supp}(\phi) \subset \bigcup_{i=1}^{p} V_{x_i}. \tag{29.58}$$

Nous prenons une partition de l'unité[8] subordonnée à ce recouvrement. C'est-à-dire des fonctions $\chi_i \in \mathscr{D}(V_{x_i})$ telles que pour tout $x \in \operatorname{supp}(\phi)$,

$$\sum_{i=1}^{p} \chi_i(x) = 1. \tag{29.59}$$

En particulier nous avons $\sum_i \chi_i(x)\phi(x) = \phi(x)$, et donc

$$\langle u, \phi \rangle = \langle u, \sum \chi_i \phi \rangle = \sum \langle u, \chi_i \phi \rangle = 0 \tag{29.60}$$

parce que $\operatorname{supp}(\chi_i \phi) \subset V_{x_i}$. □

Lemme 29.30 ([202]).
Si u est une distribution d'ordre fini N sur $\mathbb{R}$, si $\operatorname{supp}(u) = \{x_0\}$ et si

$$\phi(x_0) = \dots = \phi^{(N)}(x_0) = 0 \tag{29.61}$$

alors $\langle u, \phi \rangle = 0$.

Démonstration. Les fonctions plateaux dont nous avons parlé dans la section 14.13.1 nous permettent de considérer une fonction $\chi \in \mathscr{D}(\mathbb{R})$ vérifiant

$$\chi(x) = \begin{cases} 1 & \text{si } x \in \overline{B(0,1)} \\ 0 & \text{si } |x| > 2 \end{cases} \tag{29.62}$$

Ensuite nous posons $\chi_n(x) = \chi\big(n(x - x_0)\big)$. Par conséquent $\chi_n(x_0) = \chi(0) = 1$ et même $\chi_n(x_0 + \epsilon) = \chi(\epsilon) = 1$ tant que ϵ est plus petit que disons $\frac{1}{2}$ pour être sur. Nous en déduisons que la fonction $1 - \chi_n$ s'annule sur un voisinage de x_0 et que donc x_0 n'est pas dans le support de $1 - \chi_n$. Donc $\operatorname{supp}(1 - \chi_n) \cap \operatorname{supp}(u) = \varnothing$ et le lemme 29.29 est utilisable : $\langle u, (1 - \chi_n \phi) \rangle = 0$, ou encore :

$$\langle u, \phi \rangle = \langle u, \chi_n \phi \rangle \tag{29.63}$$

7. Définition 29.25.
8. Lemme 19.20.

pour tout n. Vu que le but est de prouver que $\langle u, \phi\rangle = 0$, nous allons prouver que

$$|\langle u, \chi_n\phi\rangle| \overset{n\to\infty}{\longrightarrow} 0. \tag{29.64}$$

Dans ce dessin nous posons

$$\|f\|_n = \sup_{x\in B(0,\frac{2}{n})} \|f(x)\| \tag{29.65}$$

et

$$\|f\|_{(p)} = \sup_{i\leqslant p} \|\partial^i f\|_\infty. \tag{29.66}$$

La distribution u est d'ordre fini N, et nous en écrivons la définition 29.26 en prenant $\operatorname{supp}(\chi_n\phi)$ en tant que K :

$$\big|\langle u, \chi_n\phi\rangle\big| \leqslant C \max_{k\leqslant N} \|\partial^k(\chi_n\phi)\|_\infty. \tag{29.67}$$

En remplaçant le maximum par une somme de $k = 0$ à $k = N$, nous majorons. De plus le support de χ_n étant contenu dans $B_n = B(x_0, 2/n)$ nous ne changeons rient en utilisant $\|.\|_n$ au lieu de $\|.\|_\infty$. Donc

$$|\langle u, \chi_n\phi\rangle| \leqslant C \sum_{k=0}^{N} \|\partial^k(\chi_n\phi)\|_n \leqslant C \sum_{k=0}^{N} \binom{k}{i} \|\partial^i\chi_n\|_n \|\partial^{k-i}\phi\|_n. \tag{29.68}$$

Notons que la seconde inégalité est une inégalité du type $\|fg\| \leqslant \|f\|\|g\|$. En dérivant un petit peu nous trouvons que

$$(\partial^i\chi_n)(x) = n^i(\partial^i\chi)\big(n(x - x_0)\big). \tag{29.69}$$

Donc [9]

$$\|\partial^i\chi_n\|_n = \sup_{x\in B_n} n^i\big|(\partial^i\chi)\big(n(x - x_0)\big)\big| = n^i \sup_{y\in[-2,2]} \big|(\partial^i\chi)(y)\big| = n^i\|\partial^i\chi\|_\infty. \tag{29.70}$$

Nous pouvons donc remplacer $\|\partial^u\chi_n\|_n$ par $n^i\|\partial^i\chi\|_\infty$.

D'autre part nous voulons majorer $\|\partial^{k-i}\phi\|_n$ par quelque chose ne dépendant ni de k ni de i. Nous faisons le théorème des accroissements finis 10.145 : $\|\partial^l\phi\|_n \leqslant \frac{2}{n}\|\partial^{l+1}\phi\|_n$. Ce n au dénominateur est salutaire parce que nous avions un n^i apparu à cause du remplacement (29.70). Nous faisons donc $i + 1$ fois le théorème des accroissements finis :

$$\|\partial^{k-i}\phi\|_n \leqslant \left(\frac{2}{n}\right)^{i+1} \|\partial^{k+1}\phi\|_n. \tag{29.71}$$

Toutes ces majorations donnent

$$\big|\langle u, \chi_n\phi\rangle\big| \leqslant C \sum_{k=0}^{N}\sum_{i=0}^{k} \binom{i}{k} n^i \underbrace{\|\partial^i\chi\|_\infty}_{\leqslant\|\chi\|_{(N)}} \left(\frac{2}{n}\right)^{i+1} \underbrace{\|\partial^{k+1}\phi\|_n}_{\leqslant\|\phi\|_{(N+1)}} \tag{29.72a}$$

$$\leqslant C\|\chi\|_{(N)}\|\phi\|_{(N+1)} \frac{1}{n} \sum_{k=0}^{N}\sum_{i=0}^{k} \binom{k}{i} 2^{i+1} \tag{29.72b}$$

$$= \frac{C'}{n} \tag{29.72c}$$

où C' est une constante qui dépend de χ, de ϕ et de N, mais pas de n. Vu que $\frac{C'}{n} \to 0$ nous avons bien

$$\langle u, \phi\chi_n\rangle = 0, \tag{29.73}$$

ce qu'il fallait démontrer. □

Proposition 29.31 ([202]).
Soit $u \in \mathscr{D}(\mathbb{R})'$ avec $\operatorname{supp}(u) = \{x_0\}$. Alors $u = \sum_{i=0}^{N} a_i\partial^i\delta_{x_0}$ où N est l'ordre de u.

9. Dans [202], la dernière égalité vient avec une inégalité, et je comprends pas pourquoi.

Démonstration. D'abord il faut préciser que l'ordre de u est fini parce que son support est compact (proposition 29.28) ; nous notons N cet ordre.

Soit $\phi \in \mathscr{D}(\mathbb{R})$. Nous considérons $\chi \in \mathscr{D}(\mathbb{R})$ telle que

$$\chi(x) = \begin{cases} 1 & \text{si } x \in \overline{B(x_0, 1)} \\ 0 & \text{si } |x - x_0| > 2. \end{cases} \tag{29.74}$$

Encore une fois, $1 - \chi$ s'annule sur un voisinage autour de x_0, ce qui fait que

$$\operatorname{supp}(u) \cap \operatorname{supp}\big((1-\chi)\phi\big) = \varnothing, \tag{29.75}$$

et donc $\langle u, (1-\chi)\phi\rangle = 0$. Au final,

$$\langle u, \phi\rangle = \langle u, \chi\phi\rangle. \tag{29.76}$$

C'est le moment de poser

$$\psi(x) = \chi(x)\big[\phi(x) - \sum_{k=1}^{N} \frac{1}{k!}(\partial^k\phi)(x_0)(x-x_0)^k\big] \tag{29.77}$$

La fonction ψ ayant un support disjoint de celui de u, nous avons aussi $\langle u, \psi\rangle = 0$, ce qui donne

$$\langle u, \phi\rangle = \langle u, \chi\phi\rangle = \langle u, \chi \sum_{k=0}^{N} \frac{1}{k!}(\partial^k\phi)(x_0)(x-x_0)^k\rangle. \tag{29.78}$$

En posant $a_k = \frac{1}{k!}\langle u, x \mapsto \chi(x)(x-x_0)^k\rangle$ nous avons alors

$$\langle u, \phi\rangle = \sum_{k=0}^{N} a_k(\partial^k\phi)(x_0) = \sum_k (-1)^k a_k(\partial^k\delta_{x_0})(\phi). \tag{29.79}$$

□

29.4 Distributions tempérées

L'espace de Schwartz[10] $\mathscr{S}(\Omega)$ est défini dans la définition 26.143 ; sa topologie y est discutée.

Définition 29.32.
*Une **distribution tempérée** est une forme linéaire continue sur $\mathscr{S}(\mathbb{R}^d)$. L'ensemble des distributions tempérées est noté $\mathscr{S}'(\mathbb{R}^d)$. Si T est une telle distribution, nous notons $\langle T, \varphi\rangle$ l'image de φ par T.*

29.33.
Pour rappel, la topologique sur $\mathscr{S}(\mathbb{R}^d)$ est celle de la définition 26.148.

Si f est une fonction sur $\mathbb{R}^d$ telle que $f\varphi \in L^1(\mathbb{R}^d)$ pour tout $\varphi \in \mathscr{S}(\mathbb{R}^d)$, alors nous définissons la distribution $T_f \in \mathscr{S}'(\mathbb{R}^d)$ par

$$\langle T_f, \varphi\rangle = \int_{\mathbb{R}^d} f(x)\varphi(x)dx. \tag{29.80}$$

Cette définition ne fonctionne pas pour toutes les fonctions. Par exemple pour $f(x) = e^{x^2}$, et $\varphi(x) = e^{-x^2} \in \mathscr{S}(\mathbb{R})$ nous avons $f\varphi = 1$ qui n'est pas du tout intégrable sur $\mathbb{R}$.

10. Attention : ce Schwartz (avec un t) est le Schwartz des distributions dont le prénom est Laurent. À ne pas confondre avec Schwarz (sans t) dont le prénom est Cauchy.

Lemme-définition 29.34.
L'application

$$\begin{aligned}\delta\colon \operatorname{Fun}(\mathbb{R}^d) &\to \mathbb{C}\\ \varphi &\mapsto \varphi(0).\end{aligned} \tag{29.81}$$

est

(1) une distribution,

(2) une distribution tempérée.

Cette distribution est nommée ***distribution de Dirac****.*

Démonstration. Juste pour rappel, $\operatorname{Fun}(X)$ est l'ensemble de toutes les fonctions sur X. Pour prouver que δ est une distribution, nous devons démontrer que $\delta\colon \mathscr{D}(\mathbb{R}) \to \mathbb{C}$ est continue. Et pour qu'elle soit une distribution tempérée, il faut démontrer que $\delta\colon \mathscr{S}(\mathbb{R}) \to \mathbb{C}$ est continue.

Nous utilisons le théorème 29.16(2) pour démontrer la continuité de δ sur $\mathscr{D}(\mathbb{R})$. Soit un compact $K \subset \mathbb{R}$ et $\varphi \in \mathscr{D}(K)$. En prenant $m = 0$ nous devons avons la majoration

$$|\delta(\varphi)| = \varphi(0) \leqslant \|\varphi\|_{K,\infty} = p_{0,K}(\varphi). \tag{29.82}$$

Pour la continuité de δ sur $\mathscr{S}(\mathbb{R}^d)$, nous utilisons les résultats de 26.149. Soit une suite $\varphi_n \xrightarrow{\mathscr{S}} 0$. En particulier, $p_{0,0}(\varphi_n) = \sup_x |\varphi_n(x)| \to 0$. Donc $\varphi_n(0) \to 0$ comme il le faut. □

Exemple 29.35.
La **valeur principale** de la fonction $x \mapsto \frac{1}{x}$ est la distribution

$$\begin{aligned}T\colon \mathscr{S}(\mathbb{R}) &\to \mathbb{R}\\ \varphi &\mapsto \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \int_{|x|>\epsilon} \frac{\varphi(x)}{x}.\end{aligned} \tag{29.83}$$

Montrons que cela définit bien une distribution tempérée.

D'abord l'intégrale existe pour tout ϵ, par exemple parce que pour les grands $|x|$ nous avons par exemple $|\varphi(x) \leqslant x^3|$ et donc $\varphi(x)/x \leqslant 1/x^2$ dont l'intégrale converge. Nous devons maintenant regarder la limite.

Nous considérons une suite $\epsilon_n \to 0$ et la suite

$$a_n = \int_{|x|\geqslant\epsilon_n} \frac{\varphi(x)}{x} dx. \tag{29.84}$$

Nous montrons que cette suite converge dans $\mathbb{R}$ en montrant qu'elle est de Cauchy. Pour cela nous travaillons un peu la forme de φ :

$$\varphi(x) = \varphi(0) + \int_0^x \varphi'(t)dt = \varphi(0) + \int_0^1 x\varphi'(x\theta)d\theta. \tag{29.85}$$

Ce qui est dans l'intégrale est borné par $K = \|M_x\varphi'\|_\infty$ qui est parfaitement fini parce que φ est à décroissance rapide. Lorsque nous calculons $|a_m - a_n|$, le terme $\varphi(0)/x$ donne une intégrale nulle parce que le domaine d'intégration $\epsilon_n \leqslant |x| \leqslant \epsilon_n$ est symétrique alors que la fonction $1/x$ est impaire.

$$|a_m - a_n| \leqslant \Big|\int_{\epsilon_m<|x|<\epsilon_n} K\Big| = 2|\epsilon_n - \epsilon_m|K \tag{29.86}$$

Tout cela nous dit que T est bien définie. Nous devons encore étudier sa continuité.

Soit χ une fonction dans $C_c^\infty(\mathbb{R})$ telle valant 1 sur $[-1,1]$, paire et à valeurs dans $[0,1]$. Pour tout $\epsilon > 0$ nous avons $\int_{|x|>\epsilon} \frac{\chi(x)}{x} dx = 0$.

Nous avons aussi $\varphi = \chi\varphi + (1-\chi)\varphi$, et donc

$$\int_{|x|>\epsilon} \frac{\varphi(x)}{x} dx = \int_{|\epsilon|>0} \chi(x) \frac{\varphi(x)-\varphi(0)}{x} dx + \int_{|\epsilon|>0} \big(1-\chi(x)\big) \frac{\varphi(x)}{x} dx \tag{29.87a}$$

$$= \int_{|\epsilon|>0} \chi(x) \int_0^1 \underbrace{\varphi'(\theta x)}_{\leqslant \|\varphi'\|_\infty} d\theta + \int_{|x|\geqslant 1} \big(1-\chi(x)\big) \frac{\varphi(x)}{x} dx \tag{29.87b}$$

$$\leqslant \|\varphi'\|_\infty \int_{|x|\geqslant\epsilon} \chi(x) dx + \|\varphi\|_{L^1} \tag{29.87c}$$

$$= C\|\varphi'\|_\infty + \|\varphi\|_1. \tag{29.87d}$$

Cela est valable pour toute fonction $\varphi \in \mathscr{S}(\mathbb{R})$. Mais nous savons que si $\varphi_n \overset{\mathscr{S}(\mathbb{R})}{\to} 0$, alors $\|\varphi_n\|_\infty \to 0$, $\|\varphi_n'\|_\infty \to 0$ et $\|\varphi_n\|_1 \to 0$; donc si $\varphi_n \overset{\mathscr{S}(\mathbb{R})}{\to} 0$, alors

$$T(\varphi_n) = \lim_{\substack{\epsilon\to 0\\ \epsilon>0}} \int_{|x|>\epsilon} \frac{\varphi(x)}{x} \leqslant C\|\varphi_n'\|_\infty + \|\varphi_n\|_1 \to 0. \tag{29.88}$$

△

29.4.1 Topologie

La topologie que nous mettons sur l'espace $\mathscr{S}'(\mathbb{R}^d)$ est le même type que celle que nous mettons sur $\mathscr{D}'(\mathbb{R}^d)$, c'est-à-dire celle des semi-normes $p_\varphi(T) = |T(\varphi)|$. La définition 29.17 et la proposition 29.18 restent.

Proposition 29.36.
Nous avons $T_n \overset{\mathscr{S}'(\mathbb{R}^d)}{\longrightarrow} T$ si et seulement si pour tout $\varphi \in \mathscr{S}(\mathbb{R}^d)$ nous avons $T_n(\varphi) \to T(\varphi)$.

29.4.2 Distributions associées à des fonctions

Si $f \in L^1_{loc}(\mathbb{R}^d)$ alors nous lui associons une distribution $T_f \in \mathscr{D}'(\mathbb{R}^d)$ définie par la formule

$$T_f(\varphi) = \int_{\mathbb{R}^d} f(x)\varphi(x) dx. \tag{29.89}$$

Proposition 29.37.
L'application ainsi définie

$$\begin{aligned} L^1_{loc}(\mathbb{R}^d) &\to \mathscr{D}'(\mathbb{R}^d) \\ f &\mapsto T_f \end{aligned} \tag{29.90}$$

est injective.

Démonstration. Si $T_f = 0$ alors pour tout $\varphi \in \mathscr{D}$ nous avons $\int_{\mathbb{R}^d} f\varphi = 0$. En vertu de la proposition 26.133 cela implique $f = 0$ presque partout. □

29.4.3 Composition avec une fonction

Proposition 29.38 ([478], page 113 et 32)**.**
Soit $T \in \mathscr{S}'(\Omega)$ et $f \in C^k(A \times \Omega)$ où A est ouvert dans $\mathbb{R}^d$. Nous posons

$$\begin{aligned} F\colon A &\to \mathbb{R} \\ \lambda &\mapsto T\big(f(\lambda, .)\big). \end{aligned} \tag{29.91}$$

Alors $F \in C^k(A)$.

29.4.4 Transformée de Fourier d'une distribution tempérée

Définition 29.39.
*La **transformée de Fourier** de la distribution tempérée $T \in \mathscr{S}'(\mathbb{R}^d)$ est la distribution $\hat{T}$ définie par*

$$\hat{T}(\varphi) = T(\hat{\varphi}) \tag{29.92}$$

pour tout $\varphi \in \mathscr{S}(\mathbb{R}^d)$.

Lemme 29.40.
Si $f \in \mathscr{S}(\mathbb{R}^d)$, nous avons $\hat{T}_f = T_{\hat{f}}$.

Démonstration. En utilisant les définitions,

$$\hat{T}_f(\varphi) = T_f(\hat{\varphi}) = \int_{\mathbb{R}^d} f(x)\hat{\varphi}(x)dx = \int_{\mathbb{R}^d} f(x)\left[\int_{\mathbb{R}^d} \varphi(y)e^{-iyx}dy\right]dx \tag{29.93}$$

où nous avons noté xy le produit scalaire $x \cdot y$. Nous permutons les intégrales en utilisant le théorème de Fubini 13.262 avec la fonction

$$(x, y) \mapsto f(x)\varphi(y)e^{-ixy} \tag{29.94}$$

qui est parfaitement dans $L^1(\mathbb{R}^d \times \mathbb{R}^d)$. Nous écrivons alors

$$\hat{T}_f(\varphi) = \int_{\mathbb{R}^d}\left[\int_{\mathbb{R}^d} f(x)\varphi(y)e^{-iyx}dx\right]dy = \int_{\mathbb{R}^d} \varphi(y)\hat{f}(y)dy = T_{\hat{t}}(\varphi). \tag{29.95}$$

□

29.4.5 Convolution d'une distribution par une fonction

Nous savons que si $\psi \in \mathscr{S}(\mathbb{R}^d)$ et si $x \in \mathbb{R}^d$ alors la fonction $y \mapsto \psi(x - y)$ est encore une fonction dans $\mathscr{S}(\mathbb{R}^d)$. Donc si $T \in \mathscr{S}'(\mathbb{R}^d)$ nous pouvons considérer la fonction $T * \psi = \psi * T$ définie par

$$(T * \psi)(x) = T\big(y \mapsto \psi(x - y)\big). \tag{29.96}$$

Notons que $T * \psi$ est bien une fonction et non une distribution.

Le but de la définition est d'avoir

$$T_f * \psi = f * \psi. \tag{29.97}$$

En effet

$$(T_f * \psi)(x) = T_f\big(y \mapsto \psi(x - y)\big) = \int_{\mathbb{R}^d} f(y)\psi(x - y)dy = (f * \psi)(x). \tag{29.98}$$

Exemple 29.41.
La distribution de Dirac est le neutre pour le produit de convolution. En effet

$$(\delta * \psi)(x) = \delta\big(y \mapsto \psi(x - y)\big) = \psi(x), \tag{29.99}$$

c'est-à-dire $\delta * \psi = \psi$. △

Proposition 29.42 ([4]).
Si $T \in \mathscr{S}'(\mathbb{R}^d)$ et $\psi \in \mathscr{S}(\mathbb{R}^d)$, alors la distribution associée à la fonction $T * \psi$ est tempérée.

Démonstration. En agissant sur $\varphi \in \mathscr{D}(\mathbb{R}^d)$ nous avons

$$T_{T*\psi}(\varphi) = \int_{\mathbb{R}^d} T\big(y \mapsto (t_x\psi)(y)\big)\varphi(x)dx \tag{29.100a}$$
$$= \int_{\mathbb{R}^d} T\big(y \mapsto \varphi(x)\psi(x-y)\big)dx \tag{29.100b}$$
$$= T\left(y \mapsto \int_{\mathbb{R}^d} \varphi(x)\psi(x-y)dx\right) \tag{29.100c}$$
$$= T\big(y \mapsto (\varphi * \check{\psi})(y)\big) \tag{29.100d}$$
$$= T(\varphi * \check{\psi}). \tag{29.100e}$$

Attention : **Problèmes et choses à faire**

Le passage à la ligne (29.100c) n'est pas justifié. □

29.4.6 Approximation de la distribution de Dirac

Lemme 29.43 ([1]).
Soient des fonctions $j_n \colon \mathbb{R} \to \mathbb{R}^+$ de classe C^∞ telles que

(1) Pour chaque n, la fonction $x \mapsto j_n\big(|x|\big)$ est strictement décroissante et converge ponctuellement vers zéro.

(2) Pour chaque x, la suite $n \mapsto j_n(x)$ est décroissante et converge vers 0.

(3) Pour tout $M > 0$, la suite j_n converge vers zéro uniformément sur $B(0,M)^c$.

(4) Pour tout δ et ϵ, il existe un $N \in \mathbb{N}$ tel que $|\int_{B(0,\delta)} j_n(x)dx - 1| \leqslant \epsilon$.

(5) Pour tout n, nous avons $\int_{\mathbb{R}} j_n = 1$.

Alors si $u \in \mathscr{S}(\mathbb{R})$ nous avons

$$\lim_{n\to\infty} \int_{\mathbb{R}} u(x)j_n(x)dx = u(0). \tag{29.101}$$

Démonstration. Nous posons

$$I_n = \int_{\mathbb{R}} j_n u \tag{29.102a}$$
$$I_{\delta,n} = \int_{B(0,\delta)} j_n u \tag{29.102b}$$
$$Z_{\delta,n} = \int_{B(0,\delta)} u(0)j_n \tag{29.102c}$$

Nous allons progressivement montrer qu'en prenant δ assez petit et n assez grand, les quantités $|I_n - I_{\delta,n}|$, $|I_{\delta,n} - Z_{\delta,n}|$ et $|Z_{\delta,n} - u(0)|$ peuvent être simultanément majorées par ϵ.

Soient $\delta > 0$ et $\epsilon > 0$; vu que $u \in \mathscr{S}(\mathbb{R})$, il existe M tel que $\int_{|x|>M} |u| < \epsilon$. Soit $N_1 \in \mathbb{N}$ tel que pour tout $n > N_1$ nous avons $|j_n(x)| < 1$ dès que $|x| > M$ (hypothèse (3)). Alors

$$\int_{|x|>M} |j_n(x)u(x)| < \epsilon. \tag{29.103}$$

De plus en posant $s = \max\{|u(x)|$ tel que $\delta \leqslant |x| \leqslant M\}$ (qui existe parce que u est continue et prise sur un compact) nous pouvons considérer N_2 tel que $j_n(x) < \epsilon/s$ pour tout $|x| > \delta$.

Avec $n > \max\{N_1, N_2\}$ nous avons

$$|\int_{B(0,\delta)} j_n u - \int_{\mathbb{R}} j_n u| = |\int_{B(0,\delta)^c} j_n u| \tag{29.104a}$$
$$\leqslant \int_{\delta\leqslant|x|\leqslant M} |j_n u| + \int_{|x|\geqslant M} |j\nu| \tag{29.104b}$$
$$\leqslant \epsilon(1 + |M - \delta|). \tag{29.104c}$$

En redéfinissant le ϵ nous avons donc montré que pour tout ϵ et δ, il existe un $N \in \mathbb{N}$ tel que

$$|I_{\delta,n} - I_n| \leqslant \epsilon \tag{29.105}$$

dès que $n \geqslant N$.

La fonction u est uniformément continue sur tout $\overline{B(0,\delta)}$, et nous pouvons donc choisir δ tel que $|u(0) - u(x)| \leqslant \epsilon$ pour tout $x \in \overline{B(0,\delta)}$. Pour ce δ, nous avons déjà trouvé un N tel que $|I_{\delta,n} - I_n| \leqslant \epsilon$ dès que $n > N$. Nous avons :

$$|I_{\delta,n} - Z_{\delta,n}| \leqslant \int_{B(0,\delta)} |u(x) - u(0)| j_n(x) dx \tag{29.106a}$$

$$\leqslant \epsilon \int_{B(0,\delta)} j_n \tag{29.106b}$$

$$\leqslant \epsilon. \tag{29.106c}$$

Nous avons donc prouvé que pour tout $\epsilon > 0$, il existe un δ et un N tels que

$$\begin{cases} |I_{\delta,n} - I_n| \leqslant \epsilon & (29.107a) \\ |I_{\delta,n} - Z_{\delta,n}| \leqslant \epsilon & (29.107b) \end{cases}$$

dès que $n \geqslant N$.

Enfin nous avons

$$|z_{\delta,n} - u(0)| = u(0) \left(\int_{B(0,\delta)} j_n - 1 \right), \tag{29.108}$$

et par l'hypothèse (4) nous pouvons choisir n assez grand pour que la parenthèse soit plus petite que ϵ.

Pour ϵ donné, nous avons donc trouvé un δ et un N tels que

$$|I_n - u(0)| \leqslant |I_n - I_{\delta,n}| + |I_{\delta,n} - Z_{\delta,n}| + |Z_{\delta,n} - u(0)| \leqslant 3\epsilon. \tag{29.109}$$

En passant à la limite nous avons bien $I_n \to u(0)$ dans $\mathbb{R}$. □

Il va sans dire que nous connaissons de telles fonctions. Nous en donnons une maintenant.

Exemple 29.44 ([480]).
Nous introduisons la fonction f_ϵ ($\epsilon > 0$) donnée par

$$f_\epsilon(x) = e^{-\epsilon|x|}. \tag{29.110}$$

Nous calculons la transformée de Fourier de f_ϵ en divisant le domaine d'intégration :

$$\hat{f}_\epsilon(k) = \int_{\mathbb{R}} e^{-ikx} e^{-\epsilon|x|} dx = \int_{-\infty}^{0} e^{\epsilon x} e^{-ikx} dx + \int_{0}^{\infty} e^{-\epsilon x}\, e^{-ikx} dx \tag{29.111}$$

En décomposant les parties imaginaires et réelles, et avec un peu de changement de variables, nous pouvons utiliser les intégrales (16.487) et (16.488) pour obtenir

$$\hat{f}_\epsilon(k) = \frac{2\epsilon}{k^2 + \epsilon^2}. \tag{29.112}$$

Sachant que $\operatorname{arctg}(x)$ est une primitive de $\frac{1}{x^2+1}$ et avec encore un peu de changement de variables, nous avons [11]

$$\int_{\mathbb{R}} \hat{f}_\epsilon(k) dk = \int_{-\infty}^{\infty} \frac{2\epsilon}{k^2 + \epsilon^2} = 2[\operatorname{arctg}(x/\epsilon)]_{-\infty}^{\infty} = 2\pi. \tag{29.113}$$

11. Et en écrivant correctement l'intégrale sur $\mathbb{R}$ comme une limite, etc.

Cela montre que si nous introduisons la fonction δ_ϵ donnée par

$$\delta_\epsilon(k) = \frac{1}{\pi}\frac{\epsilon}{\epsilon^2 + k^2}, \tag{29.114}$$

alors nous avons une fonction qui tout en même temps ressemble à $\hat{f}_\epsilon$ et vérifie

$$\int_{\mathbb{R}} \delta_\epsilon(k)dk = 1 \tag{29.115}$$

pour tout ϵ.

Jusqu'ici nous avons montré que

$$\int_{\mathbb{R}} e^{-ikx}e^{-\epsilon|x|}dx = 2\pi\delta_\epsilon(k). \tag{29.116}$$

Pour chaque $\epsilon > 0$ nous avons $\delta_\epsilon \in L^1(\mathbb{R})$. △

Proposition 29.45 ([1]).
Soit $g \in \mathscr{S}(\mathbb{R})$. Alors nous avons

$$\int_{\mathbb{R}}\int_{\mathbb{R}} g(x)e^{-ixy}dx\,dy = 2\pi g(0). \tag{29.117}$$

Démonstration. Soit $u \in \mathscr{S}(\mathbb{R})$; nous multiplions l'équation (29.116) par $u(k)$ et nous intégrons par rapport à k :

$$\int_{\mathbb{R}} u(k)\left[\int_{\mathbb{R}} e^{-ikx}e^{-\epsilon|x|}dx\right]dk = 2\pi\int_{\mathbb{R}} u(k)\delta_\epsilon(k)dk. \tag{29.118}$$

Il s'agit de passer à la limite dans l'équation (29.118). Les intégrales à gauche peuvent effectuées séparément parce qu'elles respectent le théorème de Fubini. En effet soit la fonction

$$f(k,x) = u(k)e^{-ikx}e^{-\epsilon|x|} \tag{29.119}$$

qui est dans $L^1(\mathbb{R}\times\mathbb{R})$ en vertu du critère du corolaire 13.261 et du fait que à la fois $k \mapsto |u(k)|$ et $x \mapsto e^{-\epsilon|x|}$ sont dans $L^1(\mathbb{R})$.

Nous pouvons donc grouper et dégrouper les intégrales et en particulier les inverser. Si nous effectuons d'abord l'intégrale sur k nous trouvons

$$\int_{\mathbb{R}} u(k)\left[\int_{\mathbb{R}} e^{-ikx}e^{-\epsilon|x|}dx\right]dk = \int_{\mathbb{R}} e^{-\epsilon|x|}\int_{\mathbb{R}} u(k)e^{-ikx}dk\,dx = \int_{\mathbb{R}} e^{-\epsilon|x|}\hat{u}(x)dx. \tag{29.120}$$

La fonction $x \mapsto |e^{-\epsilon|x|}\hat{u}(x)|$ est majorée (uniformément en ϵ) par $x \mapsto \hat{u}(x)$ qui est intégrable parce que la transformée de Fourier d'une fonction de $\mathscr{S}$ est dans $\mathscr{S}$ par la proposition 28.20. Le théorème de la convergence dominée de Lebesgue 13.184 nous permet de permuter la limite $\epsilon \to 0$ avec l'intégrale et obtenir

$$\lim_{\epsilon\to 0}\int_{\mathbb{R}} u(k)\int_{\mathbb{R}} e^{-ikx}e^{-\epsilon|x|}dx\,dk = \int_{\mathbb{R}}\hat{u}(x)dx = \int_{\mathbb{R}}\int_{\mathbb{R}} u(k)e^{-ikx}dk\,dx. \tag{29.121}$$

Notons qu'en passant à la limite nous avons perdu le droit de permuter les intégrales.

Nous devons encore prouver que

$$\lim_{\epsilon\to 0}\int_{\mathbb{R}} u(k)\delta_\epsilon(k)dk = u(0). \tag{29.122}$$

Cela n'est rien d'autre que le lemme 29.43 appliqué à la suite de fonctions $j_n = \delta_{1/n}$. □

29.46.
Notons que les intégrales dans (29.117) ne peuvent pas être permutées parce que $\int_{\mathbb{R}} e^{-ixy}dy$ n'existe pas. Il faut avouer que, malgré tous les conseils du type « attention : permuter des intégrales doit être fait avec prudence », ce n'est pas tous les jours que nous trouvons des intégrales qui ne peuvent pas être permutées, autrement que dans des exemples fait exprès.

29.4.7 Peigne de Dirac

Proposition 29.47.
La formule

$$\Delta_a = \sum_{k\in\mathbb{Z}} \delta_{ka} \tag{29.123}$$

définit un élément de $\mathscr{D}'(\mathbb{R})$.

La forme linéaire Δ_a est le **peigne de Dirac** de pas a.

Démonstration. Nous utilisons le critère de continuité séquentielle en zéro du théorème 29.16. Soit une suite $\varphi_n \to 0$ dans $\mathscr{D}(\mathbb{R})$. Par le théorème 29.13 il existe un compact K de $\mathbb{R}$ pour lequel $\varphi_n \in \mathscr{D}(K)$ pour tout n et $\varphi_n \to 0$ dans $\mathscr{D}(K)$. La somme 29.123 est donc finie et nous pouvons la permuter avec une limite :

$$\lim_{n\to\infty} \Delta_a(\varphi_n) = \sum_{k\in\mathbb{Z}} \lim_{n\to\infty} \varphi_n(ka). \tag{29.124}$$

La limite $\varphi_n \to 0$ dans $\mathscr{D}(K)$ signifie que nous avons convergence uniforme de la fonction et de toutes ses dérivées vers 0. En particulier $\|\varphi_n\|_\infty \to 0$; disons que la somme (qui est finie) fasse s termes :

$$\sum_{k\in\mathbb{Z}} \varphi_n(ka) \leqslant s\|\varphi_n\|_\infty. \tag{29.125}$$

Le terme de droite tend vers zéro lorsque n tend vers l'infini. □

Donc Δ_a est bien une distribution au sens de la définition 29.15.

Lemme 29.48 ([455])**.**
Le peigne de Dirac vérifie la relation

$$\Delta_a = \frac{1}{a}\Delta_1 \circ D_a \tag{29.126}$$

où D_a *est l'application* $D_a\colon \mathscr{D}(\mathbb{R}) \to \mathscr{D}(\mathbb{R})$,

$$(D_a f)(x) = af(ax). \tag{29.127}$$

Démonstration. Pour $\varphi \in \mathscr{D}(\mathbb{R})$ nous avons

$$\Delta_a(\varphi) = \sum_{k\in\mathbb{Z}} \varphi(ka) = \frac{1}{a}\sum_{k\in\mathbb{Z}} (D_a\varphi)(k) = \frac{1}{a}\Delta_1(D_a\varphi). \tag{29.128}$$

□

Proposition 29.49.
Le peigne de Dirac est une distribution tempérée.

Notez qu'il y a plus de fonctions dans $\mathscr{S}(\mathbb{R})$ que dans $\mathscr{D}(\mathbb{R})$; il est donc plus difficile de rentrer dans $\mathscr{S}'(\mathbb{R})$ que dans $\mathscr{D}'(\mathbb{R})$: il est plus compliqué d'avoir existence de $T(\varphi)$ pour tout $\varphi \in \mathscr{S}(\mathbb{R})$ que pour tout $\varphi \in \mathscr{D}(\mathbb{R})$.

Démonstration. Soit $\varphi \in \mathscr{S}(\mathbb{R})$. Nous avons

$$|\Delta_a(\varphi)| = |\sum_k \varphi(ak)| = \left|\sum_k \frac{(1+a^2k^2)\varphi(ak)}{1+a^2k^2}\right| \leqslant_{x\in\mathbb{R}} \big|(1+x^2)\varphi(x)\big| \sum_k \frac{1}{1+a^2k^2}. \tag{29.129}$$

La somme $\sum_k \frac{1}{1+a^2k^2}$ est une somme convergente, et et supremum est borné par la proposition 26.146 en prenant $Q(x) = 1 + x^2$. En effet sur $\overline{B(0,r)}$ la fonction $x \mapsto (1+x^2)\varphi(x)$ est bornée par ce que c'est une fonction continue sur un compact, et à l'extérieur de $B(0,r)$ cette fonction est alors bornée par 1. □

Si aucune ambigüité n'est à craindre, nous noterons f la distribution T_f.

Exemple 29.50.
La transformée de Fourier de la distribution de Dirac est la fonction constante : $\hat{\delta} = 1$. En effet si nous agissons sur une fonction test,

$$\hat{\delta}(\varphi) = \delta(\hat{\varphi}) = \hat{\varphi}(0) = \int_{\mathbb{R}^d} \varphi(x)dx. \tag{29.130}$$

△

29.5 L'espace $C^\infty(\mathbb{R}, \mathscr{D}'(\mathbb{R}^d))$

D'abord parlons un peu de continuité en recopiant la proposition 7.336 dans notre contexte.

Proposition 29.51.
Soient I un intervalle ouvert de $\mathbb{R}$ et $u\colon I \to \mathscr{D}'(\mathbb{R}^d)$ une fonction continue. Alors

(1) Pour tout $\varphi \in \mathscr{D}(\mathbb{R}^d)$, l'application $t \mapsto u_t(\varphi)$ est continue.

(2) Pour tout $\varphi \in \mathscr{D}(\mathbb{R}^d)$, nous avons la limite

$$\lim_{t\to t_0} u_t(\varphi) = u_{t_0}(\varphi). \tag{29.131}$$

(3) Nous avons la limite dans $\mathscr{D}'(\mathbb{R}^d)$

$$\lim_{t\to t_0} u_t = u_{t_0}. \tag{29.132}$$

En ce qui concerne la définition de l'espace $C^\infty(I, \mathscr{D}'(\mathbb{R}^d))$, c'est la définition 7.337. Grâce au point (1) de la proposition 29.51, nous retenons que la propriété fondamentale d'une application $T \in C^k\big(I, \mathscr{D}'(\Omega)\big)$ est que pour tout $\varphi \in (\Omega)$, l'application

$$\begin{aligned} I &\to \mathbb{C} \\ t &\mapsto T_t(\varphi) \end{aligned} \tag{29.133}$$

est de classe C^k.

Proposition 29.52.
Soit $(T_t) \in C^0\big(I, \mathscr{D}'(\Omega)\big)$ et $\psi \in \mathscr{D}(I \times \Omega)$. Alors l'application

$$t \mapsto T_t\big(\psi(t, .)\big) \tag{29.134}$$

est continue sur I.

Démonstration. La fonction dont nous voulons prouver la continuité est une fonction $\mathbb{R} \to \mathbb{R}$; il est donc loisible de se contenter de la continuité séquentielle. Soient $t_0 \in I$ et (t_j) une suite dans I convergeant vers t_0. Nous posons $U_j = T_{t_j}$ et $\psi_j = \psi\big(t_j, .\big)$. Par hypothèse de continuité de (T_t) nous avons $U_j \xrightarrow{\mathscr{D}'(\Omega)} T_{t_0}$. D'autre part le support de ψ étant compact nous avons $\operatorname{supp}(\psi) \subset [c, d] \times K$ où $[c, d] \subset I$ et K est compact dans Ω. Par conséquent nous avons aussi $\operatorname{supp}(\psi_j) \subset K$.

Affin d'alléger les notations notons $\tilde{\psi}(x) = \psi(t_0, x)$. Pour tout multiindice α et pour tout j nous avons

$$p_\alpha(\psi_i - \tilde{\psi}) = \Big|\partial^\alpha \psi(t_j, x) - \partial^\alpha \psi(t_0, x)\Big| \leqslant |t_j - t_0| \sup_{\substack{t\in[c,d]\\ x\in K}} |\partial_t \partial^\alpha \psi(t, x)| \to 0. \tag{29.135}$$

Nous avons donc la convergence

$$\psi_j \xrightarrow{\mathscr{D}(K)} \psi(t_0, .). \tag{29.136}$$

Étant donné que $U_j \xrightarrow{\mathscr{D}'(\Omega)} T_{t_0}$ et $\psi_j \xrightarrow{\mathscr{D}(\Omega)} \tilde{\psi}$, le point (3) du corolaire 26.6 nous donne la convergence

$$U_j(\psi_j) \to T_{t_0}(\tilde{\psi}) \tag{29.137}$$

dans $\mathbb{C}$. Cela est bien la continuité de la fonction $t \mapsto T_t\big(\psi(t,.)\big)$. □

Proposition 29.53 ([478]).
Soit $(T_t) \in C^0\big(I, \mathscr{D}'(\Omega)\big)$. Nous définissons l'application $T\colon \mathscr{D}(\Omega) \to \mathbb{C}$ par la formule

$$T(\psi) = \int_I T_t\big(\psi(t,.)\big)\, dt \tag{29.138}$$

pour tout $\psi \in \mathscr{D}(I \times \Omega)$. Alors $T \in \mathscr{D}'(I \times \Omega)$.

Démonstration. La proposition 29.52 nous indique que la fonction $t \mapsto T_t\big(\psi(t,.)\big)$ est continue. Étant donné qu'elle est seulement non nulle sur un compact, l'intégrale

$$\int_I T_t\big(\psi(t,.)\big) dt \tag{29.139}$$

a un sens et est finie. L'application $T\colon \mathscr{D}(I \times \Omega) \to \mathbb{C}$ ainsi définie est linéaire. Il reste à voir qu'elle est continue. Pour cela nous allons utiliser le théorème 29.16(2) qui nous dit que nous pouvons nous fixer un compact $[c,d] \times K \subset I \times \Omega$ et considérer $\psi \in \mathscr{D}\big([c,d] \times K\big)$.

Soit, pour commencer, donnée une application $\varphi \in \mathscr{D}(K)$. L'application $t \mapsto T_t(\varphi)$ est continue et non nulle sur le et il existe donc $C_\varphi > 0$ tel que

$$|T_t(\varphi)| \leqslant C_\varphi \tag{29.140}$$

pour tout $t \in [c,d]$.

Nous voulons utiliser le théorème de Banach-Steinhaus dans sa version 26.7 sur la famille d'applications paramétrée par $u \in [c,d]$:

$$\begin{aligned} U_u\colon \mathscr{D}\big([c,d] \times K\big) &\to \mathbb{R} \\ \psi &\mapsto T_u\big(\psi(u,.)\big). \end{aligned} \tag{29.141}$$

Commençons par prouver que cela est une application continue pour chaque u. Ce sera le cas si la projection

$$\begin{aligned} \texttt{proj}\colon \mathscr{D}\big([c,d] \times K\big) &\to \mathscr{D}(K) \\ \psi &\mapsto \psi(u,.) \end{aligned} \tag{29.142}$$

est continue. Pour cela nous notons P_{kl} la semi-norme sur $\mathscr{D}\big([c,d] \times K\big)$ donnée par

$$P_{k,l}(\psi) = \sum_{n \leqslant k} \sum_{|\alpha| \leqslant l} \sup_{\substack{t \in [c,d] \\ x \in K}} \big|\partial_t^n \partial^\alpha \psi(t,x)\big|. \tag{29.143}$$

Nous montrons que $\texttt{proj}$ est séquentiellement continue ; étant donné que les topologies sur $\mathscr{D}(K)$ et $\mathscr{D}\big([c,d] \times K\big)$ sont données par des métriques (proposition 29.14), cela suffit pour assurer la continuité grâce à la proposition 7.268. Montrons que si $\psi_n \xrightarrow{\mathscr{D}\left([c,d]\times K\right)} 0$, alors $\texttt{proj}(\psi_n) \xrightarrow{\mathscr{D}(K)} 0$. Pour cela nous remarquons que

$$p_j\big(\texttt{proj}(\psi)\big) = \sum_{|\alpha| \leqslant j} \sup_{x \in K} |\partial^\alpha \psi(u,x)| \tag{29.144a}$$

$$\leqslant \sum_{|\alpha| \leqslant j} \sup_{t \in [c,d]} \sup_{x \in K} |\partial^\alpha \psi(t,x)| \tag{29.144b}$$

$$= P_{0,j}(\psi). \tag{29.144c}$$

Par conséquent

$$p_j\big(\texttt{proj}(\psi_n)\big) \leqslant P_{0,j}(\psi_n) \to 0 \tag{29.145}$$

où nous avons utilisé la proposition 7.329. Utilisant cette même proposition à l'envers, nous déduisons que $\texttt{proj}(\psi_n) \overset{\mathscr{D}(K)}{\longrightarrow} 0$. Les applications U_u sont donc continues ; elles sont également bornées parce que si $\psi \in \mathscr{D}\big([c,d]\times K\big)$ nous avons

$$\sup_{u\in[c,d]} \big|U_u(\psi)\big| = \sup_u \big|T_u\big(\psi(u,.)\big)\big|, \tag{29.146}$$

et la continuité déjà évoquée, sur le compact $[c,d]$, nous dit que cette quantité est finie. Le théorème de Banach-Steinhaus peut maintenant être appliqué et il existe $C>0$ et $k,l\in\mathbb{N}$ tels que pour tout $\psi\in\mathscr{D}\big([c,d]\times K\big)$,

$$\big|U_u(\psi)\big| \leqslant CP_{k,l}(\psi) = C\sum_{|\alpha|\leqslant k}\sum_{n\leqslant l}\sup_{t,x}\big|\partial_t^n\partial^\alpha\psi(t,x)\big| \leqslant C\sum_{|\alpha|+n\leqslant k+l}\sup_{t,x}\big|\partial_t^n\partial^\alpha\psi(t,x)\big|. \tag{29.147}$$

Quelques remarques

— Nous n'avons pas mis de maximum devant le supremum (alors que la conclusion (26.5) en demande) parce que dans le cas des semi-normes P_{kl}, c'est toujours celle avec k et l le plus grand possible qui sont les plus grandes parce qu'elles sont des sommes emboitées les unes les autres.

— La fusion de deux sommes est bien une majoration parce qu'il y a plus de termes dans la seconde que dans la première.

— La quantité la plus à droite est (à part le C) ce que nous pouvons noter $P_{k+l}(\psi)$: c'est bien une des semi-normes associées à l'espace de dimension $d+1$.

Nous majorons maintenant $T(\psi)$ par

$$\big|T(\psi)\big| \leqslant \int_c^d \big|T_t\big(\psi(t,.)\big)\big|dt = \int_c^d \big|U_t(\psi)\big|dt \leqslant C|d-c|P_{k+l}(\psi). \tag{29.148}$$

Maintenant le théorème 29.16(2) appliqué à l'ouvert $I\times\Omega$ et avec ψ au lieu de φ nous informe que $T\in\mathscr{D}(I\times K)$. □

29.5.1 Dérivation

Quelques propriétés de dérivation des fonctions $I\to\mathscr{D}(\Omega)$ seront directement énoncées et démontrées dans le cas des distributions tempérées. Les résultats 29.60 et 29.61 seront a fortiori valables si nous remplaçons $\mathscr{S}$ par $\mathscr{D}$.

29.6 L'espace $C^\infty(\mathbb{R},\mathscr{S}'(\mathbb{R}^d))$

Dans cette section nous notons I un ouvert de $\mathbb{R}$ et Ω un ouvert de $\mathbb{R}^d$; si ψ est une fonction sur $I\times\Omega$ nous allons noter $\psi_t\colon\Omega\to\mathbb{R}$ la fonction $\psi_t(x)=\psi(t,x)$. C'est une notation plus légère que $\psi(t,.)$.

29.6.1 Propriétés générales

La définition de l'espace $C^\infty(I,\mathscr{S}'(\Omega))$ est encore la définition 7.337 et les propriétés énoncées dans la proposition 29.51 sont encore bonnes ici.

D'abord parlons un peu de continuité en recopiant la proposition 7.336 dans notre contexte.

Proposition 29.54.
Soient I un intervalle ouvert de $\mathbb{R}$ et $u\colon I\to\mathscr{S}'(\mathbb{R}^d)$ une fonction continue. Alors

(1) Pour tout $\varphi\in\mathscr{S}(\mathbb{R}^d)$, l'application $t\mapsto u_t(\varphi)$ est continue.

(2) Pour tout $\varphi \in \mathscr{S}(\mathbb{R}^d)$, nous avons la limite

$$\lim_{t\to t_0} u_t(\varphi) = u_{t_0}(\varphi). \tag{29.149}$$

(3) Nous avons la limite dans $\mathscr{S}'(\mathbb{R}^d)$

$$\lim_{t\to t_0} u_t = u_{t_0}. \tag{29.150}$$

Lemme 29.55.
Nous avons $C^\infty(I, \mathscr{S}'(\Omega)) \subset C^\infty(I, \mathscr{D}'(\Omega))$.

Démonstration. Soit $(T_t) \in C^\infty(I, \mathscr{S}'(\Omega))$. Pour chaque t nous avons

$$T_t \in \mathscr{S}'(\Omega) \subset \mathscr{D}'(\Omega). \tag{29.151}$$

Ensuite il suffit de dire que pour tout $\varphi \in \mathscr{D}(\Omega)$ la fonction

$$t \mapsto T_t(\varphi) \tag{29.152}$$

est de classe C^∞ parce que c'est le cas pour toute fonction dans $\mathscr{S}(\Omega)$. La proposition 29.51 (en changeant $\mathscr{D}$ en $\mathscr{S}$) conclut que $(T_t) \in C^\infty(I, \mathscr{D}'(\Omega))$. □

Proposition 29.56.
L'espace $\mathscr{S}(\Omega)$ est complet et métrisable.

Démonstration. En ce qui concerne le métrisable nous reprenons la formule de l'écart (7.405). Dans notre cas pour l'écrire explicitement il faudrait une énumération de $\mathbb{N}^2$ à partir de 1 (et non de zéro). Cette formule donne bien une distance parce que si $d(\varphi_1 - \varphi_2) = 0$ alors en particulier $p_{00}(\varphi_1 - \varphi_2) = \|\varphi_1 - \varphi_2\|_\infty = 0$ et donc $\varphi_1 = \varphi_2$.

Nous montrons maintenant que $\mathscr{S}(\Omega)$ est complet en y considérant une suite de Cauchy (φ_n). Soit $\epsilon > 0$ et $\alpha, \beta \in \mathbb{N}$ ainsi que k, l assez grands pour que $\varphi_k - \varphi_l \in B_{\alpha\beta}(0, \epsilon)$. En particulier pour $\alpha = \beta = 0$ nous avons $\|\varphi_k - \varphi_l\|_\infty \leqslant \epsilon$, ce qui signifie que nous avons une suite vérifiant le critère de Cauchy uniforme 11.293. Elle converge donc uniformément vers une certaine fonction φ que la proposition 11.294 nous assure être continue. Il existe donc $\varphi \in C(\Omega)$ telle que

$$\varphi_k \xrightarrow{unif} \varphi. \tag{29.153}$$

Nous devons montrer que $\varphi \in \mathscr{S}(\Omega)$. Le fait que φ soit de classe C^∞ s'obtient en utilisant les semi-normes $p_{0,\alpha}(\varphi) = \|\partial^\alpha \varphi\|_\infty$ de la même façon que dans la preuve que $\mathscr{D}(\Omega)$ était complet (proposition 29.14). Nous obtenons en particulier que

$$\partial^\alpha \varphi_k \xrightarrow{unif} \partial^\alpha \varphi \tag{29.154}$$

pour tout multiindice α. Montrons encore que φ est à décroissance rapide : nous devons montrer que pour tout α et β nous avons

$$p_{\alpha\beta}(\varphi) = \sup_{x\in\Omega} \left| x^\beta (\partial^\alpha \varphi)(x) \right| < \infty. \tag{29.155}$$

Étant donné que (φ_n) est de Cauchy dans $\mathscr{S}(\Omega)$ nous avons (pour ϵ fixé et k, l assez grands) :

$$\left| x^\beta (\partial^\alpha \varphi_k - \partial^\alpha \varphi_l)(x) \right| \leqslant \epsilon \tag{29.156}$$

pour tout $x \in \Omega$. En considérant l fixé et en prenant la limite $k \to \infty$ et en utilisant la convergence uniforme (29.154) nous trouvons que

$$\left| x^\beta (\partial^\alpha \varphi - \partial^\alpha \varphi_l)(x) \right| \leqslant \epsilon \tag{29.157}$$

Du coup nous pouvons faire la majoration

$$\sup_{x\in\Omega} \left| x^\beta (\partial^{\alpha\varphi})(x) \right| \leqslant \sup_x \left| x^\beta (\partial^\alpha \varphi - \partial^\alpha \varphi_l)(x) \right| + \sup_x \left| (\partial^\alpha \varphi_l)(x) \right| \leqslant \epsilon + p_{\alpha\beta}(\varphi_l) < \infty \tag{29.158}$$

du fait que $p_{\alpha\beta}(\varphi_l) < \infty$ parce que $\varphi_l \in \mathscr{S}(\Omega)$.

Donc $\varphi \in \mathscr{S}(\Omega)$ et ce dernier est alors complet. □

Proposition 29.57.
Soit $(T_t)\in C^0\big(I,\mathscr{S}'(\Omega)\big)$ et $\psi\in\mathscr{S}(I\times\Omega)$. Alors la fonction

$$t\mapsto T_t(\psi_t) \tag{29.159}$$

est continue sur I.

Démonstration. Soient $t_0\in I$ et une suite convergente vers t_0 : $t_j\to t_0$ dans $\mathbb{R}$. Vu que (T_t) est continue en t, elle est en particulier séquentiellement continue et nous avons

$$T_{t_j}\xrightarrow{\mathscr{S}'(\Omega)}T_{t_0}. \tag{29.160}$$

Montrons que nous avons aussi $\psi_{t_j}\xrightarrow{\mathscr{S}(\Omega)}\psi_{t_0}$. Pour cela nous utilisons les semi-normes [12] $p_{\alpha\beta}$ définies en (26.567) :

$$p_{\alpha\beta}(\psi_{t_j}-\psi_{t_0})=\sum_{x\in\Omega}\Big|x^\beta\big(\partial^\alpha\psi(t_j,x)-\partial^\alpha\psi(t_0,x)\big)\Big| \tag{29.161a}$$

$$\leqslant\sup_{x\in\Omega}\Big|x^\beta|t_0-t_j|\sup_{t\in[t_0,t_j]}\big|\partial_t\partial^\alpha\psi(t,x)\big|\Big| \tag{29.161b}$$

$$\leqslant|t_0-t_j|\sup_{x\in\Omega}\sup_{t\in I}\big|x^\beta\partial_t\partial^\alpha\psi(t,x)\big| \tag{29.161c}$$

$$\leqslant|t_0-t_j|P_{(\alpha t),\beta}(\psi). \tag{29.161d}$$

Pour la première majoration nous avons utilisé le théorème des accroissements finie 11.257. Pour la dernière ligne nous avons noté $P_{\alpha\beta}$ les semi-normes de $\mathscr{S}(I\times\Omega)$ et $(t\alpha)$ est le multiiindice qui commence par la variable t et qui continue par α. Étant donné que $P_{(\alpha t)\beta}(\psi)<\infty$ nous avons bien

$$p_{\alpha\beta}(\psi_{t_j}-\psi_{t_0})\to 0 \tag{29.162}$$

et donc $\psi_{t_j}\xrightarrow{\mathscr{S}(\Omega)}\psi_{t_0}$.

Étant donné que $\mathscr{S}(\Omega)$ est métrisable et complet, le corolaire 26.6 nous dit que

$$T_{t_j}(\psi_{t_j})\to T_{t_0}(\psi_{t_0}), \tag{29.163}$$

ce qui est bien le critère de continuité séquentielle de la fonction (29.159). □

Remarque 29.58.
La proposition 29.53 nous dit, a fortiori, que si $(T_t)\in C^\infty\big(I,\mathscr{S}'(\Omega)\big)$ alors la formule

$$\tilde{T}(\psi)=\int_I T_t(\psi_t) \tag{29.164}$$

donne un élément $\tilde{T}\in\mathscr{D}'(I\times\Omega)$. Au cas où aucune confusion n'est à craindre, nous pourrons noter également T l'élément de $\mathscr{D}'(I\times\Omega)$ déduit de $T\in C^\infty\big(I,\mathscr{S}'(\Omega)\big)$.

Notons que ce T ne sera pas toujours une distribution tempérée comme le montre l'exemple suivant.

Exemple 29.59.
En posant $T_t(\varphi)=e^{t^2}\varphi(0)$ avec $I=\mathbb{R}$, l'intégrale

$$T(\psi)=\int_{\mathbb{R}}T_t(\psi_t)=\int_{\mathbb{R}}e^{t^2}\psi(t,0)dt \tag{29.165}$$

ne converge pas pour tout $\psi\in\mathscr{S}(\mathbb{R}\times\Omega)$. En effet par rapport à t, la fonction $\psi(t,0)$ décroît rapidement mais pas spécialement assez rapidement pour compenser e^{t^2}. △

12. Pas parce que nous en avons envie, mais bien parce qu'elles font partie de la définition de la convergence et de tous ces trucs.

29.6.2 Dérivation

Proposition 29.60 ([478]).
Soit $T \in C^k\big(I, \mathscr{S}'(\Omega)\big)$ et $0 \leqslant l \leqslant k$. Pour tout $t_0 \in I$ l'application

$$\begin{aligned} T_{t_0}^{(l)} \colon \mathscr{S}(\Omega) &\to \mathbb{C} \\ \varphi &\mapsto \left(\frac{d^l}{dt} T_t(\varphi)\right)(t_0) \end{aligned} \tag{29.166}$$

est bien définie, est une distribution et de plus

$$t \mapsto T_t^{(l)} \in C^{k-l}\big(I, \mathscr{S}'(\Omega)\big). \tag{29.167}$$

Attention que la formule (29.166) est bonne si $\varphi \in \mathscr{S}(\Omega)$. Si par contre $\psi \in \mathscr{S}(I \times \Omega)$ et qu'on veut regarder $u_t^{(1)}(\psi_t)$ alors il faut regarder la proposition 29.61 et utiliser la formule (29.172) dans laquelle se trouve $u_t^{(1)}(\psi_t)$.

Démonstration. Pour $k = 0$ nous avons $T_t^{(0)} = T_t$ et c'est bon. Pour le cas $k = 1$ et $l = 0$ c'est encore $T_t^{(0)} = T_t$ qui fonctionne.

Le premier cas non trivial à traiter est $k = 1$ et $l = 1$. Nous considérons $t_0 \in I$; par définition de la dérivée, pour tout $\varphi \in \mathscr{S}(\Omega)$, nous avons (pour peu que les limites existent) :

$$T_{t_0}^{(1)}(\varphi) = \frac{d}{dt}\Big[T_t(\varphi)\Big]_{t=t_0} = \lim_{j\to\infty} \frac{T_{t_0+\epsilon_j}(\varphi) - T_{t_0}(\varphi)}{\epsilon_j} = \lim_{j\to\infty} U_j(\varphi) \tag{29.168}$$

où

$$U_j = \frac{1}{\epsilon_j}\big(T_{t_0+\epsilon_j} - T_{t_0}\big). \tag{29.169}$$

et (ϵ_j) est une suite de réels tendant vers zéro.

Vu que $(T_t) \in C^k(I, \mathscr{S}'(\Omega))$, l'application $t \mapsto T_t(\varphi)$ est de classe C^k et en particulier l'expression (29.168) a une limite lorsque $j \to \infty$. Donc $T_{t_0}^{(1)}(\varphi)$ est bien définie. Le point (1) du corolaire 26.6 nous dit que $\lim_{j\to\infty} U_j(\varphi) = T_{t_0}^{(1)}(\varphi)$ et $T_{t_0}^{(1)}$ est une distribution (linéaire et continue).

Nous devons encore voir que $t \mapsto T_t^{(1)}$ est une application $C^0\big(I, \mathscr{S}'(\Omega)\big)$. Cela est une conséquence du fait que (T_t) soit de classe C^1, ce qui se traduit par le fait que l'application

$$t \mapsto \frac{d}{dt}\Big(T_t(\varphi)\Big) \tag{29.170}$$

est continue (définition de la dérivée et point (1) de la proposition 29.54 appliquée à la dérivée).

Les cas $k \geqslant 1$ se traitent par récurrence. □

Proposition 29.61 ([478]).
Soit $(T_t) \in C^1\big(I, \mathscr{S}'(\Omega)\big)$ et $\psi \in \mathscr{S}(I \times \Omega)$. Alors la fonction

$$t \mapsto T_t\big(\psi(t, .)\big) \tag{29.171}$$

est de classe C^1 sur I et

$$\frac{d}{dt}\Big(T_t\big(\psi(t, .)\big)\Big) = T_t^{(1)}\big(\psi(t, .)\big) + T_t\left(\frac{\partial\psi}{\partial t}(t, .)\right) \tag{29.172}$$

Démonstration. Soient $t_0 \in I$ et $\epsilon_j \to 0$ une suite réelle. Le membre de gauche de (29.172), écrit en t_0, donne

$$\spadesuit = \lim_{j\to\infty} \frac{T_{t_0+\epsilon_j}\big(\psi(t_0+\epsilon_j, .)\big) - T_{t_0}\big(\psi(t_0, .)\big)}{\epsilon_j} \tag{29.173}$$

Afin d'alléger les notations nous allons écrire $\psi_t = \psi(t,.)$. Dans le numérateur de (29.173) nous ajoutons et soustrayons la quantité $T_{t_0+\epsilon_j}(\psi_{t_0})$ et nous découpons la limite en deux morceaux :

$$\spadesuit = \lim_{j\to\infty} \frac{T_{t_0+\epsilon_j}(\psi_{t_0+\epsilon_j} - \psi_{t_0})}{\epsilon_j} + \lim_{j\to\infty} \frac{(T_{t_0+\epsilon_j} - T_{t_0})(\psi_{t_0})}{\epsilon_j} \tag{29.174}$$

Le second terme vaut

$$\frac{d}{dt}\Big(T_t(\psi_{t_0})\Big)_{t=t_0} = T_{t_0}^{(1)}(\psi_{t_0}) \tag{29.175}$$

par la proposition 29.60. Occupons nous de l'autre morceau de $\spadesuit$. Nous posons $U_j = T_{t_0+\epsilon_j}$ et

$$\varphi_j = \frac{1}{\epsilon_j}(\psi_{t_0+\epsilon_j} - \psi_{t_0}). \tag{29.176}$$

Nous voulons utiliser le corolaire 26.6(3) pour obtenir

$$\lim_{j\to\infty} U_j(\varphi_j) = T_{t_0}\Big(\frac{\partial\psi}{\partial t}(t_0,.)\Big). \tag{29.177}$$

D'une part (T_t) est de classe C^∞ en t et nous avons donc la convergence $U_j \overset{\mathscr{S}'(\Omega)}{\longrightarrow} T_{t_0}$. Reste à prouver que

$$\varphi_j \overset{\mathscr{S}(\Omega)}{\longrightarrow} \frac{\partial\psi}{\partial t}(t_0,.). \tag{29.178}$$

Cela en remarquant bien que la variable de dérivation n'est pas celle par rapport à laquelle nous voulons la convergence Schwartz [13]. Soient α et β des naturels et calculons un peu :

$$p_{\alpha\beta}\big(\varphi_j - \frac{\partial\psi}{\partial t}(t_0,.)\big) = \sup_{x\in\Omega}\Big|x^\beta\partial^\alpha\Big(\frac{1}{\epsilon_j}\big(\psi(t_0+\epsilon_j,x) - \psi(t_0,x)\big) - \frac{\partial\psi}{\partial t}(t_0,x)\Big)\Big| \tag{29.179}$$

Il est à présent l'heure d'utiliser un développement de Taylor avec le reste de la proposition 11.386 :

$$\psi(t_0+\epsilon_j,x) = \psi(t_0,x) + \epsilon_j\frac{\partial\psi}{\partial t}(t_0,x) + \frac{\epsilon_j^2}{2}\frac{\partial^2\psi}{\partial t^2}(\bar{t},x) \tag{29.180}$$

pour un certain $\bar{t}\in[t_0,t_0+\epsilon_j]$. En mettant ça dans le calcul (29.179) nous restons avec

$$p_{\alpha\beta}\big(\varphi_j - \frac{\partial\psi}{\partial t}(t_0,.)\big) = \sup_{x\in\Omega}\Big|x^\beta\partial^\alpha\Big(\epsilon_j\frac{\partial^2\psi}{\partial t^2}(\bar{t},x)\Big)\Big| \leqslant \epsilon_j P_{\alpha,2;\beta,0}(\psi) \tag{29.181}$$

où $P_{\alpha,k;\beta,l}$ sont les semi-normes de $\mathscr{S}(I\times\Omega)$ avec la notation plus ou moins évidente de prendre α dérivations sur x, k sur t puis de multiplier par $x^\beta t^l$. Au final nous avons bien

$$\lim_{j\to\infty} p_{\alpha\beta}\big(\varphi_j - \frac{\partial\psi}{\partial t}(t_0,.)\big) = 0 \tag{29.182}$$

et donc la convergence $\varphi_j \overset{\mathscr{S}(\Omega)}{\longrightarrow} \frac{\partial\psi}{\partial t}(t_0,.)$. □

Lemme 29.62.
Soit $(T_t)\in C^1\big(I,\mathscr{S}'(\Omega)\big)$ alors si $\mathcal{F}$ dénote la transformée de Fourier nous avons

$$\mathcal{F}\big(T_t^{(1)}\big) = (\mathcal{F}T)_t^{(1)} \tag{29.183}$$

où $(\mathcal{F}T)$ est la famille de distributions $(\mathcal{F}T)_t = \mathcal{F}T_t$.

Démonstration. Pour la preuve il suffit de tester l'égalité sur une fonction $\varphi\in\mathscr{S}(\Omega)$:

$$(\mathcal{F}T_t^{(1)})(\varphi) = T_t^{(1)}(\mathcal{F}\varphi) = \frac{d}{dt}\Big(T_t(\mathcal{F}\varphi)\Big) = \frac{d}{dt}\Big((\mathcal{F}T_t)(\varphi)\Big) = (\mathcal{F}T)_t^{(1)}(\varphi). \tag{29.184}$$

□

13. Je ne sais pas si je me suis bien fait comprendre là.

29.7 Une équation de distribution

Nous allons étudier l'équation

$$(x-x_0)^\alpha u = 0 \tag{29.185}$$

pour $u \in \mathscr{D}'(\mathbb{R})$ et $\alpha \in \mathbb{N}$ est donné fixé. Notons tout de suite que (29.185) est un petit abus de notation pour dire qu'en vertu de la définition 29.21 du produit d'une distribution par une fonction, pour tout $\phi \in \mathscr{D}(\mathbb{R})$, nous avons $u\Big(x \mapsto (x-x_0)\phi(x)\Big) = 0$.

Lemme 29.63 ([202]).
Soit $\alpha \in \mathbb{N}$. Une solution à l'équation

$$(x-x_0)^\alpha u = 0 \tag{29.186}$$

est une distribution à support dans $\{x_0\}$ et d'ordre fini.

Démonstration. Nous commençons par prouver que u est une solution de (29.186) si et seulement si [14] $\langle u, \phi\rangle = 0$ pour tout $\phi \in \mathscr{D}$ telle que

$$\phi(x_0) = \ldots = \partial^{\alpha-1}\phi(x_0) = 0. \tag{29.187}$$

Condition nécéssaire Supposons que u soit une solution. Alors le corolaire 16.29 du théorème de Hadamard donne $\psi \in \mathscr{D}(\mathbb{R})$ telle que $\phi(x) = (x-x_0)^\alpha\psi(x)$. Dans ce cas, si u est solution de (29.186), alors

$$0 = \langle (x-x_0)^\alpha u, \psi\rangle = \langle u, (x-x_0)^\alpha\psi(x)\rangle = \langle u, \phi\rangle. \tag{29.188}$$

Nous avons vu que si u est solution, alors $\langle u, \phi\rangle = 0$ pour tout ϕ satisfaisant la condition (29.187).

Condition suffisante Supposons maintenant l'inverse : u est une distribution s'annulant sur toute fonction $\phi \in \mathscr{D}'$ satisfaisant (29.187). Nous allons alors prouver que u est une solution. Soit donc $\psi \in \mathscr{D}$ et calculons

$$\langle (x-x_0)u, \psi\rangle = \langle u, (x-x_0)\psi\rangle = 0 \tag{29.189}$$

parce que la fonction $(x-x_0)\psi(x)$ vérifie la condition (29.187).

Nous passons maintenant au cœur de la preuve : nous supposons que u est une solution. Si le support de ϕ est contenu dans $\mathbb{R}\backslash\{x_0\}$ alors ϕ est nulle dans un voisinage de x_0 (et donc $\partial^k\phi = 0$ pour tout k) et $\langle u, \phi\rangle = 0$. Autrement dit, pour tout $\phi \in \mathscr{D}\big(\mathbb{R}\backslash\{x_0\}\big)$ nous avons $\langle u, \phi\rangle = 0$, ce qui signifie que $\operatorname{supp}(u) \cap \big(\mathbb{R}\backslash\{x_0\}\big) = \varnothing$ ou encore que $\operatorname{supp}(u) = \{x_0\}$.

Maintenant que u a un support compact, la proposition 29.28 nous indique qu'elle est d'ordre fini. □

Théorème 29.64 ([202]).
Soit $\alpha \in \mathbb{N}$ et l'équation

$$(x-x_0)^\alpha u = 0 \tag{29.190}$$

pour $u \in \mathscr{D}'(\mathbb{R})$. Les solutions sont les combinaisons linéaires des dérivées de δ_{x_0} jusqu'à la α^eexclue.

Démonstration. D'abord montrons que les $\partial^i\delta_{x_0}$ sont des solutions. Avec les définition 29.21 et 29.22 des dérivées de distributions et de leur produits avec des fonctions [15],

$$(x-x_0)^\alpha\partial^i\delta_{x_0}(\phi) = \delta_{x_0}\Big(\partial^i\big((x-x_0)^\alpha\phi(x)\big)\Big) \tag{29.191}$$

14. En réalité nous n'aurons besoin que de la condition nécessaire, en particulier pour le théorème 29.64.

15. Comme souvent, dans l'expression suivante, il y a un abus de notation parce que x est une variable muette : il faudrait écrire « $x \mapsto$ » au début de la grade parenthèse.

Si $i < \alpha$ alors dans chaque terme de Leibnitz, il y aura un facteur $(x - x_0)$, et la prise de δ_{x_0} annulera. Si par contre $i \geqslant \alpha$ alors il y aura le terme

$$\binom{i}{\alpha}\partial^\alpha\big((x-x_0)^\alpha\big)\partial^{i-\alpha} = \binom{i}{\alpha}\alpha!(\partial^{i-\alpha}\phi)(x_0) \tag{29.192}$$

qui est le seul terme contenant $(\partial^{i-\alpha}\phi)(x_0)$. Il suffit alors de choisir $\phi \in \mathscr{D}(\mathbb{R})$ de sorte que

$$(\partial^k\phi)(x_0) = \begin{cases} 0 & \text{si } k \neq i-\alpha \\ 1 & \text{si } k = i-\alpha \end{cases} \tag{29.193}$$

et alors on est certain que le tout n'est pas nul, et donc que $(x - x_0)^\alpha(\partial^i\delta_{x_0}) \neq 0$.

Jusqu'ici nous avons prouvé que $\partial^i\delta_{x_0}$ est solution si et seulement si $0 \leqslant i < \alpha$.

Il faut encore prouver que les solutions sont toutes des combinaisons linéaires de dérivées de delta de Dirac centrées en x_0. Pour cela nous invoquons d'abord le lemme 29.63 qui nous assure que u est d'ordre fini et de support $\{x_0\}$. Ensuite la proposition 29.31 nous indique que u doit alors être une combinaisons linéaire de dérivées de Dirac. □

Chapitre 30

Espaces de Sobolev, équations elliptiques

30.1 Espaces de Sobolev

Rappel : la définition de la dérivée faible est 29.2.

30.1.1 Sur un intervalle de $\mathbb{R}$

Sauf mention du contraire dans cette section I est un intervalle borné ouvert $I =]a, b[$ de $\mathbb{R}$.

Définition 30.1.
*Soit $I =]a, b[$ un ouvert borné de $\mathbb{R}$. L'**espace de Sobolev** $H^1(I)$ est l'ensemble*

$$H^1(I) = \Big\{ u \in L^2(I) \text{ tel que } \exists g \in L^2(I) \text{ tel que } \forall \varphi \in C_c^\infty(I), \int_I u\varphi' = -\int_I g\varphi \Big\}. \tag{30.1}$$

L'unique élément g de $L^2(I)$ vérifiant $\int_I u\varphi' = -\int_I g\varphi$ est noté u' est est nommé **dérivée** ; nous verrons dans les prochaines pages pourquoi.

L'espace H^1 accepte le produit scalaire suivant :

$$\langle u, v \rangle = \int_I uv + \int_I u'v', \tag{30.2}$$

et nous notons $\|.\|_{H^1}$ la norme correspondante qui n'est autre que

$$\|u\|_{H^1} = \langle u, u \rangle = \|u\|_{L^2}^2 + \|u'\|_{L^2}. \tag{30.3}$$

Nous introduisons l'espace $L^1_{loc}(I)$ des fonctions étant L^1 sur tout compact de I.

Corolaire 30.2.
Si $u \in H^1(I)$ et si $u' = 0$ alors il existe une constante C telle que $u = C$ presque partout.

Démonstration. L'hypothèse $u' = 0$ signifie que pour toute fonction $\varphi \in C_c^\infty(I)$,

$$\int_I u\varphi' = \int_I u'\varphi = 0. \tag{30.4}$$

La proposition 26.133 nous dit alors qu'il existe une constante C telle que $u = C$ presque partout. □

Lemme 30.3.
Tout élément de $H^1(I)$ admet un unique représentant continu.

Nous verrons dans le corolaire 30.5 que ce représentant pourra être prolongé par continuité sur $\bar{I}$.

Démonstration. Soient $y_0 \in I$ et $u \in H^1(I)$. Nous considérons la fonction

$$\bar{u}(x) = \int_{y_0}^{x} u'(t)dt. \tag{30.5}$$

Notons que par définition, $u' \in L^2$ donc l'intégrale ne pose pas de problèmes. Montrons que $\bar{u}$ est continue sur $\bar{I}$. Pour cela nous considérons $x \in \bar{I}$ et h tel que $x + h \in \bar{I}$. Alors

$$\big|\bar{u}(x+h) - \bar{u}(x)\big| = \Big|\int_{x}^{x+h} u'\Big| \leqslant \int_{x}^{x+h} |u'|. \tag{30.6}$$

Mais la fonction $|u'|$ est dans $L^1_{loc}(I)$ par le lemme 26.35 ; elle est en particulier intégrable sur un ouvert contenant x et par conséquent la dernière intégrale tend vers zéro lorsque h tend vers 0.

Nous prouvons à présent que $\bar{u}$ est dans $H^1(I)$ et que sa dérivée est égale à u' ; pour cela nous allons montrer que pour tout $\varphi \in C_c^\infty(I)$,

$$\int_I \bar{u}\varphi' = -\int_I u'\varphi. \tag{30.7}$$

Nous avons

$$\int_I \bar{u}\varphi' = \int_I \left(\int_{y_0}^{x} u'(t)dt\right)\varphi'(x)dx = \int_a^{y_0}\left(\int_{y_0}^{x} u'(t)dt\right)\varphi'(x)dx + \int_{y_0}^{b}\left(\int_{y_0}^{x} u'(t)dt\right)\varphi'(x)dx. \tag{30.8}$$

Pour faire plus court, nous notons $f(t,x) = u'(t)\varphi'(x)$. La première intégrale vaut

$$\begin{aligned}
\int_a^{y_0}\left(\int_{y_0}^{x} u'(t)\varphi'(x)\right) &= \int_a^{y_0}\left(\int_{y_0}^{a} f(t,x)\mathbb{1}_{t<x}(t,x)dt\right)dx && (30.9a)\\
&= \int_{y_0}^{a}\int_a^{y_0} f(t,x)\mathbb{1}_{t>x}dxdt && (30.9b)\\
&= \int_{y_0}^{a}\int_a^{t} f(t,x)dxdt && (30.9c)\\
&= -\int_a^{y_0}\int_a^{t} u'(t)\varphi'(x)dx\,dt && (30.9d)
\end{aligned}$$

La permutation d'intégrales pour obtenir (30.9b) est due au théorème de Fubini 13.262(3). Par le même petit jeu, la seconde intégrale vaut

$$\int_{y_0}^{b}\int_t^{b} u'(t)\varphi'(x)dx\,dt. \tag{30.10}$$

En refaisant la somme,

$$\begin{aligned}
\int_I \bar{u}\varphi' &= -\int_a^{y_0} u'(t)\left(\int_a^{t}\varphi'(x)dx\right)dt + \int_{y_0}^{b} u'(t)\left(\int_t^{b}\varphi'(x)dx\right)dt && (30.11a)\\
&= -\int_a^{y_0} u'(t)\big(\varphi(t)-\varphi(a)\big)dt + \int_{y_0}^{b} u'(t)\big(\varphi(b)-\varphi(t)\big) && (30.11b)\\
&= -\int_a^{b} u'\varphi && (30.11c)\\
&= -\int_I u'\varphi. && (30.11d)
\end{aligned}$$

Notons que $\varphi(a) = \varphi(b) = 0$ parce que φ est à support compact dans $]a,b[$. Nous avons donc prouvé que $\bar{u}$ est dans $H^1(I)$ et que $\bar{u}' = u'$. Par le corolaire 30.2, nous avons une constante C telle que $\bar{u} = u + C$ presque partout, c'est-à-dire $u = \bar{u} + C$ dans $H^1(I)$.

En résumé, $\tilde{u}\tilde{u} = \bar{u} + C$ est un représentant continu de u dans $L^2(I)$.

L'unicité du représentant continu est simplement le fait que deux fonctions continues égales presque partout sont égales (proposition 19.149).

□

Proposition 30.4.
Si $u \in H^1(I)$, alors

$$u(x) - u(y) = \int_y^x u' \tag{30.12}$$

pour tout $x, y \in I$.

Démonstration. Pour fixer les idées, nous supposons $x < y$. Nous considérons une suite $\varphi_n \in C_c^\infty(I)$ convergeant uniformément sur I vers $\mathbb{1}_{[x,y]}$. Nous exigeons de plus que

— φ_n' est positive sur $[a, x + \frac{1}{n}]$
— φ_n' est négative sur $[y - \frac{1}{n}, b]$
— $\varphi_n = 1$ sur $[x + \frac{1}{n}, y - \frac{1}{n}]$.
— $\varphi_n = 0$ sur $[a, x - 1/n]$ et sur $[y + 1/n, b]$.

Pour chaque n, nous découpons l'intégrale comme

$$-\int_I u'\varphi_n = \int_I u\varphi_n' = \int_a^{a-1/n} u\varphi_n' + \int_{x-1/n}^{x+1/n} u\varphi_n' + \int_{x+1/n}^{y-1/n} u\varphi_n' + \int_{y-1/n}^{y+1/n} u\varphi_n' + \int_{y+1/n}^{b} u\varphi_n'. \tag{30.13}$$

Par construction de φ_n, de ces 5 morceaux, il n'en reste que deux de non nulles :

$$\int_I u\varphi' = \underbrace{\int_{x-1/n}^{x+1/n} u(t)\varphi_n'(t)dt}_{A} + \underbrace{\int_{y-1/n}^{y+1/n} u(t)\varphi_n'(t)dt}_{B} \tag{30.14}$$

Soit $\epsilon > 0$ et n suffisamment grand pour avoir $u(t) \in B\big(u(x), \epsilon\big)$ pour tout $t \in B(x, \frac{1}{n})$ et (en même temps) $u(t) \in B\big(u(y), \epsilon\big)$ pour tout $t \in B(y, \frac{1}{n})$. C'est la continuité de u qui permet de trouver un tel n. Pour cette valeur de n, en tenant compte des hypothèses sur la positivité de φ_n' nous avons

$$\int_{x-1/n}^{x+1/n} \big(u(x) - \epsilon\big)\varphi_n'(t)dt \leqslant \int_{x-1/n}^{x+1/n} u(t)\varphi_n'(t)dt \leqslant \int_{x-1/n}^{x+1/n} \big(u(x) + \epsilon\big)\varphi_n'(t)dt, \tag{30.15}$$

mais par hypothèse sur φ_n nous trouvons

$$\int_{x-1/n}^{x+1/n} \varphi_n'(t)dt = \varphi_n(x + \frac{1}{n}) - \varphi(x + \frac{1}{n}) = 1. \tag{30.16}$$

donc

$$u(x) - \epsilon \leqslant \int_{x-1/n}^{x+1/n} u(t)\varphi_n'(t)dt \leqslant u(x) + \epsilon. \tag{30.17}$$

Pour encadrer la seconde, il faut être plus prudent avec les signes parce que φ_n' y est négative. En posant $\psi_n = -\varphi_n$ nous avons

$$-B = \int_{y-1/n}^{y+1/n} u(t)\psi_n(t)dt, \tag{30.18}$$

et donc

$$u(y) - \epsilon \leqslant -B \leqslant u(y) + \epsilon \tag{30.19}$$

ou encore

$$-\epsilon - u(y) \leqslant B \leqslant \epsilon - u(y). \tag{30.20}$$

En additionnant avec (30.17) nous voyons que pour tout $\epsilon > 0$ il existe un $N(\epsilon)$ tel que nous ayons

$$u(x) - u(y) - 2\epsilon \leqslant \int_I u'\varphi_n \leqslant u(x) - u(y) + 2\epsilon \tag{30.21}$$

pour tout $n \geqslant N$. Nous voulons évidemment prendre la limite $\epsilon \to 0$, c'est-à-dire $n \to \infty$. Étant donné que $\varphi_n(t) < 1$ pour tout t et pour tout n, la fonction $t \mapsto u'(t)\varphi_n(t)$ est dominée par u', qui est dans $L^1(I)$ par le lemme 26.35. Le théorème de la convergence dominée nous permet donc d'affirmer que

$$\lim_{n\to\infty} \int_I u'\varphi_n = \int_I u' \mathbb{1}_{[x,y]} = \int_x^y u', \tag{30.22}$$

et donc les inégalités (30.21) donnent le résultat, grâce au signe dans (30.13). □

Corolaire 30.5.
Si $[u] \in H^1(I)$, le représentant continu $u \in C^0(I)$ peut être prolongé par continuité en $u \in C^0(\bar{I})$.

Démonstration. Soit (x_n) une suite strictement croissante dans $]a, b[$ convergeant vers b. Nous voulons montrer que la suite $\big(u(x_n)\big)$ est de Cauchy dans $\mathbb{R}$, ce qui nous permettra de définir

$$u(b) = \lim_{n\to\infty} u(x_n). \tag{30.23}$$

qui sera évidemment continue. Cette construction ne dépendra pas du choix de la suite (x_n) parce que deux fonctions continues sur $\bar{I}$ et égales sur I sont égales sur $\bar{I}$.

En notant u' la dérivée de u dans H^1, nous avons par construction du représentant continu : $u(x) = \int_{y_0}^x u'(t)dt$. Et donc

$$\big|u(x_n) - u(x_{n+p})\big| = \left|\int_{y_0}^{x_n} u' - \int_{y_0}^{x_{n+p}} u'\right| = \left|\int_{x_n}^{x_{n+p}} u'\right|. \tag{30.24}$$

Vu que la suite (x_n) est de Cauchy et que u' est intégrable (même sur $\bar{I}$), la limite $n \to \infty$ de cela est zéro, quelle que soit la valeur de p. Donc $\big(u(x_n)\big)$ est ce Cauchy dans $\mathbb{R}$ et est donc convergente. □

Proposition 30.6 ([54, 1])**.**
Quelques propriétés de l'espace de Sobolev $H^1(I)$ où $I =]a, b[$ est un ouvert borné de $\mathbb{R}$.

(1) $H^1(I)$ est un espace de Hilbert.

(2) $H^1(I)$ s'injecte de façon compacte dans $C^0(\bar{I})$.

(3) $H^1(I)$ s'injecte de façon continue dans $L^2(I)$.

Démonstration. Nous prouvons point par point.

(1) Le seul critère à vérifier est la complétude. Pour cela nous considérons une suite de Cauchy (u_n) dans $H^1(I)$. Si $\epsilon > 0$, alors il existe $N > 0$ tel que pour tout $p \geqslant 0$ nous ayons $\|u_{n+p} - u_n\|_{H^1}^2 \leqslant \epsilon$, c'est-à-dire

$$\|u_{n+p} - u_n\|_{L^2}^2 + \|u'_{n+p} - u'_n\|_{L^2}^2 + \tag{30.25}$$

En particulier les suites (u_n) et (u'_n) sont de Cauchy dans L^2 qui est complet par le théorème de Fischer-Riesz 26.41. Nous notons donc

$$u_n \xrightarrow{L^2} u \tag{30.26a}$$

$$u'_n \xrightarrow{L^2} v. \tag{30.26b}$$

Nous allons maintenant montrer quelques limites.

$\underline{u_n\varphi \xrightarrow{L^2} u\varphi}$ Si M est une constante qui majore φ alors $\|u_n\varphi - u\varphi\|_2 \leqslant M\|u_n - u\|_2 \to 0$.

$\underline{u'_n\varphi \xrightarrow{L^2} v\varphi}$ C'est la même chose avec $\|u'_n\varphi - v\varphi\|_2 \leqslant M\|u'_n - v\|_2 \to 0$.

$\underline{u \in H^1(I) \textbf{ avec } u' = v}$ Attendu le corolaire 26.36 qui permet de permuter intégrale et limite dans $L^2(I)$ et les limites que nous venons de prouver,

$$\int_I u\varphi' = \lim_{n\to\infty} \int_I u_n\varphi' = -\lim_{n\to\infty} \int_I u_n'\varphi = -\int_I v\varphi. \tag{30.27}$$

Cela signifie que v est la dérivée faible de u : $u' = v$.

$\underline{u_n \overset{H^1}{\to} u}$ Nous pouvons alors prouver que $u_n \to u$ dans $H^1(I)$:

$$\|u_n - u\|_{H^1(I)}^2 = \|u_n - u\|_{L^2}^2 + \|u_n' - u'\|_{L^2}^2. \tag{30.28}$$

Mais nous savons déjà que $u_n \to u$ dans L^2 (d'ailleurs c'est la définition de u) et que $u' = v$ alors que par définition de v, nous avons $u_n' \to v$ dans L^2.

Tout cela donne que $u_n \to u$ dans $H^1(I)$ et donc que $H^1(I)$ est un espace complet.

(2) L'application que nous allons prouver être compacte entre $H^1(I)$ et $C^0(\bar{I})$ est

$$\begin{aligned} \psi\colon H^1(I) &\to C^0(\bar{I}) \\ [u] &\mapsto \tilde{u} \end{aligned} \tag{30.29}$$

où $[u]$ désigne une classe de fonction dans $H^1(I)$ et $\tilde{u}$ est son représentant continu prolongé par continuité à $\bar{I}$ [1], qui existe par le lemme 30.3 et le corolaire 30.5. Cette application est une injection par l'unicité du représentant continu. Nous allons prouver que c'est une application compacte en utilisant le critère (2) de la proposition 26.3. Pour cela nous allons commencer par utiliser le théorème d'Ascoli sur l'ensemble $\tilde{\mathcal{B}}$ des représentants continus des éléments de $\mathcal{B}$, prolongés par continuité sur $\bar{I}$; c'est-à-dire $\tilde{B} \subset C^0(\bar{I})$.

Soit $u \in \tilde{\mathcal{B}}$; par la proposition 30.4, nous avons

$$\begin{aligned} \big|u(x) - u(y)\big| &= \Big|\int_y^x u'(t)dt\Big| && (30.30a)\\ &= \Big|\int_I \mathbb{1}_{[x,y]}(t)u'(t)dt\Big| && (30.30b)\\ &\leqslant \|\mathbb{1}_{[x,y]}\|_{L^2}\|u'\|_{L^2} && (30.30c)\\ &\leqslant \sqrt{|x-y|}\|u'\|_{H^1} && (30.30d)\\ &\leqslant \sqrt{|x-y|}. && (30.30e) \end{aligned}$$

où nous insistons sur le fait que la continuité n'impliquant pas la dérivabilité, le u' ici est la dérivé au sens de H^1, et non la dérivée usuelle. Quoi qu'il en soit, l'ensemble $\tilde{\mathcal{B}}$ est équicontinu [2]. Nous montrons à présent qu'il est également borné pour la norme uniforme. Soit $u \in \tilde{\mathcal{B}}$; vu la construction du représentant continu au lemme 30.3, nous avons

$$\begin{aligned} \big|u(x)\big| &= \Big|\frac{1}{b-a}\int_a^b u(x)dy\Big| && (30.31a)\\ &= \Big|\frac{1}{b-a}\int_a^b \Big(\int_y^x u'(t)dt - u(y)\Big)dy\Big| && (30.31b)\\ &= \Big|\frac{1}{b-a}\int_a\int_y^x u'(t)dtdy - \frac{1}{b-a}\int_a^b u(y)dy\Big| && (30.31c)\\ &\leqslant \frac{1}{b-a}\int_a^b\int_a^b |u'(t)|dt\,dy + \frac{1}{b-a}\int_a^b |u(y)|dy. && (30.31d) \end{aligned}$$

1. Encore que par souci d'économie d'encre nous n'allons pas écrire toujours les tildes et noter u le représentant continu prolongé à $\bar{I}$ par le corolaire 30.5.
2. Définition 7.321.

À ce niveau, il faut remarquer que dans la première intégrale, le passage de la valeur absolue à l'intérieur de l'intégrale en même temps que l'élargissement des bornes n'a rien d'innocent. Si $x < y$, les bornes ne sont pas « dans le bon ordre » et nous ne pouvons pas faire la majoration usuelle en entrant simplement la valeur absolue. Ici nous tenons compte de cela en élargissant les bornes, et en les mettant dans le bon ordre. Le passage exact est le suivant : si $x, y \in]a, b[$, nous avons

$$\left|\int_y^x f(t)dt\right| \leqslant \left|\int_y^x |f(t)|dt\right| \leqslant \left|\int_a^b |f(t)|dt\right| = \int_a^b |f(t)|dt. \tag{30.32}$$

Notons en particulier que dans le cas du passage vers l'équation (30.31d), le nombre x est fixé alors que y est une variable d'intégration. Donc l'ordre des deux est certainement de temps en temps le « mauvais ».

Quoi qu'il en soit, la première intégrale se réduit à une multiplication par $b - a$ et le calcul continue :

$$\begin{aligned}
|u(x)| &\leqslant \int_I |u'(t)|dt + \frac{1}{b-a}\int_I |u| && (30.33a)\\
&\leqslant \sqrt{b-a}\|u'\|_{L^2} + \frac{1}{\sqrt{b-a}}\|u\|_{L^2} && (30.33b)\\
&\leqslant \left(\sqrt{b-a} + \frac{1}{\sqrt{b-a}}\right)\left(\|u'\|_{L^2} + \|u\|_{L^2}\right) && (30.33c)\\
&\leqslant \left(\sqrt{b-a} + \frac{1}{\sqrt{b-a}}\right)\|u\|_{H^1} && (30.33d)\\
&= \sqrt{b-a} + \frac{1}{\sqrt{b-a}}. && (30.33e)
\end{aligned}$$

Donc $\tilde{\mathcal{B}}$ est borné pour la norme L^∞. Et c'est même borné par un nombre facilement calculable connaissant I. En particulier l'ensemble

$$\{u(x) \text{ tel que } u \in H^1\} \tag{30.34}$$

est pour, tout x, contenu dans la boule de rayon $\sqrt{a-b} + \frac{1}{\sqrt{a-b}}$ et donc est relativement compact dans $\mathbb{R}$. Par conséquent le théorème d'Ascoli 26.5 nous dit que l'ensemble $\tilde{B}$ est relativement compact dans $C^0(I)$.

Par conséquent nous avons montré que l'image par ψ de la boule unité fermée $\mathcal{B}$ de $H^1(I)$ est relativement compacte dans $C^0(\bar{I})$, ce qui signifie que ψ est une application compacte.

(3) Les éléments de $H^1(I)$ sont des éléments de $L^2(I)$; donc l'identité est une injection. Nous devons seulement étudier la continuité. Si (u_n) est une suite dans H^1 convergeant dans H^1 vers u, alors

$$\|u_n - u\|_{L^2} \leqslant \|u_n - u\|_{L^2} + \|u_n' - u'\|_{L^2} = \|u_n - u\|_{H^1} \to 0. \tag{30.35}$$

Donc la suite des images (par l'identité) converge dans L^2. L'identité est donc continue.

□

30.1.2 Sur un ouvert de $\mathbb{R}^n$

Soit Ω, un ouvert de $\mathbb{R}^n$ et $v \in L^2(\Omega)$ (voir 26.72). Les fonctions considérées sont à valeurs réelles.

30.1.2.1 Définition

Définition 30.7 (Espace de Sobolev $H^1(\Omega)$)**.**
*Soit Ω une partie de $\mathbb{R}^n$. L'espace de **Sobolev** $H^1(\Omega)$ est :*

$$H^1(\Omega) = \{v \in L^2(\Omega) \text{ tel que } \forall i = 1, \dots, n, \partial_i v \in L^2(\Omega)\}. \tag{30.36}$$

Nous munissons cet espace d'un produit scalaire

$$(u,v)_{H^1} = \langle u,v\rangle_{L^2} + \langle \nabla u, \nabla v\rangle_{L^2}, \tag{30.37}$$

où $\nabla u = \sum_i \partial_i u \in L^2$.

L'existence des intégrales dans le produit scalaire est assurée par le fait que u, v, ∇u et ∇v sont dans $L^2(\Omega)$. La définition du produit scalaire dans L^2 est la définition 26.200 (mais sans la conjugaison complexe).

Pour la même raison, $(u,u)_{H^1} = 0$ demande que chacun des deux termes est séparément nul, et nous avons $u = 0$ dans L^2, et donc aussi dans H^1.

Théorème 30.8 ([481]).
L'espace $H^1(\Omega)$ est un espace de Hilbert[3].

Démonstration. Nous devons nous assurer que l'espace H^1 est complet. Pour cela nous considérons une suite de Cauchy (u_n) dans H^1. Soit $\epsilon > 0$; il existe $N > 0$ tel que si $n, m > N$ alors $\|u_n - u_m\|_{H^1} < \epsilon$. Dans ce cas nous avons en particulier

$$\|u_m - u_n\|_{H^1}^2 = (u,u)_{H^1} = \langle u,u\rangle + \langle \nabla u, \nabla u\rangle = \|u\|_{L^2}^2 + \|\nabla u\|_{L^2}^2, \tag{30.38}$$

et en particulier les suites (u_n) et (∇u_n) sont de Cauchy dans L^2. Vu que L^2, lui, est complet (théorème 26.40), il existe $u \in L^2$ et $v_i \in L^2$ tels que

$$u_n \xrightarrow{L^2} u \tag{30.39a}$$

$$\partial_i u_n \xrightarrow{L^2} v_i. \tag{30.39b}$$

Nous savons que l'injection $i\colon L^2 \to \mathscr{D}'$ est continue par la proposition 29.20. Nous avons donc aussi les limites

$$T_{u_n} \xrightarrow{\mathscr{D}'} T_u \tag{30.40a}$$

$$T_{\partial_i u_n} \xrightarrow{\mathscr{D}'} T_{v_i}. \tag{30.40b}$$

La dérivée étant une opération continue sur $\mathscr{D}'$ nous avons de plus

$$\partial_i(T_{u_n}) \xrightarrow{\mathscr{D}'} \partial_i(T_u) \tag{30.41}$$

En utilisant le lemme 29.23 nous avons alors

$$T_{\partial_i u_n} = \partial_i(T_{u_n}) \xrightarrow{\mathscr{D}} \partial_i(T_u) = T_{\partial_i u}. \tag{30.42}$$

En comparant avec (30.40b) et par l'unicité de la limite, nous avons $T_{v_i} = T_{\partial_i u}$. Cela implique $v_i = \partial_i u$.

Vu que $v_i \in L^2$ nous avons aussi $\partial_i u \in L^2$. Par conséquent $u \in H^1(\Omega)$ parce que ses dérivées sont dans L^2.

Nous devons maintenant prouver que $u_n \xrightarrow{H^1} u$. Nous avons

$$\|u_n - u\|_{H^1} = \|u_n - u\|_{L^2} + \|\nabla u_n - \nabla u\|_{L^2} \tag{30.43}$$

Le premier terme tend vers zéro parce que $u_n \xrightarrow{L^2} u$ et le second parce que $\partial_i u_n \xrightarrow{L^2} \partial_i u$. □

3. Définition 24.1.

30.1.3 Espace de Sobolev fractionnaire

Définition 30.9.
*Pour $m \in \mathbb{N}$ et un ouvert Ω de $\mathbb{R}^d$ nous définissons l'**espace de Sobolev***

$$H^m(\Omega) = \{u \in L^2(\Omega) \text{ tel que } \partial^\alpha u \in L^2(\Omega)\, \forall |\alpha| \leqslant m\}. \tag{30.44}$$

Nous définissons également un produit scalaire sur H^m par

$$(u, v)_{H^m} = \sum_{|\alpha| \leqslant m} \langle \partial^\alpha u, \partial^\alpha v \rangle_{L^2}. \tag{30.45}$$

En particulier la topologie est celle de la norme dérivée du produit scalaire :

$$\|u\|_{H^m(\Omega)} = \sum_{|\alpha| \leqslant m} \|\partial^\alpha u\|_{L^2(\Omega)}. \tag{30.46}$$

Le lemme suivant montre que la proposition 28.19 fonctionne encore avec L^2 au lieu de $\mathscr{S}$.

Lemme 30.10 (Lemme de transfert[482], thème 69)**.**
Soit $f \in H^m(\mathbb{R}^d)$. Alors pour tout multiindice α avec $|\alpha| \leqslant m$ nous avons

$$\mathcal{F}(\partial^\alpha f) = \big[\xi \mapsto i^{|\alpha|} \xi^\alpha \hat{f}(\xi)\big]. \tag{30.47}$$

Lemme 30.11.
Il existe des constantes c_1 et c_2 telles que pour tout $x \in \mathbb{R}^d$,

$$c_1(1 + \|x\|^2)^m \leqslant \sum_{|\alpha| \leqslant m} (x^\alpha)^2 \leqslant c_2(1 + \|x\|^2)^m. \tag{30.48}$$

Lemme 30.12.
Soit $u \in L^2(\mathbb{R}^d)$. Nous avons $u \in H^m(\mathbb{R}^d)$ si et seulement si l'application

$$\xi \mapsto \big(1 + |\xi|^2\big)^{k/2} \hat{u} \tag{30.49}$$

est dans $L^2(\mathbb{R}^d)$ pour tout $k \leqslant m$. Ici $|\xi|$ est la norme euclidienne de ξ dans $\mathbb{R}^d$.

Démonstration. Vu le lemme 30.10, il suffit de montrer que

$$\big(1 + |\xi|^2\big)^{k/2} \hat{u} \tag{30.50}$$

est dans L^2 pour tout $k \leqslant m$ si et seulement si

$$\xi^\alpha \hat{u} \tag{30.51}$$

l'est pour tout α avec $|\alpha| \leqslant m$.

L'expression (30.50) est une somme d'expressions du type (30.51). Donc l'implication dans un sens est montrée. Pour l'autre sens, nous savons que

$$\xi^\alpha = \xi_1^{\alpha_1} \dots \xi_n^{\alpha_n}, \tag{30.52}$$

et donc

$$|\xi^\alpha| \leqslant |\xi_1|^{\alpha_1} \dots |\xi_n|^{\alpha_n}. \tag{30.53}$$

Or $|\xi|^{|\alpha|} = |\xi|^{\sum_i \alpha_i} = |\xi|^{\alpha_1} \dots |\xi|^{\alpha_n}$ et $|\xi| \geqslant |\xi_i|$ pour tout i, donc

$$|\xi^\alpha| \leqslant |\xi|^{|\alpha|}. \tag{30.54}$$

D'autre part pour tout $x \in \mathbb{R}^+$ et tout k positif nous avons

$$(1 + x^2)^{k/2} \geqslant x^k \tag{30.55}$$

qui est facile à vérifier en prenant le carré des deux membres.

En remettant tout ensemble,

$$|\xi^\alpha \hat{u}| \leqslant |\xi^\alpha||\hat{u}| \leqslant |\xi|^{|\alpha|}|\hat{u}| \leqslant \big(1 + |\xi|^2\big)^{|\alpha|/2}|\hat{u}|. \tag{30.56}$$

Donc si le membre de droite est de carré intégrable, celui de gauche l'est également. □

Définition 30.13 (Espace de Sobolev H^s[483])**.**
Pour $s > 0$ nous définissons l'espace de Sobolev $H^s(\mathbb{R}^d)$ par

$$H^s(\mathbb{R}^d) = \{u \in L^2(\mathbb{R}^d) \text{ tel que } \big(1 + \|\xi\|^2\big)^{s/2}\hat{u} \in L^2(\mathbb{R}^d)\}. \tag{30.57}$$

Nous y mettons le produit scalaire

$$(u, v)_{H^s} = \int_{\mathbb{R}^d} \hat{u}(\xi)\overline{\hat{v}(\xi)}(1 + \|\xi\|^2)^s d\xi. \tag{30.58}$$

30.14.
Vu que $\mathscr{D}(\mathbb{R}^d)$ est dense dans $L^2(\mathbb{R}^d)$ (théorème 26.47), on pourrait croire à la densité a fortiori dans $H^s(\mathbb{R}^d)$. Mais attention : $\mathscr{D}(\mathbb{R}^d)$ est dense dans L^2 pour la norme L^2. Nous n'avons encore rien dit pour la norme $H^s(\mathbb{R}^d)$.

Proposition 30.15 ([484])**.**
La partie $\mathscr{S}(\mathbb{R}^d)$ est dense dans $H^s(\mathbb{R}^d)$.

Démonstration. Soit $u \in H^s(\mathbb{R}^d)$. Par définition l'application

$$\xi \mapsto (1 + \|\xi\|^2)^{s/2}\hat{u} \tag{30.59}$$

est dans $L^2(\mathbb{R}^d)$. Elle peut donc être approximée au sens L^2 par des fonctions dans $\mathscr{D}(\mathbb{R}^d)$ (théorème 26.47(5)), c'est-à-dire qu'il existe des fonctions $\phi_n \in \mathscr{D}(\mathbb{R}^d)$ telles que

$$\phi_n \xrightarrow{L^2(\mathbb{R}^d)} (1 + \|\xi\|^2)^{s/2}\hat{u}. \tag{30.60}$$

Nous posons

$$\psi_n = \frac{\phi_n}{(1 + \xi^2)^{s/2}} \tag{30.61}$$

Cela est encore une fonction de $\mathscr{D}(\mathbb{R}^d)$, et donc de $\mathscr{S}(\mathbb{R}^d)$. Vu que la transformée de Fourier est une bijection de $\mathscr{D}(\mathbb{R}^d)$ (proposition 28.26), nous pouvons considérer une suite $\varphi_n \in \mathscr{D}(\mathbb{R}^d)$ telle que $\hat{\varphi}_n = \psi_n$, et nous allons montrer que $\varphi_n \xrightarrow{H^s(\mathbb{R}^d)} u$.

Nous avons :

$$\begin{aligned}
\|\varphi_n - u\|_{H^s}^2 &= \int_{\mathbb{R}^d} |\hat{\varphi}_n - \hat{u}|^2(1 + \xi^2)^s d\xi && (30.62a)\\
&= \int_{\mathbb{R}^d} \Big|\frac{\phi_n(\xi)}{(1 + \xi^2)^{s/2}} - \hat{u}(\xi)\Big|^2(1 + \xi^2)^s d\xi && (30.62b)\\
&= \int_{\mathbb{R}^d} |\phi_n(\xi) - \hat{u}(\xi)(1 + \xi^2)^{s/2}|^2 d\xi && (30.62c)\\
&= \|\phi_n - (1 + \xi^2)^{s/2}\hat{u}\|_{L^2}^2. && (30.62d)
\end{aligned}$$

Par définition de la suite ϕ_n nous avons donc bien

$$\|\varphi_n - u\|_{H^s}^2 = \|\phi_n - (1 + \xi^2)^{s/2}\hat{u}\|_{L^2}^2 \to 0. \tag{30.63}$$

Notons que même si ϕ_n est dans $\mathscr{D}(\mathbb{R}^d)$, nous n'avons pas prouvé la convergence $\phi_n \xrightarrow{H^s} u$, mais bien $\varphi_n \xrightarrow{H^s} u$. Or les fonctions φ_n sont dans $\mathscr{S}(\mathbb{R}^d)$, et rien n'assure qu'elles soient à support compact. Nous avons donc bien prouvé la densité de $\mathscr{S}$ et non celle de $\mathscr{D}$. □

Remarque 30.16.
Pour qui a tout compris, cela peut sembler une évidence, mais nous précisions que nous parlons de densité de $\mathscr{S}(\mathbb{R}^d)$ dans $H^s(\mathbb{R}^d)$, à aucun moment la topologie de $\mathscr{S}(\mathbb{R}^d)$ n'entre en compte.

Un peu moins évident : ce que nous avons réellement montré est la densité de $\iota\big(\mathscr{D}(\mathbb{R}^d)\big)$ dans $H^s(\mathbb{R}^d)$ où ι est l'application « prise de classe ». Nous n'avons pas insisté là-dessus, mais il faut

dire que dans la preuve de la proposition 30.15, u est un représentant d'un élément choisi dans $H^s(\mathbb{R}^d)$.

Nous avons ensuite prouvé la convergence $\|\varphi_n - u\|_{H^s(\mathbb{R}^d)} \to 0$ qui est une convergence d'une suite dans $\mathbb{R}$, et dans laquelle l'opération $\|.\|_{H^s}$ est définie sur un espace de fonctions et n'est pas une norme (c'est pour que cela devienne une norme que l'on prend les classes).

Nous en avons déduit la convergence $\varphi_n \overset{H^s(\mathbb{R}^d)}{\longrightarrow} u$ où maintenant φ_n et u sont des classes dans $H^s(\mathbb{R}^d)$.

Proposition 30.17.
La partie $\mathscr{D}(\mathbb{R}^d)$ est dense dans $\big(H^s(\mathbb{R}^d), \|.\|_{H^s(\mathbb{R}^d)}\big)$.

Démonstration. Nous savons déjà que $\mathscr{D}(\mathbb{R}^d)$ est dense dans $H^s(\mathbb{R}^d)$ par la proposition 30.15. Nous devons seulement prouver que $\mathscr{D}(\mathbb{R}^d)$ est dense dans $\big(\mathscr{S}(\mathbb{R}^d), \|.\|_{H^2(\mathbb{R}^d)}\big)$. Pour cela nous utilisons la densité de $\mathscr{D}(\mathbb{R}^d)$ dans $\mathscr{S}(\mathbb{R}^d)$ de la proposition 26.153. Soit donc $f \in \mathscr{D}(\mathbb{R}^d)$ et une suite f_k dans $\mathscr{D}(\mathbb{R}^d)$ telle que

$$f_k \overset{\mathscr{S}(\mathbb{R}^d)}{\longrightarrow} f. \tag{30.64}$$

Vu que la transformée de Fourier est continue sur $\mathscr{S}(\mathbb{R}^d)$ (proposition 28.20) nous avons aussi

$$\hat{f}_k \overset{\mathscr{S}(\mathbb{R}^d)}{\longrightarrow} \hat{f}, \tag{30.65}$$

et en particulier pour tout polynôme P nous avons la convergence uniforme

$$P\hat{f}_k \overset{unif}{\longrightarrow} P\hat{f}. \tag{30.66}$$

D'autre part la fonction $\xi \mapsto |\hat{f}_k(\xi) - \hat{f}(\xi)|^2(1+\xi^2)^s$ est Schwartz et en tout point décroissante en k. Soientt $\epsilon > 0$ et $r > 0$ choisis de telle sorte à avoir

$$\int_{\|\xi\|>r} |\hat{f}_k(\xi) - \hat{f}(\xi)|(1+\xi^2)^s d\xi < \epsilon. \tag{30.67}$$

pour tout k. La convergence uniforme (30.66) permet de considérer k_0 tel que pour tout $k > k_0$,

$$|\hat{f}_k - \hat{f}|(1+\xi^2)^s < \frac{\epsilon}{\mathrm{Vol}\big(B(0,r)\big)} \tag{30.68}$$

dans $B(0,r)$. Avec tout cela, dès que $k > k_0$ nous avons

$$\|f_k - f\|_{H^s(\mathbb{R}^d)} = \int_{\mathbb{R}} |\hat{f}_k - \hat{f}|(1+\xi^2)^s d\xi = \int_{B(0,r)} \dots + \int_{\|\xi\|>r} \dots \leqslant 2\epsilon. \tag{30.69}$$

Donc nous avons bien $\|f_k - f\|_{H^s(\mathbb{R}^d)} \to 0$ et convergence de f_k vers f dans $H^s(\mathbb{R}^d)$. □

30.2 Trace

Définition 30.18 ([483]).
*Nous définissons la **trace** d'une fonction par*

$$\begin{aligned} \gamma_0 \colon \mathscr{D}(\mathbb{R}^d) &\to \mathscr{D}(\mathbb{R}^{d-1}) \\ (\gamma_0 v)(x_1, \dots, x_{d-1}) &= v(x_1, \dots, x_{d-1}, 0). \end{aligned} \tag{30.70}$$

Théorème 30.19 ([485, 483]).
Si $s > \frac{1}{2}$, alors γ_0 accepte une unique extension en opérateur linéaire borné

$$\gamma_0 \colon H^s(\mathbb{R}^d) \to H^{s-\frac{1}{2}}(\mathbb{R}^{d-1}). \tag{30.71}$$

Démonstration. Nous subdivisons la preuve en plusieurs pas.

Une inégalité pour $\varphi \in \mathscr{D}(\mathbb{R}^d)$ Nous commençons par considérer $v \in \mathscr{D}(\mathbb{R}^d)$ (fonction C^∞ à support compact). Nous allons alors prouver que

$$\|\gamma_0\varphi\|_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})} \leqslant K\|\varphi\|_{H^s(\mathbb{R}^d)} \tag{30.72}$$

pour une certaine constante K (qui ne dépend en particulier pas de φ).
Nous avons

$$\begin{aligned}
\|\gamma_0\varphi\|^2_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})} &= (\gamma_0\varphi, \gamma_0\varphi)_{H^{s-\frac{1}{2}}} && (30.73a)\\
&= \int_{\mathbb{R}^{d-1}} |\widehat{\gamma_0\varphi(\xi)}|^2(1+\|\xi\|^2)^{s-\frac{1}{2}}d\xi && (30.73b)\\
&= \int_{\mathbb{R}^{d-1}} \Big|\int_{\mathbb{R}^{d-1}} (\gamma_0\varphi)(x)e^{-i\xi x}dx\Big|^2(1+\|\xi\|^2)^{s-\frac{1}{2}}d\xi && (30.73c)\\
& && (30.73d)
\end{aligned}$$

Nous appliquons la trace en appliquant la formule du corolaire 28.27,

$$(\gamma_0\varphi)(x) = \varphi(x,0) = \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{-iky}\varphi(x,y)dy\,dk \tag{30.74}$$

En remplaçant dans (30.73c) nous avons

$$\|\gamma_0\varphi\|^2_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})} = \frac{1}{2\pi}\int_{\mathbb{R}^{d-1}} \Big|\int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{-iky}e^{-i\xi x}\varphi(x,y)dy\,dk\,dx\Big|^2(1+\|\xi\|^2)^{s-\frac{1}{2}}d\xi. \tag{30.75}$$

Nous voudrions permuter les intégrales en k et en x. Pour cela nous étudions la fonction $u\colon \mathbb{R}\times\mathbb{R}^{d-1}\to\mathbb{C}$ donnée par

$$u(k,x) = e^{-ikx}\int_{\mathbb{R}} e^{-i\xi x}\varphi(x,y)dy \tag{30.76}$$

Effectuer l'intégrale par rapport à y revient à calculer la transformée de Fourier partielle dont nous parlons dans la proposition 28.21 [4]. Elle est donc une fonction Schwartz de k et de x (conjointement et non seulement séparément) et est donc dans $L^1(\mathbb{R}\times\mathbb{R}^{d-1})$. Les intégrales sur k et sur x peuvent donc être réunies et permutées par le théorème de Fubini 13.262 (n'oubliez tout de même pas de vous convaincre que la condition (2) est remplie).
Nous avons donc

$$\|\gamma_0\varphi\|^2_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})} = \frac{1}{2\pi}\int_{\mathbb{R}^{d-1}} \Big|\int_{\mathbb{R}}\int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}} e^{-iky}e^{-i\xi x}\varphi(x,y)dy\,dx\,dk\Big|^2(1+\|\xi\|^2)^{s-\frac{1}{2}}d\xi. \tag{30.77}$$

Étant donné que φ est à support compact, les intégrales sur x et sur y peuvent se réunir en utilisant encore le théorème de Fubini ; ces intégrales donnent :

$$\int_{\mathbb{R}^{d-1}\times\mathbb{R}} e^{-iky}e^{-i\xi x}\varphi(x,y)dx\otimes dy = \int_{\mathbb{R}^{d-1}\times\mathbb{R}} e^{-i(\xi,k)\,\cdot\,(x,y)}\varphi(x,y)dx\otimes dy = \hat{\varphi}(\xi,k). \tag{30.78}$$

Nous restons avec

$$\|\gamma_0\varphi\|^2_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})} = \frac{1}{2\pi}\int_{\mathbb{R}^{d-1}} \Big|\int_{\mathbb{R}} \hat{\varphi}(\xi,k)dk\Big|^2(1+\|\xi\|^2)^{s-\frac{1}{2}}d\xi. \tag{30.79}$$

Nous allons maintenant traiter la partie du milieu :

$$\clubsuit = \Big|\int_{\mathbb{R}} \hat{\varphi}(\xi,k)dk\Big| = \Big|\int_{\mathbb{R}} \hat{\varphi}(\xi,k)(1+\xi^2+k^2)^{s/2}\frac{1}{(1+\xi^2+k^2)^{s/2}}dk\Big| = |\langle f_1, f_2\rangle_{L^2(\mathbb{R}^d)}| \tag{30.80}$$

4. Dont une relecture de la preuve ne serait vraiment pas de trop, ainsi que la preuve de 26.147.

Ici ξ est vu comme une constante et les fonctions f_1 et f_2 sont

$$f_1\colon k \to \hat{\varphi}(\xi,k)(1+\xi^2+k^2)^{s/2} \tag{30.81a}$$

$$f_2\colon k \to \frac{1}{(1+\xi^2+k^2)^{s/2}} \tag{30.81b}$$

Nous pouvons utiliser l'inégalité de Cauchy-Schwarz 7.191 :

$$\clubsuit \leqslant \left(\int_{\mathbb{R}} |\hat{\varphi}(\xi,k)|^2(1+\xi^2+k^2)^s dk\right)^{1/2} \left(\int_{\mathbb{R}} \frac{1}{(1+\xi^2+k^2)^s} dk\right)^{1/2} \tag{30.82}$$

Nous notons $g(\xi)$ ce qui se trouve dans la seconde parenthèse (après intégration sur k). Avec cela nous continuons :

$$\|\gamma_0\varphi\|^2_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})} \leqslant \frac{1}{2\pi}\int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}} |g(\xi)||\hat{\varphi}(\xi,k)|^2(1+\xi^2+k^2)^s(1+\|\xi\|^2)^{s-\frac{1}{2}} dk\, d\xi. \tag{30.83}$$

Vu que $\hat{\varphi}$ est Schwartz, la fonction qui est à l'intérieur des deux intégrales est dans $L^1(\mathbb{R}^{d-1}\times\mathbb{R})$ et nous pouvons réunir les deux intégrales :

$$\|\gamma_0\varphi\|^2_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})} \leqslant \frac{1}{2\pi}\int_{\mathbb{R}\times\mathbb{R}^{d-1}} |g(\xi)||\hat{\varphi}(\xi,k)|^2(1+\xi^2+k^2)^s(1+\|\xi\|^2)^{s-\frac{1}{2}} dk\otimes d\xi. \tag{30.84}$$

À ce point nous démontrons qu'en réalité la combinaison $g(\xi)(1+\xi^2)^{s-\frac{1}{2}}$ ne dépend pas de ξ. En effet

$$g(\xi)(1+\xi^2)^{s-\frac{1}{2}} = (1+\xi^2)^{s-\frac{1}{2}}\int_{\mathbb{R}} \frac{1}{(1+\xi^2+k^2)} dk \tag{30.85a}$$

$$= \frac{1}{(1+\xi^2)^{1/2}}\int_{\mathbb{R}} \left(\frac{1+\xi^2}{1+\xi^2+k^2}\right)^s dk \tag{30.85b}$$

$$= \frac{1}{(1+\xi^2)^{1/2}}\int \left(\frac{1}{1+\frac{k^2}{1+\xi^2}}\right)^s dk. \tag{30.85c}$$

Nous effectuons le changement de variables $t=\frac{k}{\sqrt{1+\xi^2}}$, $dk=(1+\xi^2)^{1/2}dt$, et le tout vaut

$$\int_{\mathbb{R}} \left(\frac{1}{1+t^2}\right)^s dt, \tag{30.86}$$

qui est effectivement indépendant de ξ. Nous nommons cela K (auquel nous ajoutons le $\frac{1}{2\pi}$) :

$$\|\gamma_0\varphi\|^2_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})} \leqslant K\int_{\mathbb{R}\times\mathbb{R}^{d-1}} |\hat{\varphi}(\xi,k)|^2(1+\xi^2+k^2)^s dk\otimes d\xi = K\|\varphi\|^2_{H^s(\mathbb{R}^d)}. \tag{30.87}$$

Nous avons donc prouvé pour tout $\varphi\in\mathscr{D}(\mathbb{R}^d)$ (avec redéfinition du K) :

$$\|\gamma_0\varphi\|_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})} \leqslant K\|\varphi\|_{H^s(\mathbb{R}^d)}. \tag{30.88}$$

À propos de classes Il serait tentant de conclure en disant que $\mathscr{D}(\mathbb{R}^d)$ est dense dans $H^s(\mathbb{R}^d)$. Hélas, techniquement, l'ensemble $\mathscr{D}(\mathbb{R}^d)$ n'est même pas un sous-ensemble de $H^s(\mathbb{R}^d)$ parce que ce dernier est un ensemble de *classes* de fonctions. Ce petit détail a ici son importance parce que γ_0 n'est pas une application qui descend aux classes. En effet, $\mathbb{R}^{d-1}$ étant de mesure nulle dans $\mathbb{R}^d$, deux fonctions de la même classe peuvent différer en *tous* les points de $\mathbb{R}^{d-1}$ en même temps.

Si nous notons ι l'application qui consiste à prendre la classe ce qui est dense dans $H^s(\mathbb{R}^d)$, c'est $\iota(\mathscr{D}(\mathbb{R}^d))$. Or chaque classe contient au maximum une seule fonction continue (qui sera même de classe C^∞ à support compact pour les éléments de $\iota(\mathscr{D})$).

L'application γ_0 considérée est l'application composée entre le γ_0 classique et le choix du représentant continu dans la classe. La formule (30.88) que nous venons de prouver est valide pour l'application γ_0 vue comme

$$\gamma_0\colon \iota\big(\mathscr{D}(\mathbb{R}^d)\big) \to \mathscr{D}(\mathbb{R}^{d-1}). \tag{30.89}$$

Densité et conclusion Ce que la majoration (30.88) prouve est la continuité de l'application

$$\gamma_0 \colon \big(\mathscr{D}(\mathbb{R}^d), \|.\|_{H^s(\mathbb{R}^d)}\big) \to \big(H^{s-\frac{1}{2}}(\mathbb{R}^{d-1}), \|.\|_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})}\big). \tag{30.90}$$

Mais la proposition 30.17 nous donne la densité de la partie $\mathscr{D}(\mathbb{R}^d)$ dans $H^s(\mathbb{R}^d)$. La proposition 16.121 nous donne alors une extension

$$\gamma_0 \colon \big(H^s(\mathbb{R}^d), \|.\|_{H^s(\mathbb{R}^d)}\big) \to \big(H^{s-\frac{1}{2}}(\mathbb{R}^{d-1}), \|.\|_{H^{s-\frac{1}{2}}(\mathbb{R}^{d-1})}\big). \tag{30.91}$$

□

Remarque 30.20.
L'extension n'est pas évidente parce que les éléments de $H^s(\mathbb{R}^d)$ sont en général des classes de fonctions dont les valeurs sur le bord ne sont pas du tout fixées du fait que le bord soit de mesure nulle.

30.3 Théorème de plongement

L'objet des théorèmes de plongement de Sobolev est de montrer que si $s > \frac{d}{2} + k$ alors les éléments de $H^s(\mathbb{R}^d)$ possèdent des représentants de classe C^k. Avant de démontrer le théorème, pour alléger, nous allons donner deux lemmes.

Lemme 30.21.
Soit (u_j) une suite dans $\mathscr{S}(\mathbb{R}^d)$ telle que

$$u_j \xrightarrow{H^s(\mathbb{R}^d)} u \tag{30.92}$$

avec $s > 0$. Alors nous avons aussi la convergence

$$u_j \xrightarrow{L^2(\mathbb{R}^d)} u. \tag{30.93}$$

Démonstration. Vu que $s > 0$ nous avons $(1 + k^2)^s > 1$ (ici nous écrivons k^2 pour $\|k\|^2$). Par conséquent

$$(u, v)_{H^s(\mathbb{R}^d)} = \int_{\mathbb{R}^d} \hat{u}\bar{\hat{v}}(1 + k^2)^s dk \geqslant \int_{\mathbb{R}^d} \hat{u}\bar{\hat{v}} dk = \langle \hat{u}, \hat{v} \rangle_{L^2(\mathbb{R}^d)}. \tag{30.94}$$

Nous avons alors

$$\begin{aligned}
\|u_j - u\|_{L^2} &= \frac{1}{(2\pi)^d} \|\hat{u}_j - \hat{u}\|_{L^2} && (30.95a)\\
&= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\hat{u}_j - \hat{u}|^2 && (30.95b)\\
&\leqslant \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\hat{u}_j - \hat{u}|^2 (1 + k^2)^s dk && (30.95c)\\
&= \frac{1}{(2\pi)^d} \|u_j - u\|_{H^s(\mathbb{R}^d)}. && (30.95d)
\end{aligned}$$

□

Lemme 30.22.
Soient des fonctions $u_j \in \mathscr{S}(\mathbb{R}^d)$ telles que

$$u_j \xrightarrow{\left(C_0^0(\mathbb{R}^d), \|.\|_\infty\right)} v. \tag{30.96}$$

Alors nous avons la convergence

$$\int_{\mathbb{R}^d} u_j \varphi \to \int_{\mathbb{R}^d} v\varphi \tag{30.97}$$

pour tout $\varphi \in \mathscr{S}(\mathbb{R}^d)$.

Démonstration. La suite (u_j) est équibornée. En effet il existe une queue de suite pour laquelle $\|u_j - v\|_\infty < \epsilon$; cette queue de suite est alors équibornée par $\|v\|_\infty + \epsilon$. Le début de la suite est un nombre fini de fonctions, toutes bornées. Le maximum des bornes donne alors une borne.

Soit donc $M > 0$ tel que $|u_j(x)| < M$ pour tout $x \in \mathbb{R}^d$ et pour tout $j \in \mathbb{N}$. Nous avons alors $|u_j\varphi| < M|\varphi|$ pour tout j et les fonctions $|u_j\varphi|$ sont majorées par la fonction $M|\varphi|$ qui est intégrable. Nous pouvons donc utiliser le théorème de la convergence dominée de Lebesgue 13.184 nous donne

$$\lim_{j\to\infty} \int_{\mathbb{R}^d} u_j\varphi = \int_{\mathbb{R}^d} v\varphi. \tag{30.98}$$

□

Nous pouvons écrire la conclusion du lemme 30.22 sous la forme

$$\langle u_j, \varphi\rangle_{L^2(\mathbb{R}^d)} \to \langle v, \varphi\rangle_{L^2(\mathbb{R}^d)} \tag{30.99}$$

pour tout $\varphi \in \mathscr{S}(\mathbb{R}^r)$ (et non pour tout $\varphi \in L^2(\mathbb{R}^d)$.

Théorème 30.23 (Théorème de Sobolev avec $k = 0$[486])**.**
Soit $s > \frac{d}{2}$ et $u \in H^s(\mathbb{R}^d)$. Alors u possède un représentant dans $C^0_0(\mathbb{R}^d)$ (les fonctions continues et qui s'annulent à l'infini). Nous écrivons cela $H^s(\mathbb{R}^d) \subset C^0_0(\mathbb{R}^d)$.

Démonstration. Nous commençons par supposer que $u \in H^s(\mathbb{R}^d) \cap \mathscr{S}(\mathbb{R}^d)$, et dans ce cas nous notons u le représentant dans $\mathscr{S}(\mathbb{R}^d)$. Nous allons prouver l'inégalité

$$\|u\|_\infty \leqslant c\|u\|_{H^s(\mathbb{R}^d)}. \tag{30.100}$$

La formule d'inversion de Fourier 28.26 appliquée à u_j donne

$$u(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{ikx}\hat{u}(k)dk, \tag{30.101}$$

nous avons alors

$$\begin{aligned}
(2\pi)^d|u(x)| &\leqslant \int_{\mathbb{R}^d} |\hat{u}(k)|dk && (30.102a)\\
&= \int_{\mathbb{R}^d} (1+k)^{s/2}|\hat{u}(k)|(1+k^2)^{-s+2}dk && (30.102b)\\
&= \int_{\mathbb{R}^d} \underbrace{\Big(|\hat{u}(k)|^2(1+k^2)^s\Big)^{1/2}}_{f} \underbrace{\Big((1+k^2)^{-s}\Big)^{1/2}}_{g} dk && (30.102c)\\
&= \langle f, g\rangle_{L^2(\mathbb{R}^d)}. && (30.102d)
\end{aligned}$$

Ici il convient nous arrêter un instant pour nous convaincre que f et g sont réellement des éléments de L^2. En ce qui concerne f, c'est facile : $\hat{u}$ est une fonction Schwartz. En ce qui concerne g il faut l'intégrabilité de $|g|^2$, c'est-à-dire de $k \mapsto (1+k^2)^{-s}$. Cela a lieu si et seulement si $2s > n$ et donc a lieu dans les hypothèses du théorème. Nous utilisons le théorème de Cauchy-Schwarz [5] pour continuer :

$$\begin{aligned}
(2\pi)^d|u(x)| &\leqslant \|f\|_{L^2}\|g\|_{L^2} && (30.103a)\\
&= c\left(\int_{\mathbb{R}^d} |\hat{u}(k)|^2(1+k^2)^s dk\right)^{1/2} && (30.103b)\\
&= c\|u\|_{H^s(\mathbb{R}^d)}. && (30.103c)
\end{aligned}$$

Donc en introduisant le facteur $(2\pi)^d$ dans la constante c nous avons

$$\|u\|_\infty \leqslant c\|u\|_{H^s(\mathbb{R}^d)}. \tag{30.104}$$

5. Formule 7.165.

Cela est tout ce que nous voulions faire avec $u \in \mathscr{S}(\mathbb{R}^d)$.

Nous considérons maintenant $u \in H^s(\mathbb{R}^d)$. Vu que la densité des fonctions Schwartz dans H^s (proposition 30.15) nous considérons une suite (u_j) dans $\mathscr{S}(\mathbb{R}^d)$ telle que

$$u_j \xrightarrow{H^s(\mathbb{R}^d)} u \tag{30.105}$$

Ici u est une classe, mais nous identifions u_j avec sa classe (parce qu'il ne faut pas exagérer non plus). La suite (u_j) est de Cauchy dans H^s, donc si $\epsilon > 0$ est donné, il existe N tel que si $n, m > N$, $\|u_m - u_n\| \leqslant \epsilon$. Nous avons alors aussi

$$\|u_m - u_n\|_\infty \leqslant c\epsilon, \tag{30.106}$$

ce qui signifie que (u_j) est également une suite de Cauchy dans $\big(C^0(\mathbb{R}^d), \|.\|_\infty\big)$ qui est un espace complet par la proposition 11.296.

Il existe donc une fonction $v \in C^0_0(\mathbb{R}^d)$ telle que

$$u_j \xrightarrow{\left(C^0_0(\mathbb{R}^d), \|.\|_\infty\right)} v. \tag{30.107}$$

La question est de savoir si nous pouvons déduire que v est un représentant de u.

Par le lemme 30.21 nous avons également la convergence

$$u_j \xrightarrow{L^2(\mathbb{R}^d)} u. \tag{30.108}$$

Pour tout $\varphi \in \mathscr{S}(\mathbb{R}^d)$ nous avons alors

$$\langle u_j, \varphi \rangle_{L^2} \to \langle u, \varphi \rangle_{L^2}. \tag{30.109}$$

Mais en même temps, la convergence (30.107) couplée au lemme 30.22 donne également

$$\langle u_j, \varphi \rangle_{L^2} \to \langle v, \varphi \rangle_{L^2}. \tag{30.110}$$

Par unicité de la limite (dans $\mathbb{R}$) nous avons

$$\langle v, \varphi \rangle_{L^2} = \langle u, \varphi \rangle_{L^2} \tag{30.111}$$

pour tout $\varphi \in \mathscr{S}(\mathbb{R}^d)$. La proposition 29.1 appliquée à $u - v$ montre alors que $u - v = 0$ presque partout, c'est-à-dire que v est bien un représentant de u.

Le représentant v de u est non seulement continu (comme limite uniforme de fonctions continues), mais également bornée, comme limite uniforme de fonctions Schwartz. □

Proposition 30.24 ([486]).
Si $u \in H^s(\mathbb{R}^d)$ ($s \in \mathbb{R}$) alors

(1) $\partial^\alpha u \in H^{s-|\alpha|}(\mathbb{R}^d)$,

(2) l'application

$$\partial^\alpha \colon H^s(\mathbb{R}^d) \to H^{s-|\alpha|}(\mathbb{R}^d) \tag{30.112}$$

est continue.

Note : ici ∂ est l'opération de dérivée faible.

Démonstration. Nous allons seulement prouver que $\partial_j \colon H^s(\mathbb{R}^d) \to H^{s-1}(\mathbb{R}^d)$ est bien définie[6] et continue. Par composition, la thèse suivra.

Soit $u \in H^s(\mathbb{R}^d)$ par le lemme 30.10 nous avons

$$\widehat{\partial_j u} = i\xi_j \hat{u}. \tag{30.113}$$

6. Au sens où l'espace d'arrivée est bien celui-là.

D'autre part, la fonction

$$\begin{aligned} f\colon \mathbb{R}^n &\to \mathbb{R} \\ x &\mapsto \frac{x_i}{1+\|x\|^2} \end{aligned} \tag{30.114}$$

est bornée (et même indépendamment de i) par une constante K. Donc nous avons pour tout[7] s :

$$k_i(1+\|k\|^2)^{-s} < K(1+\|k\|^2)^{-s+1}. \tag{30.115}$$

Avec cela nous pouvons calculer un peu : si $u \in H^s(\mathbb{R}^d)$, nous avons

$$\begin{aligned} \|\partial_j u\|_{H^{s-1}(\mathbb{R}^d)} &= \int_{\mathbb{R}^d} |\widehat{\partial_j u}|(1+k^2)^{s-1}dk & (30.116a)\\ &= \int_{\mathbb{R}^d} k_j|\hat{u}|(1+k^2)^{s-1}dk & (30.116b)\\ &\leqslant \int_{\mathbb{R}^d} K|\hat{u}|(1+k^2)^{s}dk & (30.116c)\\ &= K\|u\|_{H^s(\mathbb{R}^d)}. & (30.116d) \end{aligned}$$

Nous avons donc que $\|\partial_j u\|_{H^{s-1}(\mathbb{R}^d)}$ est fini lorsque $u \in H^s(\mathbb{R}^d)$.

La majoration $\|\partial_j u\| \leqslant K\|u\|$ donne la majoration suivante pour la norme de l'opérateur ∂_j :

$$\|\partial_j\| = \sup_{\|u\|_{H^s}=1} \|\partial_j u\|_{H^{s-1}} \leqslant K. \tag{30.117}$$

Le fait d'être borné implique d'être continu par la proposition 10.18. □

Théorème 30.25 (Théorème de plongement de Sobolev [486])**.**
Soient $k \in \mathbb{N}$ et $m > \frac{d}{2} + k$. Alors

$$H^s(\mathbb{R}^d) \subset C_0^k(\mathbb{R}^d). \tag{30.118}$$

Remarques :

— L'espace $C_0^k(\mathbb{R}^d)$ est l'ensemble des fonctions de classe C^k qui s'annulent à l'infini.

— L'inclusion (30.118) signifie que tout élément dans H^s possède un représentant dans $C_0^k(\mathbb{R}^d)$.

Démonstration. Pour $k = 0$, c'est le théorème 30.23. Si $|\alpha| < k$ nous savons que $\partial^\alpha u \in H^{s-k} \subset C_0^0(\mathbb{R}^d)$. Cela signifie que les dérivées faibles sont continues, mais pas qu'il existe un représentant qui est réellement k fois continument dérivable.

Soit $u \in H^s(\mathbb{R}^d)$ et une suite (u_j) dans $\mathscr{S}(\mathbb{R}^d)$ telle que

$$u_j \xrightarrow{H^s(\mathbb{R}^d)} u. \tag{30.119}$$

Vu que l'espace topologique $\big(C_0^k(\mathbb{R}^d), \|.\|_\infty\big)$ est complet il existe $v \in C_0^k$ tel que

$$u_j \xrightarrow{C_0^k} v. \tag{30.120}$$

Il reste à montrer que v est un représentant de u. Cela se fait comme plus haut en montrant que $u_j \xrightarrow{L^2} u$. □

7. Question : dans [486], il faut dépendre cette constante de s. Je ne comprends pas pourquoi.

Chapitre 31

Équations différentielles ordinaires

Une équation différentielle ordinaire est la recherche de toutes les fonctions définie sur une partie de $\mathbb{R}$ satisfaisant à une certaine égalité, faisant intervenir les dérivées de la fonction recherchée.

Dans la suite, I désignera un intervalle de $\mathbb{R}$. Une fonction sera **dérivable sur** I si elle est dérivable au sens usuel sur l'intérieur de I, et si elle est dérivable à droite (resp. à gauche) sur l'éventuel bord gauche (resp. droit) de I.

Définition 31.1.
Une ***équation différentielle ordinaire d'ordre*** n ***sur*** I *est la recherche d'une fonction* $y : I \to \mathbb{R}$ *dérivable* n *fois, satisfaisant à une équation du type*

$$F(t, y(t), y'(t), \dots, y^{n\prime}(t)) = 0 \quad \textit{pour tout } t \in I \tag{31.1}$$

où I *est un intervalle de* $\mathbb{R}$ *et* $F : (I \times D) \subset (\mathbb{R} \times \mathbb{R}^{n+1}) \to \mathbb{R}$ *est une fonction donnée.*

Remarque 31.2.
L'équation différentielle (31.1) sera raccourcie sous la forme

$$F(t, y, y', \dots, y^{n\prime}) = 0 \tag{31.2}$$

où la dépendance en t est sous-entendue.

Exemple 31.3.
Soit $f : I \to \mathbb{R}$ une fonction continue fixée. L'équation différentielle

$$y' = f(t) \tag{31.3}$$

se ramène à la recherche des primitives de f sur l'intervalle I. △

Le lemme suivant sert de temps en temps.

Lemme 31.4 (Lemme de Grönwall)**.**
Soient ϕ *et* ψ *deux fonctions telles que pour tout* $t \in [t_0, t_1]$, $\phi(t) \geqslant 0$, $\psi(t) \geqslant 0$ *et*

$$\phi(t) \leqslant K + L \int_{t_0}^{t} \psi(s)\phi(s)ds \tag{31.4}$$

où K *et* L *sont des constantes positives. Alors*

$$\phi(t) \leqslant K \exp\big(L \int_{t_0}^{t} \psi\big). \tag{31.5}$$

Lemme 31.5 (Lemme de Grönwall[487])**.**
Si $u, a, b \in C^0\big([0, T], \mathbb{R}^+\big)$ *sont telles que*

$$u(t) \leqslant b(t) + \int_0^t a(s)u(s)ds \tag{31.6}$$

pour tout $t \in [0,T]$ alors pour tout $t \in [0,T]$ nous avons aussi

$$u(t) \leqslant b(t) + \int_0^t b(s)a(s)e^{\int_s^t a(u)du}ds. \tag{31.7}$$

31.1 Équation homogène, solution particulière

Voici un petit morceau d'algèbre linéaire. Soient des espaces vectoriels V et W ainsi qu'une application linéaire $D\colon V \to W$. Nous voulons résoudre $D(u) = v$, c'est-à-dire déterminer l'ensemble

$$D^{-1}(v) = \{u \in V \text{ tel que } Du = v\}. \tag{31.8}$$

Lemme 31.6.
Soient des espaces vectoriels V et W ainsi qu'une application linéaire $D\colon V \to W$. Si $u_P \in V$ satisfait à $Du_P = v$ alors

$$D^{-1}(v) = \ker(D) + u_P. \tag{31.9}$$

Démonstration. Si $u \in \ker(D) + u_p$ alors $u = k + u_p$ avec $Dk = 0$, ce qui donne tout de suite $Du = Dk + Du_P = v$. Donc $u \in D^{-1}(v)$.

Dans l'autre sens, si $u \in D^{-1}(v)$ alors nous pouvons écrire $u = (u - u_P) + u_P$. Vu que $u - u_P \in \ker(D)$ nous avons bien $u \in \ker(D) + u_P$. □

Ce petit lemme explique pourquoi la résolution d'équation différentielles passe par le principe « générale de l'homogène plus particulière de la non-homogène ». Cela marche autant pour les équations différentielles ordinaires que pour celles aux dérivées partielles.

Exemple 31.7.
Considérons l'équation différentielle ordinaire

$$y' - y = 4. \tag{31.10}$$

L'opérateur dont nous parlons est par exemple

$$D\colon C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}) y \mapsto y' - y. \tag{31.11}$$

Nous devons résoudre $Dy = 4$ où « 4 » est l'élément fonction constante égale à 4 dans $C^\infty(\mathbb{R})$. L'ensemble $\ker(D)$ sont les éléments $y \in C^\infty(\mathbb{R})$ tels que $y' = y$:

$$\ker(D) = \{t \mapsto Ke^t \text{ tel que } K \in \mathbb{R}\}. \tag{31.12}$$

Nous devons trouver un élément quelconque y_P de $D^{-1}(4)$. Facile : $y_P(t) = -4$.

Au final,

$$D^{-1}(4) = \{t \mapsto Ke^t - 4 \text{ tel que } K \in \mathbb{R}\}. \tag{31.13}$$

△

Dans cet exemple nous avons pris $V = W = C^\infty(\mathbb{R})$. Mais souvent nous sommes amenés à considérer des espaces plus subtils, parce qu'il existe simplement pas de solutions dans C^∞, ou alors parce que beaucoup de solutions n'y sont pas.

31.2 Que faire avec $f(z)dz = g(t)dt$?

Dans de nombreux exercices d'équations différentielles, nous tombons sur $u' = f(t)$, et nous faisons formellement

$$\frac{du}{dt} = f(t) \quad \Rightarrow \quad du = f(t)dt, \tag{31.14}$$

et ensuite, il y a la formule un peu magique

$$u - u_0 = \int_{t_0}^{t} f(t)dt. \tag{31.15}$$

Voyons ce qu'il en est. Tout d'abord, il faut comprendre ce que signifie la formule

$$f(z)dz = g(t)dt. \tag{31.16}$$

Il s'agit d'une égalité entre deux formes différentielles sur $\mathbb{R}$ où z est une fonction de t. Étant donné que z est une fonction de t, il faut voir dz comme la différentielle de cette fonction. La différentielle d'une fonction à une variable est donné par la dérivée :

$$dz_t = z'(t)dt \tag{31.17}$$

Écrire l'équation (31.16) pour chaque t revient donc à écrire

$$f\big(z(t)\big)z'(t)dt = g(t)dt \tag{31.18}$$

Cela est une égalité entre deux formes différentielles. Nous avons donc égalité entre les intégrales des formes sur un chemin. Prenons un chemin tout simple de t_0 vers t :

$$\int_{t_0}^{t} f\big(z(t)\big)z'(t)dt = \int_{t_0}^{t} g(t)dt. \tag{31.19}$$

Dans le premier membre, nous faisons un changement de variable $\xi = z(t)$, $d\xi = z'(t)dt$, et nous obtenons

$$\int_{z_0}^{z(t)} f(\xi)d\xi = \int_{t_0}^{t} g(t)dt. \tag{31.20}$$

où nous avons remplacé la constante $z(t_0)$ par z_0 dans la borne d'intégration. Si F est une primitive de f et G une primitive de g, nous avons

$$F(z) - F(z_0) = G(t) - G(t_0). \tag{31.21}$$

Si aucun problème de Cauchy n'est donné, les constantes $F(z_0)$ et $G(t_0)$ sont mises en une seule et nous écrivons la solution

$$F\big(z(t)\big) = G(t) + C, \tag{31.22}$$

qui est une équation implicite pour $z(t)$.

Nous trouvons assez souvent le cas simple

$$f(z)dz = dt. \tag{31.23}$$

En remplaçant $g(t) = 1$ dans (31.20), nous trouvons la fameuse

$$t - t_0 = \int_{z_0}^{z} f(z)dz, \tag{31.24}$$

dans laquelle il y a un abus de notation terrible entre le z de la borne (que les étudiants oublient souvent) et la variable d'intégration z ! !

Le passage de (31.23) à (31.24) sera très souvent utilisé dans le cours de mécanique par exemple.

31.3 Équations linéaires du premier ordre

Une **équation différentielle linéaire** est une équation de la forme

$$y' + u(t)y = v(t). \tag{31.25}$$

Exemple 31.8.
Tant qu'il n'y a pas de second membre, c'est facile. Prenons l'exemple suivant :

$$y' + 2ty = 0. \tag{31.26}$$

Nous mettons tous les t d'un côté et tous les y et y' de l'autre :

$$\frac{y'}{y} = -2t, \tag{31.27}$$

et puis on intègre sans oublier la constante d'intégration :

$$\ln(y) = -t^2 + C, \tag{31.28}$$

et donc $y(t) = Ke^{-t^2}$. △

Exemple 31.9.
Lorsqu'il y a un second membre, il y a une astuce. Prenons par exemple

$$y' + 2ty = 4t. \tag{31.29}$$

L'astuce est de commencer par résoudre l'équation sans le second membre (l'équation homogène associée). Nous notons y_H la solution. Ici, la réponse est

$$y_H(t) = Ke^{-t^2}. \tag{31.30}$$

Ensuite le truc est d'essayer de trouver la solution de l'équation (31.29) sous la forme

$$y(t) = K(t)e^{t^2}. \tag{31.31}$$

L'idée est de prendre la même que la solution de l'équation homogène (sans second membre), mais en disant que K est une fonction. Afin de trouver la fonction K qui donne la solution, il suffit de remettre l'essai (31.31) dans l'équation (31.29) :

$$\underbrace{K'e^{-t^2} - 2tKe^{-t^2}}_{y'(t)} + \underbrace{2tKe^{-t^2}}_{2ty(t)} = 4t \tag{31.32}$$

Les deux termes avec K se simplifient et il reste

$$K'(t) = 4te^{t^2}, \tag{31.33}$$

ce qui signifie $K(t) = 2e^{t^2+C}$. Nous avons donc déterminé la fonction qui fait fonctionner l'essai, et la solution à l'équation est

$$y(t) = \big(2e^{t^2} + C\big)e^{-t^2} = 2 + Ce^{-t^2}. \tag{31.34}$$

△

La technique pour résoudre cette équation est de commencer par résoudre l'équation homogène associée. Si $U(t)$ est une primitive de $u(t)$, nous avons

$$\begin{aligned} y_H'(t) + u(t)y_H(t) &= 0 \\ \frac{y_H'}{y_H} &= -u(t) \\ \ln(y_H) &= -U(t) + C \\ y_H(t) &= e^{-U(t)+C} = Ke^{-U(t)} \end{aligned} \tag{31.35}$$

où $K = e^C$.

Cela fournit la solution générale de l'équation homogène. Il existe un truc génial qui permet d'en tirer la solution générale du système non homogène. Lorsque nous avons trouvé $y_H(t) = Ke^{-U(t)}$, le symbole K désigne une constante. La méthode de **variation des constantes** consiste à essayer la solution

$$y(t) = K(t)e^{-U(t)}, \tag{31.36}$$

c'est-à-dire à dire que la constante est en réalité une fonction. Afin de trouver quelle fonction $K(t)$ fait en sorte que l'essai (31.36) soit une solution, nous la remplaçons dans l'équation de départ $y' + uy = v$. Maintenant,

$$y'(t) = K'(t)e^{-U(t)} - K(t)u(t)e^{-U(t)}. \tag{31.37}$$

En remettant dans l'équation,

$$y' + uy = K'e^{-U} - Kue^{-U} + uKe^{-U} = K'e^{-U} = v. \tag{31.38}$$

Notez que les termes en K se sont miraculeusement simplifiés. Cela est directement dû au fait que e^{-U} est solution de l'équation homogène. Nous restons avec l'équation

$$K' = \frac{v}{e^{-U}} \tag{31.39}$$

pour $K(t)$. La solution générale du problème non homogène est donc finalement donnée par

$$y(t) = \big(W(t) + C\big)e^{-U(t)} \tag{31.40}$$

si $W(t)$ est une primitive de $v(t)e^{U(t)}$.

Tout ceci est un peu heuristique. La proposition suivante dit dans quels cas ça fonctionne.

Proposition 31.10.
Soient u et v continues sur I et U, une primitive de u sur I et W une primitive de ve^{-U} sur I. Une fonction $y\colon I \to \mathbb{R}$ est solution de $y' + u(t)y = v(t)$ si et seulement s'il existe une constante $C \in \mathbb{R}$ telle que

$$y(t) = \big(W(t) + C\big)e^{U(t)} \tag{31.41}$$

pour tout $t \in I$.

31.3.1 Pourquoi la variation des constantes fonctionne toujours ?

Prenons une équation non homogène

$$z'(t) = f(t)z(t) + g(t), \tag{31.42}$$

et supposons avoir une solution de l'homogène associée sous la forme $z_H(t) = Ch(t)$. Le coup de la variation des constates consiste à essayer une solution pour l'équation non homogène sous la forme [1]

$$z(t) = K(t)h(t). \tag{31.43}$$

Nous injectons cette solution dans l'équation de départ en utilisant le fait que $z'(t) = K'(t)h(t) + K(t)h'(t)$:

$$K'(t)h(t) + K(t)h'(t) = f(t)K(t)h(t) + g(t). \tag{31.44}$$

Le terme $K(t)h'(t)$ se récrit en utilisant la propriété de définition de h, c'est-à-dire que $h'(t) = f(t)h(t)$. Nous voyons que les termes ne contenant pas de K' se simplifient ; il reste

$$K'h = g. \tag{31.45}$$

Cette équation a comme solution

$$K = \int \frac{f}{h} + C. \tag{31.46}$$

1. Je ne sais plus qui a eu l'idée de changer le nom de la constante de C vers K au moment de la transformer en fonction, mais c'est une bonne idée.

J'insiste sur la constante d'intégration ! En réalité, celles et ceux qui auront compris l'équation (31.24) sauront que K est donné par

$$K(t) = \int_{\xi_0}^{t} \frac{f(\xi)}{g(\xi)} d\xi \tag{31.47}$$

où ξ_0 joue le rôle de la constante d'intégration.

Quoi qu'il en soit, la solution générale de l'équation non homogène est

$$z(t) = K(t)h(t) = \left(\int \frac{g}{h} + C \right) h. \tag{31.48}$$

Cette solution comprend deux termes : Ch qui est solution de l'homogène, et $\left(\int \frac{g}{h}\right) h$ qui est une particulière de l'équation non homogène.

Quelques conclusions :

(1) Si vous avez encore du K (et pas que du K') dans votre équation qui donne K, c'est que vous n'être pas dans le cadre d'une équation de type (31.42). Le plus souvent, c'est que vous avez fait une faute de calcul quelque part.

(2) La méthode des variations des constantes n'est pas en contradiction avec le principe de « SGEH+SPENH ». En effet, la SGEP et la SPENH sont toutes deux dans la solution (31.48).

(3) La variation des constantes peut être vue comme une façon cool de trouver une solution particulière de l'équation non homogène.

(4) La simplification ne se fait que après avoir remplacé Kh' par Kfh, c'est-à-dire après avoir utilisé le fait que z_H est solution de l'homogène. Sinon, la simplification n'est pas du tout évidente à priori. Il se peut même que, visuellement, les termes Kh' et Kfh ne se ressemblent pas du tout. Un exemple de cela arrivera par exemple dans l'exemple 31.13, pour arriver à l'équation (31.64).

31.4 Équations à variables séparées

Une **équation à variables séparées** est une équation de la forme

$$y' = u(t)f(y) \tag{31.49}$$

où $u\colon I \to \mathbb{R}$ et $f\colon J \to \mathbb{R}$ sont deux fonctions continues données. Les propositions 31.11 et 31.12 résolvent ce cas, mais avant de voir cela, nous allons donner quelques indications « pratiques ».

31.4.1 La méthode rapide

On peut évidemment mettre tous les y et y' d'un côté :

$$\frac{y'}{f(y)} = u(x). \tag{31.50}$$

Une fois que cela est fait, on écrit $y' = \frac{dy}{dx}$, et on envoie le dx du côté des x :

$$\frac{dy}{f(y)} = u(x)dx. \tag{31.51}$$

Maintenant il suffit de prendre l'intégrale des deux côtés : comme la position des dx et dy l'indiquent, il faut intégrer par rapport à y d'un côté et par rapport à dx de l'autre côté.

L'intégrale à gauche est facile : c'est $\ln(y)$. À droite, par contre, ça dépend tout à fait de u.

31.4.2 La méthode plus propre

$$y'(t) = u(t)f\big(y(t)\big). \tag{31.52}$$

Nous considérons U, une primitive de u sur I et G, une primitive de $1/f$ sur J. Si $I' \subseteq I$ et $y\colon I' \to J$, alors y est solution de (31.49) si et seulement s'il existe une constante C telle que

$$G\big(y(t)\big) = U(t) + C. \tag{31.53}$$

La recherche des solutions de l'équation différentielle se ramène donc à la recherche de primitives et de solutions d'une équation algébrique (il faut isoler $y(t)$ dans (31.53)). Réciproquement toute solution régulière de cette dernière relation est solution de l'équation différentielle.

Remarque : lorsque nous cherchons U et G, nous ne cherchons que *une* primitive. Il ne faut pas considérer des constantes d'intégration à ce niveau.

31.4.3 Les théorèmes

Proposition 31.11.
Nous considérons l'équation (31.49) avec $u(t)$ continue sur I et f continue sur J avec $f(\eta) \neq 0$ pour tout $\eta \in J$. Soit U, une primitive de u sur I, et G, une primitive de $1/f$ sur J.

Si $y\colon Y' \to J$ est une fonction sur un intervalle $I' \subset I$, alors y est solution de l'équation (31.49) si et seulement s'il existe $C \in \mathbb{R}$ tel que

$$G\big(y(t)\big) = U(t) + C. \tag{31.54}$$

Cette proposition dit que toutes les solutions qui ne s'annulent jamais sur un intervalle ont la forme $G\big(y(t)\big) = U(t) + C$ et peuvent donc être trouvées en calculant des primitives.

La formule (31.54) peut être obtenue de la façon heuristique suivante, en écrivant $y' = dy/dt$, et en passant le dt à droite. Nous trouvons successivement

$$\begin{aligned} y' &= u(t)f(y) \\ dy &= u(t)f(y)dt \\ \frac{dy}{f(y)} &= u(t)dt \\ \int \frac{dy}{f(y)} &= \int u(t)dt \\ G(y) &= U(t) + C. \end{aligned} \tag{31.55}$$

Proposition 31.12.
Soient u continue sur I et f continue sur J, et $f(\eta) \neq 0$ sur J. Soient $t_0 \in I$ et $y_0 \in J$. Alors il existe $I' \subset I$ avec $t_0 \in I'$ et $f \in C^1(I' \to J)$ tels que

(1) y est solution de (31.49) sur I' et vérifie $y(t_0) = y_0$,

(2) si z est une solution de (31.49) sur $I'' \subset I'$ avec $t_0 \in I''$ et $z(t_0) = y_0$, alors $I'' \subset I'$ et $z(t) = y(t)$ pour tout $t \in I''$.

Exemple 31.13.
Résoudre l'équation différentielle

$$y - \cos(t)y' = \cos(t)\big(1 - \sin(t)\big)y^2. \tag{31.56}$$

La fonction $y = 0$ est solution. En posant $z = 1/y$, nous trouvons l'équation

$$z + \cos(t)z' = \cos(t)\big(1 - \sin(t)\big) \tag{31.57}$$

à laquelle z doit satisfaire. L'équation homogène est

$$z'_H = -\frac{z_H}{\cos(t)}. \tag{31.58}$$

Ceci est une équation à variables séparées que nous résolvons en suivant les méthodes données plus haut : nous posons

$$\begin{aligned} u(t) &= \frac{1}{\cos(t)}, \\ f(z) &= -z, \\ U(t) &= \ln\left[\tan\left(\frac{\pi}{4}+\frac{t}{2}\right)\right] \quad \text{(voir formulaire)}, \\ G(z) &= \ln\left(\frac{1}{z}\right). \end{aligned} \tag{31.59}$$

La solution z_H est donnée par l'équation

$$\ln\left(\frac{1}{z}\right) = \ln\left[K\tan\left(\frac{\pi}{4}+\frac{t}{2}\right)\right], \tag{31.60}$$

c'est-à-dire

$$z_H(t) = \frac{K}{\tan\left(\frac{\pi}{4}+\frac{t}{2}\right)}. \tag{31.61}$$

Nous appliquons maintenant la méthode de variation des constantes sur cette solution afin de trouver la solution générale de l'équation (31.57). En utilisant la règle de Leibnitz, $z' = K'z_H + Kz'_H$, nous trouvons

$$\frac{K}{\tan\left(\frac{\pi}{4}+\frac{t}{2}\right)} + \cos(t)\left(\frac{K'}{\tan\left(\frac{\pi}{4}+\frac{t}{2}\right)} - \frac{K}{2\sin^2\left(\frac{\pi}{4}+\frac{t}{2}\right)}\right) = \cos(t)\big(1-\sin(t)\big). \tag{31.62}$$

Malgré leurs apparences, les deux termes en K se simplifient. En effet, en vertu de l'équation $z'_H = \frac{-z_H}{\cos(t)}$, nous avons

$$\frac{-K}{2\sin^2\left(\frac{\pi}{4}+\frac{t}{2}\right)} = \frac{-K}{\cos(t)\tan\left(\frac{\pi}{4}+\frac{t}{2}\right)}. \tag{31.63}$$

Le travail de voir quel est le lien entre $\sin^2\left(\frac{\pi}{4}+\frac{t}{2}\right)$, $\tan\left(\frac{\pi}{4}+\frac{t}{2}\right)$ et $\cos(t)$ est en réalité fait dans votre formulaire au moment où vous l'avez utilisé pour intégrer u pour obtenir le $U(t)$ de (31.59).

Après cette simplification durement méritée, nous trouvons l'équation suivante pour $K(t)$:

$$\frac{K'}{\tan\left(\frac{\pi}{4}+\frac{t}{2}\right)} = 1-\sin(t). \tag{31.64}$$

Résoudre cela revient à trouver la primitive de

$$\big(1-\sin(t)\big)\tan\left(\frac{\pi}{4}+\frac{t}{2}\right), \tag{31.65}$$

ce qui est relativement compliqué. La réponse est

$$\begin{aligned} K(t) = \ln\left(\sin\left(\frac{2\,x+\pi}{4}\right)+1\right) + \ln\left(\sin\left(\frac{2\,x+\pi}{4}\right)-1\right) \\ + 2\,\ln\sec\left(\frac{2\,x+\pi}{4}\right) + 2\,\sin^2\left(\frac{2\,x+\pi}{4}\right) \end{aligned} \tag{31.66}$$

Nous pouvons un peu simplifier en utilisant le fait que $\ln(a+b)+\ln(a-b) = \ln(a^2-b^2)$:

$$K(t) = \ln\left(-\cos^2\left(\frac{2\,x+\pi}{4}\right)\right) + 2\,\ln\sec\left(\frac{2\,x+\pi}{4}\right) + 2\,\sin^2\left(\frac{2\,x+\pi}{4}\right). \tag{31.67}$$

Il me semble toutefois qu'il faudrait prendre des valeurs absolues pour les logarithmes.

△

31.5 Équations linéaires d'ordre supérieur

31.5.1 Équations et systèmes linéaires à coefficients constants

Nous regardons l'équation

$$y^{(n)} + a_1 y^{(n-1)} + \cdots + a_{n-1} y' + a_n y = v(t) \tag{31.68}$$

où les coefficients a_k sont maintenant des constantes. Il faut commencer par résoudre le polynôme caractéristique

$$r^n + a_1 r^{n-1} + \cdots + a_n = 0. \tag{31.69}$$

Si $\lambda_1, \ldots, \lambda_k$ sont les solutions avec multiplicité $\mu_1, \ldots, \mu_k$, alors le **système fondamental** de solutions linéairement indépendantes est l'ensemble suivant de solutions à l'équation homogène :

$$\begin{gathered} e^{\lambda_1 t}, te^{\lambda_1 t}, \ldots, t^{\mu_1 - 1} e^{\lambda_1 t} \\ \vdots \\ e^{\lambda_k t}, te^{\lambda_k t}, \ldots, t^{\mu_k - 1} e^{\lambda_k t}. \end{gathered} \tag{31.70}$$

Nous notons y_i ces solutions. La solution générale de l'équation homogène est donc donnée par

$$y_H = \sum_i c_i y_i. \tag{31.71}$$

Afin de trouver la solution générale de l'équation non homogène, nous appliquons la méthode de variation des constantes, en imposant les $n-1$ conditions

$$\sum_{i=1}^{n} c_i'(t) y_i^{(l)}(t) = 0 \tag{31.72}$$

avec $l = 0, \ldots, n-2$. Ces conditions plus l'équation de départ (31.68) forment un système de n équations différentielles pour les n fonctions inconnues $c_i(t)$.

Cette condition peut paraitre mystérieuse. Il est cependant encore possible de travailler sans poser la condition (31.72) en suivant la recette, en calculant des déterminants de Wronskien. Des exemples sont donnés dans les exercices sur le second ordre.

31.5.2 Si les coefficients ne sont pas constants ?

Une équation différentielle linéaire d'ordre n sur I est une équation de la forme

$$y^{(n)} + u_1(t) y^{(n-1)} + \cdots + u_{n-1}(t) y' + u_n(t) y = v(t) \tag{31.73}$$

où v et u_k sont des fonctions continues fixées de I vers $\mathbb{R}$.

Pour résoudre cette équation, il faut commencer par résoudre l'équation homogène correspondante (c'est-à-dire celle que l'on obtient en posant $v(t) = 0$). Ensuite, nous trouvons la solution de l'équation (31.73) en appliquant la méthode de la **variation des constantes**.

Donnons un exemple du pourquoi la méthode de variations des constantes est efficace. Soit l'équation

$$u' + f(t) u = g(t), \tag{31.74}$$

et disons que u_H est une solution de l'équation homogène. La méthode de variations des constantes consiste à poser $u(t) = K(t) u_H(t)$, et donc $u'(t) = K' u_H + K u_H'$. En remettant dans l'équation de départ,

$$K' u_H + K u_H' + f K u_H = g. \tag{31.75}$$

La somme $K u_H' + f K u_H$ est nulle, par définition de u_H. Par conséquent, il ne reste que

$$K' = \frac{g(t)}{u_H}. \tag{31.76}$$

Lorsqu'on utilise la méthode de variation des constantes, nous trouvons toujours une simplification « miraculeuse ».

Dans l'immédiat, nous ne considérons que le cas où les u_i sont des constantes. Le cas où les u_i deviennent des fonctions de t sera vu plus tard.

31.6 Système d'équations linéaires

31.6.1 La magie de l'exponentielle...

Prenons l'équation différentielle très simple

$$y' = ay. \tag{31.77}$$

La solution est $y(t) = Ae^{at}$. Et si on a la donnée de Cauchy $y(t_0) = y_0$, alors

$$y(t) = Ae^{at}e^{-at_0}e^{at_0} = e^{a(t-t_0)}y(t_0). \tag{31.78}$$

Donc on a le facteur multiplicatif $e^{a(t-t_0)}$ qui sert à faire passer de $y(0)$ à $y(t)$. C'est un peu un opérateur d'évolution. Ce qui fait la magie de l'exponentielle, c'est son développement en série

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots \tag{31.79}$$

qui est tel que chaque terme est la dérivée du terme suivant.

Maintenant, si on a un système

$$\bar{y}' = A\bar{y}, \tag{31.80}$$

il n'est pas du tout étonnant d'avoir comme solution $\bar{y}(t) = e^{At}$ où l'exponentielle de la matrice est définie exactement par la série (31.79). C'est un peu longuet, mais dans le cours, c'est effectivement ce qui est prouvé. La matrice résolvante $R(t, t_0)\colon \bar{y}_0 \to \bar{y}(t; t_0, y_0)$ est donné par

$$R(t, t_0) = e^{(t-t_0)A}, \tag{31.81}$$

exactement comme dans l'équation (31.78).

31.6.2 ...mais la difficulté

Maintenant, il est suffisant de calculer des exponentielles de matrices pour résoudre des systèmes. Hélas, il est en général très difficile de calculer des exponentielles. Tu peux essayer de prouver les deux suivantes :

$$\begin{aligned} A &= \begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix} \rightsquigarrow e^A = \begin{pmatrix} \cos(a) & \sin(a) \\ -\sin(a) & \cos(a) \end{pmatrix} \\ S &= \begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix} \rightsquigarrow e^S = \begin{pmatrix} \cosh(a) & \sinh(a) \\ \sinh(a) & \cosh(a) \end{pmatrix}. \end{aligned} \tag{31.82}$$

La première, tu vas la revoir si tu fais de la géométrie différentielle ou de la mécanique quantique : l'algèbre de Lie du groupe des matrices orthogonales de déterminant 1 est l'algèbre des matrices antisymétriques.

La seconde se retrouve en relativité parce que e^S est la matrice qui préserve $x^2 - y^2$, tout comme e^A préserve $x^2 + y^2$. Quelques mots sur l'uutilisation des fonctions hyperboliques en relativité dans 41.4.8.1.

31.6.3 La recette

Afin d'éviter de devoir calculer explicitement des exponentielles de matrices, nous faisons appel à toutes sortes de trucs, dont la forme de Jordan. Le résultat final est la méthode suivante. Soit le système homogène

$$\bar{y}' = A\bar{y}. \tag{31.83}$$

(1) D'abord, nous calculons les valeurs propres de A.

(2) Ensuite les vecteurs propres.

(3) Une bonne valeur propre, c'est une valeur propre dont l'espace propre a une dimension égale à sa multiplicité. C'est-à-dire que si λ est de multiplicité m, alors on a, dans les bons cas, m vecteur propres linéairement indépendants.

Dans ce cas, si $v_1, \dots, v_m$ sont les vecteurs, alors on a les solutions linéairement indépendantes suivantes :

$$\begin{pmatrix} \vdots \\ v_1 \\ \vdots \end{pmatrix} e^{\lambda t}, \dots, \begin{pmatrix} \vdots \\ v_m \\ \vdots \end{pmatrix} e^{\lambda t}. \tag{31.84}$$

Pour chaque bonne valeur propre, ça nous fait un tel paquet de solutions linéairement indépendantes.

(4) Si λ n'est pas une bonne valeur propre, alors les choses se compliquent. Mettons que λ ait k vecteurs propres en moins que sa multiplicité. Dans ce cas, il faut chercher des solutions sous la forme

$$\begin{pmatrix} a_1^{(k)} t^k + \dots + a_1^{(0)} \\ \vdots \\ a_1^{(k)} t^k + \dots + a_n^{(0)} \end{pmatrix} e^{\lambda t}. \tag{31.85}$$

C'est-à-dire qu'on prend comme coefficient de $e^{\lambda t}$, un vecteur de polynômes de degré k. Il faut mettre cela dans l'équation de départ pour voir quelles sont les contraintes sur les constantes $a_i^{(j)}$ introduites.

(5) Nous avons un cas particulier du cas précédent. Si λ est une valeur propre de multiplicité m qui n'a que un seul vecteur propre v, alors il faut chercher des polynômes de degré $m-1$, et on peut directement fixer le coefficient de t^{m-1}, ce sera l'unique vecteur propre :

$$\left[\begin{pmatrix} \vdots \\ v \\ \vdots \end{pmatrix} + \begin{pmatrix} a_1^{(m-2)} \\ \vdots \\ a_n^{(m-2)} \end{pmatrix} t^{m-2} + \dots \right] e^{\lambda t}. \tag{31.86}$$

Cela économise quelques calculs par rapport à poser brutalement (31.85).

31.6.4 Système d'équations linéaires avec matrice constante

Nous considérons l'équation différentielle

$$y'(t) = Ay(t) \tag{31.87}$$

pour la fonction $y\colon \mathbb{R} \to \mathbb{R}^n$ et A est une matrice ne dépendant pas de t. Nous supposons que A est diagonalisable pour les vecteurs propres v_i et les valeurs propres λ_i correspondantes.

La matrice

$$R(t) = \left[e^{\lambda_1 t} v_1 \dots e^{\lambda_n t} v_n \right] \tag{31.88}$$

est la **matrice résolvante** du système. Alors la solution du système (31.87) pour la condition initiale $y(0) = y_0$ est

$$y(t) = R(t) y_0. \tag{31.89}$$

En effet

$$AR(t) = \left[A \begin{pmatrix} \uparrow \\ e^{\lambda_1 t} v_1 \\ \downarrow \end{pmatrix} \dots A \begin{pmatrix} \uparrow \\ e^{\lambda_n t} v_n \\ \downarrow \end{pmatrix} \right] = R'(t). \tag{31.90}$$

Par conséquent $y'(t) = R'(t) y_0 = AR(t) y_0 = Ay(t)$.

31.6.5 Système d'équations linéaires avec matrice non constante

Théorème 31.14 ([488]).
Soient I un intervalle de $\mathbb{R}$ et $M\colon \mathbb{R} \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ une fonction. Si les composantes M_{ij} sont des fonctions continues sur I alors :

(1) pour tout $t_0 \in I$ et pour tout $y_0 \in R^n$ le système

$$y'(t) = M(t)y(t) \tag{31.91}$$

admet une unique solution maximale définie sur I telle que $y(t_0) = y_0$;

(2) l'ensemble des solutions de l'équation (31.91) sur I est un espace vectoriel de dimension n.

31.7 Réduction de l'ordre

Afin de diminuer l'ordre d'une équation dans laquelle le paramètre n'apparaît pas, il y a deux changements de variables très utiles. Le premier, le plus simple, est simplement de poser $z(t) = y'(t)$, ce qui donne $z'(t) = y''(t)$. Le second, *qui n'est pas le même*, est $z\big(y(t)\big) = y'(t)$, qui entraîne $y''(t) = z'\big(y(t)\big)z(t)$. Dans ce second cas, il faut également changer de variable, et utiliser $y(t)$ comme variable au lieu de t.

Si ça ne marche pas, il faut suivre la procédure ci-après.

Nous supposons avoir une équation différentielle d'ordre p dans laquelle $y^{(p)}$ est isolée des autres dérivées :

$$y^{(p)}(t) = f\big(t, y(t), y'(t), \dots, y^{(p-1)}(t)\big) \tag{31.92}$$

où f est une fonction $f\colon \mathbb{R} \times \mathbb{R}^p \to \mathbb{R}$, et la fonction cherchée est $y\colon \mathbb{R} \to \mathbb{R}$.

La méthode proposée ici consiste à transformer cette équation d'ordre p en un système d'équations d'ordre 1. Pour cela nous posons

$$\begin{aligned} F\colon \mathbb{R} \times \mathbb{R}^p &\to \mathbb{R}^p \\ (t, x) &\mapsto \begin{pmatrix} x_2 \\ \vdots \\ x_p \\ f(t, x_1, \dots, x_p) \end{pmatrix}. \end{aligned} \tag{31.93}$$

Nous considérons alors l'équation différentielle

$$Y'(t) = F\big(t, Y(t)\big). \tag{31.94}$$

pour $Y\colon \mathbb{R} \to \mathbb{R}^p$. La fonction $y = Y_1$ résout l'équation (31.92) si et seulement si la fonction Y résout l'équation (31.94).

De plus si l'équation (31.92) est donnée avec les conditions initiales $y^{(k)} = a_k$ $(k = 0, \dots, p-1)$ alors l'équation (31.94) vient avec les conditions initiales

$$Y(t_0) = \begin{pmatrix} a_0 \\ \vdots \\ a_{p-1} \end{pmatrix}, \tag{31.95}$$

c'est-à-dire $Y(t_0) = A_0$ avec $A_0 \in \mathbb{R}^p$.

Le théorème de Cauchy-Lipschitz 16.40 nous donne existence et unicité locale de la solution au système 31.94. Lorsque le système est linéaire, c'est-à-dire sous la forme $Y'(t) = M(t)Y(t)$, alors il y a mieux : le théorème 31.14.

Exemple 31.15.
Nous considérons l'équation différentielle

$$\begin{cases} -u''(t) - u(t) = 1 & \text{(31.96a)} \\ u(0) = a_0 \in \mathbb{R} & \text{(31.96b)} \\ u'(0) = a_1 \in \mathbb{R}, & \text{(31.96c)} \end{cases}$$

et nous voulons montrer que ce système accepte une unique solution. Vu que l'équation différentielle se présente sous la forme $u'' = f(t,u)$ avec $f(t,u) = -1-u$ nous posons

$$\begin{aligned} F\colon \mathbb{R}\times\mathbb{R}^2 &\to \mathbb{R}^2 \\ (t,x) &\mapsto \begin{pmatrix} x_2 \\ -1-x_1 \end{pmatrix}, \end{aligned} \tag{31.97}$$

et nous considérons l'équation différentielle

$$Y' = F(t,Y) \tag{31.98}$$

pour la fonction $Y\colon \mathbb{R}\to\mathbb{R}^2$. La fonction F est Lipschitz (et même globalement) par rapport à Y. En effet,

$$\|F(x)-F(y)\| = (x_2-y_2)^2 + (x_1-y_1)^2 = \|x-y\|^2. \tag{31.99}$$

Le théorème de Cauchy-Lipschitz 16.40 s'applique et en posant $Y_0 = (a_0,a_1)$, il existe une unique solution à l'équation $Y' = F(t,Y)$ vérifiant $Y(0) = Y_0$. Nous notons $t\mapsto Y(t)$ cette solution.

En quoi cela nous aide ? Nous posons $u(t) = Y_1(t)$. Alors

$$u'(t) = Y_1'(t) = F_1'(t,Y) = Y_2(t). \tag{31.100}$$

En dérivant encore,

$$u''(t) = Y_2'(t) = F_2'(t,Y) = -1-Y_1(t) = -1-u(t), \tag{31.101}$$

ce qu'il fallait. La fonction $t\mapsto Y_1(t)$ est solution de notre équation de départ. Quid des conditions initiales ? Vu que $u = Y_1$ et $u' = Y_2$ nous avons

$$u(0) = Y_1(0) = a \tag{31.102}$$

et

$$u'(0) = Y_2(0) = b. \tag{31.103}$$

Toutes les prescriptions sont respectées.

Si vous voulez vraiment résoudre cette équation, il faudra plus de travail. D'abord résoudre l'équation homogène associée, c'est-à-dire l'équation caractéristique $r^2+1=0$, ce qui va donner

$$u_H(t) = Ae^{it} + Be^{-it}, \tag{31.104}$$

et ensuite faire le coup de la variation des constantes pour déterminer la solution générale du problème non homogène. △

Exemple 31.16.
Nous reprenons l'équation différentielle

$$-u''(t) - u(t) = 1. \tag{31.105}$$

Nous avons déjà vu dans l'exemple 31.15 que cette équation avait une solution unique pour toute condition initiale. Cette fois nous voulons étudier les solutions lorsque nous imposons les conditions aux limites $u(0) = u(\pi) = 0$. Nous allons voir qu'il n'y a pas de telles solutions.

Pour ce faire, soit une solution u. D'abord u'' existant, la fonction u est de classe au moins C^1. Mais $u'' = -1-u$, donc u'' est également C^1, ce qui donne la régularité C^3 pour u. En continuant ainsi nous trouvons que u est de classe C^∞.

Le truc est de considérer a fonction $v(t) = \sin(t)$ qui vérifie l'équation différentielle

$$\begin{cases} -v''-v = 0 & \text{(31.106a)} \\ v(0) = 0 & \text{(31.106b)} \\ v'(0) = 0. & \text{(31.106c)} \end{cases}$$

Nous calculons le produit scalaire sur $L^2\big(]0,\pi[\big)$ de (31.105) avec v :

$$\langle u'', v\rangle + \langle u, v\rangle = -\langle v, 1\rangle. \tag{31.107}$$

Le calcul de $\langle v, 1\rangle$ est simplement l'intégrale de $\sin(t)$ pour t allant de 0 à π, c'est-à-dire $\langle v, 1\rangle = 2$. Vu que u et v sont toutes deux des fonctions qui s'annulent en 0 et en π nous pouvons faire des intégrations par partie les yeux fermés et exprimer $\langle u'', v\rangle$ sans dérivées sur u :

$$\langle u'', v\rangle = -\langle u', v'\rangle = \langle u, v''\rangle = -\langle u, v\rangle \tag{31.108}$$

où la dernière égalité n'est autre que le fait que $v = \sin$, donc $v'' = -v$. Le membre de gauche de (31.107) vaut donc zéro alors que celui de droite vaut -2.

Nous concluons que le problème aux limites posé n'admet pas de solutions. △

31.8 Autour de Cauchy-Lipschitz

Dans cette section nous étudions les équations différentielles du type

$$\begin{cases} y'(t) = f\big(t, y(t)\big) & \text{(31.109a)} \\ y(t_0) = y_0. & \text{(31.109b)} \end{cases}$$

31.8.1 Fuite des compacts et explosion en temps fini

Théorème 31.17 (Fuite des compacts[489, 321])**.**
Nous considérons l'équation différentielle

$$\begin{cases} y'(t) = f\big(t, y(t)\big) & \text{(31.110a)} \\ y(t_0) = y_0, & \text{(31.110b)} \end{cases}$$

où $f\colon I \times \Omega \to \mathbb{R}^n$ est continue et Ω ouvert dans $\mathbb{R}^n$. Soit la solution maximale $y_M\colon J_M =]t_{min}, t_{max}[\to \Omega$. Si $t_{max} < \sup(I)$ alors $y_M(t)$ sort de tout compact de Ω lorsque $t \to t_{max}$.

Démonstration. Soit K un compact de Ω et nous considérons une suite (t_m) dans $]t_{min}, t_{max}[$ telle que $t_m \to t_{max}$. Si nous supposons que $y_M(t)$ ne sort pas de K alors nous avons $y_M(t_m) \in K$, c'est-à-dire une suite dans un compact. Quitte à passer à une sous-suite, nous supposons qu'elle est convergente. Soit $x_1 \in K$ la limite $\lim_{m\to\infty} y_M(t_m) = x_1$.

Vu que $t_{max} \in I$, la condition initiale $y(t_{max}) = x_1$ est valide et le théorème de Cauchy-Lipschitz 16.40 nous donne une unique solution maximale y_P définie sur un ouvert J_P autour de t_{max}.

Nous allons maintenant construire une solution au problème initial qui contredit la maximalité de y_M. Attention : il n'est pas évident à priori que $y_P(t) = y_M(t)$ sur l'intersection des domaines. Si c'était évident, la proposition serait démontrée.

Soit $\tilde{J} = J_M \cup J_P \cap]t_{min}, +\infty[$ et la fonction

$$\tilde{y}(t) = \begin{cases} y_M(t) & \text{si } t < t_{max} \\ y_P(t) & \text{si } t \geqslant t_{max}. \end{cases} \tag{31.111}$$

La fonction $\tilde{y}$ est continue par construction parce que

$$\lim_{t\to t_{max}} y_M(t) = x_1 = y_P(t_{max}). \tag{31.112}$$

Nous vérifions à présent que $\tilde{y}$ est une solution : $\tilde{y}'(t_{max}) = f\big(t_{max}, y(t_{max})\big)$:

$$\lim_{\epsilon\to 0}\frac{\tilde{y}(t_{max}-\epsilon)-\tilde{y}(t_{max})}{\epsilon} = \lim_{\epsilon\to 0}\frac{y_M(t_{max}-\epsilon)-y_P(t_{max})}{\epsilon} \tag{31.113a}$$

$$= \lim_{\epsilon\to 0}\frac{y_M(t_{max}-\epsilon)-y_P(t_{max}-\epsilon)+y_P(t_{max}-\epsilon)-y_P(t_{max})}{\epsilon} \tag{31.113b}$$

$$= \lim_{\epsilon\to 0}\frac{y_P(t_{max}-\epsilon)-y_P(t_{max})}{\epsilon} \tag{31.113c}$$

$$= y_P'(t_{max}). \tag{31.113d}$$

Donc $\tilde{y}$ est solution pour la condition initiale $\tilde{y}(t_{max}) = x_1$ et coïncide avec y_P en t_{max} et avec y_M avant t_{max}. Donc en réalité y_P, y_M et $\tilde{y}$ sont identiques et cela contredit la maximalité de y_M. □

Corolaire 31.18 (Explosion en temps fini)**.**
Soit (y_m, J) la solution maximale du problème de Cauchy (16.145) :

$$\begin{cases} y' = f(t,y) & (31.114a)\\ y(t_0) = y_0, & (31.114b)\end{cases}$$

avec $f\colon U = I\times\Omega\to\mathbb{R}^n$ où I est ouvert dans $\mathbb{R}$ et Ω ouvert dans $\mathbb{R}^n$. Nous supposons que f est continue sur U et localement Lipschitz par rapport à y.

Si la solution maximale est définie sur $J =]t_{min}, t_{max}[$ alors nous avons l'alternative suivante :

(1) Soit $t_{max} = \sup(I)$,

(2) soit $t_{max} < \sup(I)$ et $\lim_{t\to t_{max}}\|y(t)\| = \infty$.

Le résultat tient aussi mutatis mutandis *pour $t_{\min}$.*

Remarque 31.19.
Attention : ceci n'est pas une simple paraphrase de la fuite des compacts. L'information supplémentaire que ce corolaire donne est que la solution sort de tout compact *pour ne plus y retourner.*

Démonstration. L'hypothèse $t_{max} < \sup(I)$ signifie que la solution finit d'exister avant que les hypothèses sur f cessent d'être vraies. C'est-à-dire que la solution maximale est moindre que ce que nous aurions pu espérer.

Soit un compact K. Supposons que que pour tout $t_0 < t_{max}$ il existe $t\in]t, t_{max}[$ tel que $y_M(t)\in K$. Alors cela crée une suite t_k dans J telle que $y_M(t_k)$ est dans K. Comme dans le théorème de la fuite des compacts nous concluons l'impossibilité de la chose.

Donc pour tout compact K de Ω, il existe $T < t_{max}$ tel que $y_M(t)\in\Omega\backslash K$ pour tout $t\in[T, t_{max}[$. En prenant des boules fermées de plus en plus grandes en guise de compacts nous concluons que

$$\lim_{t\to t_{max}}\|y_M(t)\| = \infty. \tag{31.115}$$

□

31.20.
Notons que si $t_{max} < \infty$, si nous sommes dans l'alternative 31.18(2) et si la solution maximale y est de classe C^1 (ce qui est le cas lorsqu'on utilise Cauchy-Lipschitz 16.40) alors la dérivée de y est également non bornée dans un voisinage de t_{max}.

Mais si f est globalement bornée, alors dans l'équation $y' = f(t,y)$, la dérivée y' sera globalement bornée. Dans ce cas, la solution ne peut pas exploser en temps fini et existe donc globalement.

Problèmes et choses à faire

Êtes-vous d'accord avec 31.20 ?

Exemple 31.21.
Soit l'équation différentielle

$$\begin{cases} y' = y(y-1)\sin(yt) & (31.116a) \\ y(0) = \frac{1}{2}. & (31.116b) \end{cases}$$

La fonction $f(t,y) = y(y-1)\sin(yt)$ ayant une dérivée bornée partout, elle est localement Lipschitz et le théorème de Cauchy-Lipschitz 16.40 s'applique. Pour toute condition initiale, une solution maximale unique existe.

Si nous oublions la condition initiale, il est facile de trouver des solutions constantes : $y' = 0$ avec $y(t) = k$ donne l'équation

$$0 = k(k-1)\sin(kt). \tag{31.117}$$

Les solutions $y_1(t) = 0$ et $y_2(t) = 1$ sont des solutions existant pour tout t.

Le graphe de la solution correspondante à la condition initiale $y(0) = \frac{1}{2}$ ne pouvant pas croiser les graphes de y_1 et y_2, elle est obligée d'exister pour tout t parce qu'elle ne peut pas exploser en temps fini. △

31.8.2 Écart entre deux conditions initiales

Proposition 31.22 ([487, 490])**.**
Soit une fonction $f\colon I \times \Omega \to \mathbb{R}^n$ continue et globalement Lipschitz en sa seconde variable (Ω est un ouvert de $\mathbb{R}^n$). Soient deux solutions $y_1\colon I_1 \to \mathbb{R}^n$ et $y_2\colon I_2 \to \mathbb{R}^n$ aux problèmes

$$\begin{cases} y_i'(t) = f\big(t, y_i(t)\big) & (31.118a) \\ y_i(t_0) = a_i. & (31.118b) \end{cases}$$

Alors pour tout $t \in I_1 \cap I_2$ nous pouvons estimer l'écart entre y_1 et y_2 par la formule

$$\|y_1(t) - y_2(t)\| \leqslant e^{L|t-t_0|}\|a_1 - a_2\|, \tag{31.119}$$

où L est la constante de Lipschitz de f.

Démonstration. Nous avons d'abord les majorations suivantes, qui semblent juste jouer avec les notations, mais qui utilisent le fait (contenu dans le théorème de Cauchy-Lipschitz) que y_i soit de classe C^1 :

$$\begin{aligned} \|y_1(t) - y_2(t)\| &= \| \int_{t_0}^{t} \big(y_1'(s) - y_1'(s)\big)\| & (31.120a) \\ &\leqslant L \int_{t_0}^{t} \|f\big(s, y_1(s)\big) - f\big(s, y_2(s)\big)\| ds & (31.120b) \\ &= L \int_{t_0}^{t} \|y_1(s) - y_2(s)\| ds. & (31.120c) \end{aligned}$$

C'est à ce moment que nous utilisons le lemme de Grönwall. Vu que

$$\|y_1(t) - y_2(t)\| \leqslant L \int_{t_0}^{t} \|y_1(s) - y_2(s)\| ds, \tag{31.121}$$

nous sommes dans les hypothèses de Grönwall 31.5 en posant

$$\begin{aligned} u(t) &= \|y_1(t) - y_2(t)\| & (31.122a) \\ b(t) &= \|y_1(0) - y_2(0)\| & (31.122b) \\ a(t) &= L. & (31.122c) \end{aligned}$$

Nous avons la majoration

$$\|y_1(t) - y_2(t)\| \leqslant \|y_1(0) - y_2(0)\| + L \int_0^t \|y_1(0) - y_2(0)\| e^{L(t-s)} ds. \tag{31.123}$$

Le calcul de l'intégrale intérieure donne

$$\int_0^t e^{L(t-s)} ds = -\frac{1}{L}(e^{-Lt} - 1). \tag{31.124}$$

Avec ça, nous avons

$$\|y_1(t) - y_2(t)\| \leqslant e^{Lt} \|y_1(0) - y_2(0)\|. \tag{31.125}$$

□

31.23.
Notons que la proposition 31.22 est plutôt une mauvaise nouvelle parce que les solutions restent seulement linéairement proches l'une de l'autre lorsqu'on rapproche les conditions initiales, mais elle divergent exponentiellement vite avec le temps. Donc deux trajectoires arbitrairement proches au départ finissent assez vite par être bien séparées.

Cette proposition est cependant cruciale parce qu'elle explique que pour des petits t, les solutions ne s'écartent pas beaucoup, c'est-à-dire que pour t fixé, l'application qui à une donnée initiale fait correspondre la solution en t est continue. C'est le premier pas pour parler de régularité du flot.

31.8.3 Flot d'un champ de vecteurs

Nous reprenons l'équation différentielle du théorème de Cauchy-Lipschitz 16.40. En ce qui concerne les notations, I est un intervalle ouvert de $\mathbb{R}$ contenant 0 et l'application $f\colon I \times \mathbb{R}^n \to \mathbb{R}^n$ est continue et localement Lipschitz en sa seconde variable. Pour $a \in \mathbb{R}^n$, nous notons (J_a, y_a) la solution maximale (donc $y_a\colon J_a \to \mathbb{R}^n$) du problème

$$\begin{cases} y_a(t) = f\big(t, y_a(t)\big) & \text{(31.126a)} \\ y_a(0) = a. & \text{(31.126b)} \end{cases}$$

Nous noterons aussi de temps en temps $\varphi(t, a) = y_a(t)$.

Nous savons que $t \mapsto y_a(t)$ est de classe C^1, et cela est directement dans le théorème de Cauchy-Lipschitz. Une question d'une toute autre difficulté est la régularité de $a \mapsto y_a(t)$ pour t fixé, et encore pire : celle de $(t, a) \mapsto y_a(t)$.

Il se fait que l'application $(t, a) \mapsto y_a(t)$ a la même régularité que celle de f, mais cela va être un peu long à prouver. En ce qui concerne la régularité C^1, ce sera le théorème 31.28 dont la démonstration, comme vous pouvez le voir sera copieuse et demandera des propositions intermédiaires pas simples.

Définition 31.24.
Si t est fixé, l'application

$$\begin{aligned} \varphi_t\colon \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto \varphi(t, x) = y_x(t), \end{aligned} \tag{31.127}$$

*est le **flot** du problème de Cauchy (31.126).*

L'application $t \mapsto \varphi_t$ est ce qui est appelé le groupe à un paramètre de flot, pour des raisons qui arriverons plus tard[2].

Le but est d'étudier les propriétés du flot : est-il continu, un difféomorphisme, existe, pour quels t ? Où se cache le champ de vecteurs du titre dans l'équation différentielle ?

2. ou pas...

Nous posons

$$\mathcal{D} = \bigcup_{x\in\Omega} \big(J_x \times \{x\}\big). \tag{31.128}$$

Comme tout produit d'espaces métrique, l'ensemble $\mathcal{D}$ est muni d'une métrique via la définition 7.229.

Proposition 31.25 ([487]).
Soit un intervalle I ouvert de $\mathbb{R}$ contenant 0 et Ω un ouvert connexe de $\mathbb{R}^n$. Soit une application $f\colon I\times\mathbb{R}^n \to \mathbb{R}^n$ continue et localement Lipschitz en sa seconde variable. Pour $a\in\mathbb{R}^n$, nous notons (J_a, y_a) la solution maximale (donc $y_a\colon J_a\to\mathbb{R}^n$) du problème

$$\begin{cases} y_a(t) = f\big(t, y_a(t)\big) & (31.129a)\\ y_a(0) = a. & (31.129b)\end{cases}$$

Nous posons

$$\mathcal{D} = \bigcup_{x\in\Omega} \big(J_x \times \{x\}\big). \tag{31.130}$$

Nous définissons la fonction φ par $\varphi(t,x) = y_x(t)$ là où ça existe.

L'ensemble $\mathcal{D}$ est ouvert. L'application $\varphi\colon \mathcal{D}\to\Omega$ est localement Lipschitz.

Démonstration. Soit $(s,a)\in\mathcal{D}$ et (J_a, y_a) la solution maximale passant par a en $t=0$. Par définition de $\mathcal{D}$ nous avons $s\in J_a$. Nous considérons J, un compact inclus dans J_a et contenant 0 et s en son intérieur. Nous posons

$$K = J\times y_a(J). \tag{31.131}$$

Vu que y_a est continue, cela est un compact. Chaque point de K possède un voisinage ouvert sur lequel f est Lipschitz [3] ; nous considérons un sous recouvrement fini et le maximum des constantes de Lipschitz. Cela nous crée un voisinage V de K dans $I\times\Omega$ dans lequel f est Lipschitz.

Vu que V est ouvert et K est compact avec $K\subset V$, nous pouvons trouver un ouvert V' et un compact K' tels que

$$K\subset V'\subset K'\subset V. \tag{31.132}$$

Sur ce V', la fonction f est de plus bornée parce que continue sur le compact K'. Nous renommons V' en V. Sur V nous avons :

— $\|f\|_{\infty,V}\leqslant M$,

— f est Lipschitz en sa seconde variable, de constante de Lipschitz L.

En tant qu'espace produit, nous avons une distance sur $I\times\Omega$ donnée en 7.229 :

$$d\big((t,y),(t',y')\big) = \max\big\{|t-t'|, \|y-y'\|\big\}. \tag{31.133}$$

Nous posons

$$V_\epsilon(K) = \{z\in I\times\Omega \text{ tel que } d(z,K)<\epsilon\} \tag{31.134a}$$
$$W_\epsilon = \{(t,y)\in J\times\Omega \text{ tel que } \|y-y_a(t)\|<\epsilon\}. \tag{31.134b}$$

$\overline{W_\epsilon}\subset\overline{V_\epsilon(K)}$ Soit $(t,y)\in W_\epsilon$. Nous avons :

$$d\big((t,y),K\big) = \inf_{(t',y')\in K} d\big((t,y),(t',y')\big) \tag{31.135a}$$
$$= \inf_{(t',y')\in K} \max\{|t-t'|, \|y-y'\|\}. \tag{31.135b}$$

Mais demander $(t,y)\in W_\epsilon$ signifie que $t\in J$ et $\|y-y_a(t)\|\leqslant\epsilon$. Dans K nous avons l'élément $\big(t, y_a(t)\big)$ qui vérifie

$$d\big((t,y),(t,y_a(t))\big) = \|y-y_a(t)\|\leqslant\epsilon. \tag{31.136}$$

Donc l'infimum de (31.135b) est majoré par ϵ. Nous avons prouvé que $W_\epsilon\subset V_\epsilon(K)$ et donc même inclusion pour les fermetures.

3. Cela est à peu près la définition d'être localement Lipschitz : 11.262, voir aussi 11.263.

Il existe $\epsilon > 0$ tel que $\overline{V_\epsilon(K)} \subset V$ Supposons que $\overline{V_\epsilon(K)}$ ne soit inclus dans V pour aucun ϵ. Alors nous considérons

$$z_n \in \overline{V_{1/n}(K)} \backslash V. \tag{31.137}$$

Nous avons par définition $d(z_n, K) \leqslant \frac{1}{n}$. Vu que K est compact, il comprend (au moins) un élément réalisant la distance : soit $z'_n \in K$ tel que

$$d(z_n, z'_n) = d(z_n, K). \tag{31.138}$$

Nous avons $d(z_n, z'_n) \to 0$, de telle sorte que les valeurs d'adhérence de (z_n) et (z'_n) sont les mêmes. Et comme (z'_n) est une suite dans un compact, elle a des valeurs d'adhérence. Soit z_∞ l'une d'elles. Vu que c'est une valeur d'adhérence d'une suite contenue dans le compact K, elle est également dans K : $z_\infty \in K$. Mais en même temps, z_n est hors de l'ouvert V, et donc dans le fermé V^c. Les valeurs d'adhérences restent dans le fermé, c'est-à-dire $z_\infty \notin V$. Vu que $K \subset V$, il y a contradiction.

Donc il existe $\epsilon > 0$ tel que $\overline{V_\epsilon(K)} \subset V$.

Il existe ϵ tel que $\overline{W_\epsilon} \subset V$ Il suffit de prendre le ϵ dont nous venons de parler pour avoir

$$\bar{\epsilon} \subset \overline{V_\epsilon(K)} \subset V. \tag{31.139}$$

Soit le ϵ en question, et $T > 0$ tel que $J \subset [-T, T]$. Nous posons $r = \epsilon e^{-LT}$. Soit $b \in \overline{B(a, r)}$ et

$$X = \{\tau \in J_+ \text{ tel que }]0, \tau] \subset J_b \text{ et } \big(t, y_b(t)\big) \in \overline{W_\epsilon}\, \forall\, t \in [0, \tau]\}. \tag{31.140}$$

Nous allons prouver que $X = J_+$ en prouvant qu'il est ouvert, fermé et non vide dans $J_+ = J \cap]0, \infty[$. Nous parlons bien de la topologie de J_+, celle induite[4] de $\mathbb{R}$. Vu que $0 \in J$, l'ensemble J_+ est ouvert à gauche, mais comme il est compact, il ne va certainement pas jusqu'à $+\infty$, de telle sorte qu'il est fermé à gauche. Les ouverts de J_+ sont les ensembles de la forme $\mathcal{O} \cap J_+$ où $\mathcal{O}$ est ouvert de $\mathbb{R}$. Il y en a de la forme $]0, m]$.

X est fermé C'est parce que $\overline{W_\epsilon}$ et J_b sont fermés.

X est ouvert Soit $\tau \in X$. Si $\tau = \sup J_+$ alors $X = J_+$ est un ouvert de J_+. Supposons donc que $0 < \tau < \sup J_+$. Dans ce cas nous avons

$$\big(\tau, y_b(\tau)\big) \in \overline{W_\epsilon} \subset V, \tag{31.141}$$

et nous pouvons résoudre localement le problème de Cauchy

$$\begin{cases} y'(t) = f\big(t, y(t)\big) & \text{(31.142a)} \\ y(\tau) = \varphi(\tau, b) = y_b(\tau). & \text{(31.142b)} \end{cases}$$

Ce y existe jusqu'à $\tau + \eta$ (pour au moins un petit η), et par l'unicité de la solution, $y = y_b$ sur $[\tau, \tau + \eta[$. Ceci pour dire que le flot $\varphi(., b)$ existe au moins jusqu'à $\tau + \eta$.

Grâce à la proposition 31.22 nous pouvons évaluer

$$\|\varphi(\tau, b) - \varphi(\tau, a)\| = \|y_a(\tau) - y_b(\tau)\| \leqslant e^{L\tau}\|b - a\|. \tag{31.143}$$

Comme nous avions choisi $r = \epsilon e^{-LT}$ et $b \in \overline{B(a, r)}$ nous avons aussi $\|b - a\| \leqslant \epsilon e^{-LT}$ et donc

$$\|\varphi(\tau, b) - \varphi(\tau, a)\| \leqslant \epsilon e^{L(\tau - T)} < \epsilon \tag{31.144}$$

parce que nous avions $\tau < \sup J_+ \leqslant T$, ce qui garantit que $e^{L(\tau - T)} < 1$.

Est-ce que ceci nous garantit que $\tau + \eta \in X$? Il faudrait $\big(\tau + \eta, y_b(\tau + \eta)\big) \in \overline{W_\epsilon}$, c'est-à-dire $\|y_b(\tau + \eta) - y_a(\tau + \eta)\| \leqslant \epsilon$. L'ensemble J_+ étant fermé dans l'ouvert J_a, ce dernier déborde certainement. Prenons donc η assez petit pour que y_a existe jusqu'en $\tau + \eta$.

Vu que y_a et y_b sont continues, et qu'en τ elles sont distantes de moins de ϵ, en $\tau + \eta$, elles restent distantes de moins de ϵ (quitte à prendre encore η plus petit).

Ceci nous permet de conclure que X est ouvert.

4. Définition 7.15.

X est non vide La solution y_b au problème

$$\begin{cases} y_b'(t) = f\big(t, y_b(t)\big) & (31.145a) \\ y_b(0) = b & (31.145b) \end{cases}$$

existe au moins localement et vérifie $\|y_b(0) - y_a(0)\| = \|b - a\| \leqslant \epsilon e^{-LT} < \epsilon$. Par continuité nous avons

$$\|y_b(t) - y_a(t)\| < \epsilon \tag{31.146}$$

pour tout t dans un voisinage de 0. Donc X est non vide.

Conclusion pour X La partie X est ouverte, fermée et non vide dans J_+ qui est connexe. Donc $X = J_+$ par la proposition 7.44(3).

La conclusion $X = J_+$ nous enseigne que pour tout $t \in J_+$ nous avons $]0, t] \in J_b$ et $\big(t, y_b(t)\big) \in \overline{W_\epsilon}$. Nous pouvons faire la même chose pour J_- et au final nous avons que pour tout $\tau \in J$ nous avons d'abord $\tau \in J_b$, ce qui prouve $J \subset J_b$. De plus pour tout $t \in J$ nous avons aussi

$$\big(t, y_b(t)\big) \in \overline{W_\epsilon} \subset V. \tag{31.147}$$

Nous en concluons que

$$J \times \overline{B(a, r)} \subset V. \tag{31.148}$$

Nous savons de plus que pour tout $b \in \overline{B(a, r)}$, $J \subset J_b$. Cela signifie que

$$J \times \overline{B(a, r)} \subset \mathcal{D}. \tag{31.149}$$

Mais $J \times \overline{B(a, r)}$ est un voisinage de (s, a) qui était au début de la preuve un point générique choisi dans $\mathcal{D}$. Donc $\mathcal{D}$ est ouvert parce qu'il contient un voisinage de chacun de ses points.

Il nous reste à voir que $\varphi\colon \mathcal{D} \to \Omega$ est localement Lipschitz. Soit donc le point générique (s, a) dans $\mathcal{D}$ et l'ensemble V qui avait été construit plus haut. Nous allons montrer que φ est Lipschitz sur $J \times \overline{B(a, r)} \subset V$. D'abord sur V, l'application f est Lipschitz, donc

$$\|\varphi(t, b_1) - \varphi(t, b_2)\| \leqslant e^{Lt}\|b_1 - b_2\| \leqslant e^{LT}\|b_1 - b_2\| \tag{31.150}$$

pour tout $t \in J$ et $b_1, b_2 \in \overline{B(a, r)}$.

Ensuite, f est bornée, majorée par M sur V, donc

$$\begin{aligned} \|\varphi(t_1, b) - \varphi(t_2, b)\| &= |\int_{[t_1, t_2]} y_('s)ds| & (31.151a) \\ &= |\int_{[t_1, t_2]} f\big(s, y_b(s)\big)| & (31.151b) \\ &\leqslant \int_{[t_1, t_2]} |f\big(s, y_b(s)\big)|ds & (31.151c) \\ &\leqslant M|t_1 - t_2|. & (31.151d) \end{aligned}$$

Et enfin nous prouvons que φ est localement Lipschitz. En posant $k = \max\{e^{LT}, M\}$ nous avons

$$\begin{aligned} \|\varphi(t_1, b_1) - \varphi(t_2, b_2)\| &\leqslant \|\varphi(t_1, b_1) - \varphi(t_1, b_2)\| + \|\varphi(t_1, b_2) - \varphi(t_2, b_2)\| & (31.152a) \\ &\leqslant e^{LT}\|b_1 - b_2\| + M|t_1 - t_2| & (31.152b) \\ &\leqslant k\big(\|b_1 - b_2\| + |t_1 - t_2|\big) & (31.152c) \\ &\leqslant 2k \max\{\|b_1 - b_2\|, |t_1 - t_2|\} & (31.152d) \\ &= 2kd\big((b_1, t_2), (b_2, t_2)\big). & (31.152e) \end{aligned}$$

Le flot φ est donc Lipschitz de constante $2k$. □

Exemple 31.26.

Problèmes et choses à faire

Cet exemple doit être lu attentivement. Il me semble prouver que le flot n'est pas dérivable en la condition initiale sans que f le soit. Le document [491] semble dire le contraire. Je ne suis pas assez sûr de mon coup pour contredire.

Il n'y a pas de raisons de penser que $a \mapsto y_a(t)$ soit mieux que continue en sans hypothèses supplémentaires sur f. Pour illustrer cela nous considérons l'équation différentielle

$$\begin{cases} \dfrac{\partial X}{\partial s} = f\big(X(s), s\big) & (31.153a) \\ X(t) = x & (31.153b) \end{cases}$$

où t et x sont des paramètres fixés. Nous allons étudier la dérivabilité de X en x lorsque

$$f(x,t) = |x|. \tag{31.154}$$

Cela est un exemple typique de fonction autant Lipschitz que l'on veut sans être dérivable. L'équation différentielle est

$$\frac{\partial X}{\partial s}(s) = |X(s)|. \tag{31.155}$$

Si $x > 0$ alors $X(s) > 0$ dans un voisinage de $s = t$ et nous avons $X(s) = Ke^s$. La constante K se fixe par la condition initiale $X(t) = x$:

$$X(s) = xe^{s-t}. \tag{31.156}$$

Et cette solution tient en réalité pour tout s parce que $X(s)$ est alors toujours positif.

Si au contraire $x < 0$ nous avons la solution

$$X(s) = xe^{t-s}. \tag{31.157}$$

Au final,

$$X(s;x,t) = \begin{cases} xe^{t-s} & \text{si } x < 0 \\ 0 & \text{si } x = 0 \\ xe^{s-t} & \text{si } x > 0 \end{cases} \tag{31.158}$$

L'application $(s,x,t) \mapsto X(s;x,t)$ est continue. En ce qui concerne la dérivée partielle $\partial_x X$ en $x = 0$ nous avons :

$$\frac{\partial X}{\partial x}(s,0,t) = \lim_{\epsilon\to 0} \frac{X(s,\epsilon,t) - X(s,0,t)}{\epsilon} = \lim_{\epsilon\to 0} \frac{X(s,\epsilon,t)}{\epsilon}. \tag{31.159}$$

La limite à droite donne :

$$\lim_{\epsilon\to 0^+} \frac{X(s,\epsilon,t)}{\epsilon} = \frac{\epsilon e^{s-t}}{\epsilon} = e^{s-t}. \tag{31.160}$$

La limite à gauche donne :

$$\lim_{\epsilon\to 0^-} \frac{X(s,\epsilon,t)}{\epsilon} = e^{t-s}. \tag{31.161}$$

Les deux limites n'étant pas égales, la limite (31.159) n'existe pas[5] et l'application $(s,x,t) \mapsto X(s,x,t)$ n'est pas dérivable par rapport à x. △

Lemme 31.27 ([492]).
Soit un application $A\colon \overline{B(t_0,\tau)} \times \overline{B(a,R)} \to \mathcal{L}(\mathbb{R}^n)$ continue par rapport à sa première variable ($t_0 \in \mathbb{R}$ et $a \in \mathbb{R}^n$). Alors en posant l'équation

$$\begin{cases} \dfrac{\partial \psi}{\partial t}(t,b) = A(t,b)\psi(t,b) & (31.162a) \\ \psi(t_0,b) = \psi_0. & (31.162b) \end{cases}$$

5. Si vous comptez donner ça à manger au jury d'un concours, soyez prudent et n'écrivez pas l'équation (31.159) au tableau. Réfléchissez comment rédiger cela correctement.

Nous avons l'estimation

$$\|\psi(t,v)-\psi(t,w)\| \leqslant \|\psi_0\|\tau \max_{s\in\overline{B(t_0,\tau)}} \|A(s,v)-A(s,w)\| \times \\ \times \exp\left(\tau \max_{s\in\overline{B(t_0,\tau)}} \max\{\|A(s,v)\|,\|A(s,w)\|\}\right) \tag{31.163}$$

pour tout $t\in\overline{B(t_0,\tau)}$ *et* $v,w\in V$.

Théorème 31.28 (Régularité C^1 du flot [492])**.**
Soit un intervalle ouvert I *de* $\mathbb{R}$ *et un ouvert connexe* Ω *de* $\mathbb{R}^n$. *Soit une fonction* $f\in C^1\big(I\times\Omega,\mathbb{R}\big)$, $a\in\Omega$ *et* $t_0\in I$.

Il existe un voisinage $W\times V=\overline{B(t_0,\tau)}\times\overline{B(a,r)}$ *de* (t_0,a) *dans* $I\times\Omega$ *et une unique application* $\varphi\colon W\times V\to\Omega$ *telle que*

$$\begin{cases} \dfrac{\partial\varphi}{\partial t}(t,x)=f\big(t,\varphi(t,x)\big) & (31.164a)\\ \varphi(t_0,x)=x & (31.164b)\end{cases}$$

pour tout $x\in V$.

L'application $(t,x)\mapsto\varphi(t,x)$ *est de classe* C^1.

Problèmes et choses à faire

La preuve qui suit doit être lue avec beaucoup d'attention, en particulier sur les incohérences possibles de notations, et sur les oublis possibles de précautions oratoires type « quitte à encore réduire les voisinages V et W ».

Démonstration. En termes de notations, pour $x\in\Omega$ fixé nous écrivons $y_x(t)$ pour $\varphi(t,x)$ et pour $t\in I$ fixé nous notons $\varphi_t(x)$ pour $\varphi(t,x)$.

De plus lorsque nous écrirons des choses comme $g\colon\mathbb{R}\to\mathbb{R}$, nous n'entendrons pas que g est effectivement définie sur tout $\mathbb{R}$. La notation « $g\colon\mathbb{R}\to\mathbb{R}$ » indiquera seulement que la variable de g est réelle, et que nous comptons préciser le domaine plus tard. Cette remarque s'applique seulement à cette démonstration et non à l'ensemble du livre.

Nous considérons $R>0$ tel que $\overline{B(a,2R)}\subset\Omega$ et ensuite nous posons $V=\overline{B(a,R)}$. La fonction y_x, solution pour la condition initiale $y_x(t_0)=x$ est définie sur $W=[t_0-\tau,t_0+\tau]$ et prend ses valeurs dans $\overline{B(x,R)}$. Ceci est parce que y_x est continue, alors en prenant τ assez petit, la valeur de $y_x(t)$ ne va pas s'éloigner de x lorsque t ne s'éloigne pas de t_0.

Nous savons déjà de la proposition 31.25 que φ est C^1 en t et localement Lipschitz en sa seconde variable, avec une constante Lipschitz uniforme sur $W\times V$. Elle est donc continue en tant que fonction

$$\varphi\colon V\times W\to\mathbb{R}^d. \tag{31.165}$$

La différentielle partielle Df Pout t fixé nous notons $Df_{(t,x)}$ la différentielle de f par rapport à x. C'est-à-dire que

$$\begin{aligned} Df_{(t,x)}\colon\mathbb{R}^n&\to\mathbb{R}^n\\ u&\mapsto\frac{d}{ds}\Big[f(t,x+su)\Big]_{s=0}. \end{aligned} \tag{31.166}$$

C'est un élément de $\mathcal{L}(\mathbb{R}^n)$, l'ensemble des applications linéaires de $\mathbb{R}^n$ vers $\mathbb{R}^n$. Nous allons montrer que

$$(t,x)\mapsto Df_{(t,x)} \tag{31.167}$$

est continue en tant qu'application $\mathbb{R}\times\mathbb{R}^n\to\mathcal{L}(\mathbb{R}^n)$. Pour cela nous introduisons l'application d'inclusion $i\colon\mathbb{R}^n\to\mathbb{R}\times\mathbb{R}^n$, $i(u)=(0,u)$. Elle donne

$$Df_{(t,x)}(u)=\frac{d}{ds}\Big[f\big((t,x)+s(0,u)\big)\Big]_{s=0}=df_{(t,x)}\circ i(u). \tag{31.168}$$

Autrement dit

$$Df_{(t,x)}=df_{(t,x)}\circ i. \tag{31.169}$$

Or l'application $(t,x) \mapsto df_{(t,x)}$ est continue par hypothèse (f est de classe C^1) et l'application

$$\begin{aligned} \mathcal{L}(\mathbb{R}\times\mathbb{R}^n,\mathbb{R}^n) &\to \mathcal{L}(\mathbb{R}^n,\mathbb{R}^n) \\ A &\mapsto A\circ i \end{aligned} \tag{31.170}$$

est également continue. Donc $(t,x) \mapsto Df_{(t,x)}$ est continue[6].

L'équation aux variations Soit $x \in \Omega$. Nous introduisons l'opérateur

$$\begin{aligned} S_x\colon \mathbb{R}\times\mathcal{L}(\mathbb{R}^n) &\to \mathcal{L}(\mathbb{R}^n) \\ S_x(t,\psi) &= Df_{(t,y_x(t))}\circ\psi. \end{aligned} \tag{31.171}$$

Par ce que nous avons raconté, cela est une fonction continue en sa première variable et Lipschitz en sa seconde variable. Nous identifions $\mathcal{L}(\mathbb{R}^n)$ à $\mathbb{R}^{2^n}$.

Toujours pour chaque x considéré nous posons l'équation différentielle ordinaire

$$\begin{cases} \dfrac{\partial\psi}{\partial t}(t,x) = S_x\big(t,\psi(t,x)\big) & \text{(31.172a)} \\ \psi(t_0,x) = \mathrm{Id}\,. & \text{(31.172b)} \end{cases}$$

qui est une équation différentielle ordinaire pour $\psi\colon \mathbb{R}\times\mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n)$ rentrant dans le cadre de Cauchy-Lipschitz.

Quel est le domaine de définition de ψ pour sa première variable ? C'est un ouvert autour de t_0. Nous réduisons W de telle sorte que la solution ψ soit définie sur W. Idem pour la variable x qui est dans un voisinage de a.

L'équation (31.172) s'appelle l'**équation aux variations**. Nous allons montrer dans la douleur que ψ est continue et est la différentielle de φ_t, c'est-à-dire que

$$(d\varphi_t)_b = \psi(t,b). \tag{31.173}$$

ψ est continue en (t,x) (début) Il s'agit de majorer les deux termes de

$$\|\psi(t_1,a_1)-\psi(t_2,a_2)\| \leqslant \|\psi(t_1,a_1)-\psi(t_2,a_1)\| + \|\psi(t_2,a_1)-\psi(t_2,a_2)\|. \tag{31.174}$$

Premier terme Nous avons

$$\begin{aligned} \|\psi(t_1,b)-\psi(t_2,b)\| &= \|\int_{[t_1,t_2]} \frac{\partial\psi}{\partial t}(s,b)ds\| && \text{(31.175a)} \\ &\leqslant \int_{[t_1,t_2]} \|Df_{(s,y_b(s))}\circ\psi(s,b)\|ds && \text{(31.175b)} \\ &\leqslant \int_{[t_1,t_2]} \|Df_{(s,t_b(s))}\|\|\psi(s,b)\|ds && \text{(31.175c)} \\ &\leqslant |t_1-t_2| \max_{s\in[t_1,t_2]} \|Df_{(s,y_b(s))}\| \max_{s\in[t_1,t_2]} \|\psi(s,b)\|. && \text{(31.175d)} \end{aligned}$$

Nous allons majorer le second maximum. Prenons $t \in [0,\tau]$; et posons $A(u,b) = Df_{(u,y_b(u))}$ pour alléger les notations. Par l'équation de définition de ψ nous avons

$$\psi(t,b) = \psi(0,b) + \int_{[0,t]} A(u,b)\psi(u,b)du, \tag{31.176}$$

et donc

$$\|\psi(t,b)\| \leqslant \|\psi_0\| + \int_{[0,t]} \|A(u,b)\|\|\psi(u,b)\|du. \tag{31.177}$$

6. Si quelqu'un peut prouver ça de façon moins verbeuse, je suis preneur. Il me semble que quel que soit la façon dont on s'y prend, sous le capot, on passe par la continuité de l'application (31.170).

En y appliquant le lemme de Grönwall dans sa version 31.4 nous trouvons

$$\|\psi(s,b)\| \leqslant \|\psi_0\| \exp\left(\int_{[0,s]} \|A(u,b)\| du\right) \tag{31.178a}$$

$$\leqslant \|\psi_0\| \exp\left(s \max_{u\in[0,s]} \|A(u,b)\|\right). \tag{31.178b}$$

En retournant à (31.175d) nous avons $\psi_0 = \mathrm{Id}$ et donc $\|\psi_0\| = 1$ et

$$\max_{s\in[t_1,t_2]} \|\psi(s,b)\| \leqslant \max_{s\in[t_1,t_2]} \exp\left(s \max_{u\in[0,t]} \|Df_{(u,y_b(u))}\|\right) \tag{31.179}$$

Là dedans nous pouvons remplacer t par $\max\{|t_1|,|t_2|\}$. Posons enfin, pour alléger les expressions

$$a(t_1,t_2,b) = \max_{s\in[t_1,t_2]} \|Df_{(s,y_b(s))}\|. \tag{31.180}$$

La majoration que nous retenons est :

$$\|\psi(t_1,b) - \psi(t_2,b)\| \leqslant |t_1 - t_2| a(t_1,t_2,b) \exp\big(\max\{|t_1|,|t_2|\} a(0,t,b)\big). \tag{31.181}$$

Cela tend vers zéro lorsque $t_1 \to t_2$.

Deuxième terme En ce qui concerne le second terme,

$$\|\psi(t,b_1) - \psi(t,b_2)\| \tag{31.182}$$

nous utilisons le lemme 31.27 qui donne, pour $t \in [t_0 - \tau, t_0 + \tau]$,

$$\begin{aligned}\|\psi(t,b_1) - \psi(t,b_2)\| \leqslant \tau \max_{s\in\overline{B(0,\tau)}} \|Df_{s,y_{b_1}(s)} - Df_{s,y_{b_2}(s)}\| \times \\ \times \exp\left(\tau \max\{\|Df_{s,y_{b_1}(s)}, \|Df_{s,y_{b_2}(s)}\|\|\}\right).\end{aligned} \tag{31.183}$$

Dans notre cas, $t_0 = 0$, donc $t \in [-\tau,\tau]$. Vu la continuité de Df, nous avons

$$\max_{s\in\overline{B(0,\tau)}} \|Df_{s,y_{b_1}(s)} - Df_{s,y_{b_2}(s)}\| \to 0 \tag{31.184}$$

lorsque $b_1 \to b_2$.

ψ est continue en (t,x) (fin) Les deux bons calculs faits, nous avons, en repartant de (31.174),

$$\lim_{(t_1,b_1)\to(t_2,b_2)} \|\psi(t_1,b_1) - \psi(t_2,b_2)\| = 0, \tag{31.185}$$

ce qui signifie que ψ est une fonction continue de ses deux variables en même temps.

Différentiabilité de φ (début) Nous montrons maintenant que $D\varphi(t,x)$ existe. Pour rappel, D est la différentielle par rapport à la seconde variable. Nous sommes à étudier l'existence de $D\varphi_{(t,b)} = d(\varphi_t)_b$. Nous posons

$$\theta(t,h) = \varphi(t,b+h) - \varphi(t,b) = y_{b+h}(t) - y_b(t) \tag{31.186}$$

où b est le point où nous étudions la différentiabilité. Il est dans un voisinage du point a fixé depuis le début et autour duquel il existe un voisinage qui donne un sens à tout ce que nous avons fait jusqu'à présent. La dépendance de θ en b est implicite. Vu que φ est Lipschitz en sa seconde variable, nous avons la majoration

$$\|\theta(t,h)\| \leqslant C\|h\| \tag{31.187}$$

dès que $t \in V$ et $b, b+h \in W$.

De plus, parce que t_0 est le temps de la condition initiale nous avons

$$\theta(t_0, h) = y_{b+h}(t_0) - y_b(t_0) = a + h - a = h. \tag{31.188}$$

Et aussi, par définition de ψ :

$$\psi(t, b) = \psi_0 + \int_{t_0}^{t} \frac{\partial \psi}{\partial t}(s, b) ds = \psi_0 + \int_{t_0}^{t} Df_{(s,y_b(s))} \circ \psi(s, b) \tag{31.189}$$

En appliquant à h et en se souvenant que $\psi_0 = \mathrm{Id}$,

$$\psi(t, b)h = h + \int_{t_0}^{t} \Big(Df_{s,y_b(s)} \circ \psi(s, b) \Big) h \, ds. \tag{31.190}$$

Puis on peut faire un calcul assez classique en se souvenant que $\theta(t_0, h) = h$:

$$\theta(t, h) = \theta(t_0, h) + \int_{t_0}^{t} \big[\frac{\partial \varphi}{\partial t}(s, b + h) - \frac{\partial \varphi}{\partial t}(s, b) \big] ds \tag{31.191a}$$

$$= h + \int_{t_0}^{t} \big[f\big(s, y_{b+h}(s)\big) - f\big(s, y_b(s)\big) \big] ds. \tag{31.191b}$$

On fait la différence entre les deux :

$$\theta(t, h) - \psi(t, b)h = - \int_{t_0}^{t} \big[Df_{s,y_b(s)} \circ \psi(s, b)h - f\big(s, y_{b+h}(s)\big) + f\big(s, y_b(s)\big) \big] ds. \tag{31.192}$$

Nous y ajoutons et soustrayons $Df_{s,y_b(s)}\theta(s, h)$ et nous retenons la majoration suivante :

$$\begin{aligned} \|\theta(t, h) - \psi(t, b)h\| \leqslant & \int_{t_0}^{t} \|Df_{(s,y_b(s))}\psi(s, b) - Df_{(s,y_b(s))}\theta(s, h)\| ds \\ & + \int_{t_0}^{t} \|f(s, y_{b+h}(s)) - f(s, y_b(s)) + Df_{(s,y_b(s))}\theta(s, h)\| ds. \end{aligned} \tag{31.193}$$

Nous allons encore majorer ces deux termes séparément. Soit $\epsilon > 0$.

Premier terme Ce qui est dans la norme à majorer est

$$Df_{(s,y_b(s))}\big(\psi(s, b)h - \theta(s, h)\big). \tag{31.194}$$

Vu que Df est continue et que y_b est continue[7], l'application $s \mapsto Df_{s,y_b(s)}$ est continue et donc de norme majorée sur le compact $[t_0, t]$. Nous rapellons la notation

$$a(t_0, t, b) = \max_{s \in [t_0,t]} \|Df_{s,y_b(s)}\|, \tag{31.195}$$

et nous majorons encore et toujours. D'abord

$$\int_{t_0}^{t} \|Df_{(s,y_b(s))}\big(\psi(s, b)h - \theta(s, h)\big)\| ds \leqslant a(t_0, t, b) \int_{t_0}^{t} \|\psi(s, b) - \theta(s, h)\| ds. \tag{31.196}$$

Deuxième terme Pour traiter le deuxième terme, nous allons provisoirement noter $x = y_b(s)$ et $y = y_{b+h}(s)$; entre autres, $y - x = \theta(s, h)$. Ce qui est écrit dans le second terme de (31.193) est

$$f(s, y) - f(s, x) + Df_{(s,x)}\theta(s, h) = f(s, y) - f(s, x) + Df_{(s,x)}(y - x) \tag{31.197}$$

Comme D ne s'applique pas à la variable s, nous pouvons alléger la notation et déduire de la différentiabilité de f qu'il existe un $\eta > 0$ tel que $x, y \in W$ avec $\|y - w\| \leqslant \eta$ implique

$$\|f(y) - f(x) - Df_x(y - x)\| \leqslant \epsilon \|y - x\|. \tag{31.198}$$

Prenons $\|h\| \leqslant \eta/C$ (le C de (31.187)) ; en déballant les notations,

$$\|f\big(s, y_{b+h}(s)\big) - f\big(s, y_b(s)\big) - Df_{(s,y_b(s))}\theta(s, h)\| \leqslant \epsilon \|\theta(s, h)\| \leqslant \epsilon C \|h\|. \tag{31.199}$$

7. Il faut encore réduire les voisinages V et W pour que ceci ait un sens.

Les deux termes ensemble En remettant les deux dans (31.193) nous trouvons la majoration

$$\|\theta(t,h)-\psi(t,b)h\| \leqslant |t-t_0|\epsilon C\|h\| + a(t_0,t,v)\int_{t_0}^{t}\|\psi(s,b)h-\theta(s,h)\|ds \tag{31.200}$$

qui est encore de la graine à Grönwall avec

$$\begin{cases} u(t) = \|\theta(t,h)-\psi(t,b)\| & (31.201a)\\ b(t) = |t-t_0|\epsilon C\|h\| & (31.201b)\\ a(s) = a(t_0,t,v), & (31.201c)\end{cases}$$

la troisième étant une fonction constante. Cela donne, pour $t\in[t_0-\tau,t_0+\tau]$,

$$\|\theta(t,h)-\psi(t,b)h\| \leqslant |t-t_0|\epsilon C\|h\| + \int_{t_0}^{t}(s-t_0)\epsilon C\|h\|a(t_0,t,b)\exp\left(\int_s^t a(t_0,t,b)du\right)ds. \tag{31.202}$$

En valeur absolue, la différence $s-t_0$ est majorée par τ, l'intégrale dans l'exponentielle vaut $(t-s)a(t_0,t,b)$, et restons avec

$$\|\theta(t,h)-\psi(t,b)h\| \leqslant \tau\epsilon C\|h\| + \tau\int_{t_0}^{t}\epsilon C\|h\|a(t_0,t,b)e^{a(t_0,t,b)(t-s)}ds. \tag{31.203}$$

En supposant $t>t_0$ nous pouvons calculer l'intégrale. Si vous m'avez suivi jusqu'ici, vous devriez avoir de tels maux de tête que je vous donne la réponse :

$$\int_{t_0}^{t}a(t_0,t,b)e^{(t-s)a(t_0,t,b)}ds = e^{(t_0-t)a(t_0,t,b)}-1. \tag{31.204}$$

En remettant dans l'expression,

$$\|\theta(t,h)-\psi(t,b)h\| \leqslant \tau\epsilon C\|h\| + \epsilon C\|h\|a(t_0,t,b)\tau\big(e^{(t-t_0)}-1\big) = \tau\epsilon C\|h\|e^{(t-t_0)a(t,t_0,b)}. \tag{31.205}$$

Nous pouvons majorer $t-t_0$ par τ et $a(t,t_0,b)$ par $a(t_0-\tau,t_0+\tau,b)$ pour avoir la majoration

$$\|\theta(t,h)-\psi(t,b)h\| \leqslant \tau\epsilon C\|h\|e^{\tau a(t_0-\tau,t_0+\tau,b)}. \tag{31.206}$$

Différentiabilité de $\varphi(t,b)$ (fin) Nous écrivons la définition 10.127 de la différentiabilité : nous voulons vérifier que

$$\lim_{h\to 0}\frac{\varphi(t,b+h)-\varphi(t,b)-\psi(t,b)h}{\|h\|} = 0. \tag{31.207}$$

Nous remplaçons $\varphi(t,b+h)-\varphi(t,b)$ par $\theta(t,h)$ et prenons la norme avec les majorations données :

$$\lim_{h\to 0}\frac{\|\varphi(t,b+h)-\varphi(t,b)-\psi(t,b)h\|}{\|h\|} \leqslant \lim_{h\to 0}\tau\epsilon Ce^{\tau e(t_0-\tau,t_0+\tau,b)}. \tag{31.208}$$

Cela étant valable pour tout ϵ, nous en déduisons la nullité de la limite.

Nous avons démontré que φ était différentiable par rapport à sa deuxième variable et que

$$D\varphi_{(t,b)} = \psi(t,b). \tag{31.209}$$

Conclusion : φ est de classe C^1 Nous avons déjà prouvé que $(t,b)\mapsto\psi(t,b)$ est continue. Donc de (31.209) nous déduisons que les dérivées partielles $(t,b)\mapsto\frac{\partial\varphi}{\partial x_i}(t,b)$ sont continues. Mais comme φ est Lipschitz en t, la dérivée partielle $(t,b)\mapsto\frac{\partial\varphi}{\partial t}(t,b)$ est également continue.

La continuité de toutes les dérivées partielles de φ nous donne la classe C^1 pour φ par le théorème 11.243.

□

Proposition 31.29 (Régularité C^p du flot[493]).
Soit un intervalle ouvert I de $\mathbb{R}$ et un ouvert connexe Ω de $\mathbb{R}^n$. Soit une fonction $f \in C^p\big(I \times \Omega, \mathbb{R}\big)$, ainsi que $a \in \Omega$ et $t_0 \in I$.

Il existe un voisinage $W \times V = \overline{B(t_0, \tau)} \times \overline{B(a, r)}$ de (t_0, a) dans $I \times \Omega$ et une unique application $\varphi \colon W \times V \to \Omega$ telle que

$$\begin{cases} \dfrac{\partial \varphi}{\partial t}(t, x) = f\big(t, \varphi(t, x)\big) & (31.210a) \\ \varphi(t_0, x) = x & (31.210b) \end{cases}$$

pour tout $x \in V$.

L'application $(t, x) \mapsto \varphi(t, x)$ est de classe C^p.

Démonstration. Nous savons déjà par le théorème 31.28 que $(t, x) \mapsto \varphi(t, x)$ est de classe C^1. Nous supposons que f est de classe C^p avec $p \geqslant 2$.

Vu que φ et f sont de classe C^1, nous avons aussi que l'application $(t, x) \mapsto f\big(t, \varphi(t, x)\big)$ est de classe C^1. L'équation donne alors immédiatement le fait que

$$(t, x) \mapsto \frac{\partial \varphi}{\partial t}(t, x) \tag{31.211}$$

est de classe C^1.

En ce qui concerne la régularité par rapport aux autres variables, il faudra travailler un peu plus.

Une équation différentielle pour le flot Nous allons commencer par un habile jeu d'écriture : la formule

$$\varphi(t, x) = x + \int_{t_0}^{t} f\big(s, \varphi(s, x)\big) ds \tag{31.212}$$

devient

$$\varphi_t(x) = x + \int_{t_0}^{t} f\big(s, \varphi_s(x)\big) ds. \tag{31.213}$$

Dans le même ordre d'idée nous notons $f_s(x) = f(s, x)$, et ce qui se trouve dans l'intégrale (31.213) n'est autre que la fonction

$$g_s(x) = (f_s \circ \varphi_s)(x). \tag{31.214}$$

Tout cela pour différentier l'égalité (31.213) par la proposition 16.26 :

$$(d\varphi_t)_x = \mathrm{Id} + \int_{t_0}^{t} (dg_s)_x ds \tag{31.215a}$$

$$= \mathrm{Id} + \int_{t_0}^{t} (df_s)_{\varphi_s(x)} \circ (d\varphi_s)_s ds. \tag{31.215b}$$

Nous dérivons ensuite cela par rapport à t :

$$\frac{\partial}{\partial t}\Big((d\varphi_t)_x\Big) = (df_t)_{\varphi_t(x)} \circ (d\varphi_t)_x. \tag{31.216}$$

Cela est une égalité dans $\mathcal{L}(\mathbb{R}^n)$.

Nous introduisons la fonction

$$\begin{aligned} F \colon I \times \mathcal{L}(\mathbb{R}^n) \times \Omega &\to \mathcal{L}(\mathbb{R}^n) \\ (t, A, x) &\mapsto (df_t)_{\varphi_t(x)} \circ A. \end{aligned} \tag{31.217}$$

En fait, à la place de I et Ω il faut prendre des petits voisinages dans lesquels les choses ont un sens. Ce que dit l'équation 31.216 est que l'application

$$\begin{aligned} A\colon I\times\Omega &\to \mathcal{L}(\mathbb{R}^n)\\ (t,x) &\mapsto (d\varphi_t)_x \end{aligned} \tag{31.218}$$

vérifie l'équation différentielle

$$\begin{cases} \dfrac{\partial A}{\partial t}(t,x) = F\big(t,A(t,x),x\big) & \text{(31.219a)}\\ A(t_0,x) = \mathrm{Id}\,. & \text{(31.219b)} \end{cases}$$

Une autre équation différentielle Nous n'oublions pas l'équation différentielle pour la dérivée par rapport à t :

$$\frac{\partial\varphi}{\partial t}(t,x) = f\big(t,\varphi(t,x)\big). \tag{31.220}$$

Réécriture pour la différentielle Nous allons récrire l'équation (31.219) de façon à ce que le paramètre x soit inclus dans la condition initiale. De cette manière, la solution pourra profiter de la régularité C^1 du flot déjà prouvée dans le théorème 31.28.

Soit

$$\begin{aligned} g\colon I\times\big(\mathcal{L}(\mathbb{R}^n)\times\Omega\big) &\to \mathcal{L}(\mathbb{R}^n)\times\Omega\\ \big(t,(A,x)\big) &\mapsto \Big(F(t,a,x),0\Big). \end{aligned} \tag{31.221}$$

Nous posons $E=\mathcal{L}(\mathbb{R}^n)\times\Omega$; c'est cet espace qui va jouer le rôle de Ω. Nous considérons à présent l'équation différentielle suivante pour $z_x\colon I\to E$:

$$\begin{cases} \begin{pmatrix} z_1'(t)\\ z_2'(t)\end{pmatrix} = z'(t) = g(t,z(t)) = \begin{pmatrix} F\big(t,z_1(t),z_2(t)\big)\\ 0\end{pmatrix} & \text{(31.222a)}\\ z(t_0) = (\mathrm{Id},x). & \text{(31.222b)} \end{cases}$$

Il devrait y avoir un indice (Id,x) à z parce que c'est sa condition initiale. La fonction g est de classe C^p, donc cette équation admet une unique solution dont le flot est de classe C^1. Autrement dit, si S est dans un voisinage de Id, l'application

$$(t,x)\mapsto z(t) \tag{31.223}$$

est de classe C^1. Nous allons montrer qu'en posant $A(t,x)=z_1(t)$, nous avons une solution de (31.219) (l'unicité de la solution impose que cette solution est effectivement la différentielle de $d\varphi_t$). D'abord, la seconde ligne de l'équation différentielle est $z_2'(t)=0$, c'est-à-dire $z_2(t)=x$ pour tout t.

Sachant cela, la première équation devient

$$\begin{cases} z_1'(t) = F\big(t,z_1(t),x\big) & \text{(31.224a)}\\ z_1(t_0) = \mathrm{Id}, & \text{(31.224b)} \end{cases}$$

qui est l'équation différentielle pour A. Rappel : il y a partout une dépendance de z en sa condition initiale x que nous n'avons pas écrite pour des raisons de légèreté notionnelle. Il n'en reste pas moins que le flot de l'équation différentielle pour z est C^1, c'est-à-dire que $(t,x)\mapsto z_1(t)$ est de classe C^1.

Par conséquent, $(t,x)\mapsto A(t,x)$ est également C^1.

Régularité C^2 du flot Le fait que A soit C^1 n'implique pas que le flot le soit parce que le flot suit la même équation différentielle que A ne signifie pas que il soit égal. Il y a un raisonnement à faire.

Le fait est que si A est une solution de (31.219), alors $z(t)=\big(A(t,x),x\big)$ est solution de (31.222). C'est l'unicité de cette dernière qui permet de déduire l'unicité de la solution pour A.

Nous avons donc que l'unique solution A du système (31.219) est égale à $A(t,x)=(d\varphi_t)_x$ et est de classe C^1 par rapport à (t,x).

Donc $(t,x)\mapsto\varphi(t,x)$ est de classe C^2.

Régularité C^p Nous avons vu que le flot de $y' = f(t,y)$ est de classe C^2 dès que f est de classe C^2. Supposons que f soit de classe C^p et montrons que si le flot est de classe C^k $(k < p)$ alors il est de classe C^{k+1}.

Vu que le flot d'une équation différentielle de classe C^p est de classe C^k, en particulier celui de (31.222) est de classe C^k. Donc aussi la solution pour $A(t,x) = (d\varphi_t)_x$ est de classe C^k. Et vu que $(t,x) \mapsto (d\varphi_t)_x$ est de classe C^k, l'application φ est de classe C^{k+1}.

□

31.30.
Le théorème d'inversion locale 16.48 nous permet de dire que, pour t fixé, le flot $x \mapsto \varphi_t(x)$ est un C^p-difféomorphisme local.

Proposition 31.31 (Cauchy-Lipschitz avec paramètre, régularité C^p[494, 495, 496])**.**
Soit un intervalle ouvert I de $\mathbb{R}$, un connexe ouvert Ω de $\mathbb{R}^n$ et un intervalle ouvert Λ de $\mathbb{R}^d$. Soit une fonction $f \in C^p(I \times \Omega \times \Lambda, \mathbb{R}^n)$ localement Lipschitz en Ω. Soient $t_0 \in I$, $y_0 \in \Omega$ et $\lambda_0 \in \Lambda$. Il existe un voisinage compact de (t_0, y_0, λ_0) sur lequel le problème

$$\begin{cases} y'_\lambda(t) = f\big(t, y_\lambda(t), \lambda\big) & (31.225a) \\ y_\lambda(t_0) = y_0 & (31.225b) \end{cases}$$

possède une unique solution. De plus $(t,\lambda) \mapsto y_\lambda(t)$ est de classe C^p par rapport à ses deux variables.

Démonstration. Nous récrivons immédiatement le problème pour la fonction $y\colon I \times \Lambda \to \mathbb{R}^n$ donné par $y(t,\lambda) = y_\lambda(t)$:

$$\begin{cases} \dfrac{\partial y}{\partial t}(t,\lambda) = f\big(t, y(t,\lambda), \lambda\big) & (31.226a) \\ y(t_0) = y_0. & (31.226b) \end{cases}$$

Nous allons montrer que ce problème est en réalité équivalent à un problème sans paramètre. Nous posons $E = \Omega \times \Lambda$ et

$$\begin{aligned} g\colon I \times E &\to E \\ (t,x) &\mapsto \big(f(t,x_1,x_2), 0\big) \end{aligned} \tag{31.227}$$

où x_1 est la composante Ω de x et x_2 est la composante Λ de x. Pour une valeur $\mu \in \Lambda$ donnée nous considérons le problème au condition initiales

$$\begin{cases} x'(t) = g\big(t, x(t)\big) & (31.228a) \\ x(t_0) = (y_0, \mu). & (31.228b) \end{cases}$$

Le théorème de Cauchy-Lipschitz que nous prenons sous la forme 31.29 nous indique que ce problème admet une unique solution maximale et que le flot $\big(t, (y_0,\mu)\big) \mapsto x_{(y_0,\mu)}(t)$ est de classe C^p.

Nous passons maintenant à la résolution du problème (31.226)

Existence d'une solution C^1 Nous montrons à présent que la fonction y donnée par

$$y(t,\mu) = x_{(t_0,\mu)}(t)_1 \tag{31.229}$$

est solution de (31.226). Vu que x a deux composantes, nous pouvons un peu déballer l'équation. Afin d'éviter les notations laborieuses nous allons noter x pour $x_{(t_0,\mu)}$ et donc $x_1(t)$ pour $x_{(t_0,\mu)}(t)$. Nous avons l'équation différentielle

$$\begin{pmatrix} x'_1(t) \\ x'_2(t) \end{pmatrix} = \begin{pmatrix} f\big(t, x_1(t), x_2(t)\big) \\ 0 \end{pmatrix} \tag{31.230}$$

avec la condition initiale

$$\begin{pmatrix} x_1(t_0) \\ x_2(t_0) \end{pmatrix} = \begin{pmatrix} y_0 \\ \mu \end{pmatrix}. \tag{31.231}$$

La seconde ligne de l'équation donne immédiatement $x_2(t) = \mu$ pour tout t. En injectant dans la première ligne :

$$x_1'(t) = f\big(t, x_1(t), \mu\big). \tag{31.232}$$

Or vue la définition de y, le nombre $x_1'(t)$ n'est autre que $\frac{\partial y}{\partial t}(t,\mu)$. La fonction y que nous avons définie vérifie donc

$$\frac{\partial y}{\partial t}(t,\mu) = f\big(t, y(t,\mu), \mu\big) \tag{31.233}$$

et la condition initiale $y(t_0) = x_1(t_0) = y_0$. Elle est donc bien solution du problème initial.

De plus l'application $(t,\mu) \mapsto y(t,\mu) = x_{(t_0,\mu)}(t)_1$ est de classe C^p.

Unicité Pour l'unicité, soit on invoque la proposition 16.43 qui donne l'unicité dans les fonctions continues et a fortiori dans les fonctions C^1. Soit on fait le jeu inverse : on prouve qu'à chaque solution de (31.226) correspond une solution de (31.228), et l'unicité de la solution x donne l'unicité du côté de y.

□

Lemme 31.32.
Soit le problème

$$\begin{cases} \dfrac{\partial y}{\partial s}(s) = f\big(y(s), s\big) & (31.234a) \\ y(t) = x & (31.234b) \end{cases}$$

avec t et x fixés. Nous supposons que f est de classe C^p.

Alors l'application $t \mapsto y_x(s)$ est de classe C^p.

Démonstration. Soit t fixé, et l'équation différentielle

$$\begin{cases} \dfrac{\partial z}{\partial s}(s)a- = f\big(z(s), t-s\big) & (31.235a) \\ z(0) = x. & (31.235b) \end{cases}$$

Par le théorème 31.29, La solution z est de classe C^p en (s,x). En posant $y(s) = z(t-s)$ il est vite vérifié que y est solution de (31.234). C'est alors bien de classe C^p en t. □

31.8.4 Stabilité de Lyapunov

Définition 31.33.
Dans le cas de l'équation différentielle $y'(t) = f\big(y(t), t\big)$ pour $y\colon \mathbb{R} \to \mathbb{R}^n$, un point $a \in \mathbb{R}^n$ est un ***point d'équilibre*** *lorsque la fonction constante $y(t) = a$ est une solution.*

Le point d'équilibre $a \in \mathbb{R}^n$ est ***stable*** *si pour tout $\epsilon > 0$, il existe $\delta > 0$ tel que $\|y(0) - a\| < \delta$ implique $\|y(t) - a\| < \epsilon$ pour tout t.*

Théorème 31.34 (Théorème de stabilité de Lyapunov[1, 497, 202, 489])**.**
Soit l'équation différentielle

$$\begin{cases} y'(t) = f(y) & (31.236a) \\ y(0) = y_0 & (31.236b) \end{cases}$$

avec une fonction $f\colon \mathbb{R}^n \to \mathbb{R}^n$ de classe C^1 vérifiant $f(0) = 0$ et $y_0 \in \mathbb{R}^n$. Nous supposons que l'application linéaire df_0 n'a que des valeurs propres dont la partie réelle est strictement négative.

Alors

(1) Il existe $k > 0$ tel que si $\|y_0\| < k$ alors la solution maximale est définie sur $\mathbb{R}$,

(2) pour le même nombre $k > 0$, si $\|y_0\| < k$ alors $y(t) \xrightarrow{t\to\infty} 0$ exponentiellement vite,

(3) la solution $y = 0$ est un point d'équilibre attractif.

Démonstration. Placer ici une phrase intelligente[8].

Prolégomène Le théorème de Cauchy-Lipschitz 16.40 nous enseigne que l'équation différentielle considérée possède une unique solution maximale (entre autres parce qu'une fonction de classe C^1 est localement Lipschitz) et nous nommons J l'intervalle sur lequel elle est définie.

Système linéarisé Nous posons $A = df_0$. La fonction $y_L(t) = e^{tA}y_0$ est solution du système linéarisé

$$\begin{cases} y'(t) = Ay(t) & (31.237a) \\ y(0) = y_0. & (31.237b) \end{cases}$$

Pour évaluer la norme de y_L nous utilisons le lemme 14.119 : il existe un polynôme P tel que

$$\|y_L(t)\| \leqslant P\big(|t|\big) \sum_{i=1}^{r} e^{\operatorname{Re}\lambda_i t}\|y_0\|. \tag{31.238}$$

Mais par hypothèse, $\operatorname{Re}(\lambda_i) < 0$ et si nous posons $\lambda = \max\{\operatorname{Re}(\lambda_i)\}$ nous avons $\lambda < 0$ et

$$\|y_L(t)\| \leqslant P\big(|t|\big)e^{\lambda t}\|y_0\|. \tag{31.239}$$

Donc quel que soit y_0 nous avons $\lim_{t\to\infty}\|y_L(t)\| = 0$ c'est-à-dire $\lim_{t\to\infty} y_L(t) = 0$.

Une forme linéaire Nous définissons la forme bilinéaire suivante sur $\mathbb{R}^n$:

$$b(x, y) = \int_0^\infty \langle e^{tA}x, e^{tA}y\rangle dt. \tag{31.240}$$

D'abord cela est bien défini pour tout $x, y \in \mathbb{R}^n$ parce que

$$\big|\langle e^{tA}x, e^{tA}y\rangle\big| \leqslant \|e^{tA}x\|\|e^{tA}y\| \leqslant P_1\big(|t|\big)P_2\big(|t|\big)e^{2\lambda t}\|x\|\|y\|, \tag{31.241}$$

qui est intégrable entre 0 et ∞ à cause de la décroissance exponentielle[9]. Montrons que b est définie positive. Soit donc $x \neq 0$ et calculons

$$b(x, x) = \int_0^\infty \|e^{tA}x\|^2 dt. \tag{31.242}$$

Ce qui est dans l'intégrale est forcément (pas strictement) positif pour tout t. Mais si $x \neq 0$ alors $\|x\|^2$ est strictement positif et sur un voisinage de $t = 0$ nous avons aussi $\|e^{tA}x\|^2$ qui est strictement positif. Ergo $b(x, x) > 0$ dès que $x \neq 0$, ce qui signifie que b est strictement définie positive (lemme 9.157).

Nous notons $q\colon V \to \mathbb{R}$ la forme quadratique associée à b et aussi la norme qui va avec : $\|x\|_q = \sqrt{q(x)}$. En ce qui concerne le gradient $\nabla q\colon V \to V$, nous avons le petit calcul suivant[497] qui se base sur une des nombreuses formules du lemme 11.196[10] :

$$\begin{aligned} \nabla q(x) \cdot y &= \frac{d}{dt}\Big[q(x + ty)\Big]_{t=0} && (31.243a) \\ &= \frac{d}{dt}\Big[q(x) + t^2 q(y) + 2tb(x, y)\Big]_{t=0} && (31.243b) \\ &= 2b(x, y). && (31.243c) \end{aligned}$$

Nous avons aussi

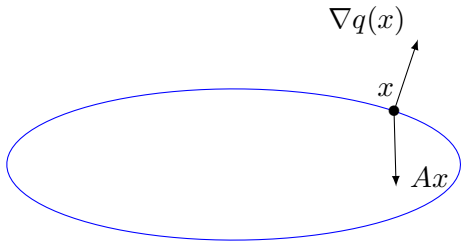
$$\begin{aligned} \nabla q(x) \cdot Ax &= 2b(x, Ax) && (31.244a) \\ &= 2\int_0^\infty \langle e^{tA}x, e^{tA}Ax\rangle && (31.244b) \\ &= \int_0^\infty \frac{\partial}{\partial t}\Big(\langle e^{tA}x, e^{tA}x\rangle\Big)(t)dt && (31.244c) \\ &= \lim_{T\to\infty}\Big[\langle e^{tA}x, e^{tA}x\rangle\Big]_{t=0}^{t=T}. && (31.244d) \end{aligned}$$

8. Parce que sinon l'environnement `description` qui suit donne un mauvais effet.
9. Proposition 14.82.
10. Le fait que q soit différentiable est simplement le fait que b soit bilinéaire.

Mais vu que $\|e^{tA}x\| \to 0$, pour $t \to \infty$ il ne reste que terme $t = 0$ de la différence, c'est-à-dire

$$\nabla q(x) \cdot Ax = 2b(x, Ax) = -\|x\|^2. \tag{31.245}$$

Étant donné que $\nabla q(x)$ est le vecteur dirigé vers l'extérieur de l'ellipsoïde de la courbe de niveau de q au point x, le vecteur Ax est dirigé vers l'intérieur.



Majoration de $q\big(y(t)\big)'$ Nous posons

$$\begin{aligned} r\colon \mathbb{R}^n &\to \mathbb{R}^n \\ x &\mapsto f(x) - Ax. \end{aligned} \tag{31.246}$$

Soit y la solution maximale au problème (31.236) que nous pouvons aussi écrire sous la forme

$$y'(t) = r\big(y(t)\big) + Ay(t). \tag{31.247}$$

Calculons un peu ...

$$\begin{aligned} q\big(y(t)\big)' &= b\big(y(t), y(t)\big)' && (31.248a)\\ &= 2b(y, y') && (31.248b)\\ &= 2b\big(y, Ay\big) + 2b\big(y, r(y)\big) && (31.248c)\\ &= -\|y\|^2 + 2b\big(y, r(y)\big) \qquad (31.245) \text{ avec } x = y(t) && (31.248d)\\ &\leqslant -\|y\|^2 + 2\|y(t)\|_q \|r\big(y(t)\big)\|_q \quad \text{Cauchy-Schwarz : } |b(a,b)| \leqslant \|a\|_q \|b\|_q. && (31.248e) \end{aligned}$$

Chacun des deux termes peut encore être majoré. En ce qui concerne le premier, par équivalence des normes [11], il existe une constante C telle que $\|y\| \geqslant C\|y\|_q$. En renommant immédiatement C^2 en C, $\|y\|^2 \geqslant C\|y\|_q^2 = Cq(y)$.

Pour le second, nous allons utiliser la différentiabilité de r et le théorème des accroissements finis. Vu que $df_0 = A$ nous avons $dr_0 = df_0 - A = 0$ et de plus r est de classe C^1 parce que f l'est. Toutes les normes étant équivalentes [12] sur $\mathbb{R}^n$ nous pouvons exprimer la continuité de dr pour la norme $\|.\|_q$: si $\epsilon > 0$ est fixé alors il existe $\alpha > 0$ tel que $\|x\| < \alpha$ implique $\|dr_x\|_q < \epsilon$. Nous pouvons écrire les accroissements finis [13] pour la fonction r :

$$\|r(x) - r(0)\|_q \leqslant \sup_{a\in[0,x]} \|df_a\| \|x\|_q. \tag{31.249}$$

La chose facile à remarquer est que $r(0) = f(0) = 0$. En ce qui concerne les choses difficiles, vu que dr est continue (parce que r est C^1) il existe un $\delta > 0$ tel que $\|dr_a\|_q < \epsilon$ dès que $a \in B_q(0, \delta)$. Si nous prenons $\|x\|_q < \delta$ alors cette majoration est valable pour tous les éléments sur lequel est pris le supremum dans la formule (31.249). Donc

$$\|r(x)\|_q \leqslant \epsilon \|x\|_q \tag{31.250}$$

11. Définition 7.221 et théorème 7.224.
12. Théorème 7.224.
13. Théorème 10.145.

tant que $\|x\|_q \leqslant \delta$. Par conséquent, tant que $\|y(t)\|_q \leqslant \delta$ nous avons $\|r\big(y(t)\big)\| \leqslant \epsilon\|y(t)\|_q$. Nous continuons le calcul (31.248) :

$$q\big(y(t)\big)' \leqslant Cq(y) + 2\epsilon\|y(t)\|_q^2 \tag{31.251a}$$
$$= -(C-2\epsilon)q(y). \tag{31.251b}$$

Si ϵ est petit on a $C - 2\epsilon > 0$ et on pose $\beta = C - 2\epsilon$ pour écrire

$$q\big(y(t)\big)' \leqslant -\beta q\big(y(t)\big) \tag{31.252}$$

tant que $\|y(t)\|_q < \delta$.

t_1 est un minimum Le nombre t_1 est bien défini et est bien un minimum. J'en veux pour preuve[14] que si $q\big(y(t_s)\big) = \delta$, on peut prendre le minimum seulement sur les $t \in [0, t_s]$; or par continuité $q\big(y(t)\big) = \delta$ définit un fermé. Bref t_1 est un infimum sur un compact (fermé borné) et donc bien un minimum atteint.

Si $q(y_0) < \delta$ alors $q\big(y(t)\big) < \delta$ Nous posons[15]

$$t_1 = \min\{t > 0 \text{ tel que } q\big(y(t)\big) = \delta\} \tag{31.253a}$$
$$t_2 = \max\{t < 0 \text{ tel que } q\big(y(t)\big) = \delta\}. \tag{31.253b}$$

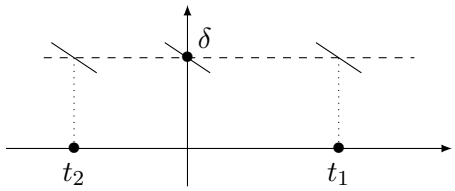
L'inégalité (31.252) est valable pour $t = 0$, $t = t_1$ et $t = t_2$; nous l'écrivons pour t_1 :

$$q\big(y(t)\big)'_{t=t_1} \leqslant -\beta q\big(y(t_1)\big) \leqslant -\beta\delta < 0 \tag{31.254}$$

Nous avons donc $q\big(y(t_1)\big) = \delta$ et $q\big(y(t)\big)'_{t=t_1} < 0$. Par conséquent pour tout t proche de t_1 avec $0 < t < t_1$ il y a $q\big(y(t)\big) > \delta$.

Pour la même raison, prise en $t = 0$ nous avons pour tout t proche de 0 avec $t > 0$ que $q\big(y(t)\big) < \delta$. Par continuité de $t \mapsto q\big(y(t)\big)$ cette fonction doit passer par la valeur δ dans $]t_2, 0[$ et $]0, t_1[$, ce qui contredit la maximalité de t_2 et la minimalité de t_1.

Ci-dessous, une partie de ce à quoi ressemble le graphe de $t \mapsto q\big(y(t)\big)$:



Deux conclusions :

— Vu que $q\big(y(t)\big)$ est borné pour tout $t \in \mathbb{R}$, nous sommes dans le cas (1) de l'alternative du théorème d'explosion en temps fini 31.18. Donc la solution $y(t)$ existe sur tout $\mathbb{R}$ pourvu que $\|y_0\|$ soit assez petit. Plus précisément par équivalence des normes, il existe un nombre $D > 0$ tel que $\|x\| \geqslant D\|x\|_q$ pour tout x. Si $\|y_0\| \leqslant D\delta$ alors

$$D\|y_0\|_q \leqslant \|y_0\| \leqslant D\delta, \tag{31.255}$$

qui donne immédiatement $\|y_0\|_q \leqslant \delta$, ce qui faut pour faire fonctionner l'existence de $y(t)$ pour tout t.

— Nous pouvons maintenant d'utiliser l'inégalité (31.252) pour tout $t \in \mathbb{R}$ sous la seule hypothèse que $q(y_0) < \delta$ au lieu de $q\big(y(t)\big) < \delta$.

La partie (1) de ce théorème est prouvée ; nous passons au reste à la partie (2). Pour cela nous supposons que $q(y_0) < \delta$.

14. J'espère que ce passage est correct. faites moi savoir si vous trouvez une erreur ou si vous pouvez me confirmer que c'est bon.

15.

À propos de $e^{\beta t}q(y)$ En sous-entendant la dépendance en t dans y nous avons

$$\Big(e^{\beta t}q(y)\Big)' = \beta e^{\beta t}q(y) + e^{\beta t}q(y)' = e^{\beta t}\big(\beta q(y) + q(y)'\big), \tag{31.256}$$

mais nous avons déjà prouvé que $q(y)' \leqslant -\beta q(y)$ (équation (31.252)), donc

$$\Big(e^{\beta t}q(y)\Big)' \leqslant 0 \tag{31.257}$$

Décroissance exponentielle Si $t \geqslant 0$, l'inégalité (31.257) donne

$$e^{\beta t}q\big(y(t)\big) \leqslant q(y_0), \tag{31.258}$$

c'est-à-dire

$$q\big(y(t)\big) \leqslant e^{-\beta t}q(y_0) \tag{31.259}$$

lorsque $t \geqslant 0$. Par équivalence des normes, nous avons des nombres D_1 et D_2 tels que

$$D_1\|x\|_q \leqslant \|x\| \leqslant D_2\|x\|_q \tag{31.260}$$

pour tout $x \in \mathbb{R}^n$. Nous avons donc pour tout $t \geqslant 0$ que

$$\|y(t)\| \leqslant D_2\|y(t)\|_q \leqslant D_2\|y_0\|_q e^{-\beta t}. \tag{31.261}$$

Pour rappel, $\beta > 0$, ce qui prouve la partie (2) du théorème.

Point d'équilibre Le point $y = 0$ est point d'équilibre (définition 31.33) parce que $f(0) = 0$, donc $y(t) = 0$ fonctionne. Dans ce cas, $y_0 = 0$.

Stabilité La stabilité est le fait que $\|y(t)\|_q \leqslant \delta$ dès que $\|y_0\|_q \leqslant \delta$.

□

31.8.5 Système proies-prédateurs de Lotka-Volterra

Le système de **Lotka-Volterra** est l'équation différentielle suivante :

$$\begin{cases} x' = ax - bxy & \text{(31.262a)} \\ y' = -cy + dxy & \text{(31.262b)} \end{cases}$$

où a, b, c, d sont des constantes positives, et avec la condition $x(t_0) > 0$, $y(t_0) > 0$.

En ce qui concerne l'interprétation des équations[498],

(1) $x(t)$ est le nombres de proies,

(2) $y(t)$ est le nombres de prédateurs,

(3) Les proies ont une reproduction rapide qui mène à une croissance exponentielle en absence de prédation (d'où le terme ax).

(4) Au contraire, les prédateurs meurent (ou migrent) rapidement lorsqu'ils n'ont pas de proies et nous supposons une décroissance exponentielle du nombre de prédateurs en l'absence de proies. D'où le terme $-cy$ avec le signe négatif.

(5) Les termes $-bxy$ et dxy sont les termes d'interaction entre les proies et les prédateurs. Ils sont proportionnels à la fréquence de leurs rencontres, lesquelles sont avantageuses pour les prédateurs et problématiques pour les proies.

Théorème 31.35 (Lotka-Volterra[202])**.**
Soient des constantes positives a, b, c, d et le système d'équations différentielles

$$\begin{cases} x' = ax - bxy & \text{(31.263a)} \\ y' = -cy + dxy & \text{(31.263b)} \\ x(t_0) > 0, y(t_0) > 0. & \text{(31.263c)} \end{cases}$$

Alors

(1) *Les solutions sont positives sur leur domaines.*
(2) *Les solutions existent sur* $\mathbb{R}$.
(3) *Les solutions sont périodiques.*

Démonstration. Nous divisons la preuve.

Comment théorème de Cauchy-Lipschitz s'applique Le théorème de Cauchy-Lipschitz [16] ne peut pas s'appliquer tel quel parce qu'il demande une condition initiale pour avoir unicité. En ce qui concerne les notations, ce qui est noté « y » dans le théorème est ici le couple x, y et la fonction f est alors

$$f\big(t, \begin{pmatrix} x \\ y \end{pmatrix}\big) = \begin{pmatrix} ax - bxy \\ -cy + dxy \end{pmatrix}. \tag{31.264}$$

C'est une fonction continue localement Lipschitz partout par le lemme 11.266 et la proposition 11.265.

Nous savons cependant que les solutions sont de classe C^1 et que moyennant la donnée d'une condition initiale, la solution est unique.

Les solutions restent positives Supposons $x(s) = 0$ pour un certain $s > t_0$. Alors le solution

$$\begin{cases} x(t) = 0 & \text{(31.265a)} \\ y(t) = \exp(-ct) & \text{(31.265b)} \end{cases}$$

est une solution pour $[t_0, s + \epsilon]$. Par unicité de la solution avec condition initiale $s(s) = 0$, nous avons aussi $x(t_0) = 0$ pour toutes les solutions, ce qui contredit notre condition.

De la même façon, avoir $y(s) = 0$ donne une solution avec $y(t) = 0$ pour tout t et donc une contradiction.

Solutions sur $\mathbb{R}$ Nous montrons maintenant que les solutions sont définies sur $\mathbb{R}$.

Nous avons $x' < ax$, donc pour tout t où la solution est définie,

$$0 < x(t) < x(t_0)e^{a(t-t_0)}, \tag{31.266}$$

c'est-à-dire que la solution ne peut pas exploser en temps fini [17] : elle est bornée par le haut et le bas. Elle doit donc exister pour tout $t \in \mathbb{R}$. Par ailleurs, $y' < dxy$ donc

$$0 < y(t) < y(t_0)e^{d\int_{t_0}^{t} x(s)ds} \tag{31.267}$$

qui est également contraire à l'explosion en temps fini.

4 zones : monotonie Nous divisons $\mathbb{R}^2$ en quatre zones d'après les signes de $a - by$ et $c - dx$. Nous montrons que dans chacune de ces zones, les solutions sont monotones. Prenons par exemple la partie

$$\{(x, y) \in \mathbb{R}^2 \text{ tel que } a - by > 0\} \times \{c - dx < 0\}. \tag{31.268}$$

Vu l'équation $x' = x(a - by)$, tant que $\big(x(t), y(t)\big)$ est dans cette zone, la fonction x' a le signe de x et est donc positive. Donc x est croissante dans cette zone.

De la même façon, $y' = -y(c - dx)$, et y' a un signe constant dans la zone.

4 zones : on bouge Nous prouvons à présent qu'une solution ne reste pas dans une zone.

(1) Supposons que $\big(x(t_0), y(t_0)\big)$ soit dans la zone

$$\{a - by > 0\} \qquad \times \qquad \{c - dx > 0\} \tag{31.269a}$$
$$x' > 0 \qquad\qquad y' < 0 \tag{31.269b}$$

16. Théorème 16.40.
17. Voir le corolaire 31.18.

et que la solution reste dans cette zone (pour les $t > t_0$). Nous avons en particulier $x' > 0$, donc x est croissante tout en ayant la borne supérieure $x < c/d$. Par conséquent x a une limite que nous appelons $x_1 \in [0, \frac{c}{d}]$.

De la même façon, ; y est décroissante et bornée vers le bas par zéro. Donc y a une limite que nous notons $y_1 \in [0, y(t_0)]$.

Vu que x est bornée et de classe C^1 nous avons forcément $\lim_{t\to\infty} x'(t) = 0$. Mais vu que $x' = ax - bxy$ nous devons avoir

$$ax_1 - bx_1y_1 = 0. \tag{31.270}$$

Mais ni $x_1 > 0$ donc $a - by_1 = 0$, ce qui donne $y_1 = \frac{a}{b}$ et aussi $x_1 = \frac{c}{d}$. Bref, y est décroissante et tend vers a/b ; donc $y(t_0) > a/b$, ce qui contredit que $y(t_0)$ soit dans la zone considérée.

Étant donné que $x' > 0$ et $y' < 0$, la solution sort de la zone pour entrer dans la zone ...

(2) Supposons que $\big(x(t_0), y(t_0)\big)$ soit dans la zone

$$\{a - by > 0\} \qquad \times \qquad \{c - dx < 0\} \tag{31.271a}$$
$$x' < 0 \qquad\qquad y' > 0 \tag{31.271b}$$

et que la solution reste dans cette zone (pour les $t > t_0$). Les fonctions x et y sont convergentes. Par conséquent $\ln(y)$ converge aussi et vu que x est croissante,

$$\frac{y'}{y} = -c + dx \geqslant -x + dx(t_0) > 0 \tag{31.272}$$

Cela signifie que $\ln(y)'$ est toujours positive et bornée par le bas. Cela est impossible si y est borné.

Donc on sort de la zone pour entrer dans ...

(3) Supposons que $\big(x(t_0), y(t_0)\big)$ soit dans la zone

$$\{a - by < 0\} \qquad \times \qquad \{c - dx < 0\} \tag{31.273a}$$
$$x' < 0 \qquad\qquad y' > 0 \tag{31.273b}$$

et que la solution reste dans cette zone (pour les $t > t_0$).

Le même type de raisonnement fait passer à la zone...

(4) Supposons que $\big(x(t_0), y(t_0)\big)$ soit dans la zone

$$\{a - by < 0\} \qquad \times \qquad \{c - dx > 0\} \tag{31.274a}$$
$$x' < 0 \qquad\qquad y' < 0 \tag{31.274b}$$

et que la solution reste dans cette zone (pour les $t > t_0$). Encore une fois, cela nous fait sortir de la zone et retourne vers la première zone.

À ce moment nous voyons déjà que la relation entre proies et prédateurs, c'est un peu le mythe de Sisyphe...

Une intégrale première Posons la fonction

$$H(x, y) = by + dx - a\ln(y) - c\ln(x). \tag{31.275}$$

Une simple dérivation montre que $x \mapsto H\big(x(t), y(t)\big)$ est constante. Nous considérons la fonction

$$\begin{aligned} f\colon \mathbb{R} &\to \mathbb{R} \\ s &\mapsto H\big(\frac{c}{d}, s\big) \end{aligned} \tag{31.276}$$

dont la dérivée n'est autre que $f'(s) = b - \frac{a}{s}$. La fonction f est donc décroissante sur l'intervalle $[\frac{a}{b}, \infty[$ et donc injective. Sur les changements de zones, il existe un t_0 tel que

$$x(t_0) = \frac{d}{c} \tag{31.277a}$$

$$y(t_0) > 0. \tag{31.277b}$$

Pour cette valeur t_0 nous avons alors $H\big(x(t_0), y(t_0)\big) = f\big(y(t_0)\big)$. En posant $s_0 = y(t_0) > 0$ nous avons

$$H(x_0, y_0) = f(s_0) \tag{31.278}$$

et f étant injective, ce s_0 est la seule valeur de s à vérifier $H(x_0, y_0) = f(s)$.

Conclusion La fonction x passant d'une zone à l'autre, il existe un $t_1 > t_0$ tel que $x(t_1) = a/b$. Nous avons évidemment

$$H\big(x(t_1), y(t_1)\big) = H(x_0, y_0) \tag{31.279}$$

parce que H est constante le long du mouvement. Cela se traduit par

$$H\big(\frac{a}{b}, y(t_1)\big) = f(s_0), \tag{31.280}$$

et donc $y(t_1) = f(s_0) = y(t_0)$. Avec tout cela nous avons

$$\begin{cases} y(t_1) = y(t_0) & \text{(31.281a)} \\ x(t_1) = x(t_2) = \dfrac{a}{b}. & \text{(31.281b)} \end{cases}$$

Cela est donc un point par lequel la solution repasse. Par unicité de la solution, elle est donc périodique.

□

31.9 Équation du second ordre

31.9.1 Wronskien

Nous considérons ici une équation différentielle de la forme

$$y''(t) + q(t)y(t) = 0 \tag{31.282}$$

Dans ce point nous allons considérer la fonction q sans hypothèse de périodicité. L'équation de Hill (sous-section 31.9.4) sera la même équation, mais en supposant que q est périodique.

Nous commençons par argumenter que si q est continue, alors l'ensemble des solutions de l'équation (31.282) est un espace vectoriel de dimension deux. Pour cela il suffit d'appliquer la méthode de réduction de l'ordre (section 31.7) puis le théorème de dimension pour les systèmes linéaires (théorème 31.14). En effet si la fonction y_1 est solution de (31.282) si et seulement si le vecteur $Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ est solution du système linéaire

$$Y'(t) = \begin{pmatrix} 0 & 1 \\ -q(t) & 0 \end{pmatrix} Y(t). \tag{31.283}$$

Soient deux solutions y_1 et y_2 de l'équation différentielle. Le **Wronskien** de ces deux solutions est le déterminant

$$W(t) = \begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix}. \tag{31.284}$$

Si nous considérons l'équation différentielle

$$y'' + py' + qy = 0, \tag{31.285}$$

le Wronskien peut être déterminé sans savoir explicitement y_1 et y_2 parce que $W = y_1y_2' - y_1'y_2$, et en dérivant,

$$\begin{aligned} W' &= y_1y_2'' + y_1'y_2' - y_1''y_2 - y_1'y_2' && (31.286a)\\ &= y_1(-py_2' - qy_2) - (-py_1' - qy_1)y_2 && (31.286b)\\ &= -p\begin{vmatrix} y_1 & y_2 \\ y_1' & y_2' \end{vmatrix}, && (31.286c) \end{aligned}$$

c'est-à-dire

$$W' = -pW. \tag{31.287}$$

Il suffit donc de savoir une condition initiale pour obtenir une équation différentielle pour W.

31.9.2 Avec second membre

Une équation différentielle du second ordre avec un second membre se présente sous la forme

$$ay''(t) + by'(t) + cy(t) = v(t) \tag{31.288}$$

où $v(t)$ est une fonction donnée. Le truc est de commencer par résoudre l'équation différentielle sans second membre, c'est-à-dire trouver la fonction $y_H(t)$ telle que

$$ay_H''(t) + by_H'(t) + cy_H(t) = 0. \tag{31.289}$$

Cela se fait en utilisant la méthode du polynôme caractéristique.

Ensuite, il faut trouver une solution particulière $y_P(t)$ de l'équation avec le second membre. Une seule. Pour y parvenir, il faut du doigté et un peu de technique. Il faut faire des essais en fonction de ce à quoi ressemble le $v(t)$:

(1) Si $v(t)$ est un polynôme, alors il faut essayer un polynôme,
(2) Si $v(t) = \cos(\omega t)$ ou bien $v(t) = \sin(\omega t)$, alors essayer $y_P(t) = A\cos(t) + B\sin(\omega t)$,
(3) Si $v(t) = e^{\omega t}$, alors essayer $y_P(t) = Ae^{\omega t}$.

31.9.3 Équation $y'' + q(t)y = 0$

Nous allons donner quelques propriétés des solutions de l'équation

$$y'' + qy = 0 \tag{31.290}$$

en fonction de telle ou telle hypothèse sur q.

Proposition 31.36.
Si $q\colon \mathbb{R}^+ \to \mathbb{R}$ est continue et si

$$\int_0^\infty |q(t)|dt \tag{31.291}$$

converge, alors

(1) toute solution bornée de $y'' + qy = 0$ vérifie $\lim_{t\to\infty} y'(t) = 0$,
(2) l'équation $y'' + qy = 0$ admet des solutions non bornées.

Démonstration. Soit y une solution bornée, et intégrons l'équation différentielle entre 0 et ∞ :

$$\int_0^\infty y''(t)dt = -\int_0^\infty q(t)y(t)dt. \tag{31.292}$$

La fonction y étant bornée, l'hypothèse sur q permet de dire que l'intégrale de droite existe. Par ailleurs,

$$\int_0^\infty y'' = \lim_{a\to\infty}\int_0^a y'' = \lim_{a\to\infty} y'(a) - y'(0). \tag{31.293}$$

Cela justifie que la limite $\lim_{t\to\infty} y'(t)$ existe. Posons $\alpha = \lim_{t\to\infty} y'(t)$ et supposons par l'absurde que $\alpha \neq 0$. Soit $\epsilon > 0$ et λ assez grand pour que

$$\|y' - \alpha\|_{[\lambda,\infty[} < \epsilon. \tag{31.294}$$

Soit aussi $x > \lambda$. Nous avons

$$\begin{aligned} y(x) &= y(\lambda) + \int_\lambda^x y'(t)dt && (31.295a)\\ &\geqslant y(\lambda) + \int_\lambda^x (\alpha - \epsilon)dt && (31.295b)\\ &= y(\lambda) + (\alpha - \epsilon)(x - \lambda). && (31.295c) \end{aligned}$$

En prenant la limite des deux côtés on voit que $y(x) \to \infty$ dès que $\alpha \neq 0$, ce qui est contraire aux hypothèses. Donc $\alpha = 0$.

Pour la seconde partie de la proposition, nous devons prouver que l'équation $y'' + qy = 0$ possède des solutions non bornées. Si l'équation a seulement des solutions bornées et si $\{u, v\}$ est une base de solutions, alors nous avons $u', v' \to 0$. Si nous reprenons l'équation (31.287) avec $p = 0$ nous savons que dans notre cas le Wronskien satisfait à $W' = 0$, c'est-à-dire qu'il est constant. Mais vu que u et v sont bornées et que les dérivées tendent vers zéro, nous avons $W(t) \to 0$ et donc $W(t) = 0$.

Or l'annulation identique du Wronskien contredit que $\{u, v\}$ serait une base de solutions. Donc il existe des solutions non bornées. □

Proposition 31.37.
Soit l'équation différentielle $y'' + qy = 0$. Si q est C^1, strictement positive et croissante, alors toutes les solutions sont bornées.

Démonstration. Soit y une solution et multiplions l'équation par $2y'$ (qui est non nulle par hypothèse) :

$$2y'y'' + 2qy'y = 0. \tag{31.296}$$

Nous allons intégrer cela en nous souvenant que $2y'y''$ est la dérivée de $(y')^2$. Pour tout $t > 0$ nous avons

$$\begin{aligned} 0 &= y'(t)^2 - y'(0)^2 + 2\underbrace{\int_0^t q(t)y'(t)y(t)dt}_{\text{par partie}} && (31.297a)\\ &= y'(t)^2 - y'(0)^2 + 2\left([qy^2]_0^t - \int_0^t q'y^2\right) && (31.297b)\\ & && (31.297c) \end{aligned}$$

Le terme qui nous intéresse est celui qui contient $y(t)$:

$$2q(t)y(t)^2 = -y'(t)^2 + y'(0)^2 + 2q(0)y(0)^2 + 2\int_0^t q'y^2 \tag{31.298}$$

Nous pouvons majorer $-y'(t)^2$ par zéro et remplacer toutes les constantes par K :

$$q(t)y(t)^2 \leqslant \int_0^t q'y^2 + K = \int_0^t \frac{q'}{q} qy^2. \tag{31.299}$$

C'est le moment d'utiliser le lemme de Grönwall 31.4 avec $\phi = qy^2$ et $\psi = q'/q$. Les hypothèses de croissance et de positivité ont été posées exprès. Bref, on a

$$qy^2 \leqslant K \exp\left(\int_0^t \frac{q'(s)}{q(s)} ds\right) \tag{31.300a}$$

$$= K \exp\left(\ln \frac{q(t)}{q(0)}\right) \tag{31.300b}$$

$$= K \frac{q(t)}{q(0)}. \tag{31.300c}$$

Notons que $q(0)$ est strictement positif. Nous déduisons que

$$y^2 \leqslant \frac{K}{q(0)} \tag{31.301}$$

et donc y est bornée. □

31.9.4 Équation de Hill

L'équation **de Hill** est une équation différentielle de la forme

$$y'' + qy = 0 \tag{31.302}$$

où

(1) $q \in C^1(\mathbb{R}, \mathbb{R})$,

(2) q est paire et π-périodique

Nous nous intéressons aux solutions complexes de cette équation différentielle.

Nous nommons $W \subset C^2(\mathbb{R}, \mathbb{C})$ l'espace des solutions complexes de l'équation (31.302). Nous savons par ce qui a été dit en 31.9.3 que cet espace est de dimension deux. De plus avec les hypothèses faites ici sur q, nous savons que les solutions sont de classe C^3 parce que si y est une solution, alors l'équation $y'' = qy$ nous indique que y est C^1 parce que y'' existe (y' est dérivable et donc continue). Mais si y est de classe C^1, alors le membre de droite qy est C^1 et donc y'' est C^1, ce qui prouve que y est de classe C^3. La récurrence ne va pas plus loin parce que q est seulement de classe C^1.

Nous considérons l'application de translation

$$\begin{aligned} T\colon C^2(\mathbb{R}, \mathbb{C}) &\to C^2(\mathbb{R}, \mathbb{C}) \\ (Ty)(x) &= y(x + \pi). \end{aligned} \tag{31.303}$$

En utilisant la règle de dérivation de fonctions composées, $(Ty)' = Ty'$ et $(Ty)'' = Ty''$, de telle sorte que si u est solution de l'équation (31.302), alors Tu est également solution. Donc W est un espace stable par T.

Le théorème 31.14 nous permet de choisir une base de W en imposant des conditions. Nous choisissons une base $\{y_1, y_2\}$ telles que

$$\begin{aligned} y_1(0) &= 1 \quad y_2(0) = 0 \\ y_1'(0) &= 0 \quad y_2'(0) = 1. \end{aligned} \tag{31.304}$$

Le théorème 31.14 nous assure que deux telles solutions existent et qu'elles forment une base de W parce que W est de dimension 2.

Lemme 31.38 ([488]).
Avec ce choix de base $\{y_1, y_2\}$ la matrice de T est donnée par

$$T = \begin{pmatrix} y_1(\pi) & y_2(\pi) \\ y_1'(\pi) & y_2'(\pi) \end{pmatrix}. \tag{31.305}$$

De plus la fonction y_1 est paire et la fonction y_2 est impaire.

Démonstration. Cherchons la matrice de T dans cette base en associant $\begin{pmatrix}1\\0\end{pmatrix}$ à y_1 et $\begin{pmatrix}0\\1\end{pmatrix}$ à y_2. Si $T=\begin{pmatrix}a&b\\c&d\end{pmatrix}$, alors

$$Ty_1=\begin{pmatrix}a&b\\c&d\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}=\begin{pmatrix}a\\c\end{pmatrix}=ay_1+cy_2. \tag{31.306}$$

En évaluant cela en $t=0$,

$$(Ty_1)(0)=ay_1(0)+cy_2(0)=a, \tag{31.307}$$

donc $a=(Ty_1)(0)=y_1(\pi)$. En dérivant (31.306), en tenant compte du fait que $(Ty_1)'=Ty_1'$ et en évaluant en $t=0$, nous trouvons de même $c=y_1'(\pi)$. Puis le même cinéma avec y_2 donne

$$T=\begin{pmatrix}y_1(\pi)&y_2(\pi)\\y_1'(\pi)&y_2'(\pi)\end{pmatrix}. \tag{31.308}$$

Passons maintenant à la parité de y_1 et y_2. Nous posons $\psi(t)=y_1(-t)$. Alors $\psi'(t)=-y_1'(-t)$ et $\psi''(t)=y_1''(t)$, tant et si bien que

$$\psi''(t)+q(t)\psi(t)=y_1''(-t)+q(t)y_1(-t)=0. \tag{31.309}$$

donc ψ est une solution de l'équation. Mais

$$\begin{cases}\psi(0)=y_1(0) & (31.310a)\\ \psi'(0)=-y_1'(0)=0, & (31.310b)\end{cases}$$

donc ψ a les mêmes conditions initiales que y_1. Par conséquent $\psi=y_1$ (par le l'unicité donnée dans le théorème de Cauchy-Lipschitz 16.40) et y_1 est paire. Nous procédons de même en partant de $\varphi(t)=-y_2(-t)$ pour trouver que $\varphi=y_2$ et que donc que y_2 est impaire.

□

Remémorons nous toutefois, pour calmer tout enthousiasme excessif, que T dépend de deux solutions et donc de la fonction q donnée dans l'équation.

Proposition 31.39 ([54]).
Nous considérons l'équation $y''+qy=0$ et sa base de solutions $\{y_1,y_2\}$ en suivant les notations données plus haut.

(1) Si $|\operatorname{Tr}(T)|<2$, alors toutes les solutions de l'équation sont bornées.

(2) Si $|\operatorname{Tr}(T)|=2$ alors nous avons une solution non bornée.

(3) Si $|\operatorname{Tr}(T)|>2$ alors toutes les solutions de l'équation sont non bornées.

(4) Le cas $|\operatorname{Tr}(T)|=2$ se présente si et seulement si $y_1'(\pi)y_2(\pi)=0$.

Démonstration. Remarquons que le déterminant de la matrice T est égal au Wronskien des solutions y_1 et y_2 calculé en $t=\pi$. Calculons sa valeur :

$$W(y_1,y_2)=\det\begin{pmatrix}y_1&y_2\\y_1'&y_2'\end{pmatrix}=y_1y_2'-y_1'y_2'. \tag{31.311}$$

En dérivant et en remplaçant y_i'' par $-qy_i$, nous trouvons tout de suite $W(y_1,y_2)'=0$. Donc le Wronskien est constant et il est facile de le calculer en $t=0$:

$$W(y_1,y_2)(0)=1-0=1. \tag{31.312}$$

Donc pour tout t nous avons $W(y_1,y_2)(t)=1$. En particulier

$$\det(T)=W(y_1,y_2)(\pi)=1, \tag{31.313}$$

et notons au passage que T est inversible.

Nous écrivons le polynôme caractéristique de T sous la forme $\chi_T = X^2 - \mathrm{Tr}(T)X + \det(T)$, c'est-à-dire

$$\chi_T = X^2 - \mathrm{Tr}(T)X + 1, \tag{31.314}$$

dont le discriminant est $\Delta = \mathrm{Tr}(A)^2 - 4$.

Nous passons à présent aux différents points de la proposition.

(1) Si $|\mathrm{Tr}(T)| < 2$, alors $\Delta < 0$ et χ_T a deux racines complexes conjuguées que nous notons ρ et $\bar{\rho}$. De plus le produit des racines étant le terme indépendant, $\rho\bar{\rho} = 1$; en particulier $|\rho| = |\bar{\rho}| = 1$. Notons $\{u, v\}$ une base de vecteurs propres : $Tu = \rho u$ et $Tv = \bar{\rho}v$. Il est vite vu que la fonction $|u|$ est π-périodique :

$$|u|(t+\pi) = |u(t+\pi)| = |(Tu)(t)| = |(\rho u)(t)| = |\rho||u|(t) = |u|(t). \tag{31.315}$$

La fonction $|u|$ est continue[18] et périodique ergo bornée. La fonction $|v|$ est bornée pour la même raison et par linéarité, toutes les fonctions de W sont bornées.

(2) Si $\mathrm{Tr}(T) = \pm 2$, alors $\Delta = 0$ et χ_T a une racine réelle double[19] qui doit être ± 1. Soit u un vecteur propre de T pour la valeur propre ± 1. Nous avons

$$|u|(t+\pi) = |Tu(t)| = |\pm u(t)|, \tag{31.316}$$

ce qui prouve encore que $|u|$ est périodique et donc bornée.

Notons que nous n'avons pas d'informations sur le fait qu'une autre solution soit ou non bornée.

(3) Si $|\mathrm{Tr}(T)| > 2$, alors χ_T a deux racines réelles distinctes r et r' avec $rr' = 1$ (toujours les relations coefficients-racines). En raison de quoi $r' = r^{-1}$ et quitte à échanger r et r' nous supposons $|r| > 1$. L'opérateur est maintenant diagonalisable et nous considérons $\{u, v\}$ une base de vecteurs propres pour les valeurs propres r et r'. Une solution non nulle de l'équation s'écrit donc sous la forme

$$y = \alpha u + \beta v \tag{31.317}$$

avec $(\alpha, \beta) \neq (0, 0)$.

— Si $\alpha = 0$, alors $\beta \neq 0$ et nous choisissons une valeur t telle que $v(t) \neq 0$. Dans ce cas,

$$y(t+n\pi) = \beta v(t+n\pi) = \beta(T^n v)(t) = \beta(r')^n v(t), \tag{31.318}$$

et en faisant $n \to -\infty$ nous obtenons $\pm\infty$ suivant le signe de β.

— Si $\alpha \neq 0$, alors nous fixons[20] t tel que $u(t) \neq 0$. Alors

$$y(t+n\pi) = \alpha r^n u(t) + \beta(r')^n(t). \tag{31.319}$$

En faisant $n \to \infty$, nous avons $(r')^n \to 0$ tandis que le premier terme tend vers $\pm\infty$ suivant le signe de α.

(4) D'abord le théorème de Cayley-Hamilton 9.115 nous indique que $\chi_T(T) = 0$, c'est-à-dire que

$$T^2 - \mathrm{Tr}(T)T + 1 = 0. \tag{31.320}$$

Nous avons déjà mentionné le fait que T était inversible. Multiplions donc (31.320) par T^{-1} :

$$T + T^{-1} = \mathrm{Tr}(T)\mathbb{1}_2. \tag{31.321}$$

18. La fonction u elle-même n'est cependant pas garantie d'être périodique.

19. Ce qui n'implique pas le fait d'avoir deux vecteurs propres pour cette valeur propre, mais tout de même au moins un, voir l'exemple 9.112.

20. Mais pas trop hein ; nous aurons encore besoin d'assigner à t d'autres valeurs dans d'autres théorèmes.

Vu que T^{-1} est l'endomorphisme $T^{-1}u(t) = u(t-\pi)$, sa matrice est donnée par

$$T^{-1} = \begin{pmatrix} y_1(-\pi) & y_2(-\pi) \\ y_1'(-\pi) & y_2'(-\pi) \end{pmatrix} = \begin{pmatrix} y_1(\pi) & -y_2(\pi) \\ -y_1'(\pi) & y_2'(\pi) \end{pmatrix} \tag{31.322}$$

où nous avons utilisé le fait que y_1 était paire et y_2 impaire (lemme 31.38). Si nous notons $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, alors $T^{-1} = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}$ et

$$T + T^{-1} = \begin{pmatrix} 2a & 0 \\ 0 & 2b \end{pmatrix}. \tag{31.323}$$

L'équation (31.321) donne alors, vu que $\mathrm{Tr}(T) = a + d$,

$$\begin{pmatrix} 2a & 0 \\ 0 & 2b \end{pmatrix} = \begin{pmatrix} a+d & 0 \\ 0 & a+d \end{pmatrix}, \tag{31.324}$$

ce qui donne immédiatement $a = d$. La matrice de T a donc comme forme $T = \begin{pmatrix} a & b \\ c & a \end{pmatrix}$ et $\mathrm{Tr}(T) = 2a$.
Donc $\mathrm{Tr}(T) = \pm 2$ si et seulement si $a = \pm 1$ et vu que $1 = \det(T) = a^2 - bc$, nous avons $a = \pm 1$ si et seulement si $bc = 0$, ce qui signifie exactement $y_1'(\pi)y_2(\pi) = 0$.

□

31.10 Différents types d'équations différentielles

31.10.1 Équation homogène

Une équation différentielle **homogène** est une équation de la forme

$$y' = f(t, y) \tag{31.325}$$

où $f(\lambda t, \lambda y) = f(t, y)$ pour tout $\lambda \neq 0$.
Elle se présente sous la forme

$$y' = \frac{\text{degré } n \text{ en } t, y}{\text{degré } n \text{ en } t, y}, \tag{31.326}$$

avec pas de y' à droite : juste du y et du t.

Lemme 31.40.
L'équation $y' = f(t, y)$ est homogène si et seulement si $f(t, y)$ est une fonction de y/t seulement.

Pour résoudre l'équation homogène, on pose

$$z(t) = \frac{y(t)}{t}, \tag{31.327}$$

donc $tz = y$, et

$$y'(t) = tv'(t) + v(t), \tag{31.328}$$

à remettre dans l'équation de départ.

31.10.2 Équation de Bernoulli

C'est une équation du type

$$y' = a(t)y + b(t)y^\alpha \tag{31.329}$$

où $\alpha \neq 0$ ou 1. Pour la résoudre, on divise l'équation par y^α, et on pose $u = y^{1-\alpha}$, et on tombe sur une équation linéaire

$$u' = (1-\alpha)\big(a(t)u + b(t)\big). \tag{31.330}$$

31.10.3 Équation de Riccati

C'est une équation de la forme

$$y' = a(t)y^2 + b(t)y + c(t). \tag{31.331}$$

En général, on ne peut pas la résoudre, mais si on en connaît *à priori* des solutions particulières, alors on peut s'en sortir.

(1) Si on sait que $y_1(t)$ est une solution, alors on pose

$$y(t) = y_1(t) + \frac{1}{u(t)}, \tag{31.332}$$

et on obtient une équation linéaire

$$u' = -\big(2y_1(t)a(t) + b(t)\big)u - a(t). \tag{31.333}$$

(2) Si y_1 et y_2 sont solutions, alors nous avons y sous forme implicite

$$\frac{y - y_1}{y - y_2} = Ke^{\int a(t)\big(y_1(t)-y_2(t)\big)dt}. \tag{31.334}$$

Pour résoudre une équation de Ricatti, il faut donc d'abord deviner une ou deux solutions.

31.10.4 Équation différentielle exacte

31.10.4.1 Résolution lorsque tout va bien

Avant de vous lancer dans les équations différentielles exacte, vous devez lire la section sur les formes différentielles 11.19. Une équation différentielle exacte est de la forme $P(t,y)+Q(t,y)y' = 0$ que nous allons écrire sous la forme

$$P(t,y)dt + Q(t,y)dy = 0. \tag{31.335}$$

Nous savons que si $\partial_y P = \partial_t Q$, alors il existe une fonction $f(t,y)$ telle que $Pdt + Qdy = df$. Pour trouver une telle fonction, nous pouvons simplement intégrer la forme $Pdt + Qdy$. En effet, si $\gamma\colon [0,1] \to \mathbb{R}^2$ est un chemin tel que $\gamma(0) = (0,0)$ et $\gamma(1) = (t,y)$, alors en définissant

$$f(t,y) = \int_\gamma [Pdt + Qdt] = \int_0^1 \big[(P\circ\gamma)(u)dt + (Q\circ\gamma)(u)\big]\big(\gamma'(u)\big)du, \tag{31.336}$$

nous avons $df = Pdt + Qdy$. N'importe quel chemin fait l'affaire. Calculons avec $\gamma(u) = (tu, yu)$. La dérivée de ce chemin est donnée par

$$\gamma'(u) = t\begin{pmatrix}1\\0\end{pmatrix} + y\begin{pmatrix}0\\1\end{pmatrix}. \tag{31.337}$$

Étant donné que $dt\begin{pmatrix}a\\b\end{pmatrix} = a$ et $dy\begin{pmatrix}a\\b\end{pmatrix} = b$, nous avons

$$\begin{aligned} f(t,y) &= \int_0^1 [Pdt + Qdy]\big(\gamma(u)\big)\left(t\begin{pmatrix}1\\0\end{pmatrix} + y\begin{pmatrix}0\\1\end{pmatrix}\right)du \\ &= \int_0^1 P\big(\gamma(t)\big)tdu + \int_0^1 Q\big(\gamma(t)\big)ydu \\ &= \int_0^1 \big[tP(tu,uy) + yQ(tu,yu)\big]du. \end{aligned} \tag{31.338}$$

Nous retrouvons exactement la formule (19.292). Si ça t'étonne, c'est que tu n'as pas compris ;) Dans le cas où nous avons la fonction f qui vérifie $P = \partial_t f$ et $Q = \partial_y f$, l'équation (31.335) devient

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial y}\frac{dy}{dt} = 0, \tag{31.339}$$

c'est-à-dire

$$\frac{d}{dt}\Big[f\big(t, y(t)\big)\Big] = 0, \tag{31.340}$$

dont la solution

$$f\big(t, y(t)\big) = C \tag{31.341}$$

donne la solution $y(t)$ sous forme implicite.

31.10.4.2 Facteur intégrant (quand tout ne va pas bien)

Si la forme $Pdt + Qdy$ n'est pas exacte, il n'existe pas de fonction f qui résolve l'affaire. Nous pouvons toutefois essayer de trouver un **facteur intégrant**. Nous cherchons une fonction M telle que

$$(MP)dt + (MQ)dy \tag{31.342}$$

soit exacte. Nous cherchons donc $M(t, y)$ telle que $\partial_y(MP) = \partial_t(MQ)$. En utilisant la règle de Leibnitz, nous trouvons l'équation suivante pour M :

$$M(\partial_y P - \partial_t Q) = Q(\partial_t M) - P(\partial_y M). \tag{31.343}$$

Cette équation est en générale extrêmement difficile à résoudre, mais dans certains cas particuliers, il est possible d'en trouver une solution à tâtons.

31.11 Distributions pour les équations différentielles

Nous commençons par définir l'espace $C^\infty\big(\mathbb{R}, \mathscr{S}'(\mathbb{R}^d)\big)$ en disant que $t \mapsto u_t$ est dans cet espace si

(1) pour tout $t \in \mathbb{R}$ nous avons $u_t \in \mathscr{S}'(\mathbb{R}^d)$,

(2) l'application $t \mapsto u_t$ est de classe C^∞.

Pour définir ce que nous entendons par une fonction de classe C^k à valeurs dans $\mathscr{S}'(\mathbb{R}^d)$ nous nous souvenons de la proposition 29.36.

31.11.1 Équation de Schrödinger

Théorème 31.41 (Équation de Schrödinger[54]).
Soit $g \in \mathscr{S}'(\mathbb{R}^d)$ et le problème

$$\begin{cases} \partial_t \tilde{u} - i\Delta\tilde{u} = 0 & \text{(31.344a)} \\ u_0 = g & \text{(31.344b)} \end{cases}$$

où $\tilde{u} \in C^\infty\big(\mathbb{R}, \mathscr{D}'(\mathbb{R}^d)\big)$ est lié à u par la remarque 29.58. Alors

(1) Il existe une unique solution dans $C^\infty\big(\mathbb{R}, \mathscr{S}'(\mathbb{R}^d)\big)$.

(2) Cette solution u vérifie de plus $\tilde{u} \in \mathscr{S}'(\mathbb{R} \times \mathbb{R}^d)$.

Démonstration. Nous allons donner explicitement une fonction $u \in C^\infty\big(\mathbb{R}, \mathscr{S}'(\mathbb{R}^d)\big)$ et nous allons vérifier l'équation (31.344a) en testant sur une fonction $\psi \in \mathscr{S}'(\mathbb{R} \times \mathbb{R}^d)$. Cela prouvera le point (2) ainsi que la partie existence de (1). Dans ce qui suit toutes les transformées de Fourier seront par rapport à la variable $x \in \mathbb{R}^d$ ou par rapport à ξ. Jamais par rapport à $t \in \mathbb{R}$.

Existence Pour $t \in \mathbb{R}$ nous posons[21]

$$u_t = \mathcal{F}^{-1}(f_t \hat{g}) \tag{31.345}$$

où $f_t \in \mathscr{S}(\mathbb{R}^d)$ est la fonction $f_t(x) = e^{-it\|x\|^2}$. Pour toute fonction $\varphi \in \mathscr{S}(\mathbb{R}^d)$ nous avons

$$u_t(\varphi) = (f\hat{g})\big(\mathcal{F}^{-1}(\varphi)\big) = \hat{g}\big(f\mathcal{F}^{-1}(\varphi)\big) = g\Big(\mathcal{F}\big(f\mathcal{F}^{-1}(\varphi)\big)\Big). \tag{31.346}$$

Le fait que $\mathcal{F}^{-1}(\varphi)$ soit une fonction Schwartz fait partie de la proposition 28.26. Pour chaque t nous avons bien $u_t \in \mathscr{S}'(\Omega)$.
De plus la fonction $h(t,x) = e^{-it\|x\|^2}(\mathcal{F}^{-1}\varphi)(x)$ est dans $C^\infty(\mathbb{R} \times \mathbb{R}^d)$, et par conséquent l'application

$$t \mapsto \hat{g}\big(h(t,.)\big) \tag{31.347}$$

est également C^∞ par la proposition 29.38. Ceci pour dire que $u \in C^\infty\big(\mathbb{R}, \mathscr{S}'(\mathbb{R}^d)\big)$. Il faut encore vérifier que cette fonction est bien une solution de notre problème. Nous testons cette équation sur $\psi \in \mathscr{S}(\mathbb{R} \times \mathbb{R}^d)$. Pour alléger les notations nous posons $\psi_t \colon x \mapsto \psi(t,x)$ et par conséquent aussi $(\partial_t \psi_t)(x) = (\partial_t \psi)(t,x)$. Nous avons :

$$\begin{aligned}
\heartsuit &= (\partial_t \tilde{u} - i\Delta\tilde{u})(\psi) && (31.348a)\\
&= -\tilde{u}(\partial_t\psi) - i\tilde{u}(\Delta\psi) && (31.348b)\\
&= -\int_{\mathbb{R}} u_t\big((\partial_t\psi_t) + i(\Delta\psi_t)\big)dt && (31.348c)
\end{aligned}$$

Ici nous nous souvenons du lemme 28.29 qui nous dit que nous pouvons permuter $\mathcal{F}^{-1}$ et ∂_t. Et pour l'autre terme il faut utiliser le lemme 28.19 avec $|\alpha| = 2$ et une somme pour obtenir que

$$\widehat{\Delta\varphi}(x) = -\|x\|^2\hat{\varphi}(x), \tag{31.349}$$

qui dans notre cas s'écrit sous la forme

$$\mathcal{F}^{-1}\Big((\Delta\psi_t)\Big)(x) = -\|x\|^2\mathcal{F}^{-1}\psi(t,x). \tag{31.350}$$

En remettant bout à bout,

$$\begin{aligned}
\heartsuit &= -\int_{\mathbb{R}} (f_t\hat{g})\Big((\partial_t - i\|.\|^2)\mathcal{F}^{-1}\psi_t\Big)dt && (31.351a)\\
&= -\int_{\mathbb{R}} \hat{g}\Big(x \mapsto e^{-it\|x\|^2}(\partial_t - i\|x\|^2)(\mathcal{F}^{-1}\psi)(t,x)\Big)dt && (31.351b)
\end{aligned}$$

Pour alléger les notations nous notons $\check{\psi}_t(x) = (\mathcal{F}^{-1}\psi)(t,x)$. Nous avons

$$\partial_t\left(e^{-it\|x\|^2}\check{\psi}_t(x)\right) = -i\|x\|^2 e^{-it\|x\|^2}\check{\psi}_t(x) + e^{-it\|x\|^2}(\partial_t\check{\psi}_t), x) = e^{-it\|x\|^2}\big(\partial_t - i\|x\|^2\big)\check{\psi}_t(x); \tag{31.352}$$

cela nous permet d'un peu factoriser une dérivée dans $\heartsuit$:

$$\begin{aligned}
\heartsuit &= -\int_{\mathbb{R}} \hat{g}\left(\partial_t\left(e^{-it\|.\|^2}\check{\psi}_t(.)\right)\right)dt && (31.353a)\\
&= -\int_{\mathbb{R}} \partial_t\hat{g}\left(e^{-it\|.\|^2}\check{\psi}_t(.)\right)dt && (31.353b)\\
&= -\lim_{N\to\infty}\left[\hat{g}\left(e^{-i\|.\|^2}\check{\psi}_t(.)\right)\right]_{t=-N}^{t=N}. && (31.353c)
\end{aligned}$$

Histoire de bien comprendre les notations, il ne s'agit pas de calculer $\hat{g}\big(e^{-it\|.\|^2}\check{\psi}_t\big)$ pour un t général et de remplacer ensuite t par N et $-N$. En effet la valeur de $\hat{g}\big(e^{-it\|.\|^2}\check{\psi}_t\big)$ pour un t

21. En utilisant la définition (29.46) du produit d'une distribution par une fonction.

donné est celle qu'on obtient en calculant $\hat{g}(\dots)$ après avoir remplacé t par ce que l'on veut. Par conséquent, en posant $\varphi(t,\xi) = e^{-i\|\xi\|^2}\check{\psi}_t(\xi)$ nous avons :

$$\heartsuit = \lim_{N\to\infty}\left[g\left(x\mapsto\int_{\mathbb{R}} e^{-ix\xi}\varphi(t,\xi)d\xi\right)\right]_{t=-N}^{t=N} \tag{31.354a}$$

$$= \lim_{N\to\infty} g\left(x\mapsto\int_{\mathbb{R}} e^{-ix\xi}\varphi(N,\xi)d\xi\right) - \lim_{N\to\infty} g\left(x\mapsto\int_{\mathbb{R}} e^{-ix\xi}\varphi(-N,\xi)d\xi\right) \tag{31.354b}$$

La limite commute avec g parce que cette dernière est une distribution (continue). De plus la limite commute avec l'intégrale parce que ce qui est dedans est Schwartz. La fonction φ étant Schwartz, la limite est nulle. Donc

$$\heartsuit = 0. \tag{31.355}$$

Cela signifie que la fonction u proposée est bien une solution de l'équation de Schrödinger dans $C^\infty(\mathbb{R},\mathscr{S}'(\mathbb{R}^d))$.

Unicité Nous considérons deux solutions $u_1, u_2 \in C^\infty\big(\mathbb{R},\mathscr{S}'(\mathbb{R}^d)\big)$ et la fonction $u = u_1 - u_2$ doit satisfaire au problème

$$\begin{cases} (\partial_t\tilde{u} - i\Delta\tilde{u})(\psi) = 0 & (31.356a)\\ u_0 = 0. & (31.356b)\end{cases}$$

Nous allons montrer que seule la fonction $u_t = 0$ peut satisfaire à cela pour tout $\psi \in \mathscr{S}(\mathbb{R}\times\mathbb{R}^d)$. Nous allons même montrer qu'en imposant ces équations seulement sur la partie de $\mathscr{S}(\mathbb{R}\times\mathbb{R}^d)$ qui est à support compact par rapport à $\mathbb{R}$, la seule solution est $u_t = 0$. Soit donc $\psi \in \mathscr{S}(\mathbb{R}\times\mathbb{R}^d)$ à support compact vis-à-vis de sa variable t. Alors

$$0 = -\tilde{u}(\partial_t\psi + i\Delta\psi) = -\int_{\mathbb{R}} u_t\Big((\partial_t\psi_t) + i(\Delta\psi_t)\Big)dt \tag{31.357}$$

où encore une fois $\partial_t\psi_t$ est la fonction $x\mapsto(\partial_t\psi)(t,x)$. Maintenant nous utilisons la proposition 29.61 pour dire que

$$\frac{d}{dt}\Big(u_t(\psi_t)\Big) = u_t^{(1)}(\psi_t) + u_t\left(\frac{\partial\psi}{\partial t}(t,.)\right) \tag{31.358}$$

pour écrire

$$0 = -\int_{\mathbb{R}}\frac{d}{dt}\big(u_t(\psi_t)\big) - u_t^{(1)}(\psi_t) + u_t\big(i(\Delta\psi)(t,.)\big)dt \tag{31.359}$$

Le premier terme est facile :

$$\int_{\mathbb{R}}\frac{d}{dt}\Big(u_t(\psi_t)\Big)dt = \lim_{N\to\infty}\Big[u_t(\psi_t)\Big]_{t=-N}^{t=N} = 0 \tag{31.360}$$

parce que ψ est à support compact par rapport à t. Nous restons donc avec

$$\int_{\mathbb{R}} u_t^{(1)}(\psi_t) - iu_t\big((\Delta\psi)(t,.)\big)dt = 0 \tag{31.361}$$

Nous traitons le terme en $u_t^{(1)}$ en utilisant le fait évident $T(\varphi) = (\mathcal{F}T)(\mathcal{F}^{-1}\varphi)$ et en remarquant le lemme 29.62 :

$$u_t^{(1)}(\psi_t) = (\mathcal{F}u_t^{(1)})(\mathcal{F}^{-1}\psi_t) = (\mathcal{F}u)_t^{(1)}(\mathcal{F}^{-1}\psi_t). \tag{31.362}$$

Pour l'autre terme on fait un peu la même chose en nous souvenant ce que fait la transformée de Fourier en traversant le laplacien :

$$u_t(\Delta\psi_t) = (\mathcal{F}u_t)(\mathcal{F}^{-1}\Delta\psi_t) = (\mathcal{F}u_t)\big(x\mapsto -\|x\|^2(\mathcal{F}^{-1}\psi_t)(x)\big). \tag{31.363}$$

En recollant encore :

$$\int_{\mathbb{R}} (\mathcal{F}u)_t^{(1)}(\mathcal{F}^{-1}\psi_t) + i(\mathcal{F}u_t)\big(\|.\|^2\mathcal{F}^{-1}\psi_t\big)dt = 0. \tag{31.364}$$

Cette équation est valable tant que $\psi \in \mathscr{S}(\mathbb{R}\times\mathbb{R}^d)$ avec support compact en t. Nous allons nous en créer une super cool. D'abord nous choisissons $\varphi \in \mathscr{S}(\mathbb{R}^d)$ et $\chi \in \mathscr{D}(\mathbb{R})$ et nous considérons [22]

$$\psi(t,x) = \mathcal{F}\Big(\xi \mapsto e^{it\|\xi\|^2}\varphi(\xi)\chi(t)\Big)(x). \tag{31.365}$$

Notons que la transformée de Fourier conserve le fait qu'une fonction soit Schwartz [23], mais pas le fait d'avoir support compact. Cependant nous ne prenons que la transformée de Fourier par rapport à x. Le résultat est donc une fonction ψ qui est Schwartz par rapport à x et support compact par rapport à t. Nous pouvons donc écrire (31.364) en utilisant la fonction (31.365) :

$$0 = \int_{\mathbb{R}} (\mathcal{F}u)_t^{(1)}\Big(x \mapsto e^{it\|x\|^2}\varphi(x)\chi(t)\Big) + i(\mathcal{F}u_t)\Big(x \mapsto \|x\|^2 e^{it\|x\|^2}\varphi(x)\chi(t)\Big)dt. \tag{31.366}$$

Là dedans, $\chi(t)$ peut sortir à la fois de la transformée de Fourier et de l'application des distributions ; il doit seulement rester dans l'intégrale. Dans le second terme nous allons utiliser l'égalité (due entre autre à la proposition 29.61) :

$$\begin{aligned}
\frac{d}{dt}\big(\hat{u}_t(e^{it\|.\|^2}\varphi)\big) &= \frac{d}{dt}\Big(u_t\big(\mathcal{F}e^{it\|.\|^2}\varphi\big)\Big) && (31.367a)\\
&= u_t^{(1)}\big(\mathcal{F}e^{it\|.\|^2}\varphi\big) + u_t\Big(\frac{\partial}{\partial t}\mathcal{F}e^{it\|.\|^2}\varphi\Big) && (31.367b)\\
&= (\mathcal{F}u_t^{(1)})\big(x \mapsto e^{it\|x\|^2}\varphi(x)\big) + (\mathcal{F}u_t)\big(x \mapsto i\|x\|^2 e^{it\|x\|^2}\varphi(x)\big) && (31.367c)\\
&= (\mathcal{F}u)_t^{(1)}\big(x \mapsto e^{it\|x\|^2}\varphi(x)\big) + (\mathcal{F}u_t)\big(x \mapsto i\|x\|^2 e^{it\|x\|^2}\varphi(x)\big). && (31.367d)
\end{aligned}$$

Et là, magie c'est exactement ce qui est dans (31.366). Donc

$$\int_{\mathbb{R}} \frac{d}{dt}\hat{u}_t\big(x \mapsto e^{it\|x\|^2}\varphi(x)\big)\chi(t)dt = 0 \tag{31.368}$$

pour toute fonctions à support compact χ. Donc la proposition 29.1 nous dit que

$$\partial_t\hat{u}_t\big(x \mapsto e^{it\|x\|^2}\varphi(x)\big) = 0. \tag{31.369}$$

C'est zéro partout et non seulement presque partout parce qu'en plus nous avons la continuité. Par conséquent pour tout $t \in \mathbb{R}$ nous avons

$$\hat{u}_t\big(x \mapsto e^{it\|x\|^2}\varphi(x)\big) = \hat{u}_0\big(x \mapsto \varphi(x)\big) = 0. \tag{31.370}$$

Et cela est vrai pour toute fonction $\varphi \in \mathscr{S}(\mathbb{R}^d)$. Nous considérons donc $t_0 \in \mathbb{R}$ et une fonction $\theta \in \mathscr{S}(\mathbb{R}^d)$ pour construire

$$\varphi(x) = e^{-it_0\|x\|^2}\theta(x). \tag{31.371}$$

Nous avons alors $\hat{u}_{t_0}\big(x \mapsto \theta(x)\big) = 0$, ce qui signifie que $\hat{u}_{t_0} = 0$. Du coup pour tout $\theta \in \mathscr{S}(\mathbb{R}^d)$ nous avons $u_{t_0}(\mathcal{F}\theta) = 0$, mais comme la transformée de Fourier est une bijection de $\mathscr{S}(\mathbb{R}^d)$ (proposition 28.26) nous avons en fait $u_{t_0}(\theta) = 0$ pour tout $\theta \in \mathscr{S}(\mathbb{R}^d)$, c'est-à-dire $u_{t_0} = 0$ pour tout $t_0 \in \mathbb{R}$ et au final $u = 0$.

□

22. Le candidat qui parvient à effectivement présenter ça comme développement, il est fort.
23. Proposition 28.20.

31.12 Équations différentielles du premier ordre

Définition 31.42 (Équation différentielle du premier ordre)**.**
*Une **équation différentielle du premier ordre** est une équation qui, sur un intervalle donné, I, décrit la relation entre une variable réelle, notée x ou t dans I, une fonction $y : I \to \mathbb{R}$, et la dérivée première de y qui on note y'.*

Souvent on écrit «$y'(x) =$ une formule contenant x et $y(x)$», c'est à dire

$$y'(x) = f(x, y(x)), \quad \text{pour } x \in I, \tag{31.372}$$

où f est une fonction de deux variables réelles.

Remarque 31.43.
La théorie des fonctions de deux variables ne sera pas abordée dans ce cours, nous allons nous contenter de prendre f dans (31.372) comme une simple notation.

On peut presque toujours omettre d'écrire la dépendance de y en x et écrire simplement (31.372) sous la forme $y' = f(x, y)$.

Définition 31.44 (Solution particulière d'une équation différentielle du premier ordre)**.**
*Une **solution particulière** de l'équation (31.372) sur l'intervalle I est une fonction $z : I \to \mathbb{R}$ telle que :*

(1) z est dérivable sur I ;

(2) $z'(x) = f(x, z(x))$, pour tout $x \in I$.

Définition 31.45 (Solution générale d'une équation différentielle du premier ordre)**.**
*Résoudre une équation différentielle veut dire trouver l'ensemble qui contient toutes ses solutions particulières. Cet ensemble s'appelle **solution générale** de l'équation.*

Exemple 31.46. (1) Résoudre une équation du type $y'(x) = f(x)$<++> revient à trouver l'ensemble des primitives de la fonction f, qui est donc la solution générale de cette équation. Il y a donc une infinité de solutions particulières, déterminées par une constante additive.
Si $f(x) = \sin(x)$ alors la solution générale sera $\mathcal{Y} = \{-\cos(x) + C \,:\, C \in \mathbb{R}\}$.

(2) L'équation

$$y' = y, \qquad x \in \mathbb{R}, \tag{31.373}$$

a peut-être été abordée dans votre cours de terminale lors de la définition de la fonction exponentielle. Sa solution générale est $\mathcal{Y} = \{Ce^x \,:\, C \in \mathbb{R}\}$. Ici aussi il y a une infinité de solutions particulières.

△

Remarque 31.47.
La solution générale d'une équation différentielle du premier ordre est une famille à un paramètre de fonctions.

Définition 31.48 (Équation différentielle du second ordre)**.**
*Une **équation différentielle du second ordre** est une équation qui, sur un intervalle donne, I, décrit la relation entre une variable réelle, notée x ou t dans I, une fonction $y : I \to \mathbb{R}$, et les dérivées première et seconde de y qui on note y' et y'' respectivement.*

On utilise la forme générale

$$y'' = f(x, y, y'), \quad \textit{pour } x \in I. \tag{31.374}$$

où f est une fonction de trois variables réelles.

On peut définir de manière analogue les équations différentielles d'ordre supérieur. Les définitions de solution particulière et de solution générale se généralisent aux équations différentielles d'ordre supérieur à un.

Définition 31.49 (Trajectoire)**.**
La trajectoire tracée par une solution particulière y de l'équation (31.372) *est le graphe de y en tant que fonction de x.*

Exemple 31.50.
Nous allons regarder de plus près l'équation (31.373), $y' = y$, pour tout $x \in \mathbb{R}$. Soient y_1 et y_2 deux solutions distinctes de cette équation. S'il existe un point $\bar{x}$ tel que $y_1(\bar{x}) = y_2(\bar{x})$ alors forcement $y_1(\bar{x})/y_2(\bar{x}) = 1$. Or, la solution générale de l'équation est $\mathcal{Y} = \{Ce^x : C \in \mathbb{R}\}$, donc $y_i(x) = C_i e^x$, $i = 1, 2$, où les C_i sont des constantes. Le rapport $y_1(\bar{x})/y_2(\bar{x})$ vaut C_1/C_2 et par conséquent $C_1 = C_2$. Ce résultat contredit l'hypothèse que les deux solutions soient distinctes. On a donc montré que *deux trajectoires distinctes de cette équation ne se croisent jamais.*

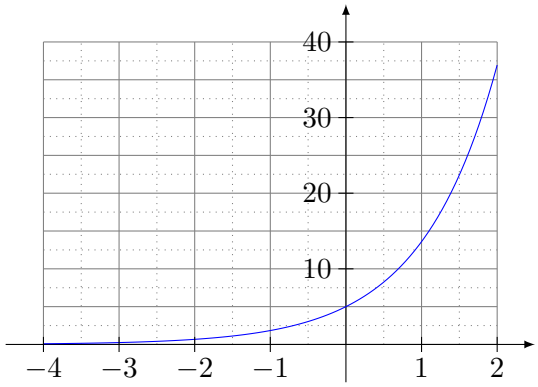


FIGURE 31.1 – Quelques trajectoires de l'équation $y' = y$.

La figure 31.1 représente quelques trajectoires de l'équation. Si on les avait tracées toutes elles recouvriraient tout le plan x-y. Cela veut dire que *par tout point (x, y) passe une et une seule trajectoire de l'équation* (31.373).

△

Définition 31.51 (Condition initiale)**.**
*Une **condition initiale** pour l'équation* (31.372) *sur l'intervalle I est un point $(\bar{x}, \bar{y}) \in I \times \mathbb{R}$.*

On dit que la solution particulière z de (31.372) *satisfait la condition initiale $(\bar{x}, \bar{y}) \in I \times \mathbb{R}$ si $z(\bar{x}) = \bar{y}$.*

Définition 31.52 (Problème de Cauchy)**.**
*L'association d'une équation différentielle et d'une condition initiale est appelée **problème de Cauchy***

$$\begin{cases} y' = f(x, y), & x \in I, \\ y(\bar{x}) = \bar{y}. \end{cases} \tag{31.375}$$

Remarque 31.53.
Sous des conditions assez générales qui seront toujours vérifiées dans ce cours, tout problème de Cauchy admet une et une seule solution.

Pour passer de la solution générale d'une équation différentielle de premier ordre à une solution particulière il faut choisir une valeur du paramètre. Comme il y a un seul paramètre une seule condition (la trajectoire de la solution doit passer par un point fixe du plan) peut suffire. Pour une équation différentielle de second ordre comme (31.374), nous aurons besoin de plus de conditions. Sans rentrer dans les détails, nous allons constater ce fait dans l'exemple suivant.

Exemple 31.54.
La solution générale de l'équation

$$y'' = -y, \tag{31.376}$$

est $\mathcal{Y} = \{C_1 \cos(x) + C_2 \sin(x) \ : \ C_1, C_2 \in \mathbb{R}\}$. Remarquez que l'équation est du second ordre et que sa solution générale est une famille d'équations à deux paramètres réels. Ce sera toujours les cas pour les équations abordées dans la section 31.15. Pour déterminer une solution particulière de (31.376) il faut fixer les valeurs des deux paramètres et donc, en général, il sera nécessaire de donner deux conditions. △

Remarque 31.55.
Une condition comme $y(0) = 4$ nous dit que la constante $C_1 = 4$ mais elle ne nous permet pas de trouver C_2. Il y a donc une infinité de solutions de (31.376) qui satisfont à la condition $y(0) = 4$.

On peut fixer les deux conditions de deux manières différentes.

(1) Problème de Cauchy : on fixe une terne de valeurs réels $\bar{x}, \bar{y}, \bar{y}'$ et on cherche la solution telle que $y(\bar{x}) = \bar{y}$, $y'(\bar{x}) = \bar{y}'$.

Exemple 31.56.
Les conditions $y(0) = 4$, $y'(0) = 15$ permettent de trouver la solution $z(x) = 4\cos(x) + 15\sin(x)$. △

(2) Problème aux bords : on fixe deux points dans le plan x-y, $A = (\bar{x}, \bar{y})$ et $B = (\tilde{x}, \tilde{y})$, et on cherche la solution dont la trajectoire passe par A et B, c'est à dire, on impose $y(\bar{x}) = \bar{y}$, $y(\tilde{x}) = \tilde{y}$.

Exemple 31.57.
Les conditions $y(0) = 4$, $y(\pi/2) = 15$ permettent de trouver la solution $z(x) = 4\cos(x) + 15\sin(x)$. △

31.13 Premier ordre, variables séparables

Pour certaines équations différentielles la recherche d'une solution particulière se réduit à une recherche de primitive moyennant un changement de variables.

Définition 31.58 (Équation différentielle du premier ordre à variables separables)**.**
*Une **équation différentielle du premier ordre à variables séparables** est une équation qui, pour tout les x dans un intervalle donné, I, peut se mettre sous la forme*

$$f(y)y' = g(x), \tag{31.377}$$

où f et g sont deux fonctions de $\mathbb{R}$ dans $\mathbb{R}$.

Nous pouvons intégrer les deux côtés de l'égalité par rapport à x et obtenir

$$\int f(y(x))y'(x)\,dx = G(x) + C,$$

où G est une primitive de g et C une constante réelle. Il est facile à ce point d'effectuer un changement de variable dans le membre de gauche de l'équation en posant (sans surprise) $y = y(x)$ et donc $y'(x)\,dx = dy$.

$$\int f(y(x))y'(x)\,dx = \int f(y)\,dy = F(y(x)) + C,$$

où F est une primitive de f et C une constante réelle. En somme nous avons

$$F(y(x)) = G(x) + C,$$

et, si F admet une fonction réciproque, alors

$$y(x) = F^{-1}(G(x) + C). \tag{31.378}$$

Remarque 31.59.
L'expression de F^{-1} peut être difficile à calculer. Il sera alors préférable de garder y dans la forme implicite.

Exemple 31.60.
L'équation

$$3y^2y' = x, \qquad \text{pour tout } x \in \mathbb{R}, \tag{31.379}$$

est une équation à variables séparables. Pour reprendre les notations du début du chapitre, ici $f(y) = 3y^2$ et $g(x) = x$. En intégrant de deux côtés on trouve

$$y^3 = \frac{x^2}{2} + C.$$

La fonction $F(y) = y^3$ est une bijection de $\mathbb{R}$ dans $\mathbb{R}$, donc nous pouvons écrire la solution générale de l'équation (31.379) dans la forme

$$\mathcal{Y} = \left\{ \left(\frac{x^2}{2} + C \right)^{1/3} \text{ tel que } C \in \mathbb{R} \right\}.$$

△

Exemple 31.61.
En intégrant de deux côtés l'équation à variables séparables

$$2yy' = x, \text{pour tout } x \in \mathbb{R}, \tag{31.380}$$

on trouve

$$y^2 = \frac{x^2}{2} + C.$$

La fonction $F(y) = y^2$ est *n'est pas inversible* sur tout $\mathbb{R}$, et on sait que $\sqrt{y^2} = |y|$. Au moment de rendre y explicite on doit choisir entre

$$y = \left(\frac{x^2}{2} + C \right)^{1/2} \qquad \text{ou} \qquad y = -\left(\frac{x^2}{2} + C \right)^{1/2}.$$

Ce choix se fait suivant la condition initiale, si elle est donnée. S'il n'y a pas de condition initiale nous pouvons écrire que la solution générale est l'ensemble

$$\mathcal{Y} = \left\{ y : \mathbb{R} \to \mathbb{R} \text{ tels que } y^2 = \frac{x^2}{2} + C \text{ et } C \in \mathbb{R} \right\}.$$

△

Exemple 31.62.
On considère le problème de Cauchy

$$\begin{cases} e^y y' = \frac{1}{x+3}, & x \in]-\infty, -3[, \\ y(-4) = 0. \end{cases} \tag{31.381}$$

En intégrant des deux côtés nous trouvons

$$e^y = \ln(|x+3|) + C.$$

Nous pouvons alors imposer la condition initiale et obtenir $e^0 = \ln(|-4+3|) + C$, c'est à dire $C = 1 - \ln(1) = 1$.

Remarque 31.63.
L'énoncé du problème de Cauchy dit que x peut varier dans $]-\infty, -3[$, mais nous voyons maintenant que la solution n'est pas définie sur toute la demi-droite, parce que e^y est toujours positif et $\ln(|x+3|)+1$ est positif seulement pour $x < -(1/e+3) \approx -3,3679$.

Donc la solution du problème de Cauchy est $y(x) = \ln(|x+3|)+1$ pour tout $x \in]-\infty, -(1/e+3)[$. △

Exemple 31.64.
Attention, cet exemple est le plus important de la section !
On considère l'équation à variables séparables

$$y' = \sin(x)y, \qquad x \in \mathbb{R}. \tag{31.382}$$

Dans ce cas, pour pouvoir écrire l'équation dans la forme (31.377) il faut pouvoir multiplier les deux côtés par $1/y$. Il faut donc éliminer tout de suite le cas où $y = 0$.

Si $y = 0$ alors $y' = 0$ et on a une solution constante (on dit souvent : une solution stationnaire) de l'équation. Par ailleurs les trajectoires des solutions ne peuvent pas se croiser ; donc si y_G est une solution non nulle de l'équation (31.382) alors $y_G(x) \neq 0$ pour tout x[24]. Il n'y a donc aucun danger à diviser par y dans la recherche d'une solution non identiquement nulle.

Supposons maintenant que $y \neq 0$ et écrivons $y'/y = \sin(x)$. En intégrant des deux côtés on trouve

$$\ln(|y|) = -\cos(x) + C,$$

d'où

$$|y| = e^{-\cos(x)+C} = e^C e^{-\cos(x)}.$$

Si on avait impose une condition initiale alors on pourrait déterminer une solution particulière de l'équation en choisissant une valeur de la constante C. Nous pouvons observer cependant que la fonction exponentielle est bijective de $\mathbb{R}$ dans $\mathbb{R}^{+,\star}$ et par conséquent il n'y a pas de perte de généralité en disant que la solution générale de l'équation est

$$\mathcal{Y} = \left\{ y \,:\, |y| = Ke^{-\cos(x)}, \text{ pour } K \in \mathbb{R}^{+,\star} \right\} \cup \{y \equiv 0\}.$$

Il n'empêche qu'il serait plus élégant d'écrire la solution générale de l'équation sous une forme plus explicite, sans valeur absolue. Nous pouvons le faire en nous nous rappelant que

$$|x| = \begin{cases} x & \text{si } x \geqslant 0, \\ -x & \text{si } x < 0, \end{cases}$$

Il suffit alors d'autoriser K dans $\mathbb{R}^\star$ pour éliminer la valeur absolue.

Pour écrire la solution générale de façon encore plus compacte nous observons que si $K = 0$ alors $y \equiv 0$, c'est à dire, on retrouve la solution constante nulle.

Finalement, la solution générale de cette équation sera toujours écrite sous la forme suivante

$$\mathcal{Y} = \left\{ y = Ke^{-\cos(x)}, \text{ pour } K \in \mathbb{R} \right\}. \tag{31.383}$$

△

31.14 Équations différentielles linéaires du premier ordre

Définition 31.65 (Équation différentielle linéaire du premier ordre)**.**
Soit $I \subset \mathbb{R}$ un intervalle .

24. Ça vaut la peine de prendre un peu de temps pour bien comprendre cela.

Une ***équation différentielle linéaire du premier ordre*** *est une équation différentielle de la forme*

$$a(x)y' + b(x)y = c(x), \quad \text{pour } x \in I, \tag{31.384}$$

où a, b, c sont des fonctions de $\mathbb{R}$ dans $\mathbb{R}$ et $a \neq 0$ pour tout $x \in I$.

On dit que a, b, c sont les coefficients de l'équation (31.384).

Remarque 31.66.
Une fonction $f : \mathbb{R} \to \mathbb{R}$ est dite *linéaire* si pour tout x_1, x_2 dans $\mathbb{R}$ et pour tout couple de constantes λ et μ on a

$$f(\lambda x_1 + \mu x_2) = \lambda f(x_1) + \mu f(x_2). \tag{31.385}$$

Ces équations différentielles sont dites linéaires parce que la partie de l'équation qui contient y (le membre de gauche) satisfait la propriété (31.385) par rapport à y. En effet par les propriétés de la dérivée nous avons que

$$a(x)(\lambda y_1 + \mu y_2)' + b(x)(\lambda y_1 + \mu y_2) = \lambda(a(x)y_1' + b(x)y_1) + \mu(a(x)y_2' + b(x)y_2).$$

Définition 31.67.
L'équation (31.384) est dite ***homogène*** *quand c est la fonction nulle. Si (31.384) n'est pas homogène on dit que l'équation*

$$a(x)y' + b(x)y = 0, \tag{31.386}$$

est son ***équation homogène associée****.*

Toute équation linéaire du premier ordre homogène est une équation du premier ordre à variables séparables, comme nous en avons vu l'exemple 31.64. Nous n'allons pas répéter les détails du procédé pour trouver sa solution générale, qui aura la forme suivante

À retenir 31.68

$$\mathcal{Y}_h = \left\{ K e^{-\int \frac{b(x)}{a(x)}\,dx} \ : \ K \in \mathbb{R} \right\}. \tag{31.387}$$

Proposition 31.69. *(1) Soit y_p une solution particulière de l'équation (31.384) et y_h une solution particulière de l'équation homogène associé (31.386). Alors la fonction somme $z = y_p + y_h$ est encore une solution particulière de l'équation (31.384).*

(2) Soient y_1 et y_2 deux solutions particulières de (31.384). Alors la fonction différence $w = y_1 - y_2$ est une solution particulière de (31.386).

Démonstration. (1)

$$a(x)\left(y_p + y_h\right)' + b(x)\left(y_p + y_h\right) - c(x) = \big(a(x)y_p' + b(x)y_p - c(x)\big) + \big(a(x)y_h' + b(x)y_h\big) = 0. \tag{31.388}$$

(2)

$$a(x)\left(y_1 - y_2\right)' + b(x)\left(y_1 - y_2\right) = \big(a(x)y_1' + b(x)y_1 - c(x)\big) - \big(a(x)y_2' + b(x)y_2 - c(x)\big) = 0. \tag{31.389}$$

□

Cette proposition permet de démontrer le théorème suivant, qui est le plus important de cette section.

Théorème 31.70.
Soit y_p une solution particulière de l'équation (31.384) et $\mathcal{Y}_h$ la solution générale de l'équation (31.386), alors la solution générale de l'équation (31.384) est l'ensemble

$$\mathcal{Y} = \mathcal{Y}_h + y_p = \{z = y_h + y_p \,:\, y = h \in \mathcal{Y}_h\}\,. \tag{31.390}$$

À retenir 31.71
La résolution d'une équation différentielle linéaire du premier ordre comporte trois étapes :

(1) résolution de l'équation homogène associée ;
(2) recherche d'une solution particulière de l'équation non homogène ;
(3) somme de la solution générale de l'équation homogène et de la solution particulière trouvée au point précédent.

La partie qui nous manque encore est de savoir comment trouver une solution particulière de l'équation non homogène (31.384). Si la fonction c dans (31.384) est une constante ou un polynôme simple, ou une exponentielle alors on peut essayer de deviner. Cette méthode cependant n'est pas la plus sure pour des débutants.

Exemple 31.72.
On considère l'équation

$$y' - 5y = 10, \qquad x \in \mathbb{R}. \tag{31.391}$$

Comme tous les coefficients de l'équation sont constants on peut essayer de trouver une solution constante.

Toutes les fonctions constantes ont une dérivée nulle, par conséquent, si une solution constante existe elle doit satisfaire $-5y = 10$, ce qui veut dire que la solution constante est $y(x) \equiv -2$. △

Exemple 31.73.
On considère l'équation

$$xy' + y = x + 1, \qquad x \in \mathbb{R}^{+,\star}. \tag{31.392}$$

Comme le membre de droite de l'équation est un polynôme de degré un on cherche une solution de la forme $y(x) = Ax + B$ avec A et B dans $\mathbb{R}$.

Par substitution on obtient $Ax + (Ax + B) = x + 1$, c'est-à-dire que une solution particulière de l'équation est $y(x) = x/2 + 1$. △

Exemple 31.74.
L'équation

$$xy' - y = x + 1, \qquad x \in \mathbb{R}^{+,\star}. \tag{31.393}$$

ressemble beaucoup à celle de l'exemple précédent, cependant il n'existe pas un polynôme de degré un qui en soit solution.

Dans un cas comme celui-ci, il faut rapidement abandonner la divination et replier sur la méthode, plus technique mais plus sure, dite *variation de la constante*. △

31.14.1 Méthode de variation de la constante

— Soit $\mathcal{Y}_h$ la solution générale de l'équation homogène associé à (31.384). Il s'agit d'une famille à un paramètre de fonctions. La première étape de cette méthode consiste à construire un candidat solution particulière y_p en remplaçant le paramètre dans $\mathcal{Y}_h$ par une fonction $C : \mathbb{R} \to \mathbb{R}$ à déterminer.

Exemple 31.75.
L'équation homogène associée à $y' - y = \cos(x)$ est $y' - y = 0$, dont la solution générale

est $\mathcal{Y}_h = \{Ce^x \ : \ C \in \mathbb{R}\}$. Le candidat solution sera alors $y_p = C(x)e^x$, avec C fonction à déterminer. △

— La deuxième étape de cette méthode consiste à injecter y_p dans l'équation. Cela permet de trouver une équation différentielle à variables séparables pour C, en principe plus facile à résoudre que l'équation de départ.

Exemple 31.76.
On continue avec l'exemple précédent. On a $y_p' = C'(x)e^x + C(x)e^x$, d'où

$$(C'(x)e^x + C(x)e^x) - C(x)e^x = \cos(x),$$

c'est-à-dire

$$C'(x) = \cos(x)e^{-x}.$$

△

— La troisième étape de la méthode consiste à trouver une solution particulière de l'équation différentielle pour C et, par conséquent déterminer une y_p.

Exemple 31.77.
La solution générale de

$$C'(x) = \cos(x)e^{-x}.$$

est $\mathcal{C} = \left\{e^{-1}\frac{(\sin(x)-\cos(x))}{2} + K \ : \ K \in \mathbb{R}\right\}$. Il nous suffit une solution particulière, nous pouvons donc choisir $K = 0$ et alors la solution particulière de (31.384) sera $y_p(x) = \frac{\sin(x)-\cos(x)}{2}$. △

Remarque 31.78.
Le plus souvent en intégrant l'équation pour C on en trouvera la solution générale. Dans ce cas on peut remplacer C par cette solution générale et obtenir d'un seul coup la solution générale de l'équation (31.384) , c'est-à-dire sans faire la somme entre la solution générale de l'homogène associée et la solution particulière.

Exemple 31.79.
Dans l'exemple qu'on vient de voir la solution générale de (31.384) est

$$\mathcal{Y} = \mathcal{Y}_h + y_p = \left\{Ce^x + \frac{(\sin(x) - \cos(x))}{2} \ : \ C \in \mathbb{R}\right\}. \tag{31.394}$$

On obtient le même résultat est écrivant $\mathcal{Y} = \left\{e^{-x}\left(e^{-1}\frac{(\sin(x)-\cos(x))}{2} + K\right) \ : \ K \in \mathbb{R}\right\}$. Notez qu'on a changé le nom du paramètre de C à K seulement pour souligner qu'on obtient de même résultat par deux chemins différents, sinon les deux expressions sont équivalentes ! △

31.15 Équations différentielles linéaires du second ordre

Définition 31.80 (Équation différentielle linéaire du second ordre)**.**
*Une **équation différentielle linéaire du second ordre** est une équation différentielle de la forme*

$$a(x)y'' + b(x)y' + c(x)y = d(x), \quad \textit{pour } x \in I, \tag{31.395}$$

où a, b, c et d sont des fonctions de $\mathbb{R}$ dans $\mathbb{R}$ et $a \neq 0$ pour tout $x \in I$.

On dit que a, b, c et d sont les coefficients de l'équation (31.395).

Dans ce cours nous allons étudier exclusivement le cas où a, b et c sont des fonctions constantes.

Définition 31.81 (Équation différentielle linéaire du second ordre homogène)**.**
*Une **équation différentielle linéaire du second ordre homogène** est une équation différentielle de la forme* (31.395)*, telle que le coefficient d est nul.*

À toute équation de la forme (31.395) on peut associer une équation homogène exactement comme on a fait dans la section précédente pour les équations linéaires du premier ordre.

31.15.1 Équations différentielles linéaires du second ordre homogènes à coefficients constants

Remarque 31.82.
L'application qui à la fonction y fait correspondre $a(x)y'' + b(x)y' + c(x)y$ est linéaire, au sens de la remarque 31.66.

Cela nous dit en particulier, que si y_1 et y_2 sont deux solutions de l'équation homogène alors toute leur combinaison de la forme $z = \lambda y_1 + \mu y_2$, avec λ et μ dans $\mathbb{R}$, est encore une solution.

Jusqu'ici nous avons toujours travaillé avec des fonctions définies sur $\mathbb{R}$ et à valeurs dans $\mathbb{R}$. Dans cette section nous nous autorisons à passer par des fonctions définies sur $\mathbb{R}$ et à valeurs dans $\mathbb{C}$, mais cela sera uniquement une étape dans nos calculs. Au final toutes les solutions que nous allons considérer sont des fonctions à valeurs dans $\mathbb{R}$.

La solution générale **à valeurs dans les complexes** d'une équation de ce type a la forme

$$\mathcal{Y}_h^{\mathbb{C}} = \{C_1 e^{r_1 x} + C_2 e^{r_2 x} \,:\, C_1,\, C_2 \in \mathbb{C},\, x \in I\}\,, \tag{31.396}$$

où r_1 et r_2 sont aussi des nombres complexes. Remarquez que la solution générale est une famille à deux paramètres. Il faut aussi observer que en tout cas l'intervalle I dans lequel varie x est un intervalle dans $\mathbb{R}$, parce que I est une des données du problème.

À partir de cette information nous pouvons, pour toute équation donnée, chercher la solution générale **complexe** par substitution. Il suffit de remplacer y dans l'équation par e^{rx} et chercher les valeurs de r qui nous conviennent.

Si notre équation de départ est

$$ay'' + by' + cy = 0, \quad \text{pour } x \in I, \tag{31.397}$$

alors la substitution nous donne

$$e^{rx}\left(ar^2 + br + c\right) = 0.$$

Il est connu que la fonction exponentielle ne prend pas la valeur 0, par conséquent ce qui s'annule est le polynôme de degré deux $ar^2 + br + c$. Il est donc très facile de trouver les valeurs de r qu'on pourra utiliser comme r_1 et r_2 dans la solution générale **complexe**.

Si $b^2 - 4ac > 0$ **:** le polynôme admet deux solutions réelles et distinctes, r_1 et r_2 ;

Si $b^2 - 4ac < 0$ **:** le polynôme admet deux solutions complexes conjuguées, $r_1 = \alpha + i\beta$ et $r_2 = \alpha - i\beta$;

Si $b^2 - 4ac = 0$ **:** le polynôme admet une solution réelle double $r = r_1 = r_2$.

Il faut maintenant écrire la solution générale **réelle** de l'équation, qui est celle que nous intéresse vraiment. La façon de l'obtenir est différente dans les trois cas.

Si $b^2 - 4ac > 0$ **:** la solution générale réelle a la même forme que la solution complexe, (31.396), il suffit de prendre les paramètres C_1 et C_2 dans $\mathbb{R}$ plutôt que dans $\mathbb{C}$.

$$\mathcal{Y}_h = \{C_1 e^{r_1 x} + C_2 e^{r_2 x} \,:\, C_1,\, C_2 \in \mathbb{R},\, x \in I\}\,, \tag{31.398}$$

Si $b^2 - 4ac < 0$ **:** le polynôme admet deux solutions complexes conjuguées, $r_1 = \alpha + i\beta$ et $r_2 = \alpha - i\beta$; Il faut alors utiliser les formules suivantes

$$\begin{aligned} e^{\alpha+i\beta} &= e^{\alpha}(\cos(\beta) + i\sin(\beta)) \\ e^{\alpha-i\beta} &= e^{\alpha}(\cos(\beta) - i\sin(\beta)). \end{aligned} \tag{31.399}$$

La somme $e^{r_1 x} + e^{r_2 x}$, où x est dans $I \in \mathbb{R}$, vaut

$$e^{(\alpha+i\beta)x} + e^{(\alpha-i\beta)x} = e^{\alpha x}(\cos(\beta x) + i\sin(\beta x)) + e^{\alpha x}(\cos(\beta x) - i\sin(\beta x)) = 2e^{\alpha x}\cos(\beta x)$$

et la différence $e^{r_1 x} - e^{r_2 x}$ vaut

$$e^{(\alpha+i\beta)x} - e^{(\alpha-i\beta)x} = e^{\alpha x}(\cos(\beta x) + i\sin(\beta x)) - e^{\alpha x}(\cos(\beta x) - i\sin(\beta x)) = 2e^{\alpha x}\sin(\beta x).$$

Par ces deux calculs élémentaires nous avons trouvé deux fonctions à valeurs dans $\mathbb{R}$ qui n'ont pas de zéros en commun. Elles sont les génératrices de la famille des solutions réelles de l'équation différentielle (la solution générale)

$$\mathcal{Y}_h = \{e^{\alpha x}\left(C_1\cos(\beta x) + C_2\sin(\beta x)\right) \,:\, C_1,\, C_2 \in \mathbb{R},\, x \in I\}\,, \tag{31.400}$$

Si $b^2 - 4ac = 0$ **:** le polynôme admet une solution réelle double $r = r_1 = r_2$. Dans ce cas la solution générale de l'équation est la famille

$$\mathcal{Y}_h = \{(C_1 + C_2 x)e^{rx} \,:\, C_1,\, C_2 \in \mathbb{R},\, x \in I\}\,. \tag{31.401}$$

Pour justifier cette formule nous observons d'abord que toute fonction $x \mapsto Ce^{rx}$, pour $C \in \mathbb{R}$ est une solution de l'équation différentielle (par construction). Ensuite nous utilisons la méthode de variation de la constante. On trouve rapidement que si une fonction de la forme $x \mapsto C(x)e^{rx}$ est une solution alors $C(x)$ est un polynôme de degré au plus 1, c'est-à-dire $C(x) = C_1 + C_2 x$ avec C_1 et C_2 dans $\mathbb{R}$.

31.15.2 Linéaires du second ordre à coefficients constants, non homogènes

Nous ne présentons pas une méthode générale pour la résolution de ces équations. Comme dans le cas des équations différentielles linéaires du premier ordre non homogènes, la solution générale de (31.395) est donnée par la somme d'une solution particulière et de la solution générale de l'équation homogène associée. La recherche d'une solution particulière est facilitée par le fait que les coefficients de (31.395) sont supposés constants, c'est-à-dire que a, b et c sont des fonctions constantes. Il faut essayer de deviner la forme d'une solution particulière à partir de la forme du second membre de l'équation, la fonction d. Si d est un polynôme il faut essayer avec un polynôme du même degré, si d est une exponentielle, par exemple $d(x) = e^{5x}$, on pourra essayer avec un multiple de la même fonction exponentielle, dans l'exemple $f(x) = ke^{5x}$, avec k à determiner. Si d est une combinaison linéaire de sinus et cosinus, comme par exemple $12\cos(x) + 2\sin(x)$, on peut essayer avec $k_1\cos(x) + k_2\sin(x)$.

Exemple 31.83.
On considère l'équation différentielle

$$y'' + 12y' + 36y = -192e^{2x}, \quad x \in \mathbb{R}. \tag{31.402}$$

Son équation homogène associée est

$$y'' + 12y' + 36y = 0, \tag{31.403}$$

dont le polynôme caractéristique est $r^2 + 12r + 36$. Ce polynôme admet une racine double, qui est -6, par conséquent la solution générale de (31.403) est

$$\mathcal{Y}_h = \left\{(C_1 + C_2 x)e^{-6x} \,:\, C_1,\, C_2 \in \mathbb{R},\, x \in \mathbb{R}\right\}.$$

Le membre de droite de (31.402) est une fonction exponentielle, nous allons donc chercher une solution particulière de (31.402) de la forme $f(x) = ke^{2x}$. Par substitution nous trouvons

$$ke^{2x}(4 + 12 \times 2 + 36) = -192e^{2x},$$

ce qui veut dire que k doit être -3.

La solution générale de l'équation (31.402) est donc

$$\mathcal{Y} = \left\{(C_1 + C_2x)e^{-6x} - 3e^{2x} \,:\, C_1,\, C_2 \in \mathbb{R},\, x \in \mathbb{R}\right\}.$$

△

Exemple 31.84.
Nous allons résoudre l'équation

$$y'' + 12y' + 36y = 12\cos(x) + 2\sin(x), \quad x \in \mathbb{R}. \tag{31.404}$$

Cette équation a comme homogène associée l'équation (31.403), comme dans l'exemple précédent. Il nous suffit donc de trouver une solution particulière de (31.402).

Nous pouvons essayer avec $f(x) = k_1\cos(x) + k_2\sin(x)$. Par substitution on trouve

$$\begin{aligned} -\left(k_1\cos(x) + k_2\sin(x)\right) + 12\left(-k_1\sin(x) + k_2\cos(x)\right) + 36\left(k_1\cos(x) + k_2\sin(x)\right) \\ = 12\cos(x) + 2\sin(x) \end{aligned}$$

Cette équation doit être satisfaite pour toute valeur de x, en particulier pour $x = 0$ et $x = \pi/2$. Cela revient à considère séparément les coefficients des fonctions sinus et cosinus. Il faut alors que k_1 et k_2 soient solutions du système

$$\begin{cases} -k_1 + 12k_2 + 36k_1 & = 12, \\ -k_2 - 12k_1 + 36k_2 & = 2. \end{cases}$$

On trouve $k_1 = 396/1369$ et $k_2 = 214/1369$, et la solution générale de notre équation est

$$\mathcal{Y} = \left\{(C_1 + C_2x)e^{-6x} + \frac{396}{1369}\cos(x) + \frac{214}{1369}\sin(x) \,:\, C_1,\, C_2 \in \mathbb{R},\, x \in \mathbb{R}\right\}.$$

△

Exemple 31.85.
Nous allons résoudre l'équation

$$y'' + 12y' + 36y = 10x^2 + 3, \quad x \in \mathbb{R}. \tag{31.405}$$

Cette équation a comme homogène associée l'équation (31.403), comme dans l'exemple précédent. Il nous suffit donc de trouver une solution particulière de (31.402).

Nous pouvons essayer avec $f(x) = k_1x^2 + k_2x + k_3$. Par substitution on trouve

$$(2k_1) + 12\left(2k_1x + k_2\right) + 36\left(k_1x^2 + k_2x + k_3\right) = 10x^2 + 3.$$

Pour trouver les bonnes valeurs des coefficients nous devons résoudre le système

$$\begin{cases} 36k_1 & = 10, \\ 24k_1 + 36k_2 & = 0, \\ 2k_1 + 12k_2 + 36k_3 & = 3, \end{cases}$$

ce qui donne $k_1 = 5/18$, $k_2 = -5/27$ et $k_3 = 7/54$. La solution générale de notre équation est

$$\mathcal{Y} = \left\{(C_1 + C_2x)e^{-6x} + \frac{5}{18}x^2 - \frac{5}{27}x + \frac{7}{54} \,:\, C_1,\, C_2 \in \mathbb{R},\, x \in \mathbb{R}\right\}.$$

△

31.16 Fonction de Green

Soit l'équation différentielle

$$\begin{cases} y''(x) = g(x) & (31.406a) \\ y(0) = y(1) = 0 & (31.406b) \end{cases}$$

pour $x \in]0,1[$ et où g est continue sur $]0,1[$.

Nous définissons la fonction de Green

$$G(x,t) = \begin{cases} t(x-1) & \text{si } 0 \leqslant t \leqslant x \leqslant 1 \\ x(t-1) & \text{si } 0 \leqslant x \leqslant t \leqslant 1, \end{cases} \tag{31.407}$$

et nous allons montrer que

$$y(x) = \int_0^1 G(x,t)g(t)dt \tag{31.408}$$

est l'unique solution.

Unicité Si y_1 et y_2 sont des solutions, alors $y_1'' = y_2''$ et donc $y_1(x) = y_2(x) + ax + b$. Les conditions aux bords donnent alors $0 = y_1(0) = y_2(0) + b = b$. D'où $b = 0$. En imposant $y_1(1) = 0$ nous trouvons alors immédiatement $a = 0$, ce qui donne $y_1 = y_2$.

Existence Il est vite vérifié qu'avec (31.408) nous avons $y(0) = y(1) = 0$ parce que $G(0,t) = G(1,t) = 0$ pour tout t. Nous fixons une valeur pour $x \in]0,1[$ et nous découpons l'intégrale :

$$y(x) = \int_0^x G(x,t)g(t)dt + \int_x^1 G(x,t)g(t)dt. \tag{31.409}$$

Pour calculer $y'(x)$, il faut dériver à la fois à travers l'intégrale et dans la borne. Si vous connaissez une formule pour faire cela, c'est bien pour vous. Nous allons faire ça à la main et poser

$$I(x,y) = \int_0^y t(x-1)g(t)dt. \tag{31.410}$$

La dérivation de I par rapport à x se fait en utilisant le théorème 16.25 :

$$\frac{\partial I}{\partial x}(x,y) = \int_0^y tg(t)dt. \tag{31.411}$$

Pour la dérivation par rapport à y, il s'agit du théorème fondamental de l'analyse, plus précisément le lien primitive et intégrale de la proposition 13.235 :

$$\frac{\partial I}{\partial y}(x,y) = y(x-1)g(y). \tag{31.412}$$

Maintenant nous considérons la fonction $\varphi_I(x) = I(x,x)$. Elle satisfait à

$$\varphi_I'(x) = \frac{\partial I}{\partial x}(x,x) + \frac{\partial I}{\partial y}(x,x) = \int_0^x tg(t) + x(x-1)g(x). \tag{31.413}$$

Le même jeu avec $J(x,y) = \int_y^1 x(t-1)g(t)dt$ donne

$$\varphi_J'(x) = \int_0^x fg(t)dt + x(x-1)g(x). \tag{31.414}$$

En remettant les bouts ensemble,

$$y(x) = \int_0^x tg(t)dt + \int_1^x (1-t)g(t)dt. \tag{31.415}$$

Le calcul de la dérivée seconde donne alors

$$y''(x) = xg(x) + (1-x)g(x) = g(x). \tag{31.416}$$

Nous pouvons aussi, sur cette équation, estimer la variation de la solution en termes d'une variation de g. Soit donc une fonction continue δ_g sur $[0,1]$ et $\tilde{g} = g + \delta_g$. Nous considérons l'équation différentielle

$$\begin{cases} \tilde{y}''(x) = \tilde{g}(x) & (31.417a) \\ \tilde{y}(0) = \tilde{y}(1) = 0. & (31.417b) \end{cases}$$

Par ce que nous venons de faire, l'unique solution est

$$\tilde{y}(x) = \int_0^1 G(x,t)\tilde{g}(t)dt = \int_0^1 G(x,t)g(t)dt + \int_0^1 G(x,t)\delta_g(t)dt = y(x) + \delta_y(x) \tag{31.418}$$

où δ_y est une fonction continue ainsi définie :

$$\delta_y(x) = \int_0^1 G(x,t)\delta_g(t)dt. \tag{31.419}$$

Supposons que $\|\delta_g\|_\infty = \epsilon$. Alors des majorations donnent

$$|\delta_y(x)| \leqslant \epsilon \int_0^1 |G(x,t)|dt = \epsilon(1-x)\int_0^x tdt + \epsilon x \int_x^1 (1-t)dt = \frac{\epsilon}{2}x(1-x). \tag{31.420}$$

Mais la fonction $x \mapsto x(1-x)$ a son maximum en $x = \frac{1}{2}$, donc nous pouvons donner une majoration indépendante de x :

$$\|\delta_y\|_\infty \leqslant \frac{1}{8}\|\delta_g\|_\infty. \tag{31.421}$$

Notons que la majoration (31.421) en norme uniforme a l'air plus impressionnante, mais la majoration (31.420) donnant une majoration séparée pour chaque x est en réalité plus précise.

Chapitre 32

Équations aux dérivées partielles

32.1 Symbole principal, équation des caractéristiques

Soit l'équation différentielle semi-linéaire d'ordre k

$$\sum_{|\alpha|=k} a_\alpha(x)(\partial^\alpha u)(x) + F\Big(x, u(x), (Du)(x), \dots, (D^{k-1}u)(x)\Big) = 0 \tag{32.1}$$

pour la fonction $u\colon \mathbb{R}^d \to \mathbb{R}$.

Définition 32.1.
*Le **symbole principal** de l'équation (32.1) est l'application*

$$\begin{aligned} \sigma\colon \mathbb{R}^d \times \mathbb{R}^d &\to \mathbb{R} \\ (x,\xi) &\mapsto \sum_{|\alpha|=k} a_\alpha(x)\xi^\alpha \end{aligned} \tag{32.2}$$

où si $\alpha = (\alpha_1, \dots, \alpha_d)$ et $\xi = (\xi_1, \dots, \xi_d)$ alors $\xi^\alpha = \xi_1^{\alpha_1} \dots \xi_d^{\alpha_d}$.

Définition 32.2.
*Les **caractéristiques** de l'équation (32.1) est une surface S de $\mathbb{R}^d$ donné par une équation de la forme $\phi(x) = 0$ où ϕ satisfait à*

$$\sigma\big(x, \nabla\phi(x)\big) = 0 \tag{32.3}$$

et $\nabla\phi(x) \neq 0$ pour tout $x \in S$.

32.2 Méthode des caractéristiques pour l'ordre 1

Nous[499, 500] voulons étudier l'équation d'ordre 1

$$a(x,y)\frac{\partial u}{\partial x}(x,y) + b(x,y)\frac{\partial u}{\partial y}(x,y) + c(x,y)u(x,y) = f(x,y) \tag{32.4}$$

Le champ de vecteurs associé à cette équation est

$$v = \begin{pmatrix} a \\ b \end{pmatrix}, \tag{32.5}$$

et l'équation peut être écrite sous la forme

$$(v \cdot \nabla) + cu = f. \tag{32.6}$$

Définition 32.3.
*Le **flot** de ce champ de vecteurs sont les courbes paramétriques $\gamma(t) = \big(x(t), y(t)\big)$ vérifiant $\gamma' t(t) = v\big(\gamma(t)\big)$.*

Les équations du flot pour l'équation (32.4) sont

$$\begin{cases} x'(t) = a\big(x(y), y(t)\big) & (32.7a) \\ y'(t) = b\big(x(y), y(t)\big). & (32.7b) \end{cases}$$

Ce sont des équations différentielles ordinaires. Un système de deux équations couplées du premier ordre.

Quel est l'intérêt du flot ? Nous allons voir que sur la ligne $t \mapsto \gamma(t)$, la fonction u est constante. Or des solutions γ au système (32.7), il y en aura plusieurs : une pour chaque valeur des constantes d'intégration. Pour peu que ces lignes recouvrent tout le plan, nous pourrons résoudre l'équation de départ ligne par ligne.

Nous posons

$$\tilde{u}(t) = u\big(x(t), y(t)\big) \tag{32.8a}$$
$$\tilde{c}(t) = c\big(x(t), y(t)\big) \tag{32.8b}$$
$$\tilde{f}(t) = f\big(x(t), y(t)\big). \tag{32.8c}$$

La fonction $\tilde{u}$ est une fonction $\mathbb{R} \to \mathbb{R}$ normale qui se dérive normalement, en suivant la règle de dérivation des fonctions composées :

$$\begin{aligned} \tilde{u}'(t) &= \frac{\partial u}{\partial x}\big(x(t), y(t)\big)x'(t) + \frac{\partial u}{\partial y}\big(x(t), y(t)\big)y'(t) && (32.9a) \\ &= a\frac{\partial u}{\partial x} + b\frac{\partial u}{\partial y} && (32.9b) \\ &= f\big(x(t), y(t)\big) - c\big(x(t), y(t)\big)u\big(x(t), y(t)\big) && (32.9c) \\ &= \tilde{f}(t) - \tilde{c}(t)\tilde{u}(t). && (32.9d) \end{aligned}$$

Nous avons pour $\tilde{u}$ l'équation différentielle ordinaire

$$\tilde{u}' + \tilde{c}\tilde{u} = \tilde{f} \tag{32.10}$$

qui est résolue par la proposition 31.10.

32.2.1 Un exemple complet un peu minimal

Nous considérons l'équation différentielle[499]

$$\frac{\partial u}{\partial x} - \frac{\partial u}{\partial y} - (x - y)u = 0. \tag{32.11}$$

Et nous allons la résoudre.

Les équations du flot, sont simples parce que les coefficients sont des constantes : $x'(t) = 1$, $y'(t) = -1$. Donc

$$x(t) = t + C_1 \tag{32.12a}$$
$$y(t) = -t + C_2. \tag{32.12b}$$

À priori nous avons une caractéristique pour chaque choix de (C_1, C_2) et nous espérons que le tout recouvre le plan $\mathbb{R}^2$. En fait seule une des deux constantes doit être laissée libre, l'autre consiste seulement en décaler le paramètre t. Nous posons donc $C_1 = 0$ et nous considérons les courbes caractéristiques

$$\gamma_C(t) = \begin{pmatrix} t \\ -t + C \end{pmatrix}. \tag{32.13}$$

Ces courbes recouvrent bien tout le plan. Pour savoir les valeurs de u sur la courbe γ_C, nous devons résoudre l'équation différentielle ordinaire

$$\tilde{u}_C' + \tilde{c}\tilde{u}_C = \tilde{f}, \tag{32.14}$$

en sachant que $\tilde{c}(t) = c\big(x(t), y(t)\big) = -\big(x(t) - y(t)\big) = 2t - C$. Cela se fait en suivant la méthode décrite dans l'exemple 31.8 et résumée dans la proposition 31.10.

En termes de notations, $\tilde{u}_C(t) = u\big(\gamma_C(t)\big)$. Récrivons l'équation :

$$\tilde{u}'(t) - (2t - C)\tilde{u}(t) = 0. \tag{32.15}$$

La méthode pour la résoudre est de mettre les $\tilde{u}$ d'un côté et les t de l'autre :

$$\frac{\tilde{u}'}{u} = 2t - C. \tag{32.16}$$

En intégrant par rapport à t des deux côtés,

$$\ln(\tilde{u}) = t^2 - Ct + K_C, \tag{32.17}$$

c'est-à-dire (avec redéfinition de K_C)

$$\tilde{u}(t) = K_C e^{t^2 - Ct} \tag{32.18}$$

ou encore

$$u\big(\gamma_C(t)\big) = K_C e^{t^2 - Ct} \tag{32.19}$$

où C est le paramètre que nous déterminons en sachant sur quelle caractéristique se trouve le point (x, y) où nous voulons calculer $u(x, y)$ et K est une constante (strictement positive parce que si vous avez suivi le mouvement, c'est une exponentielle) qui doit être déterminée par les conditions initiales. Dès que K est fixé pour un des points de la courbe γ_C, alors il est fixé pour tous les points.

Ce que nous avons obtenu est qu'il existe un K_C tel que pour tout t nous avons

$$u\big(\gamma_C(t)\big) = K_C e^{t^2 - Ct}. \tag{32.20}$$

Soit donc un point $(x_0, y_0) \in \mathbb{R}^2$. Nous devons d'abord déterminer où ce point se trouve par rapport aux caractéristiques, c'est-à-dire quelle est la valeur de C pour laquelle (x_0, y_0) est sur la courbe γ_C, et ensuite déterminer pour quelle valeur de t nous aurons $\gamma_C(t) = (x_0, y_0)$. À résoudre :

$$\gamma_C(t_0) = \begin{pmatrix} t_0 \\ -t_0 + C \end{pmatrix} = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}. \tag{32.21}$$

Donc $t_0 = x_0$ et $C = x_0 + y_0$. En reprenant (32.19) nous avons

$$u\big(\gamma_C(t_0)\big) = K e^{x_0^2 - (x_0 + y_0)x_0} = K e^{-x_0 y_0}. \tag{32.22}$$

Pour peu que des conditions soient donnée sur chaque caractéristique, nous pouvons déterminer K. Attention : ce K est une constante d'intégration de l'équation différentielle ordinaire pour $\tilde{u}$. Donc elle n'est valable que sur chaque caractéristique séparément. Cela n'est donc pas du tout une constante sur $\mathbb{R}^2$.

Nous pouvons maintenant écrire la solution générale de l'équation de départ. L'équation cartésienne de la courbe γ_C est

$$x + y = C. \tag{32.23}$$

Donc K est une fonction de $x + y$, pas de x et y séparément. Cela est important à comprendre. À priori nous avons

$$u(x, y) = K(x, y)e^{-xy} \tag{32.24}$$

où $K(x, y)$ est constante sur la courbe γ_C contenant (x, y). Nous avons

— si $x_1 + y_1 = x_2 + y_2$,
— alors il existe C tel que (x_1, y_1) et (x_2, y_2) sont sur γ_C,
— alors $K(x_1, y_2) = K(x_2, y_2)$.

Donc il existe une fonction $\mathbb{R} \to \mathbb{R}$ telle que $K(x, y) = f(x + y)$.

Au final, la solution générale de l'équation est

$$u(x, y) = f(x + y)e^{-xy} \tag{32.25}$$

où f est une fonction à déterminer par les conditions initiales qui peuvent être données. Typiquement nous espérons que les conditions imposent une et une seule valeur de u sur chacune des courbes γ_C.

32.2.2 Un théorème d'existence et d'unicité

La méthode des caractéristiques donne essentiellement une preuve de l'unicité des solutions aux équations de transport, et une méthode pour construire cette solution. En effet, la procédure suivante permet de construire $u(x_0, y_0)$.

— Trouver la caractéristique passant par le point (x_0, y_0)

— Calculer en quel point elle passe par une condition initiale donnée.

— Attribuer à $u(x_0, y_0)$ la valeur trouvée sur la caractéristique là où elle passe par une condition initiale.

Rien ne permet à priori de savoir que cette procédure construit effectivement une solution. En particulier, comment calculer $\partial_x u$? Le quotient différentiel serait

$$\frac{\partial u}{\partial x}(x, y) = \lim_{\epsilon \to 0} \frac{u(x + \epsilon, y) - u(x, y)}{\epsilon}, \tag{32.26}$$

mais la caractéristique donnant la valeur de $u(x + \epsilon, y)$ est différente pour chaque ϵ. Rien à priori ne permet d'affirmer que le calcul soit simple, ni qu'il arrive à une solution du problème donné.

D'où la nécessité d'avoir un résultat un peu rigoureux donnant des conditions sous lesquelles les choses vont bien.

Proposition 32.4 (Équation de transport à coefficients variables[500, 491])**.**
Soit une fonction $c\colon \mathbb{R}^2 \to \mathbb{R}$ de classe C^2 en ses deux variables et uniformément Lipschitziennes en sa première variable[1] *et $g \in C^1(\mathbb{R})$. Alors l'équation aux dérivées partielles de premier ordre*

$$\begin{cases} \dfrac{\partial u}{\partial x}(x, t) + c(x, t)\dfrac{\partial u}{\partial t}(x, y) = 0 & \text{(32.27a)} \\ u(x, 0) = h(y) & \text{(32.27b)} \end{cases}$$

admet une unique solution de classe C^1.

Cette solution est construite de la façon suivante[2]*. D'abord nous considérons la solution X au problème*

$$\begin{cases} \dfrac{\partial X}{\partial s}(s; x, t) = c\big(X(s; x, t), s\big) & \text{(32.28a)} \\ X(t; x, t) = x, & \text{(32.28b)} \end{cases}$$

et ensuite le problème (32.27) a pour unique solution

$$u(x, t) = h\big(X(0; x, t)\big). \tag{32.29}$$

Démonstration. Nous commençons par étudier l'existence et l'unicité de la fonction X définie par le problème 32.28. La fonction c ici est dans les hypothèses de la fonction f du théorème de Cauchy-Lipschitz global 16.41. D'où l'existence et l'unicité de la fonction $s \mapsto X(s; x, t)$ sur $\mathbb{R}$ pour chaque (x, y) donné[3].

Le lemme 31.32 nous dit que X est de classe C^2 en (s, x, t). Donc nous pourrons dériver et permuter les dérivées autant que nous voudrons (sans exagérer : ordre 2 au maximum).

Unicité Nous montrons que u doit être constante sur le chemin

$$\gamma_{(x,t)}(s) = \begin{pmatrix} X(s; x, t) \\ s \end{pmatrix}. \tag{32.30}$$

1. Dans [491], on ne demande que continue puis uniformément Lipschitz. Moi je crois que ce n'est pas assez pour assurer la dérivabilité de X par rapport à x, et encore moins pour permuter les dérivées dans $\partial^2_{tx} Y$.

2. Le fait que la construction ait un sens fait partie des choses à prouver

3. Avec l'énoncé tel que donné dans [500], il faut utiliser la technique de 31.20 pour l'existence globale, parce que la fonction b là-dedans n'est pas dans les mêmes hypothèses.

En effet, en posant

$$\varphi(s) = u\big(\gamma(s)\big) = u\big(X(s;x,t),s\big), \tag{32.31}$$

et en dérivant nous obtenons

$$\varphi'(s) = \frac{\partial u}{\partial t}\big(X(s;x,t),s\big)\frac{\partial X}{\partial s}(s;x,t) + \frac{\partial u}{\partial t}\big(X(s;x,t)\big) = 0. \tag{32.32}$$

Par conséquent la valeur commune de tous les $u\big(\gamma_{(x,t)}(s)\big)$ doit être celle en $\gamma_{(x,t)}(0) = h(X(0;x,t))$

Cela prouve l'unicité parce que la valeur de u est fixée en tout point. Nous devons encore vérifier que la fonction u ainsi construite est bien une solution du problème. C'est l'objet de la partie « existence » de la preuve.

Existence Même l'existence est divisée en plusieurs étapes.

Mise en place Nous prouvons que la fonction u donné par (32.29) est une solution du problème. Nous avons :

$$\frac{\partial u}{\partial t}(x,t) = h'\big(X(0;x,t)\big)\frac{\partial X}{\partial t}(0;x,t) \tag{32.33}$$

et

$$\frac{\partial u}{\partial x}(x,t) = h'\big(X(0;x,t)\big)\frac{\partial X}{\partial x}(0;x,t), \tag{32.34}$$

de sorte qu'en posant

$$g(s;x,t) = \frac{\partial X}{\partial t}(s;x,t) + c(x,t)\frac{\partial X}{\partial x}(s;x,t) \tag{32.35}$$

nous avons

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = h'\big(X(0;x,t)\big)g(0;x,t). \tag{32.36}$$

Une équation différentielle pour g Nous allons prouver que

$$\frac{\partial g}{\partial s}(s;x,t) = \alpha_{(x,t)}(s)g(s;x,t) \tag{32.37}$$

avec [4]

$$\alpha_{(x,t)}(s) = \frac{\partial c}{\partial x}\big(X(s;x,t),s\big). \tag{32.38}$$

D'abord nous avons

$$\frac{\partial g}{\partial s}(s;x,t) = \frac{\partial^2 X}{\partial s\partial t}(s;x,t) + c(x,t)\frac{\partial^2 X}{\partial s\partial x}(s;x,t). \tag{32.39}$$

Nous permutons les dérivées et nous tenons compte de (32.28) :

$$\frac{\partial g}{\partial s}(s;x,t) = \frac{\partial}{\partial t}\Big(c\big(X(s;x,t),s\big)\Big) + c(x,t)\frac{\partial}{\partial x}\Big(c\big(X(s;x,t),s\big)\Big). \tag{32.40}$$

Nous dérivons maintenant plus en profondeur. D'une part

$$\frac{\partial}{\partial t}\Big(c\big(X(s;x,t),s\big)\Big) = \frac{\partial c}{\partial x}\Big(X(s;x,t),s\Big)\frac{\partial X}{\partial t}(s;x,t) \tag{32.41}$$

et d'autre part,

$$\frac{\partial}{\partial x}\Big(c\big(X(s;x,t),s\big)\Big) = \frac{\partial c}{\partial x}\Big(X(s;x,t),s\Big)\frac{\partial X}{\partial x}(s;x,t), \tag{32.42}$$

de telle sorte que

$$\frac{\partial g}{\partial s}(s;x,t) = \frac{\partial c}{\partial x}\big(X(s;x,t),s\big)\Big[\frac{\partial X}{\partial t}(s;x,t) + c(x,t)\frac{\partial X}{\partial x}(s;x,t)\Big] \tag{32.43a}$$

$$= \alpha_{(x,t)}(s)g(s;x,t). \tag{32.43b}$$

4. La ligne suivante est une de celles qui me font penser qu'il manque des hypothèses dans [491]. Il faut bien pouvoir dériver c.

Une condition initiale pour g Nous montrons maintenant que $g(t;x,t)=0$. La condition initiale pour X est $X(t;x,t)=x$ pour tout $t,x\in\mathbb{R}$. Nous dérivons cette dernière par rapport à x et à t :

$$\frac{\partial X}{\partial x}(t;x,t)=1 \tag{32.44a}$$

$$\frac{\partial X}{\partial s}(t;x,t)+\frac{\partial X}{\partial t}(t;x,t)=0. \tag{32.44b}$$

Mais $\partial_s X(t;x,t)=c\big(X(t;x,t),t\big)$, donc la relation (32.44b) donne

$$\frac{\partial X}{\partial t}(t;x,t)=-c\big(X(t;x,t),t\big)=-c(x,t) \tag{32.45}$$

où nous avons tenu compte du fait que $X(t;x,t)=x$.
Voyons à présent ce que (32.44a) et (32.45) donnent pour $g(t;x,t)$:

$$g(t;x,t)=\frac{\partial X}{\partial t}(t;x,t)+c(x,t)\frac{\partial X}{\partial x}(t;x,t)=-c(x,t)+c(x,t)=0. \tag{32.46}$$

Conclusion pour g La fonction g vérifie l'équation différentielle

$$\begin{cases}\dfrac{\partial g}{\partial s}(s)=\alpha(s)g(s) & (32.47a)\\ g(t)=0. & (32.47b)\end{cases}$$

Bien entendu, $g(s)=0$ est une solution. Mais la solution à ce système est unique par Cauchy-Lipschitz 16.40. Ici nous utilisons le fait que

$$(s,y)\mapsto\alpha(s)y \tag{32.48}$$

est continue. C'est-à-dire entre autres que

$$s\mapsto\frac{\partial c}{\partial x}\big(X(s;x,t),s\big) \tag{32.49}$$

doit être continue. C'est le cas parce que c est de classe C^1 en sa première variable.

□

Les hypothèses de la proposition 32.4 sont loin d'être optimales. Voici un exemple dans lequel c n'est même pas dérivable par rapport à t et qui se passe très bien quand même.

Exemple 32.5.
Soit l'équation différentielle

$$\begin{cases}\dfrac{\partial u}{\partial t}(x,t)+|t-1|\dfrac{\partial u}{\partial x}(x,t)=0 & (32.50a)\\ u(x,0)=h(x) & (32.50b)\end{cases}$$

où h est une fonction bien régulière ; mettons C^p. En suivant la méthode de la proposition nous devrions poser l'équation différentielle

$$\begin{cases}\dfrac{\partial X}{\partial s}(s;x,t)=|s-1| & (32.51a)\\ X(t;x,t)=x. & (32.51b)\end{cases}$$

Cela est la caractéristique passant par (x,t). Cependant il sera plus simple de chercher les caractéristiques en demandant qu'elles passent par $(x_0,0)$. Nous allons donc plutôt résoudre pour X_{x_0} l'équation différentielle

$$\begin{cases}\dfrac{\partial X}{\partial s}(s)=|s-1| & (32.52a)\\ X_{x_0}(0)=x_0 & (32.52b)\end{cases}$$

et les courbes caractéristiques seront les chemins

$$\gamma_{x_0}(s) = \begin{pmatrix} X_{x_0}(s) \\ s \end{pmatrix}. \tag{32.53}$$

La résolution donne d'abord

$$X_{x_0}(s) = \begin{cases} -\frac{s^2}{2} + s + K_1 & \text{si } s < 1 \\ \frac{s^2}{2} - s + K_2 & \text{si } s > 1. \end{cases} \tag{32.54}$$

Vu que la condition initiale est donnée pour $s = 0$, nous fixons K_1 pour la condition initiale et K_2 pour la continuité :

$$X_{x_0}(s) = \begin{cases} -\frac{s^2}{2} + s + x_1 & \text{si } s < 1 \\ \frac{1}{2} + x_0 & \text{si } s = 1 \\ \frac{s^2}{2} - s + 1 + x_0 & \text{si } s > 1. \end{cases} \tag{32.55}$$

Soit $(x, t) \in \mathbb{R}^2$. Quelle caractéristique passe par là ? Nous allons déterminer la fonction $x_0(x, t)$ qui donne le x_0 tel que la caractéristique γ_{x_0} passe par (x, t). Nous devons résoudre

$$\begin{pmatrix} X_{x_0}(s) \\ s \end{pmatrix} = \begin{pmatrix} x \\ t \end{pmatrix}. \tag{32.56}$$

Directement : $s = t$. Et ensuite $X_{x_0}(t) = x$. Nous avons

$$x_0(x, t) = \begin{cases} x + \frac{t^2}{2} - t & \text{si } t < 1 \\ x - \frac{1}{2} & \text{si } t = 1 \\ x - \frac{t^2}{2} + t - 1 & \text{si } t > 1. \end{cases} \tag{32.57}$$

Le truc presque étonnant est que x_0 est de classe C^1. En effet le calcul de

$$\frac{\partial x_0}{\partial t}(1) = \lim_{\epsilon \to 0} \frac{x_0(x, 1 + \epsilon) - x_0(x, 1)}{\epsilon} \tag{32.58}$$

se fait en séparant les limites $\epsilon \to 0^+$ et $\epsilon \to 0^-$. Le résultat est que $\partial_t x_0(1) = 0$. Nous avons donc

$$\frac{\partial x_0}{\partial t}(t) = \begin{cases} t - 1 & \text{si } t < 1 \\ 0 & \text{si } t = 1 \\ -t + 1 & \text{si } t > 1. \end{cases} \tag{32.59}$$

Cela étant continu, la fonction x_0 est de classe C^1 en t, et la dérivée en x étant toujours 1, elle est de classe C^1.

En ce qui concerne la solution de l'équation de départ,

$$u(x, t) = h\big(X_{x_0(x,t)}(0)\big) = h\big(x_0(x, t)\big). \tag{32.60}$$

Pourvu que h soit assez régulière, la fonction u est facilement de classe C^1. △

32.3 Méthode des caractéristiques pour l'ordre 2

32.3.1 Principe général

Soit l'opérateur différentiel agissant sur $C^2(\mathbb{R}^2)$:

$$D = a(x, y)\frac{\partial^2}{\partial x^2} + b(x, y)\frac{\partial^2}{\partial x \partial y} + c(x, y)\frac{\partial^2}{\partial y^2}. \tag{32.61}$$

Nous voulons résoudre des équations du type $Du = 0$ pour $u \colon \mathbb{R}^2 \to \mathbb{R}$.

Pour commencer[499], et c'est le point crucial, nous voyons D comme un polynôme en ∂_x et ∂_y et nous le factorisons : si

$$aX^2 + bXY + cY = (\alpha X + \beta Y)(\gamma X + \delta Y) \tag{32.62}$$

alors nous avons

$$D = \left(\alpha\frac{\partial}{\partial x} + \beta\frac{\partial}{\partial y}\right)\left(\gamma\frac{\partial}{\partial x} + \delta\frac{\partial y}{\partial}\right) + \text{ termes d'ordre inférieurs.} \tag{32.63}$$

Les « termes d'ordre inférieurs » sont ceux de la forme $\alpha(x,y)\frac{\partial\delta}{\partial x}\frac{\partial}{\partial y}$.

L'astuce est de poser

$$v = (\gamma\partial_x + \delta\partial_y)u, \tag{32.64}$$

et de résoudre le système

$$\begin{cases} (\alpha\partial_x + \beta\partial_y)v = 0 & (32.65a)\\ (\gamma\partial_x + \delta\partial_y)u = v. & (32.65b)\end{cases}$$

Cela sont deux équations différentielles du premier ordre pour lesquelles nous avons déjà des techniques décrites en la section 32.2.

Afin que les fonctions α, β, γ et δ soient réelles, il faut que $b^2 - 4ac \geqslant 0$. Sachant que $a = \alpha\gamma$, $b = \alpha\delta + \beta\gamma$ et $c = \beta\delta$ cette condition sur a, b et c donne

$$(\alpha\delta + \beta\gamma)^2 - 4\alpha\gamma\beta\delta \geqslant 0. \tag{32.66}$$

Cela revient à

$$(\alpha\delta - \beta\gamma)^2 \geqslant 0. \tag{32.67}$$

Nous supposons à présent que l'inégalité soit stricte (cas hyperbolique). Nous avons en particulier que

$$\alpha\delta - \beta\gamma \neq 0. \tag{32.68}$$

Cette condition implique que les équations

$$\frac{dx}{dt} = \alpha(x,y) \quad \frac{dy}{dt} = \beta(x,y) \tag{32.69}$$

sont indépendantes des équations

$$\frac{dx}{dt} = \gamma(x,y) \quad \frac{dy}{dt} = \delta(x,y) \tag{32.70}$$

Ce sont les équations caractéristiques des équations (32.65).

32.3.2 Exemple : l'équation d'onde

Nous considérons l'équation aux dérivées partielles

$$\frac{\partial^2 u}{\partial t^2} - c^2\frac{\partial^2 u}{\partial x^2} = 0 \tag{32.71}$$

où c est une constante réelle. Nous en cherchons des solutions de classe C^2.

L'opérateur différentiel est donné par le polynôme $P(T,X) = T^2 - c^2X^2$ qui se factorise en

$$P = (T + cX)(T - cX), \tag{32.72}$$

c'est-à-dire que nous pouvons récrire l'équation des ondes sous la forme

$$(\partial_t + c\partial_x)(\partial_t - c\partial_x)u = 0. \tag{32.73}$$

Nous posons donc $v = (\partial_t - c\partial_x)u$ et nous avons le système[490]

$$\begin{cases} (\partial_t + c\partial_x)v = 0 & (32.74a) \\ (\partial_t - c\partial_x)u = v. & (32.74b) \end{cases}$$

La méthode des caractéristiques est efficace pour résoudre la première, et pour trouver la solution générale de l'homogène associée à la seconde.

Nous nous lançons dans la résolution de (32.74a). Le flot est $v = \begin{pmatrix} 1 \\ c \end{pmatrix}$, et nous cherchons ses courbes intégrales sous la forme $\varphi(t) = \big(t, x(t)\big)$. Immédiatement, $x'(t) = c$, ce qui donne

$$\gamma_C(t) = \begin{pmatrix} t \\ ct + C \end{pmatrix}. \tag{32.75}$$

Cela donne une caractéristique pour chaque valeur de C. En posant $\tilde{v}_C(t) = v(t, ct + C)$ nous avons

$$\tilde{v}_C'(t) = \frac{\partial v}{\partial t}\big(\gamma_C(t)\big) + c\frac{\partial v}{\partial t}\big(\gamma_C(t)\big) = 0. \tag{32.76}$$

Donc $\tilde{v}_C$ est une fonction constante. Donc u est constant sur la courbe γ_C dont l'équation cartésienne est $x - ct = C$. Cela implique que

$$v(t, x) = f(x - ct) \tag{32.77}$$

où f est une fonction de classe C^1. En effet si (t_1, x_1) et (t_2, x_2) vérifient $x_1 - ct_1 = x_2 - ct_2$ alors $v(t_1, x_2) = v(t_2, x_2)$. Le fait que f soit C^1 est une demande que u soit au final dans C^2.

Nous devons maintenant résoudre l'équation (32.74b)

$$(\partial_t - c\partial_x)u = v. \tag{32.78}$$

Nous allons agir conformément à la stratégie expliquée par le lemme 31.6. Nous devons résoudre $Du = v$ avec

$$\begin{aligned} D\colon C^2(\mathbb{R}) &\to D(C^2(\mathbb{R})) \\ u &\mapsto (\partial_t - c\partial_x)u. \end{aligned} \tag{32.79}$$

Par la même méthode des caractéristiques que celle déjà menée plus haut nous trouvons $\ker(D)$ comme solution générale de $(\partial_t - c\partial_x)u_G = 0$. C'est-à-dire

$$u_G = g(x + ct) \tag{32.80}$$

où g est une fonction quelconque de classe C^2.

Il nous faut maintenant une solution particulière de

$$(\partial_t - c\partial_x)u_P(t, x) = f(x - ct). \tag{32.81}$$

Si F est une primitive de f alors

$$u_P(t, x) = -\frac{1}{2c}F(x - ct) \tag{32.82}$$

fonctionne. Vu que f est quelconque dans $C^1(\mathbb{R})$, la fonction F est un élément quelconque de $C^2(\mathbb{R})$. Au final, la solution générale de l'équation des ondes est

$$u(t, x) = g_1(x + ct) + g_2(x - ct) \tag{32.83}$$

où g_1 et g_2 sont des éléments de $C^2(\mathbb{R})$.

32.4 Classification des équations du second ordre

Soit une équation générale d'ordre 2 sur $\mathbb{R}^2$:

$$a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} + d\frac{\partial u}{\partial x} + e\frac{\partial u}{\partial y} + \beta u = f. \tag{32.84}$$

En ce qui concerne son symbole principal nous avons

$$\sigma(x, y, \xi_1, \xi_2) = a(x, y)\xi_1^2 + b(x, y)\xi_1\xi_2 + c(x, y)\xi_2^2, \tag{32.85}$$

ce qui donne l'équation des caractéristiques

$$a\left(\frac{\partial \phi}{\partial x}\right)^2 + b\frac{\partial \phi}{\partial x}\frac{\partial \phi}{\partial y} + c\left(\frac{\partial \phi}{\partial y}\right)^2 = 0. \tag{32.86}$$

Si nous nous posons sur un point (x_0, y_0) tel que $\nabla\phi(x_0, y_0) = 0$ et $\partial_x\phi(x_0, y_0) \neq 0$ alors, via le théorème de la fonction implicite [5], la condition $\phi(x, y) = 0$ définit une fonction $y \mapsto x(y)$ vérifiant

$$\phi\big(x(y), y\big) = 0 \tag{32.87}$$

pour tout y dans un voisinage de y_0.

Nous pouvons obtenir une équation différentielle ordinaire pour x de la façon suivante. D'abord nous posons $\varphi(y) = \phi\big(x(y), y\big)$ et ensuite nous calculons la dérivée de φ (qui est nulle par construction) :

$$0 = \varphi'(y) = \frac{\partial \phi}{\partial x}x' + \frac{\partial \phi}{\partial y}. \tag{32.88}$$

Nous pouvons donc remplacer $\partial_y\phi$ par $x'\partial_x\phi$ dans l'équation des caractéristiques (32.86) :

$$a\left(\frac{\partial \phi}{\partial x}\right)^2 + bx'\left(\frac{\partial \phi}{\partial x}\right)^2 + c\left(\frac{\partial \phi}{\partial x}\right)^2 (x')^2 = 0. \tag{32.89}$$

Vu que nous avons supposé $(\partial_x\phi) \neq 0$ sur un voisinage de (x_0, y_0) nous pouvons simplifier par $(\partial_x\phi)^2$ et avoir l'équation différentielle ordinaire

$$a\big(x(y), y\big) + b\big(x(y), y\big)x'(y) + c\big(x(y), y\big)x'(y)^2 = 0. \tag{32.90}$$

Notons que, conformément à ce que raconte le théorème des fonctions implicites, nous avons pris un voisinage de y_0 suffisamment petit pour que $x(y)$ reste dans un voisinage de x_0. Ce voisinage étant, nous pouvons le restreindre pour nous assurer du signe de a, b et c. Cela est évidemment très théorique parce que le théorème de la fonction implicite parle de l'existence de voisinages, mais pas de façon de les construire.

Nous donnons la classification suivante.

Définition 32.6.
*Si $b^2 - 4ac < 0$ alors l'équation est **elliptique**.*
*Si $b^2 - 4ac > 0$ alors l'équation est **hyperbolique**.*
*Si $b^2 - 4ac = 0$ alors l'équation est **parabolique**.*

Exemple 32.7.
Un exemple d'équation parabolique est l'équation de la chaleur

$$\frac{\partial u}{\partial t} - \alpha\frac{\partial^2 u}{\partial x^2} = 0 \tag{32.91}$$

où $\alpha > 0$ est une constante. Cette équation est avec $a = c = 0$, donc elle est parabolique. △

5. Théorème 16.49.

Une équation aux dérivées partielles peut changer de nature selon le point.

Exemple 32.8.
Soit l'équation

$$\frac{\partial^2 u}{\partial x^2} - (x^2 - y^2)\frac{\partial^2 u}{\partial y^2} = 0. \tag{32.92}$$

Son $b^2 - 4ac$ vaut $4(x^2 - y^2)$. Elle peut donc être hyperbolique, parabolique ou elliptique selon le point où l'on se trouve. △

32.4.1 Problème au limite

Dans les définitions qui suivent nous considérons un ouvert $\Omega \subset \mathbb{R}^d$ assez régulier et possédant en particulier un vecteur normal extérieur $n(x)$ pour tout point $x \in \partial\Omega$.

Définition 32.9.
Un problème aux limites ***de Dirichlet*** *est d'imposer la condition*

$$u(x) = g(x) \tag{32.93}$$

pour tout $x \in \Gamma \subset \partial\Omega$. C'est-à-dire imposer la valeur de u sur une partie du bord du domaine.

Définition 32.10.
Un problème aux limites de ***Von Neumann*** *est d'imposer*

$$\frac{\partial u}{\partial n} \cdot x = g(x) \tag{32.94}$$

pour tout $x \in \Gamma \subset \partial\Omega$. C'est-à-dire imposer les valeurs de la dérivée normale de u sur une partie du bord.

Exemple 32.11.
Lorsqu'on veut imposer un flux de chaleur aux bords d'un domaine pour l'équation de la chaleur, il s'agit de poser des conditions de type Von Neumann. △

Définition 32.12.
Soit un domaine Ω de $\mathbb{R}^d$ et un opérateur différentiel L sur une partie de $\mathrm{Fun}(\Omega)$*. Soit une fonction g sur $\partial\Omega$. Un problème* ***aux limites stationnaires*** *est un problème du type : trouver u définie sur Ω telle que*

$$\begin{cases} L(u) = f & (32.95a) \\ u|_{\partial\Omega} = g. & (32.95b) \end{cases}$$

Définition 32.13.
Un problème ***aux limites d'évolution*** *est du type : trouver $u \in \mathrm{Fun}\left(]0,\infty[\times \Omega\right)$ tel que*

$$\begin{cases} \frac{\partial^m u}{\partial t^m} + L(u) = f & \textit{sur }]0,\infty[\times \Omega \\ u(t,.) = g(t,.) & \textit{sur }]0,\infty[\times \partial\Omega \\ u(0,.) = u_0 & \textit{sur } \Omega \end{cases} \tag{32.96}$$

où u_0 est une fonction sur Ω.

L'opérateur L ne doit pas opérer sur la partie « t » de u.

Définition 32.14 (Problème bien posé au sens de Hadamard)**.**
Un problème aux limites est ***bien posé au sens de Hadamard*** *si*

(1) Il admet une unique solution.

(2) La solution dépend de façon continue en les données du problème.

La continuité est au sens des normes sur les fonctions sur $\partial\Omega$ et sur Ω en ce qui concerne les fonctions « données » du problème et des normes pour les fonctions sur Ω ou $\bar{\Omega}$ en ce qui concerne la solution.

Exemple 32.15 (Un problème de Dirichlet bien posé)**.**
Trouver la fonction u définie sur $\Omega = [0,1]^2$ telle que

$$\begin{cases} -\Delta u = 0 & \text{sur } \Omega \\ u = g & \text{sur } \partial\Omega. \end{cases} \tag{32.97}$$

△

32.5 Principe du maximum

Lemme 32.16 ([389])**.**
Soit un ouvert borné $\Omega \subset \mathbb{R}^n$. Soit une matrice symétrique strictement définie positive A telle que $A_{ij} \in C^O(\bar{\Omega})$ pour laquelle il existe $\lambda > 0$ minorant toutes les valeurs propres de toutes les matrices $A(x)$ pour $x \in \Omega$[6]*.*

Nous posons $L' = -\sum_{ij} A_{ij}\partial_{ij}$. Si $u \in C^2(\Omega)$ atteint un minimum local en $x_0 \in \Omega$, alors

$$(L'u)(x_0) \leqslant 0. \tag{32.98}$$

Démonstration. Nous allons bien entendu diagonaliser A. Si T est une matrice nous avons, en posant $u(x) = v(Tx)$:

$$\frac{\partial u}{\partial x_i}(x) = \sum_k \frac{\partial v}{\partial x_k}(Tx)T_{ki} \tag{32.99}$$

et

$$\frac{\partial^2 u}{\partial x_j \partial x_i}(x) = \sum_{kl} T_{ki}T_{lj}\frac{\partial^2 v}{\partial x_l \partial x_k}(Tx). \tag{32.100}$$

Si T est en particulier une matrice orthogonale diagonalisant A (théorème 9.150) nous avons $T^{-1} = T^t$ et

$$\sum_{ij} T_{ki}A_{ij}T_{jl}^{-1} = D_{kl} = \lambda_k\delta_{kl} \tag{32.101}$$

où les λ_k sont les valeurs propres de A. Notons que partout ici, tout est fonction de x sur Ω : tant A que T que les λ_k. Avec tous ces résultats nous calculons vite que

$$(Lu)(x_0) = -\sum_k \lambda_k(\partial_k^2 v)(Tx_0). \tag{32.102}$$

Si $x_0 \in \Omega$ est un minimum local de u, alors l'application v a un minimum local en $T^{-1}x_0$. Et donc

$$\frac{\partial^2 v}{\partial x_k^2}(Tx_0) \geqslant 0. \tag{32.103}$$

Du coup,

$$(Lu)(x_0) \leqslant 0. \tag{32.104}$$

□

Lemme 32.17 ([389])**.**
Soit un ouvert borné $\Omega \subset \mathbb{R}^n$. Soit une matrice symétrique strictement définie positive A telle que $A_{ij} \in C^O(\bar{\Omega})$ pour laquelle il existe $\lambda > 0$ minorant toutes les valeurs propres de toutes les matrices $A(x)$ pour $x \in \Omega$.

6. Cela est plus que dire que toutes les $A(x)$ son symétriques strictement définie positive.

Nous posons

$$L = -\sum_{ij} A_{ij}\partial_{ij}^2 + \sum_i b_i\partial_i + c \tag{32.105}$$

où $b_i, c \in C^0(\bar{\Omega})$.

Soit $u \in C^0(\bar{\Omega}) \cap C^2(\Omega)$ *telle que* $Lu \geqslant 0$ *sur* Ω. *Alors*

(1) Si $c = 0$ *alors*

$$\min_{\bar{\Omega}}(u) = \min_{\partial\Omega}(u). \tag{32.106}$$

(2) Si $c \geqslant 0$ *alors*

$$\min_{\bar{\Omega}}(u) \geqslant \min_{\partial\Omega}(-u_-) \tag{32.107}$$

où u_- *est défini par*

$$u_-(x) = \begin{cases} 0 & si\ u(x) \geqslant 0 \\ -u(x) & si\ u(x) \leqslant 0. \end{cases} \tag{32.108}$$

Démonstration. D'abord, vu que Ω est borné, la fermeture $\bar{\Omega}$ est compacte et u y atteint son minimum. De plus $\partial\Omega$ est également compact (borné et le complémentaire est ouvert parce que Ω est ouvert). Donc u y atteint également son minimum. Cela pour dire que les minimums écrits dans (32.106) et (32.107) ont un sens.

Pour (1).

<u>$Lu \geqslant \eta > 0$</u> Nous supposons qu'il existe $\eta > 0$ tel que $(Lu)(x) \geqslant \eta$ pour tout $x \in \Omega$. Soit x_0 le point minimum sur $\bar{\Omega}$. Si $x_0 \in \Omega$ alors il est intérieur et $\partial_i u(x_0) = 0$ par la proposition 16.69. Dans ce cas nous avons

$$(Lu)(x_0) = (L'u)(x_0) > 0, \tag{32.109}$$

ce qui contredit le lemme 32.16. Nous en déduisons que le minimum de u sur $\bar{\Omega}$ n'est pas atteint dans Ω, mais sur $\partial\Omega$. Pour la définition de la frontière, voir 7.239. Ici nous avons $\bar{\Omega}\backslash\Omega = \partial\Omega$.

Cela prouve (32.106) dans ce cas.

<u>$Lu \geqslant 0$ **sur** Ω</u> Nous prenons maintenant le cas général. Nous posons

$$u_{\gamma,\epsilon}(x) = u(x) - \epsilon e^{\gamma x_1}. \tag{32.110}$$

Nous avons[7]

$$L(e^{\gamma x_1}) = e^{\gamma x_1}\big(-A_{11}(x)\gamma^2 + b_1(x)\gamma\big). \tag{32.111}$$

Soit γ suffisamment grand pour que

$$\lambda\gamma^2 - \|b_1\|_{\bar{\Omega}}\gamma > 0. \tag{32.112}$$

Ici λ minore toutes les valeurs propres des $A(x)$ et nous notons que γ ne dépend pas de ϵ. En utilisant l'inégalité du lemme 9.1579.157, pour tout $\xi \in \mathbb{R}^n$ nous avons

$$\sum_{ij} A_{ij}\xi_i\xi_j \geqslant \lambda|\xi|^2, \tag{32.113}$$

ce qui donne avec $\xi = e_1$: $A_{11} \geqslant \lambda$, et même pour être plus précis : $A_{11}(x) \geqslant \lambda$ pour tout x. Ces inégalités donnent

$$-A_{11}(x)\gamma^2 + b_1(x)\gamma \leqslant -\lambda\gamma^2 + \|b_1\|_{\bar{\Omega}}\gamma < 0. \tag{32.114}$$

Nous avons donc

$$L(e^{\gamma x_1}) = e^{\gamma x_1}\big(-A_{11}(x)\gamma^2 + b_1(x)\gamma\big) \leqslant -\lambda\gamma^2 + \|b_1\|_{\bar{\Omega}}\gamma < 0. \tag{32.115}$$

7. Nous abusons un peu de l'écriture parce que ce que nous calculons vraiment est $L\big(x \mapsto e^{\gamma x_1}\big)$.

Nous posons

$$\eta = \epsilon\big(\lambda\gamma^2 - \|b_1\|_{\bar{\Omega}}\gamma\big)\min_{\bar{\Omega}}(e^{\gamma x_1}), \tag{32.116}$$

où le minimum a un sens parce que Ω est borné. Nous avons $\eta > 0$.

C'est le moment de calculer ce que u_ϵ peut pour nous :

$$\begin{aligned}
L(u_\epsilon) &= Lu - \epsilon L(e^{\gamma x_1}) && (32.117a)\\
&\geqslant -\epsilon L(e^{\gamma x_1}) && (32.117b)\\
&= -\epsilon e^{\gamma x_1}\big(-A_{11}(x)\gamma^2 + b_1(x)\gamma\big) && (32.117c)\\
&\geqslant -\epsilon e^{\gamma x_1}\big(\lambda\gamma^2 - \|b_1\|_{\bar{\Omega}}\gamma\big) && (32.117d)\\
&\geqslant \eta > 0. && (32.117e)
\end{aligned}$$

Justification :

— (32.117b) parce que $Lu \geqslant 0$.

— (32.117d) par (32.115).

La fonction u_ϵ est donc dans le cas précédent et nous avons

$$\min_{x\in\bar{\Omega}}\big(u(x) - \epsilon e^{\gamma x_1}\big) = \min_{x\in\partial\Omega}\big(u(x) - \epsilon e^{\gamma x_1}\big). \tag{32.118}$$

Mentionnons le fait que le choix fait de γ ne dépend pas de ϵ. Nous pouvons donc encore faire varier ϵ sans toucher à γ et en maintenant toutes les inégalités prouvées jusqu'ici.

Vu que l'expression $e^{\gamma x_1}$ est majorable sur $\bar{\Omega}$ nous avons convergence uniforme

$$u_\epsilon \xrightarrow{\|\cdot\|_{\bar{\Omega}}} u \tag{32.119}$$

pour $\epsilon \to 0$.

Supposons qu'aucun point de $\partial\Omega$ ne réalise le minimum de u. Alors il existe $x_0 \in \Omega$ tel que $u(x_0) > u(x)$ pour tout $x \in \partial\Omega$. Mais comme $\partial\Omega$ est compact, il existe $\eta > 0$ tel que

$$u(x_0) < u(x) + \eta. \tag{32.120}$$

Soit ϵ tel que $\|u - u_\epsilon\|_{\bar{\Omega}} < \eta/2$. Nous avons

$$u(x_0) < u(x) - \eta \tag{32.121}$$

et donc aussi

$$u_\epsilon(x_0) < u(x_0) + \frac{\eta}{2} < u(x) - \frac{\eta}{2} < u_\epsilon(x), \tag{32.122}$$

ce qui signifierait que u_ϵ prend son minimum dans Ω. Or nous savons que ce n'est pas le cas. Donc il existe un point de $\partial\Omega$ qui réalise le minimum de u.

Pour (2).

Supposons pour commencer que $u \geqslant 0$ sur Ω. Alors par continuité $u \geqslant 0$ sur $\bar{\Omega}$. Alors $u_- = 0$ et

$$\min_{\bar{\Omega}} u \geqslant 0, \tag{32.123}$$

ce qui fait que l'inégalité (32.107) est évidente.

Nous supposons donc que l'ensemble

$$\Omega_- = \{x \in \Omega \text{ tel que } u(x) < 0\} \tag{32.124}$$

est non vide. Notons que c'est également un ouvert.

Soit $\bar{L}u = Lu - cu$. Vu que $c \geqslant 0$ et que $Lu \geqslant 0$ nous avons $\bar{L}u \geqslant 0$ sur Ω_-. Par le point (1) nous avons

$$\min_{x\in\bar{\Omega}} u(x) = \min_{x\in\partial\Omega} u(x). \tag{32.125}$$

Mais le minimum de u est certainement atteint dans Ω_-, donc

$$\min_{x\in\bar{\Omega}_-} u(x) = \min_{x\in\bar{\Omega}} u(x). \tag{32.126}$$

Cela nous donne une première bonne égalité :

$$\min_{x\in\partial\Omega_-} u(x) = \min_{x\in\bar{\Omega}} u(x). \tag{32.127}$$

Nous pouvons la prolonger :

$$\begin{aligned}
\min_{x\in\bar{\Omega}} u(x) &= \min_{x\in\partial\Omega_-} u(x) && (32.128a)\\
&= \min_{x\in\partial\Omega_-} \big(-u_-(x)\big) && (32.128b)\\
&= \min_{x\in\partial\Omega_-\cap\partial\Omega} \big(-u_-(x)\big) && (32.128c)\\
&= \min_{x\in\partial\Omega} \big(-u_-(x)\big) && (32.128d)
\end{aligned}$$

Justifications :

— Pour (32.128c) nous avons la décomposition

$$\partial\Omega_- = \big(\partial\Omega_-\cap\Omega\big)\cup\big(\partial\Omega_-\cap\partial\Omega\big). \tag{32.129}$$

Or sur $\partial\Omega_-\cap\Omega$ nous avons $u(x)=0$ et donc pas le minimum.

— Pour (32.128d). Sur $\partial\Omega$, le minimum est atteint dans la partie $\partial\Omega_-$ parce que le reste ne contient que des valeurs positives de u.

Cela prouve ce que nous voulions. □

Théorème 32.18.
(Principe du maximum fort) Soit un ouvert borné Ω de $\mathbb{R}^n$. Soit une matrice A dont

— *A_{ij} est dans $C^0(\bar{\Omega})$*

— *$A(x)$ est symétrique strictement définie positive pour tout x.*

— *Il existe $\lambda>0$ minimisant toutes les valeurs propres des toutes les matrices $A(x)$ sur $\bar{\Omega}$.*

Soit $b_i, c\in C^0(\bar{\Omega})$ avec $c(x)\geqslant 0$ sur $\bar{\Omega}$.

Soit $u\in C^0(\bar{\Omega})\cap C^2(\Omega)$ telle que

$$\begin{cases}
-\sum_{ij} A_{ij}\partial_{ij}u(x) + \sum_i b_i\partial_i u(x) \geqslant 0 & (32.130a)\\
u(x)\geqslant 0\ \forall x\in\partial\Omega. & (32.130b)
\end{cases}$$

Alors $u\geqslant 0$ sur $\bar{\Omega}$.

Démonstration. Nous appliquons le lemme 32.17(2) :

$$\min_{\bar{\Omega}} u \geqslant \min_{\partial\Omega}(-u_-). \tag{32.131}$$

Vu que $u(x)\geqslant 0$ sur $\partial\Omega$, nous avons $u_-=0$ sur $\partial\Omega$ et donc $\min_{\bar{\Omega}} u\geqslant 0$. □

32.6 Quelques exemples

32.6.1 Un changement de variables

Soit l'équation différentielle

$$\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} = 0 \tag{32.132}$$

sur $\Omega =]0, \infty[\times \mathbb{R}$. Nous imposons la condition aux bords

$$u(0,x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 & \text{si } x > 0. \end{cases} \tag{32.133}$$

Nous cherchons les solutions sous la forme

$$u(t,x) = f\left(\frac{x}{\sqrt{t}}\right). \tag{32.134}$$

Nous avons

$$\frac{\partial u}{\partial t} = -\frac{1}{2} x t^{-3/2} f'\left(\frac{x}{\sqrt{t}}\right)) \tag{32.135}$$

ainsi que

$$\frac{\partial u}{\partial x} = \frac{1}{\sqrt{t}} f'\left(\frac{x}{\sqrt{t}}\right) \tag{32.136}$$

et

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{t} f''\left(\frac{x}{\sqrt{t}}\right). \tag{32.137}$$

En remettant le tout dans l'équation de départ et en simplifiant par $1/t$ (qui est permis parce que $t > 0$) :

$$-\frac{1}{2} f'\left(\frac{x}{\sqrt{t}}\right) \frac{x}{\sqrt{t}} - f''\left(\frac{x}{\sqrt{t}}\right) = 0. \tag{32.138}$$

Nous résolvons l'équation différentielle

$$\frac{z}{2} g'(z) + g''(z) \tag{32.139}$$

pour la fonction g de la variable réelle z. Cela fait, la réponse sera $f = g \circ z$ où z serait la fonction

$$z(x,t) = x/\sqrt{t}, \tag{32.140}$$

Pour résoudre (32.139) nous commençons par résoudre pour la dérivée $h = g'$, c'est-à-dire l'équation différentielle

$$\frac{z}{2} h(z) + h'(z) = 0 \tag{32.141}$$

qui donne

$$\frac{g'(z)}{g'z} = -z/2. \tag{32.142}$$

Une intégration fournit $\ln\big(h(z)\big) = -z^2/4 + K$ et donc

$$h(z) = K e^{-z^2/4} \tag{32.143}$$

et

$$g(z) = C + K \int_0^z e^{-s^2/4} ds. \tag{32.144}$$

Nous ne pouvons pas aller plus loin parce que nous ne sommes pas capables de calculer la primitive demandée. Nous laissons donc g sous cette forme et nous posons

$$u(x,t) = g(x/\sqrt{t}), \tag{32.145}$$

en pleine confiance du fait que cela soit une solution de l'équation aux dérivées partielles (32.132).

Nous devons fixer K et C de telle façon à respecter les conditions aux bords (32.133). D'abord écrivons aussi explicitement que possible la fonction u :

$$u(t,x) = K + C \int_0^{x\sqrt{t}} e^{-s^2/4} ds. \tag{32.146}$$

Il faut calculer $u(0,x)$ en termes de C et K. Pour cela nous calculons, pour un x fixé :

$$\lim_{t\to 0^+} u(t,x) = K + C \lim_{t\to 0^+} \int_0^{x/\sqrt{t}} e^{-s^2/4} ds. \tag{32.147}$$

En utilisant un petit changement de variables sur l'intégrale gaussienne de l'exemple 13.265, et en remarquant que la fonction est symétrique,

$$\lim_{t\to 0^+} u(t,x) = \begin{cases} K + C\sqrt{\pi} & \text{si } x > 0 \\ K - C\sqrt{\pi} & \text{si } x < 0. \end{cases} \tag{32.148}$$

À résoudre :

$$\begin{cases} K + C\sqrt{\pi} = 1 & \text{(32.149a)} \\ K - C\sqrt{\pi} = 0. & \text{(32.149b)} \end{cases}$$

Solution : $C = 1/2\sqrt{\pi}$ et $K = 1/2$. Au final,

$$u(t,x) = \frac{1}{2} + \frac{1}{2\sqrt{\pi}} \int_0^{x/\sqrt{t}} e^{-s^2/4}. \tag{32.150}$$

Notons que cela donne une valeur pour $u(t,0)$:

$$u(t,0) = \frac{1}{2}. \tag{32.151}$$

Chapitre 33

Numérique

D'autres lectures agréables dans [501].

33.1 Introduction

À quels types de problèmes peut-on s'attendre lorsqu'on se lance dans du calcul numérique, et en particulier dans la résolution numérique d'équations (algébrique, différentielles ou aux dérivées partielles, etc) ?

Quelques réflexions en vrac sur ce sujet.

(1) Les erreurs de représentation de nombres : troncature et propogation de décalales (*drift*),

(2) Erreur de compensation (*cancellation*),

(3) Conditionnement, stabilité : les réponses peuvent fortement dépendre des paramètres,

(4) Si on utilise une méthode itérative, comment savoir à quel moment on s'arrête ? Calculer la différence $|x_k - x_{k-1}|$ mène t-il à une erreur de cancellation ?

(5) Lors d'une implémentation, les matrices des systèmes à résoudre sont souvent très grandes et/ou très creuses. Cela pose la question de la manière de les enregistrer.

(6) Pour la parallélisation, il faut faire attention au fait que parfois créer un nouveau processus demande plus de ressources que le mini-calcul qu'on voulait faire. Donc il ne faut pas toujours paralléliser tout ce qui est théoriquement parallélisable.

(7) Le fait que certaines méthodes sont non-déterministes (Monté-Carlo) mène à des problèmes pour les tests unitaires des implémentations.

33.2 Représentations numériques

Dans cette section, les séquences de chiffres écrites entre crochet sont à comprendre comme des séquences de chiffres qui représentent une quantité suivant un codage donné.

33.2.1 Entier relatif en complément à deux (binaire)

Si nous avons m bits pour coder un entier relatif, une idée serait de prendre le premier bit pour le signe (0 pour positif et 1 pour négatif) et les autres pour la valeur absolue. Deux inconvénients :

(1) Il y a deux codages pour le zéro, donc gaspillage.

(2) L'algorithme pour faire la somme passe mal. Par exemple pour faire $1 + (-1)$, le 1 est codé comme [001] et le -1 par [101] et la somme se ferait naïvement comme

$$\begin{array}{ccc} 0 & 0 & 1 \\ 1 & 0 & 1 \\ \hline 1 & 1 & 0 \end{array}$$

Donc le résultat est [110] qui s'interprète comme -2. Complètement faux.

Une solution est d'utiliser le **complément à deux**, qui est la façon usuelle de représenter des entiers signés.

Les entiers positifs se codent normalement, en laissant à zéro le premier bit (donc si nous disposons de m bits, nous codons sur $m-1$ bits).

Les entiers négatifs se codent en trois étapes.

— coder la valeur absolue

— inverser tous les bits (d'où le nom de « complément à deux »)

— soustraire 1.

Exemple 33.1.
Pour coder -1 nous faisons

— Nous codons 1 : [001]

— Nous inversons tous les bits : [110]

— Nous faisons -1 : [101].

△

Avec ce système, la somme passe bien : calculer $1+(-1)$ donne

$$\begin{array}{ccc} 0 & 0 & 1 \\ 1 & 0 & 1 \\ \hline 1 & 1 & 0 \\ \hline \end{array}$$

La réponse est donc [110] qu'il faut interpréter via le complément à deux.

$$110 \xrightarrow{+1} 111 \xrightarrow{\text{complément}} 000. \tag{33.1}$$

Et ce dernier [000] s'interprète comme zéro.

Définition 33.2 (Entier signé en complément à deux[502])**.**
La suite de bits $[a_{m-1}\dots a_0]$ *s'interprète via la formule*

$$-a_{m-1}2^{m-1}+\sum_{i=0}^{m-2}a_i2^i. \tag{33.2}$$

Le premier bit donne effectivement le signe du nombre, mais l'interprétation d'un nombre n'est pas aussi simple que ce que l'on pourrait croire de prime abord.

Exemple 33.3 (Entier signé en 8 bits)**.**
Que pouvons nous faire avec 8 bits ? Le plus grand nombre est codé par [01111111] qui vaut $\sum_{k=0}^{6}2^k=2^7-1=127$. (avez-vous utilisé la somme (10.125) ?)

Le plus petit nombre codable en 8 bits n'est pas [11111111] mais bien [10000000] (cela est plus clair en regardant la formule (33.2) qu'en tentant de suivre la construction du complément à deux) qui signifie $-2^7=-128$.

Nous pouvons donc coder tous les nombres de -128 à 127. △

Plus généralement un système qui codes des entiers signés en N bits utilisant le complément à deux peut coder de $-(2^{N-1})$ à $2^{N-1}-1$.

33.4 (Le dépassement)**.**
Que se passe-t-il lorsque nous commettons un dépassement ? Calculons sur 3 bits la somme [011] + [001] qui revient à ajouter 1 au nombre le plus grand :

$$\begin{array}{ccc} 0 & 1 & 1 \\ 0 & 0 & 1 \\ \hline 1 & 0 & 0 \end{array}$$

qui signifie $-2^2 = -4$. Lors d'un dépassement, nous retombons automatiquement sur le plus petit.

Ce phénomène est bien connu des personnes qui programment sans faire attention dans certains languages de programmation qui ne font pas attention à votre place.

Définition 33.5 (Représentation en virgule fixe)**.**
*Soit x un réel. On définit sa **représentation en virgule fixe** par*

$$x = \{[x_n x_{n-1} ... x_0, x_{-1} ... x_{-m}], b, s\} \tag{33.3}$$

avec $b \in \mathbb{N}, b \geqslant 2$, $s \in \{0, 1\}$ et $x_j \in \mathbb{N}, x_j < b$ suivant la formule

$$x = (-1)^s \sum_{j=-m}^{n} x_j . b^j. \tag{33.4}$$

33.2.2 Représentation en virgule flottante

Définition 33.6 (Représentation en virgule flottante)**.**
*La **représentation en virgule flottante normalisée** en base b d'un nombre est la donnée de*

(1) Un bit s pour le signe

(2) Un entier non signé *q de e chiffres pour l'exposant*

(3) Une suite de chiffres $[a_1 \dots a_m]$ pour la mantisse.

Ces données s'interprètent via la formule

$$\mathrm{fl}(s, q, [a_1, \dots, a_m]) = (-1)^s \sum_{j=1}^{m} b^j a_j \times b^{q-d} \tag{33.5}$$

*où $d = b^{e-1}$ est le **décalage**.*

Une idée à retenir est que l'exposant est un entier non signé parce qu'il est plus simple d'introduire un décalage dans la formule (33.5) que de compliquer l'écriture de l'exposant.

33.2.3 Simple précision, IEEE-754

En écriture binaire, la représentation en virgule flottante est un peu différente parce qu'il y a une idée supplémentaire ; la simple précision que nous allons voir maintenant n'est donc pas un cas particulier de 33.6 avec $b = 2$.

Nous commençons par une description informelle de la précision simple avant de donner la définition. La représentation en **précision simple** d'un nombre se fait sur 32 bits répartis comme suit :

(1) 1 bit pour le signe,

(2) 8 bits pour l'exposant interprété comme nombre entier non signé

(3) 23 bits pour la mantisse

Soit le triple

$$\big(s, q, [a_1, \dots, a_{23}]\big) \tag{33.6}$$

Dans le cas générique, l'idée est de donner 24 bits pour la mantisse, mais en gardant en tête le fait que de toutes façons, le premier bit doit être 1, sinon il suffirait de décaler, c'est-à-dire changer l'exposant. Par conséquent la mantisse ne reçoit que 23 bits ; il y a un « 1 » sous-entendu en première position. Donc la mantisse $[a_1, \dots, a_{23}]$ est à lire comme le nombre

$$1, a_1 \dots a_{23} = 1 + \sum_{j=1}^{23} a_j 2^{-j}. \tag{33.7}$$

Exemple 33.7.
La mantisse $[011100 \dots 0]$ signifie $1,0111 = 1 + 2^{-2} + 2^{-3} + 2^{-4} = 1 + \frac{1}{4} + \frac{1}{8} + \frac{1}{16}$. △

Cela pour justifier la formule

$$\mathrm{sp}\left(s, q, [a_1, \dots, a_{23}]\right) = (-1)^s \big(1 + \sum_{j=1}^{23} a_j 2^{-j}\big) 2^{q-127}. \tag{33.8}$$

Notons :

(1) Le « 1+ » dans la parenthèse correspond au 1 implicite en première position de la mantisse.

(2) Il y a un décalage de 127 dans l'exposant, parce que q est un entier non signé.

Notons que cette règle du 1 implicite dans la mantisse empêche d'écrire le nombre 0, et ne permet pas d'écrire des nombres franchement petits parce que le 1 implicite est en *première* position dans la mantisse.

D'où l'idée de donner une règle particulière lorsque l'exposant vaut 0. Lorsque l'exposant est $q = 0$, alors nous ne considérons pas de 1 implicite dans la mantisse, et le décalage de l'exposant est -126 au lieu de -127. D'où la formule

$$\mathrm{sp}(s, q = 0, [a_1 \dots a_{23}]) = (-1)^s 2^{-216} \sum_{j=1}^{23} a_j 2^{-j}. \tag{33.9}$$

En particulier, si $q = 0$ et $a = [0 \dots 0]$, nous avons le nombre zéro exact (il y a deux possibilités pour le code).

Enfin, nous avons des cas particuliers lorsque l'exposant est maximum, c'est-à-dire lorsque $q = [1111\,1111] = 2^8 - 1 = 255$. Dans ce cas, le nombre codé est soit $+\infty$ soit NaN. Nous posons

$$\mathrm{sp}(s, q = 255, a = 0) = +\infty \tag{33.10}$$

et

$$\mathrm{sp}(s, q = 255, a \neq 0) = NaN. \tag{33.11}$$

Il y a en réalité plusieurs valeurs différentes de NaN, mais nous n'entrons pas dans ces détails[503].

Définition 33.8 (Représentation en simple précision (binaire))**.**
*La représentation en **précision simple** d'un nombre se fait sur* 32 *bits répartis comme suit :*

(1) 1 *bit pour le signe,*

(2) 8 *bits pour l'exposant interprété comme nombre entier non signé*

(3) 23 *bits pour la mantisse*

Un nombre est représenté par un triple

$$\left(s, q, [a_1, \dots, a_{23}]\right) \tag{33.12}$$

Selon que l'exposant $q - d$ soit égal à 0, $2^8 - 1 = 255$ *ou autre chose, les règles d'interprétation sont différentes. Il y a donc trois cas.*

Exposant q générique[504] *Si $q \neq 0$ et $q \neq 255$ alors le nombre est **normalisée**. La règle de lecture est alors*

$$\mathrm{sp}\left(s, q, [a_1, \dots, a_{23}]\right) = (-1)^s \big(1 + \sum_{j=1}^{23} a_j 2^{-j}\big) 2^{q-127}. \tag{33.13}$$

Exposant q égal à 0 *Le nombre est dit **dénormalisé** et la règle de lecture est*

$$\mathrm{sp}\left(s, q, [a_1, \dots, a_{23}]\right) = (-1)^s 2^{-126} \sum_{j=1}^{23} a_j 2^{-j}. \tag{33.14}$$

Exposant q égal à 255 *La règle de lecture est alors au cas pas cas ou à peu près.*

(1) $\mathrm{sp}(s, q = 255, a = 0) = +\infty$.

(2) $\mathrm{sp}(s, q = 255, a \neq 0) = \mathrm{NaN}$.

Vous pouvez jouer avec la simple précision dans [505].

Exemple 33.9 (Plus petit normalisé)**.**
Pour faire un nombre normalisé, il faut au minimum $q = 1$. En prenant $a_j = 0$ nous obtenons le plus petit nombre normalisé possible en simple précision. La formule (33.13) donne

$$\mathrm{sp}(1, q = 1, a = 0) = 2^{1-127} = 2^{-126} \simeq 1.17549435082229 \times 10^{-38}. \tag{33.15}$$

△

Exemple 33.10 (Plus grand normalisé)**.**
L'exposant q ne peut pas être maximum, sous peine de tomber dans les règles spéciales de $+\infty$ ou NaN. Donc $q = [1111\,1110] = 2^8 - 2 = 254$. En ce qui concerne la mantisse, il faut la prendre maximale, c'est-à-dire $a_j = 1$ pour tout j. Nous avons alors le nombre

$$\mathrm{sp}(1, q = 254, a = [1 \dots 1]) = \big(1 + \sum_{j=1}^{23} 2^{-j}\big)2^{254-127} = (1 - \frac{1}{2^{24}})2^{128} \tag{33.16a}$$

$$= 3.40282346638528859811704183484516925440 \times 10^{38} \tag{33.16b}$$

où nous avons utilisé la somme (10.125) (et Sage pour le dernier calcul). △

Notons ceci avec Sage :

```
 SageMath Version 7.0, Release Date: 2016-01-19
 Type "notebook()" for the browser-based notebook interface.
 Type "help()" for help.

sage: A=(1- (1/2**24)  )*2**(128)
sage: type(A)
<type 'sage.rings.rational.Rational'>
```

tex/sage/sageSnip003.sage

La précisions du nombre donné en (33.16b) aurait été embarrassante si le type avait été un nombre en simple précision. Précision technique : en Python, le type `int` n'a pas de limite supérieure à part la mémoire.

Exemple 33.11 (Plus petit non nul dénormalisé)**.**
Pour être dénormalisé il faut $q = 0$ (ce qui est toutefois assez logique si nous voulons un petit nombre), et pour ne pas être nul, il faut une mantisse non nulle. Donc $a = [0 \dots 01]$. La formule (33.14) donne alors

$$\mathrm{sp}(s = 0, q = 0, a = [0 \dots 01]) = 2^{-126}2^{-23} = 2^{-149} \simeq 1.40129846432482 \times 10^{-45}. \tag{33.17}$$

△

Exemple 33.12 (Plus grand dénormalisé)**.**
Pour être dénormalisé il faut toujours $q = 0$, mais cette fois nous prenons la plus grande mantisse possible :

$$\mathrm{sp}(s = 0, q = 0, a = [1 \dots 1]) = 2^{-126} \sum_{j=1}^{23} 2^{-j} = 2^{-216}(1 - 2^{-23}) = 1.17549421069244 \times 10^{-38} \tag{33.18}$$

△

Notons ceci avec Sage :

```
sage: B=2**(-126)*(1-2**(-23))
sage: A=2**(-126)
sage: n(A-B)
1.40129846432482e-45
```

tex/sage/sageSnip004.sage

Vu que $2^{-23} \simeq 1.2 \times 10^{-7}$, approximer la parenthèse par 1 donne une faute sur la septième décimale, ce qui est visible en simple précision.

33.3 Problèmes pour écrire des nombres

Définition 33.13.
*L'**erreur relative** commise en remplaçant un nombre réel x par une valeur approchée $\hat{x}$ est définie par*

$$\epsilon_x := \left| \frac{x - \hat{x}}{x} \right|. \tag{33.19}$$

L'erreur relative n'est pas influencée par l'ordre de grandeur de x. En effet, l'ordre de grandeur de $\hat{x}$ est certainement la même que celle de x, dans la majorité des cas sans problèmes. Du coup si $x' = 200x$ alors $\hat{x'} \simeq 200\hat{x}$ et le 200 se simplifie.

Le nombre de chiffres significatifs correct dans l'approximation est donné par $-\log_{10}(\epsilon_x)$. La partie entière de ce nombre est le nombre de chiffres tout à fait exacts et la partie décimale donne une idée sur le fait que le chiffre suivant est plus ou moins bien.

Remarque 33.14.
Si nous voulons donner $x \in \mathbb{R}$ à un ordinateur, nous sommes soumis à deux erreurs :

(1) D'abord, vu que nous ne pouvons pas taper sur le clavier toutes les décimales de x, nous faisons une **erreur de troncature**.

(2) L'ordinateur devant convertir cela en base deux, il commet une seconde erreur, dite **erreur d'assignation**.

33.3.1 Troncature : la base

Supposons que nous voulions écrire le nombre (écrit ici en base 10)

$$0.4567894251 \tag{33.20}$$

de façon plus facile à lire, on peut demander de ne laisser que t chiffres significatifs. Disons $t = 3$.

Technique de troncature On garde 3 chiffres significatifs : 0.456. Facile.

Technique d'arrondi Vu que le premier qu'on supprime est un 7, le dernier qu'on garde est majoré de 1 : on écrit 0.457.

Que faire si le premier chiffre rejeté est un 5 ? En première approximation, nous pouvons prendre la règle suivante : si le premier chiffre rejeté est un 5, il faut augmenter de 1 de dernier chiffre gardé parce qu'il y a presque certainement encore un chiffre non nul derrière.

Remarque 33.15.
Les ordinateurs travaillent tous en mode d'arrondi.

Exemple 33.16.
Si on doit entrer le nombre 0.38358546 dans un ordinateur qui ne garde que 3 chiffres significatifs, il faut taper 0.384 au clavier (erreur classique dans les exercices). △

33.3.2 Troncature : le drift

Soit une machine ne pouvant retenir que 3 chiffres significatifs et effectuant les arrondis vers le haut lorsque le chiffre à éliminer est un 5. Nous notons $\oplus$ et $\ominus$ les opérations d'addition et soustraction avec arrondis[506]. Les égalités comprenant plus de trois chiffres significatifs sont des égalités au sens de la machine. Nous écrirons donc sans états d'âme :

$$1 \oplus 0.555 = 1.555 = 1.56. \tag{33.21}$$

Considérons la suite numérique

$$\begin{cases} x_0 = 1.00 & (33.22a) \\ x_n = (x_{n-1} \ominus y) \oplus y & (33.22b) \end{cases}$$

avec $y = -0.555$.

Nous avons

$$x_1 = (1 \oplus 0.555) \ominus 0.555 = 1.56 \ominus 0.555 = 1.005 = 1.01 \tag{33.23}$$

et ensuite

$$x_2 = (1.01 \oplus 0.555) \ominus 0.555 = 1.565 \ominus 0.555 = 1.57 \ominus 0.555 = 1.015 = 1.02. \tag{33.24}$$

Et ainsi de suite. La suite est donc croissante alors que la définition nous donnerait envie d'avoir $x_n = x_0$ pour tout n.

Remarque 33.17.
En réalité, cette suite se stabilise à $x_n = 10$ pour tout n à partir de $n = 845$. En effet,

$$(10 \oplus 0.555) \ominus 0.555 = 10.555 \ominus 0.555 = 10.6 \ominus 0.555 = 10.045 = 10. \tag{33.25}$$

Le fait est qu'à ce moment, l'erreur de troncature est assez loin dans les décimales pour que le premier chiffre négligé soit un "0" au lieu d'un "5".

Notons toutefois que cette stabilité n'est pas là pour nous rassurer parce qu'elle n'en est pas moins complètement fausse.

La règle de troncature adoptée dans Sage est d'arrondir au nombre pair le plus proche lorsque le premier nombre à négliger est un 5. Donc 12.5 s'arrondit à 12 plutôt que 13.

Exemple 33.18.
Soient les expressions (algébriquement égales) :

(1) $A = x(x+1)$

(2) $B = x^2 + x$

Nous savons que

$$x = \mathrm{fl}(x) = 10^{-30} \tag{33.26}$$

et

$$1 = \mathrm{fl}(1) \tag{33.27}$$

parce que pour 1 et 10^{-30}, il n'y a pas d'erreurs d'assignation.

En précision simple, $10^{-30} + 1 = 1$ parce qu'en précision simple, il n'y a que 7 ou 8 chiffres significatifs[1].

Nous avons $A = 10^{-30}$, mais x^2 donne un `underflow` parce que 10^{-60} ne peut pas être représenté en précision simple. En pratique, beaucoup de logiciels en font 0. Dans ce cas, en réalité B donne effectivement 10^{-30} après avoir fait $x^2 + x = 0 + x = 10^{-30}$. △

1. Erreur de « relation normale ».

33.3.3 Quelques bonnes règles

(1) Si on a plusieurs nombres à additionner ou soustraire, il vaut mieux commencer par sommer ou soustraire ceux dont on sait qu'ils ont le même ordre de grandeur. Il n'y a donc pas tout à fait « associativité » des erreurs.

(2) Les opérations délicates sont l'addition et la soustraction. La multiplication et la division sont sans dangers, à part l'erreur de dépassement du maximum. Dans une multiplication, on perd au pire quelques chiffres significatifs, mais certainement les derniers, pas les premiers.

33.3.4 Erreur de "cancellation"

Lorsque deux nombres sont de même ordre de grandeur, avec plusieurs nombres significatifs identiques. La cancellation est le fait que, suite à la soustraction, tous les chiffres significatifs ou presque se sont simplifiés et qu'il ne reste plus que des chiffres non significatifs.

Exemple 33.19 ([507])**.**
Sur une machine ne gardant que 4 chiffres significatifs, faire

$$0.5678 \times 10^6 - 0.5677 \times 10^6 = 0.0001 \times 10^6 = 0.1000 \times 10^3. \tag{33.28}$$

Le fait est que les trois derniers zéros ne sont pas significatifs, mais maintenant la machine nous fait croire qu'ils le sont.

Une autre façon de voir ce problème est d'imaginer qu'il faille calculer la différence

$$0.5678\,289798 \times 10^6 - 0.5677\,3136907 \tag{33.29}$$

sur cette machine. Certes la machine nous autorise à avoir 4 chiffres significatifs, donc au moment d'entrer les nombres nous perdons un beau paquet de chiffres. Mais au moment de faire la différence, nous perdons (presque) tout le reste. Donc là où nous pouvions espérer avoir 4 chiffres significatifs de la différence, nous n'en avons que 1. Les trois derniers zéros de la réponse (0.1000×10^3) sont faux. △

Remarque 33.20.
L'erreur de cancellation provoque des chiffres significatifs faux, mais ne provoque pas de faute dans l'*ordre de grandeur* des réponses[2]. Donc si nous voulons nous assurer que a et b sont égaux « à erreur numérique près », le test

$$|a - b| < \epsilon \tag{33.30}$$

est valide, malgré l'erreur de cancellation qui ne manquera pas de se produire dans le calcul de la différence.

Exemple 33.21.
Soit à résoudre l'équation $ax^2 + bx + c = 0$ avec $a, b, c \neq 0$ et $b^2 - 4ac > 0$. Solution :

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \tag{33.31}$$

Supposons que $|4ac| \ll b^2$ avec tout de même pas tellement petit qu'on se perd dans la précision. Bref, on suppose que seules quelques dernières décimales de $b^2 - 4ac$ sont différentes de zéro.

On a :

$$\sqrt{b^2 - 4ac} = \sqrt{\tilde{b}} = |\tilde{b}| \tag{33.32a}$$

$$x_1 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \tag{33.32b}$$

$$x_2 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{33.32c}$$

2. Est-ce bien vrai, cela ?

Si $b > 0$, nous avons une erreur de cancellation dans x_2 parce qu'on fait la différence entre deux nombres presque égaux. Donc x_2 mal calculé. Par contre x_1 est bien calculé.

Si par contre $b < 0$, c'est le contraire.

Avec $a = 10^{-3}$, $b = 0.8$, $c = -1.2 \times 10^{-5}$. À la main nous obtenons : $x_1 = -800$, $x_2 = 1.5 \times 10^{-5}$, et un ordinateur se tromperait ...

```

 SageMath Version 7.0, Release Date: 2016-01-19
 Type "notebook()" for the browser-based notebook interface.
 Type "help()" for help.

sage: f(x)=10**(-3)*x**2+0.8*x-1.2*10**(-5)
sage: solve(f(x)==0,x)
[x == -1/50*sqrt(400000030) - 400, x == 1/50*sqrt(400000030) - ↩
   400]
sage: numerical_approx(-1/50*sqrt(400000030))
-400.000015000000
sage: numerical_approx( 1/50*sqrt(400000030) - 400  )
0.0000149999996779115
```

tex/sage/sageSnip001.sage

Donc Sage ne tombe pas dans le piège. △

Comment résoudre ce problème ? Ou, autre façon de poser la question : comment Sage a fait pour résoudre le problème ?

Utilisons les relations coefficients-racines :

$$x_1 + x_2 = -b/a \tag{33.33a}$$
$$x_1 x_2 = c/a \tag{33.33b}$$

La première lie les deux racines par des opérations de addition et soustractions, et donc n'est pas intéressantes. La seconde est bien. Si nous connaissons x_1, nous calculons

$$x_2 = \frac{c}{a x_1}. \tag{33.34}$$

Quitte à redéfinir x_1 et x_2, la solution bien calculée est :

$$x_1 = \frac{-b - \operatorname{sgn}(b)\sqrt{b^2 - 4ac}}{2a}. \tag{33.35}$$

Exemple 33.22.
Nous considérons :

$$f(x) = cos(x + \delta) - \cos(x). \tag{33.36}$$

Cela a une erreur de cancellation lorsque $|\delta| \ll |x|$. On élimine l'erreur de cancellation par

$$f(x) = -2\sin(\delta/2)\sin\left(x + \frac{\delta}{2}\right). \tag{33.37}$$

Problèmes et choses à faire

Pourquoi la condition pour avoir l'erreur est $\delta \ll x$ et non simplement $\delta \ll 1$?

△

Exemple 33.23.
Pour

$$f(x) = \sqrt{x + \delta} - \sqrt{x}. \tag{33.38}$$

On fait la coup du binôme conjugué :

$$f(x) = \frac{\delta}{\sqrt{x+\delta}+\sqrt{x}}. \tag{33.39}$$

Plus d'erreur de cancellation, vu qu'au dénominateur nous avons une somme de deux positifs. △

Les erreurs de cancellation ne se résolvent pas en augmentant la précision des nombres donnés.

Exemple 33.24 (Dans la vie réelle)**.**
La préparation de l'exemple 16.45 nous a porté à calculer la différence entre $\exp(x)$ et $f_{30}(x)$ où f_{30} est censée être une bonne approximation de l'exponentielle. Des erreurs de cancellation sont donc à craindre.

Et en effet, le code suivant produit un résultat non déterministe :

```
f=1/152444172305856930250752000000*x^28 + ↩
   1/108888694504183521607680000000*x^27 + ↩
   1/155112100433309859840000000*x^25 + ↩
   1/31022420086661971968000*x^24 + 1/2585201673888497664000*x↩
   ^23 + 1/510909421717094400000*x^21 + 1/1216451004088320000*x^20↩
    + 1/121645100408832000*x^19 + 1/355687428096000*x^17 + ↩
   1/10461394944000*x^16 + 1/1307674368000*x^15 + 1/6227020800*x↩
   ^13 + 1/239500800*x^12 + 1/39916800*x^11 + 1/362880*x^9 + ↩
   1/20160*x^8 + 1/5040*x^7 + 1/120*x^5 + 1/12*x^4 + 1/6*x^3 + x ↩
   - cos(x) + 2 -exp(x)
a=numerical_approx(10)
print(f(a))
```

tex/sage/sageSnip016.sage

Voir la question ici :
https://ask.sagemath.org/question/37946/undeterministic-numerical-approximation/
△

33.3.5 Calcul d'une dérivée

Pour calculer la dérivée de f en a, il est loisible d'utiliser la formule

$$f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a)}{h}. \tag{33.40}$$

Le numérateur est alors sujet à une erreur d'absorption dans le calcul de $a+h$ et ensuite une erreur de cancellation dans le calcul de la différence.

En utilisant la formule

$$f'(a) = \lim_{h\to 0} \frac{f(a+h)-f(a-h)}{2h} \tag{33.41}$$

nous pouvons espérer avoir une erreur de cancellation plus petite.

33.3.6 Erreur d'absorption

L'addition d'un nombre avec un nombre très différent peut faire perdre de l'information sur le plus petit. Par exemple avec 4 chiffres significatifs,

$$0.5678 \oplus 0.0001237 = 0.5679 \tag{33.42}$$

où nous avons perdu presque toute l'information du petit nombre.

Une situation particulièrement ennuyeuse est celle où justement c'est le petit nombre qui nous intéresse parce que le grand est censé se simplifier :

$$(0.0001327 \oplus 0.5678) \ominus 0.5678 = 0.5679 \ominus 0.5678 = 0.0001 \tag{33.43}$$

qui ne possède qu'un seul chiffre significatif correct alors que voyant le calcul, la réponse aurait pu être trouvée.

Moralité : si certains manipulations algébrique peuvent faire apparaitre des simplifications avant de passer le calcul à la machine, il est bon de les effectuer.

33.4 Conditionnement et stabilité

Définition 33.25.
Soit F une fonction à valeurs réelles définie sur $X \times D$ où X et D sont des espaces vectoriels réels normés. Le problème de la recherche des solutions de

$$F(x, d) = 0 \tag{33.44}$$

*est dit **stable** autour de $d_0 \in D$ si*

(1) la solution $x = x(d)$ existe et est unique pour tout d ;

(2) Pour tout $\eta > 0$, et pour tout d_0, il existe un nombre $K > 0$ tel que $\|d - d_0\| < \eta$ entraine $\|x(d) - x(d_0)\| \leqslant K \|d - d_0\|$.

La seconde condition est le fait que x soit Lipschitz[3] sur un voisinage de d_0.

Exemple 33.26 (Stabilité de la différence)**.**
Prenons le problème qui consiste à calculer la différence entre deux nombres : $x = a - b$. Cela se traduit par

$$\begin{aligned} F\colon \mathbb{R} \times \mathbb{R}^2 &\to \mathbb{R} \\ x &\mapsto x - a + b. \end{aligned} \tag{33.45}$$

Nous avons :

$$\begin{aligned} \big|x(a,b) - x(a',b')\big| &= |a - b - a' + b'| && (33.46a) \\ &\leqslant |a - a'| + |b - b'| && (33.46b) \\ &= \|(a,b) - (a',b')\|_1 && (33.46c) \end{aligned}$$

où nous avons utilisé la norme $\|.\|_1$ sur $\mathbb{R}^2$. Par la proposition 7.223 sur les équivalences de normes, le nombre $K = \sqrt{2}$ fonctionne pour toute valeurs de η.

La problème de la différence est donc un problème stable. △

Exemple 33.27 (Stabilité de la multiplication)**.**
Si a est fixé, le problème de calculer ab (b est la donnée) est stable. En effet ce problème est donné par la fonction $F(x, b) = x - ab$, dont la solution est $x(b) = ab$. Nous avons donc

$$\big|x(b) - x(b')\big| = |ab - ab'| = |a||b - b'|. \tag{33.47}$$

La constante de Lipschitz de ce problème est donc $|a|$. △

Définition 33.28.
Le nombre

$$K_{abs}(d_0, \eta) := \sup_{d \text{ tel que } |d_0 - d| < \eta} \frac{\|x(d) - x(d_0)\|_X}{\|d - d_0\|_D} \tag{33.48}$$

*est appelé le **conditionnement absolu** du problème autour de d_0.*

3. Définition 11.259.

Soit $F(x,d)=0$ un problème stable de conditionnement absolu $K_{abs}(d,\eta)$. Le conditionnement relatif est défini par

$$K_{rel}(d,\eta) := K_{abs}(d,\eta)\frac{\|d\|_D}{\|x(d)\|_X}. \tag{33.49}$$

*Le problème est dit **bien conditionné** près de d si $K_{rel}(d,\eta)$ est petit.*

Exemple 33.29 (Mauvais conditionnement de la différence)**.**
Reprenons le problème de la différence, mais en fixant a. Nous avons donc $x(b)=a-b$ et le conditionnement absolu est

$$\sup\frac{|x(b)-x(b_0)|}{|b-b_0|}=1 \tag{33.50}$$

Le conditionnement relatif est :

$$K_{rel}(b_0,\eta)=\frac{|b|}{|a-b|}. \tag{33.51}$$

Et donc le problème est mal conditionné autour de a.

Autrement dit, si a' est un nombre proche de a, calculer la différence $a-a'$ est un problème mal conditionné. △

Exemple 33.30 (Bon conditionnement de la multiplication)**.**
Pour le problème $F(x,b)=x-ab$ nous avons

$$K_{abs}=\sup_{b'}\frac{|ab-ab'|}{|b-b'|}=|a|. \tag{33.52}$$

Et aussi

$$K_{rel}=a\frac{|b|}{|ab|}=1. \tag{33.53}$$

Le conditionnement relatif du problème de la multiplication est donc toujours 1. Il est donc un toujours un problème bien conditionné. △

Ne pas confondre :

Le conditionnement provient du problème lui-même.

La stabilité provient de l'algorithme de résolution.

Exemple 33.31 (Un problème mal conditionné)**.**
Le système

$$\begin{cases} 2.1x+3.5y=8 & (33.54a)\\ 4.19x+7.0y=15 & (33.54b)\end{cases}$$

Solution : $x=100$, $y=-57.714285\ldots$ (périodique)

Perturbons : nous remplaçons 4.19 par 4.192. L'erreur relative est : 4.77×10^{-4}.

Solution : $\bar{x}=125$, $\bar{y}=-72.714285\ldots$, avec donc erreur relative de 0.26. Autrement dit : l'erreur relative sur la solution est grande même avec une petite erreur relative sur la donnée.

C'est un problème mal conditionné.

Le fait est que c'est une intersection de deux droites presque parallèles. Donc effectivement une petite perturbation d'une des deux droites donne une grande perturbation du point d'intersection.

Le fait est qu'un ordinateur effectue *toujours* une perturbation, au moins de l'ordre 10^{-16} pour ne fut-ce que représenter les nombres. C'est-à-dire une perturbation sur les six nombres définissant le système. Il n'y a donc pas d'espoir d'obtenir un algorithme donnant une bonne réponse. △

Un résultat pratique pour étudier le conditionnement d'un problème est le suivant.

Corolaire 33.32.
Soit $x=x(d)$ un problème stable. Supposons $\mathbb{D}$ de dimension finie, supposons que U est ouvert dans $\mathbb{D}$. Supposons encore $x\colon U\to\mathbb{R}$ différentiable en d_0. Alors quand η est petit, on a

$$K^{\eta}_{abs}(d_0)\sim\|\nabla x(d_0)\|. \tag{33.55}$$

Lemme 33.33.
Tout problème de la forme $x = x(d)$ avec $d \in \mathbb{R}$ et $x \in C^1(\mathbb{R})$ est stable.

Démonstration. Il faut démontrer qu'une fonction C^1 sur $\mathbb{R}$ vérifie automatiquement la condition (2) de la définition de la stabilité. Pour cela, remarquons qu'une fonction C^1 possède une dérivée continue, et donc bornée sur tout compact [4]

Prenons $\eta > 0$ et $d_0 \in \mathbb{R}$ et puis un d tel que $|d - d_0| < \eta$. Par le théorème des bornes atteintes, la fonction x' est bornée sur l'intervalle $[d_0 - \eta, d_0 + \eta]$. Appelons K un majorant de x' sur cet intervalle. La fonction

$$f(d) = x(d_0) + K|d - d_0| \tag{33.56}$$

majore $x(d)$, et donc on a

$$\big|x(d) - x(d_0)\big| \leqslant K|d - d_0|. \tag{33.57}$$

Attention : vérifier si ce raisonnement est correct avec $d_0 > d$, et adapter au besoin. □

Exemple 33.34.
Un exemple de problème stable de la forme $x = x(d)$ avec $d \in \mathbb{R}$ et $x \in C^0(\mathbb{R})\backslash C^1(\mathbb{R})$.

La fonction

$$x(d) = \begin{cases} 0 & \text{si } x \geqslant 0 \\ x & \text{si } x > 0 \end{cases} \tag{33.58}$$

est continue, mais pas C^1 (non dérivable en $x = 0$). La dérivée est partout bornée par 1, et donc le problème est stable.

Un autre exemple très classique serait de prendre $x(d) = |d|$. Dans ce cas, on peut prendre n'importe que η et $K = 1$. Le calcul est que

$$|x(d) - x(d_0)| < K|d - d_0| \tag{33.59a}$$

$$\big||d| - |d_0|\big| < |d - d_0|. \tag{33.59b}$$

Cette dernière inéquation est correcte, comme on peut le voir en mettant au carré les deux membres. △

Exemple 33.35.
Un exemple de problème instable de la forme $x = x(d)$ avec $d \in \mathbb{R}$ et $x \in C^0(\mathbb{R})$.

Un exemple assez classique de fonction dont la dérivée n'est pas bornée sans pour autant que la fonction aie un comportement immoral [5] est $x \mapsto \sqrt{x}$. Afin d'avoir une fonction définie sur $\mathbb{R}$ tout entier, nous regardons la fonction

$$x(d) = \sqrt{|d|}. \tag{33.60}$$

Si nous considérons maintenant $d_0 = 0$ et n'importe quel η, nous avons

$$\frac{|x(d) - x(d_0)|}{|d - d_0|} = \frac{\sqrt{d}}{d} = \frac{1}{\sqrt{d}}. \tag{33.61}$$

Il n'est pas possible de trouver un K qui majore ce rapport. Le problème est donc mal conditionné.

Attention : dans ce calcul nous avons supposé $d > 0$. Pensez à adapter au cas $d < 0$. △

Exemple 33.36 (Problème bien conditionné avec algorithme instable)**.**
Soit à calculer

$$I_n = \frac{1}{e}\int_0^1 x^n e^x dx \tag{33.62}$$

avec $n \geqslant 0$. Par partie, nous obtenons :

$$I_n = 1 - nI_{n-1}. \tag{33.63}$$

4. Un compact est un ensemble fermé et borné, typiquement un intervalle du type $[a, b]$.
5. Penser à $x \mapsto x\sin(1/x)$.

D'autre part, $I_0 = \frac{e-1}{e}$, $I_1 = \frac{1}{e}$. Puis par récurrence, c'est tout en main.

Du côté de l'ordinateur, nous lui donnons forcément une approximation de I_1, parce que nous lui donnons une approximation de e. Soit l'erreur ϵ_1 sur I_1.

Sans démonstration :

Lemme 33.37.
Nous avons $\lim_{n\to\infty} I_n = 0$.

Mais numériquement, il n'est pas possible de rester longtemps sous ϵ_1 parce que nous n'espérons pas avoir une erreur plus petite que ça. Donc à partir du moment où $I_n < \epsilon_1$, les valeurs sont toutes complètement fausses. Cela est le mieux que l'on puisse espérer. Mais la réalité est pire.

En réalité, en lançant le calcul sur un ordinateur, les valeurs sont même croissantes avec n à partir d'un certain moment.

On peut étudier l'erreur et montrer que l'erreur est donnée par :

$$\epsilon_n = (-1)^{n-1} n! \epsilon_1. \tag{33.64}$$

Mais comme la factorielle est tellement forte que c'est sans espoir d'aller loin en essayant très fort de donner une petite erreur sur ϵ_1.

△

Il existe heureusement un algorithme stable pour cette intégrale. La formule est :

$$I_{n-1} = \frac{1}{n}(1 - I_n). \tag{33.65}$$

Si nous savons un I_N avec N grand, cette formule donne les I_i avec $i = N, N-1, \dots, 2$. Posons donc $I_N = a \in \mathbb{R}$ n'importe comment. Donc ϵ_N est grand. Mais il se trouve que l'erreur sur ϵ_1 est donnée par

$$\epsilon_1 = \frac{(-1)^{N-1}}{N!}\epsilon_N. \tag{33.66}$$

Donc même en prenant vraiment n'importe quoi pour I_N, nous obtenons de bonnes approximations pour I_i avec les petits i. Même avec $I_{20} = 1000$ (qui est complètement faux), nous trouvons énormément de chiffres significatifs corrects pour I_1.

33.4.1 Comment choisir et penser le K ?

La formule (33.48) contient une formule qui ressemble étrangement à la dérivée. La stabilité d'un problème est très liée à la dérivée de F. La stabilité et la dérivée ne sont pas les mêmes choses, mais il n'est pas mauvais de penser au K de la stabilité comme la dérivée. Ou plus précisément : le supremum de la dérivée.

Un fil conducteur du lemme 33.33 et des exemples 33.34, 33.35 est que l'on a un K qui fonctionne lorsque la dérivée est bornée sur l'intervalle $]d_0 - \eta, d_0 + \eta[$. Dans le cas où ce supremum existe, le prendre en guise de K fonctionne souvent.

Il faut cependant parfois faire acte d'imagination. La fonction $x \mapsto |x|$ n'est pas dérivable en 0. Il n'empêche que $K = 1$ fait fonctionner la définition de la stabilité. Remarquez que $K = 1$ est le supremum de la dérivée là où elle existe.

À partir du moment où c'est clair que le K est le supremum de la dérivée, on comprend pourquoi c'est le gradient qui arrive dans le corolaire 33.32. En effet, le gradient indique la direction de plus grande pente. C'est donc bien dans cette direction qu'il faut chercher la « plus grande dérivée ».

Proposition 33.38.
Pour le problème stable $x = x(d)$ avec $x \in C^1(\mathbb{R}^n, \mathbb{R})$, on a

$$K_{abs}(d) \sim \|dx_d\| \tag{33.67}$$

où dx_d désigne la différentielle de x en d et la norme est la norme opérateur.

33.5 Un peu de points fixes

33.5.1 Choix de la fonction à point fixe

Pour l'équation $f(x) = 0$, il existe une infinité de fonctions g pour lesquelles l'équation est équivalente à $x = g(x)$.

Exemple : $f(x) = x^2 - 2 - \ln(x)$, nous pouvons faire

(1) $x = x^2 - 2 - \ln(x) + x$

(2) Poser $x^2 = 2 + \ln(x)$ et donc

$$x = -\sqrt{2 + \ln(x)} \tag{33.68a}$$

$$x = \sqrt{2 + \ln(x)}. \tag{33.68b}$$

(3) Ou encore

$$x = \frac{2 + \ln(x)}{x} \tag{33.69}$$

où nous savons déjà que $x \neq 0$ parce que $x = 0$ n'est pas dans le domaine de f.

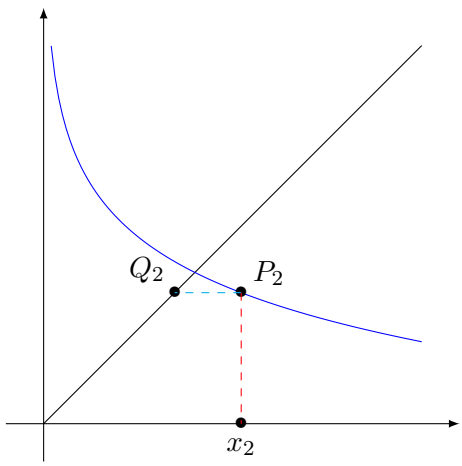
(4) Ou par l'exponentielle :

$$x = e^{x^2-2}. \tag{33.70}$$

Dans tous ces cas nous pouvons construire une suite (x_n) en posant un nombre arbitraire pour x_0 et ensuite la récurrence

$$x_{n+1} = g(x_n). \tag{33.71}$$

Graphiquement, la solution de l'équation est l'intersection entre les courbes $y = x$ et $y = g(x)$. Un petit dessin pour montrer la convergence :



Attention : cette méthode ne converge pas toujours. Parfois elle converge de façon monotone, et parfois pas. Le choix de la fonction g qui fait $x = g(x)$ peut énormément changer la vitesse de convergence.

Théorème 33.39 (Condition suffisante pour existence d'un point fixe)**.**
Une fonction continue $f\colon [a,b] \to [a,b]$ admet au moins un point fixe dans $[a,b]$.

Théorème 33.40 (Condition suffisante pour l'unicité)**.**
Soit f continue sur $[a,b]$ avec $g(x) \in [a,b]$ pour tout $x \in [a,b]$. Supposons qu'il existe $0 < k < 1$ tel que pour tout $x \in [a,b]$ nous ayons $|g'(x)| \leqslant k$ alors

(1) La fonction g possède un unique point fixe dans $[a,b]$.

(2) Pour tout $x_0 \in [a,b]$, tous les termes de la suite $x_{n+1} = g(x_n)$ sont dans $[a,b]$.

(3) Ladite suite (x_n) converge vers le point fixe.

Théorème 33.41.
Soit f continue sur $[a,b]$ avec $g(x) \in [a,b]$ pour tout $x \in [a,b]$. Supposons

(1) qu'il existe $0 < k < 1$ tel que pour tout $x \in [a,b]$ nous ayons $|g'(x)| \leqslant k$ et

(2) g est p fois dérivable sur $[a,b]$.

(3) $g'(\alpha) = g''(\alpha) = \ldots = g^{(p-1)}(\alpha)$ et $g^{(p)}(\alpha) \neq 0$ où α est l'unique point fixe.

Alors la suite (x_n) converge avec un ordre p.

Exemple 33.42.
Nous reprenons

$$f(x) = x^2 - 2 - \ln(x). \tag{33.72}$$

Et nous voulons résoudre $f(x) = 0$. Graphiquement c'est l'intersection entre $y = x^2 - 2$ et $y = \ln(x)$. Il est vite tracé de savoir qu'il y a deux solutions : $\alpha_1 \in [0,1]$ et $\alpha_2 \in [\sqrt{2}, 2]$.

Déjà un petit problème : l'intervalle $[0,1]$ ne va pas parce que f n'y est pas continue. Un petit raffinement d'analyse nous fournit $\alpha_1 \in [e^{-2}, 1]$.

Nous avons au moins les fonctions de points fixes suivantes :

$$g_1(x) = \sqrt{2 + \ln(x)} \tag{33.73a}$$

$$g_2(x) = e^{x^2-2}. \tag{33.73b}$$

Pour la première, il y avait un $\pm$ qui a été négligé parce que nous savons que les deux solutions cherchées sont positives. Travaillons avec la première. D'abord

$$g_1'(x) = \frac{1}{2x\sqrt{2 - \ln(x)}}. \tag{33.74}$$

Nous avons $\lim_{x \to e^{-2}} g_2'(x) = +\infty$. Il ne sera donc pas possible de trouver $0 < k < 1$ tel que $|g'(x)| \leqslant k$. Tentons quand même la méthode :

$$x_0 = 0.5 \tag{33.75}$$

Il se fait que cela est plus proche de α_1 que de α_2. Mais en réalité la suite converge vers α_2.

Passons à la seconde méthode.

$$g_2'(x) = 2xe^{x^2-2}. \tag{33.76}$$

Sur l'intervalle $[e^{-2}, 1]$, g_2' est croissante et prend toutes ses valeurs dans $[e^{-2}, 1]$. Nous pouvons prouver que

$$|g_2'(x)| \leqslant 2e^{-1} < 1. \tag{33.77}$$

Donc poser $k = 2e^{-1}$ fait fonctionner la proposition. Donc quel que soit le x_0 pris dans cet intervalle, nous aurons une suite convergente vers un point fixe à l'intérieur de l'intervalle. C'est-à-dire convergente vers α_1.

Cela est un exemple de problème pour lequel changer de fonction g change réellement la vie. △

33.5.2 Convergence quadratique

Définition 33.43.
*Une suite (x_n) a une convergence **quadratique** vers α si elle converge vers α et s'il existe un C tel que pour tout n nous ayons*

$$\|x_{n+1} - \alpha\| \leqslant C\|x_n - \alpha\|^2. \tag{33.78}$$

Il est bien entendu possible de parler de convergence quadratique si la relation (33.78) a lieu seulement à partir d'un certain indice.

Le lemme suivant donne l'importance du choix de point de départ lorsqu'on utilise une méthode itérative dont la convergence est quadratique.

Lemme 33.44.
Soit une suite $x_n \to \alpha$ de convergence quadratique. Si $\|x_0 - \alpha\| \leqslant r$ alors

$$\|x_n - \alpha\| \leqslant \frac{1}{C}(Cr)^{2^n} \tag{33.79}$$

Démonstration. Nous pourrions directement prouver la formule (33.79) par récurrence, mais nous allons la reconstruire un peu. Nous cherchons

$$\|x_n - \alpha\| \leqslant C^{k(n)} r^{2^n}. \tag{33.80}$$

Nous avons les inégalités

$$\begin{aligned} \|x_{n+1} - \alpha\| &\leqslant C\|x_n - \alpha\|^2 && (33.81a)\\ &\leqslant CC^{2k(n)} r^{2^{n+1}} && (33.81b)\\ &= C^{2k(n)+1} r^{2^{n+1}} && (33.81c) \end{aligned}$$

d'où nous voyons que la fonction k doit vérifier

$$\begin{cases} k(0) = 0 & (33.82a)\\ k(n+1) = 2k(n) + 1 & (33.82b) \end{cases}$$

La première équation est l'hypothèse $\|x_0 - \alpha\| \leqslant r$ comparée à la formule (33.80). Il est vite vérifié que $k(n) = 2^n - 1$. D'où le résultat. □

Si le point de départ est choisi de façon à avoir $Cr < 1$ alors nous avons là un très bon majorant parce qu'il s'agit d'un majorant convergeant très rapidement vers zéro. Si au contraire $Cr > 1$ alors ce majorant ne sert à rien.

33.45.
Le fait d'avoir une convergence quadratique signifie que le nombre décimales correctes double (environ) à chaque itération, dans n'importe quelle base. En effet supposons que x_n ait k décimales correctes ; cela signifie que $|x_n - \alpha| \sim 10^{-k}$. Donc

$$|x_{n+1} - \alpha| \lessapprox M10^{-2k}. \tag{33.83}$$

Cela est le double de décimales correctes de $|x_n - \alpha|$, moins l'ordre de grandeur de M.

Pour la méthode de bisection, le nombre de décimales augmente de 1 à chaque itération, mais seulement en base 2. En base 10, de façon générique[6] il faut entre 3 et 4 itérations pour avoir une décimale de plus.

33.46 (Condition d'arrêt[508])**.**
D'autre part, lorsqu'une méthode a une convergence quadratique, nous avons un test d'arrêt. Pour ce voir, nous avons la limite

$$\lim_{n\to\infty} \frac{|x_{n+1} - \alpha|}{|x_n - \alpha|} \leqslant \lim_{n\to\infty} \frac{C|x_n - \alpha|^2}{|x_n - \alpha|} = 0. \tag{33.84}$$

Cette limite est alors également valable sans les valeurs absolues et si nous soustrayons $x_n - \alpha$ au numérateur, la limite devient -1 :

$$-1 = \lim_{n\to\infty} \frac{x_{n+1} - x_n}{x_n - \alpha}. \tag{33.85}$$

Ou encore

$$\lim_{n\to\infty} \frac{x_n - x_{n+1}}{x_n - \alpha} = 1. \tag{33.86}$$

Cela a pour conséquence que si n est grand,

(1) x_{n+1} a le même ordre de grandeur que $x_n - \alpha$.
(2) $x_n - x_{n+1}$ et $x_n - \alpha$ ont le même signe.

Donc si nous voulons une approximation de α avec une erreur ϵ, il suffit d'arrêter le calcul lorsque $|x_{n+1} - x_n| \leqslant \epsilon$. Et ce faisant nous savons de plus si l'approximation est par excès ou par défaut.

6. C'est-à-dire sauf coup de malchance ou coup de chance.

33.5.3 Convergence

Proposition 33.47 (Convergence d'une méthode de point fixe[508]).
Soit $g\colon \mathbb{R} \to \mathbb{R}$ de classe C^1 et α un point fixe attractif[7] *de g. Soit k tel que $|g'(\alpha)| < k < 1$ et δ tel que $\|g'\|_{B(\alpha,\delta)} < k$.*

Alors

(1) La fonction g est k-contractante[8] *sur $B(\alpha,\delta)$.*

(2) Nous avons $g\big(B(\alpha,\delta)\big) \subset B(\alpha,\delta)$.

(3) Pour tout $x_0 \in B(\alpha,\delta)$ la suite $x_{n+1} = g(x_n)$ converge vers α et

$$|x_n - \alpha| \leqslant |x_0 - \alpha| k^n. \tag{33.87}$$

Si de plus $g'(\alpha) = 0$ et g est de classe C^2 alors nous avons convergence quadratique (définition 33.43).

Démonstration. Vu que α est un point fixe attractif de g nous pouvons considérer un k tel que $|g'(\alpha)| < k < 1$. Et comme g est de classe C^1, la fonction g' est continue et donc bornée sur toute boule du type $\overline{B(\alpha,\delta)}$. Soit δ le plus grand nombre tel que $\|g'\|_{\overline{B(\alpha,\delta)}} \leqslant k$. Nous notons $I = \overline{B(\alpha,\delta)}$ pour cette valeur de δ.

Pour tout $x \in I$ nous avons, en utilisant le théorème des accroissements finis 11.130(2) :

$$\begin{aligned}
|g(x) - \alpha| &= |g(x) - g(\alpha)| && (33.88a)\\
&\leqslant \sup_{t\in I} |g'(t)||x-\alpha| && (33.88b)\\
&\leqslant k|x-\alpha| && (33.88c)\\
&< \delta && (33.88d)
\end{aligned}$$

parce que $k < 1$ et $|x - \alpha| \leqslant \delta$. Par conséquent $g(x) \in B(\alpha,\delta)$. Cela prouve le point (2). Pour le point (1), soient $x, y \in B(\alpha,\delta)$ et

$$|g(x) - g(y)| \leqslant \sup_{a\in I} |g'(a)||x-y| \leqslant k|x-y|. \tag{33.89}$$

Pout le point (3) nous avons $|g(x_n) - \alpha| \leqslant k|x_n - \alpha|$, c'est-à-dire

$$|x_{n+1} - \alpha| \leqslant k|x_n - \alpha|. \tag{33.90}$$

Le résultat annoncé s'obtient par récurrence sur n.

En ce qui concerne la convergence quadratique, c'est du Taylor (proposition 11.385). Développons $g(x_n)$ autour de $g(\alpha)$:

$$g(x_n) = g(\alpha) + g'(\alpha)(x_n - \alpha) + \frac{1}{2}(x_n - \alpha)^2 \epsilon(x_n - \alpha) \tag{33.91}$$

avec $\lim_{t\to 0} \epsilon(t) = 0$. En posant $C = \frac{1}{2}\sup_{t<\delta} |\epsilon(t)|$ nous avons $|g(x_n) - g(\alpha)| \leqslant C|x_n - \alpha|^2$, c'est-à-dire

$$|x_{n+1} - \alpha| \leqslant C|x_n - \alpha|^2. \tag{33.92}$$

□

Ce corolaire est une paraphrase de la proposition 33.47. Il en retient seulement les points intéressants en pratique.

Corolaire 33.48.
Soit α une solution de l'équation $x = g(x)$, avec g continue sur un voisinage de α et dérivable dans l'intérieur. Nous supposons que

$$|g'(\alpha)| < 1. \tag{33.93}$$

Alors il existe un rayon δ tel que si $x_0 \in B(\alpha,\delta)$, la suite (x_n) converge vers α.

7. Définition 16.30.
8. Définition 16.33

Certes cette proposition demande moins d'hypothèses, mais en réalité, il ne donne pas de vrais moyens de choisir un point de départ x_0. Avec les deux théorèmes précédents, nous pouvions prendre x_0 n'importe où dans $[a, b]$. Le fait est que pour choisir x_0 nous pouvons tracer et donner à la main un x_0 proche de ce qui semble être α. Si ça ne converge pas, il faut donner un x_0 plus proche. La proposition nous assure que si nous jouons bien à choisir x_0 très proche, la suite finira par converger.

Notons que le corolaire 33.48 a encore l'inconvénient de demander de calculer $g'(\alpha)$ alors que α est inconnu. La résolution de l'inéquation $|g'(x)| < 1$ nous donne un certain nombre d'intervalles dans $\mathbb{R}$.

Soient I_n les intervalles solutions de l'inéquation. Si $\alpha \in I_n$ alors la méthode converge. Sinon, c'est pas garantit. En tout cas nous ne devons pas savoir réellement α pour appliquer le théorème. Il suffit de savoir que α est dans un des I_n.

33.6 Méthode de Newton

L'objectif de la méthode de Newton est d'évaluer une racine α de l'équation $f(x) = 0$ lorsque nous avons déjà une approximation x_0 de la racine α.

C'est la méthode de Newton qui est à l'origine de la suite de la proposition 1.127 donnant une suite dans $\mathbb{Q}$ qui converge vers $\sqrt{A}$.

Définition 33.49.
*Le nombre α est une **racine simple** de l'équation $f(x) = 0$ si $f(\alpha) = 0$ et $f'(\alpha) \neq 0$. Le nombre α est une **racine multiple** d'ordre r de $f(x) = 0$ si*

$$f(\alpha) = f'(\alpha) = \ldots = f^{(r-1)(\alpha)} = 0 \tag{33.94}$$

et $f^{(r)}(\alpha) \neq 0$.

Exemple 33.50.
La fonction $x \mapsto x^3$ en $x = 0$ est un racine d'ordre 3. △

33.6.1 « Justification » par la formule par Taylor

Soit une fonction f continue et dérivable sur $[a, b]$. Soit α une racine de f et x_n une de ses approximations. Nous notons l'erreur θ et nous avons $\alpha = x_n + \theta$. Du coup nous avons $f(x_n + \theta) = f(\alpha) = 0$.

Écrivons la série de Taylor du théorème 11.375 autour de x_n : il existe une fonction $\epsilon\colon \mathbb{R} \to \mathbb{R}$ telle que $\lim_{t\to 0} \epsilon(t) = 0$ telle que

$$f(\alpha) = f(x_n + \theta) = f(x_n) + \theta f'(x_n) + \frac{\theta^2}{2}\epsilon(\theta). \tag{33.95}$$

Nous isolons le θ du terme d'ordre 1 en nous souvenant que le membre de gauche est nul :

$$\theta = -\frac{f(x_n) - \theta^2\epsilon(\theta)}{f'(x_n)} \tag{33.96}$$

Vu que $\alpha = x_n + \theta$, nous pouvons écrire

$$\alpha = x_n - \frac{f(x_n) + \theta^2\epsilon(\theta)}{f'(x_n)}. \tag{33.97}$$

Il est donc raisonnable de poser

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{33.98}$$

en espérant que cela soit une meilleur approximation de α que x_n.

En tout cas l'erreur sur x_{n+1} est

$$\alpha - x_{n+1} = x_n + \theta - x_n + \frac{f(x_n) + \theta^2\epsilon(\theta)}{f'(x_n)} = \theta + \frac{f(x_n) + \theta^2\epsilon(\theta)}{f'(n_n)}, \tag{33.99}$$

qui ne doit pas être fondamentalement plus grand que θ dès que θ est petit, surtout que si x_n est une approximation de α, nous pouvons espérer que $f(x_n)$ soit également petit. Là où les choses peuvent déraper en grand, c'est si $f'(x_n)$ est petit.

Cette méthode de Newton ne converge pas toujours. Le pire est lorsque par malheur il y a une bosse pas loin de la racine. Alors il y a un risque de tomber sur $f'(x_{n+1}) = 0$ ou en tout cas très proche de zéro. Dans ce cas le point x_{n+2} est envoyé très loin.

33.6.2 « Justification » par points fixes

Nous savons que pour résoudre $f(x) = 0$ par une méthode de point fixe, il y a de nombreux choix possibles de fonctions g telles que $g(x) = x$ donne la même solution que $f(x) = 0$. Soit α une solution de $f(x) = 0$ et cherchons une fonction g de la forme

$$g(x) = x - kf(x). \tag{33.100}$$

Nous savons par la proposition 33.47 que la fonction g donne une convergence quadratique lorsque $g'(\alpha) = 0$. Pour la forme (33.100) nous avons $g'(\alpha) = 1 - kf'(\alpha)$, ce qui nous donne l'idée de poser $k = \frac{1}{f'(\alpha)}$.

Le fait est que $f'(\alpha)$ n'est pas connu, mais nous pouvons l'approximer par $f'(x)$ lorsque x est proche de α. D'où l'idée de considérer la fonction

$$g(x) = x - \frac{f(x)}{f'(x)}, \tag{33.101}$$

et donc la suite $x_{n+1} = g(x_n)$ c'est-à-dire

$$x_{n+1} = x - \frac{f(x_n)}{f'(x_n)}. \tag{33.102}$$

Dès que x_n est proche de α, sous l'hypothèse (raisonnable par continuité) que $f'(x_n)$ soit proche de $f'(\alpha)$, la méthode devrait donner une convergence quadratique.

Remarque 33.51.
Cette justification par points fixes n'est pas vraiment différente de celle par Taylor parce que Taylor est utilisé dans la preuve de la proposition 33.47.

Définition 33.52 (Méthode de Newton)**.**
*La **méthode de Newton** pour la fonction f est la suite définie par récurrence*

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{33.103}$$

Cette définition ne précise pas la valeur de x_0, ni de condition d'arrêt.

33.6.3 Convergence de la méthode de Newton

Théorème 33.53 (Convergence quadratique de la méthode de Newton[508])**.**
Soit f une fonction continue vérifiant $f(\alpha) = 0$ et $f'(\alpha) \neq 0$. Nous considérons la fonction

$$g(x) = x - \frac{f(x)}{f'(x)} \tag{33.104}$$

que nous supposons être de classe C^2.

Si C est une majoration de $\|g''\|$ sur un intervalle contenant α, alors en posant $\delta = 1/C$ nous avons

(1) La boule $B(\alpha,\delta)$ est préservée par g : $g\big(B(\alpha,\delta)\big)\subset B(\alpha,\delta)$.

(2) Pour tout $x_0\in B(\alpha,\delta)$ nous avons convergence quadratique vers α de la suite définie par $x_{n+1}=g(x_n)$.

(3) Nous avons l'estimation

$$|x_n-\alpha|\leqslant\frac{1}{C}\big(C|x_0-\alpha|\big)^{2^n}\tag{33.105}$$

où C est la constante de la définition de convergence quadratique.

Démonstration. Nous commençons par calculer la dérivée de g :

$$g'(x)=-\frac{f(x)f''(x)}{f'(x)^2},\tag{33.106}$$

d'où nous déduisons que $g'(\alpha)=0$. Ensuite nous utilisons abondamment la formule des accroissements finis (théorème 11.257) en commençant par

$$|g(t)-g(\alpha)|\leqslant\|g'\|_{[t,\alpha]}|t-\alpha|\tag{33.107}$$

où par $\|f\|_A$ nous entendons la norme uniforme de f sur A, c'est-à-dire $\|f\|_A=\sup_{x\in A}\|f(x)\|$. Note : nous écrivons $[t,\alpha]$, mais ça pourrait être $[\alpha,t]$.

Si $x\in[t,\alpha]$ alors

$$\begin{aligned}|g'(x)|&=|g'(x)-g'(\alpha)|&(33.108a)\\&\leqslant\|g''\|_{[x,\alpha]}|x-\alpha|&(33.108b)\\&\leqslant\|g''\|_{[x,\alpha]}|t-\alpha|&(33.108c)\\&\leqslant\|g''\|_{[t,\alpha]}|t-\alpha|.&(33.108d)\end{aligned}$$

En particulier, $\|g'\|_{[t,\alpha]}\leqslant\|g''\|_{[t,\alpha]}|t-\alpha|$, et nous pouvons continuer les majorations (33.107) :

$$|g(t)-g(\alpha)|\leqslant\|g''\|_{[t,\alpha]}|t-\alpha|^2.\tag{33.109}$$

La fonction g étant de classe C^2, la dérivée seconde g'' est bornée (nous supposons déjà travailler sur un compact contenant α). Soit C une borne. Nous sommes en mesure de prouver le point (1) avec $\delta=1/C$. En effet si $t\in B(\alpha,1/C)$ alors

$$|g(t)-\alpha|=|g(t)-g(\alpha)|\leqslant C|t-\alpha|^2\leqslant C\frac{1}{C^2}=\frac{1}{C},\tag{33.110}$$

ce qui prouve que $g(t)\in B(\alpha,1/C)$.

Le point (2) se prouve de la même manière : si $x_n\in B(\alpha,1/C)$ alors

$$|x_{n+1}-\alpha|=|g(x_n)-g(\alpha)|\leqslant C|x_n-\alpha|^2,\tag{33.111}$$

ce qui est bien la convergence quadratique.

La majoration du point (3) s'obtient par récurrence sur n. Pour $n=0$, la relation (33.105) devient $|x_0-\alpha|\leqslant|x_0-\alpha|$ qui est vraie. Ensuite par la convergence quadratique et la récurrence,

$$|x_{n+1}-\alpha|\leqslant C|x_n-\alpha|^2\leqslant C\big[\frac{1}{C}(C|x_0-\alpha|)^{2^n}\big]^2=\frac{1}{C}\big[M|x_0-\alpha|\big]^{2^{n+1}}.\tag{33.112}$$

□

33.54.

Dans le cas pratiques, nous commençons souvent par résoudre l'équation $f(x)=0$ par dichotomie. Au moment où nous sommes assez proche de la solution nous commençons Newton.

La raison est que la dichotomie fonctionne toujours : nous allons toujours nous approcher de la solution. Si par contre le point de départ est mal choisit, la méthode de Newton peut envoyer n'importe où, y compris très loin de la solution.

La proposition suivante nous indique que dans le cas d'une fonction convexe, le choix de point de départ de la méthode de Newton n'est pas tellement crucial parce que il sont tous bons. De plus la convergence se faisant de façon décroissante (si on part de la droite), nous savons que le résultat sera une approximation par excès de α.

Proposition 33.55 (Newton dans le cas convexe)**.**
Soit f de classe C^2 et une racine α telle que $f'(\alpha) > 0$. Soit $b > \alpha$ tel que f soit convexe sur $[\alpha, b]$.

Alors pour tout $x_0 \in [\alpha, b]$ la suite de la méthode de Newton est

(1) décroissante

(2) reste dans $[\alpha, b]$

(3) converge vers α.

Démonstration. Nous savons par la proposition 16.81(2) que la fonction f' est croissante, et par hypothèse $f'(\alpha) > 0$, donc sur $[\alpha, b]$ nous avons $f' > 0$. Par conséquent, nous avons aussi $f > 0$ sur $[\alpha, b]$.

Le graphe de f est au dessus de la tangente de f en $x = x_n$ (proposition 16.87). Si nous nommons t_x la fonction qui donne la tangente en x nous avons $t_{x_n}(\alpha) < 0$ parce que $f(\alpha) = 0$. Par conséquent

$$t_{x_n}(x) = 0 \tag{33.113}$$

pour $\alpha < x < x_n$. Cela prouve que $x_{n+1} \in [\alpha, b]$, et que (x_n) est une suite décroissante

Étant donné que (x_n) est une suite décroissante dans le compact $[\alpha, b]$, elle est convergente. Notons β sa limite. Nous avons la relation de récurrence

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{33.114}$$

En passant à la limite $n \to \infty$ nous avons l'équation

$$\beta = \beta - \frac{f(\beta)}{f'(\beta)}. \tag{33.115}$$

Vu que $f(x) > 0$ sur $]\alpha, b]$ nous avons automatiquement $\beta = \alpha$. □

33.6.4 Formalisation de l'algorithme

La méthode de Newton consiste a exprimer la solution x de $f(x) = 0$ avec $f \in C^1(\mathbb{R})$ comme limite d'une suite $\{x_n\}_{n\in\mathbb{N}}$ définie par récurrence par la formule

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{33.116}$$

où x_0 est arbitraire.

Si on veut exprimer cela en termes d'algorithmes, nous disons que l'algorithme de Newton est donné par la suite de problèmes

$$F_n(x_{n+1}, x_n, f) = x_{n+1} - x_n + \frac{f(x_n)}{f'(x_n)}. \tag{33.117}$$

La donnée du problème est la fonction f, et rien que elle.

Plus précisément, une fois que la fonction f est donnée, il existe une infinité de problèmes : pour chaque $a \in \mathbb{R}$ nous avons le problème

$$G_a(x_n, f) = x - a + \frac{f(a)}{f'(a)}. \tag{33.118}$$

La méthode de Newton consiste à sélectionner une partie de ces problèmes de la façon suivante :

$$\begin{cases} F_0 = G_{x_0} & (33.119a) \\ F_n = G_{x_n}. & (33.119b) \end{cases}$$

Le problème F_0 fournit un nombre x_1 qui nous permet de sélectionner le problème G_{x_1} qui va fournir le nombre x_2, etc.

Au moment de calculer le conditionnement de F_n, nous ne devons pas voir x_{n-1} comme fonction de x_0 et de la donnée f. Il ne faut donc pas dériver à travers les x_n.

Proposition 33.56.
Si une racine est multiple, alors l'ordre de convergence de la méthode de Newton est 1.

Voici un algorithme possible :

```
def Newton(f,x0,toll,maxit):
    fp=f.derivative()
    n=0
    x=x0
    diff=toll+1
    while abs(diff)>toll and n<maxit:
        n=n+1
        diff=-f(x)/fp(x)
        x=x+diff
    return x,n
```

tex/frido/codeSnip_2.py

Commentaires :

(1) Notons que dans un langage vraiment numérique comme Matlab, il faut passer f' en argument.

(2) Dans le `while` il faudrait mettre $x_{n+1} - x_n$ (en valeur absolue), mais cette différence est aussi utilisée pour calculer x_{n+1} donc on la calcule une seule fois.

(3) Il faudrait faire une vérification sur $f(x_n) \neq 0$. Il n'y a pas tellement de choix que de changer le point initial.

33.6.5 Caractéristiques

L'algorithme de Newton a les caractéristiques suivantes :

(1) Pour résoudre le problème numéro n, il faut avoir résolu le problème numéro $n-1$.

(2) Aucune des solutions x_n aux problèmes intermédiaires n'est une solution au problème de départ (à moins d'un coup de chance).

(3) Étant donné que la donnée du problème F_n est la fonction f de départ, nous avons $d_m = d_n = d$ pour tout m et n.

Théorème 33.57.
Soit f continue sur un voisinage de α, racine simple. Alors il existe un voisinage de α de rayon σ tel que pour tout x_0 dans ce voisinage, la méthode converge vers α avec ordre de convergence $p = 2$.

Donc dès qu'on a continuité autour de la solution recherchée, il suffit de prendre x_0 assez proche pour que tout se passe bien. Cela se fait par localisation des racines, par exemples en traçant la fonction avec un bon niveau de zoom. Le fait est qu'on cherche disons 3 décimales à la main (travail sur ordinateur et graphique) et Newton donne les 20 décimales suivantes à la vitesse de la lumière.

33.6.6 Exemple de la racine carrée

Nous allons nous lancer dans un exemple : le cas de la racine carrée. Soit à calculer une approximation numérique de $\sqrt{2}$. Il s'agit d'une racine de la fonction $f(x) = x^2 + 2$. La fonction de la méthode de Newton associée est :

$$g(x) = x - \frac{f(x)}{f'(x)} = \frac{x^2 - 2}{2x}. \tag{33.120}$$

Cherchons un intervalle autour de $\sqrt{2}$ sur lequel nous avons convergence de la méthode de Newton. Cela s'obtient grace à la proposition 33.47 qui nous informe qu'il suffit de trouver un intervalle autour de $\sqrt{2}$ sur lequel $|g'(x)| \leqslant 1$.

Nous avons

$$g'(x) = \frac{x^2 - 2}{2x^2}, \tag{33.121}$$

et nous cherchons à résoudre $|g'(x)| \leqslant 1$. D'abord $g'(x) = 1$ n'a aucune solutions alors que $g'(\sqrt{2}) = 0$. Donc nous avons $g'(x) \leqslant 1$ pour tout $x \in \mathbb{R}^+$. Par contre l'équation $g'(x) = -1$ a des solutions : $x = \pm\sqrt{2/3}$.

Nous avons donc convergence de la méthode de Newton pour x_0 dans un intervalle de la forme

$$[\sqrt{2/3}, \sqrt{2} + \ldots] \tag{33.122}$$

où les les trois points représentent l'expression qu'il faut pour que ce soit symétrique autour de $\sqrt{2}$. La valeur précise n'a pas tellement d'importance parce, vu que nous sommes en train de chercher $\sqrt{2}$, il est peu probable que nous ayons déjà en main une bonne approximation de nombres du type $\sqrt{2/3}$.

Proposition 33.58.
La méthode de Newton pour la fonction $f(x) = x^2 - 2$ converge vers $\sqrt{2}$ pour toute valeur de départ dans $]0, +\infty[$.

Démonstration. La fonction $f(x) = x^2 - 2$ est convexe et $f'(\sqrt{2}) = 2\sqrt{2} > 0$. Donc la méthode converge vers $\sqrt{2}$ pour tout $x_0 \geqslant \sqrt{2}$ par la proposition 33.55.

Si par contre $x_0 \in]0, \sqrt{2}[$ nous avons

$$x_1 = \frac{x_0^2 + 2}{2x_0}. \tag{33.123}$$

En posant $h(x) = (x^2 + 2)/2x$ et en résolvant $h'(x) = 0$ nous trouvons $x = \sqrt{2}$. Et là, $h(\sqrt{2}) = \sqrt{2}$. Donc $h(x)$ est toujours plus grand que $\sqrt{2}$ pour tout $x \in]0, \sqrt{2}[$.

En d'autres termes, si $x_0 \in]0, \sqrt{2}[$ alors $x_1 \geqslant \sqrt{2}$ et nous retombons dans le premier cas. □

33.6.7 Si multiplicité

Supposons que α soit de multiplicité r (définition 33.49).

Cela se remarque en voyant que la méthode de Newton demande plutôt 20 itérations que 5. Le problème que cela pose est que chaque itération, les évaluations provoquent des erreurs. Donc moins d'itérations, c'est mieux.

Nous pouvons modifier la formule avec

$$x_{n+1} = x_n - r\frac{f(x_n)}{f'(x_n)}. \tag{33.124}$$

Il est possible de prouver que cette suite est à nouveau à convergence quadratique.

Ou alors on pose $F(x) = f^{(r-1)}(x)$ et α est une racine simple pour F. Donc faire Newton pour F est à nouveau quadratique, tout en donnant la même solution parce que $F(\alpha) = 0$ et $F'(\alpha) \neq 0$.

La seconde façon est bien parce que le théorème de localisation fonctionne 33.57

Et si r n'est pas connu ?

Il est toujours possible de faire $r = 2$ puis $r = 3$ et caetera jusqu'au moment où l'on remarque que le nombre d'itérations baisse un grand coup.

Mais ça demande beaucoup de calculs. Le mieux est de changer de méthode.

33.6.8 Et la dérivée ?

Un des problèmes de la méthode de Newton est que l'on doit pouvoir calculer la dérivée. Typiquement, il faut savoir f de façon analytique. Si cela n'est pas possible, nous pouvons changer de méthode et utiliser la méthode des sécantes décrite en 33.9.

33.6.9 Méthode de Newton : le cas général

Lemme 33.59.
Soient A et B deux matrices inversibles telles que la matrice $(A+\epsilon B)$ soit inversible pour tout ϵ assez petit. Alors il existe une matrice $X(\epsilon)$ telle que

$$(A+\epsilon B)^{-1} = (A^{-1}+\epsilon X) \tag{33.125}$$

et telle que $\lim_{\epsilon\to 0} X(\epsilon) = -A^{-1}BA^{-1}$.

Démonstration. Le candidat matrice X est relativement simple à trouver en écrivant

$$(A+\epsilon B)(A^{-1}+\epsilon X) = \mathbb{1} + \epsilon AX + \epsilon BA^{-1} + \epsilon^2 BX. \tag{33.126}$$

En imposant que cela soit $\mathbb{1}$, nous trouvons

$$X(\epsilon) = -(A+\epsilon B)^{-1}BA^{-1}. \tag{33.127}$$

La matrice $X(\epsilon)$ étant un inverse à droite de $(A+\epsilon B)$, son déterminant est non nul et X est inversible. Par conséquent elle est également inversible au sens usuel. Le calcul de la limite est direct :

$$\lim_{\epsilon\to 0} -(A+\epsilon B)^{-1}BA^{-1} = A^{-1}BA^{-1} \tag{33.128}$$

parce que l'inverse est une fonction continue sur $\mathbb{M}(n,\mathbb{R})$. □

Remarque 33.60.
Un calcul naïf nous permet de trouver le même résultat de façon plus heuristique. En effet un développement usuel (dans $\mathbb{R}$) est

$$\frac{1}{a+\epsilon b} = \frac{1}{a} - \frac{\epsilon b}{a^2} + \dots \tag{33.129}$$

Si nous récrivons cela avec des matrices, nous écrivons (attention : passage heuristique !) :

$$(A+\epsilon B)^{-1} = A^{-1} - \epsilon A^{-1}BA^{-1} + \dots \tag{33.130}$$

Notons le choix de généraliser b/a^2 par $a^{-1}ba^{-1}$. Dans les réels les deux écritures sont équivalentes, mais pas dans les matrices.

Étudions si $A^{-1}-\epsilon A^{-1}BA^{-1}$ est bien un inverse à ϵ^2 près de $(A+\epsilon B)$:

$$(A+\epsilon B)(A^{-1}+\epsilon A^{-1}BA^{-1}) = 1-\epsilon BA^{-1}+\epsilon BA^{-1}-\epsilon^2 BA^{-1}BA^{-1} = 1-\epsilon^2 BA^{-1}BA^{-1}. \tag{33.131}$$

Par conséquent, à des termes en ϵ^2 près la matrice $A^{-1}-\epsilon A^{-1}BA^{-1}$ est bien un inverse de $A+\epsilon B$.

Théorème 33.61 (Méthode de Newton[509])**.**
Soit $f\colon \mathbb{R}^n \to \mathbb{R}^n$ une application de classe C^2 et un point $a\in\mathbb{R}^n$ tel que $f(a)=0$. Nous supposons que df_a est inversible.

Alors il existe un voisinage V de a tel que pour tout $x_0\in V$ la suite définie par récurrence

$$x_{n+1} = x_n - (df_a)^{-1}\big(f(x_n)\big) \tag{33.132}$$

converge vers a. De plus la vitesse est quadratique au sens où il existe $C>1$ tel que

$$\|x_n - a\| \leqslant C^{-1-2^n}. \tag{33.133}$$

Démonstration. Étant donné que df_a est inversible et que df est continue, l'application df_x est continue[9] pour tout x dans un voisinage de a. Nous prenons $r > 0$ tel que df_x est inversible pour tout $x \in B(a, r)$.

Nous considérons la fonction

$$\begin{aligned} F\colon B(a,r) &\to \mathbb{R}^n \\ x &\mapsto x - (df_x)^{-1}\big(f(x)\big). \end{aligned} \tag{33.134}$$

Cela est une application C^1. La clef est de montrer que l'application de F à un point $a+h$ rapproche de a pourvu que h soit assez petit. Nous avons la formule suivante :

$$F(a+h) - F(a) = h - \big(df_{a+h}\big)^{-1}\big(f(a+h)\big). \tag{33.135}$$

Nous allons maintenant utiliser un développement de Taylor par rapport à h en suivant la formule (11.1060). Nous avons

$$f(a+h) = f(a) + df_a(h) + \|h\|^2\xi(h) \tag{33.136}$$

où $\xi\colon \mathbb{R}^n \to \mathbb{R}^n$ est une fonction qui tend vers une constante lorsque $h \to 0$. Nous avons aussi

$$df_{a+h} = df_a + \|h\|\tau(h) \tag{33.137}$$

où $\tau\colon \mathbb{R}^n \to \mathbb{M}(n, \mathbb{R})$ est une application qui tend vers une constante lorsque $h \to 0$. En ce qui concerne l'inverse nous utilisons le lemme[10] 33.59 :

$$\big(df_a + \|h\|\tau(h)\big)^{-1} = (df_a)^{-1} + \|h\|A(h) \tag{33.138}$$

où A est une autre matrice fonction de h qui tend vers une constante lorsque h tend vers zéro. En substituant le tout dans (33.135) nous trouvons

$$F(a+h) - F(a) = \|h\|^2(df_a)^{-1}\xi(h) + \|h\|\big(A(h) \circ df_a\big)(h) + \|h\|^3 A(h)\xi(h). \tag{33.139}$$

En ce qui concerne la norme nous utilisons le fait que si T est un opérateur, $\|Tx\| \leqslant \|T\|\|x\|$. Nous trouvons

$$\|F(a+h) - F(a)\| \leqslant \|h\|^2\|(df_a)^{-1}\|\|\xi(h)\| + \|h\|^2\|A(h) \circ df_a\| + \|h\|^3\|A(h)\|\|\xi(h)\| \tag{33.140a}$$

$$= \|h\|^2\alpha(h) \tag{33.140b}$$

pour une certaine fonction $\alpha\colon \mathbb{R}^n \to \mathbb{R}$ qui tend vers une constante lorsque $h \to 0$.

En posant $C = \lim_{h\to 0} \alpha(h)$ nous avons la majoration

$$\|F(x) - a\| \leqslant C\|x - a\|^2. \tag{33.141}$$

Nous pouvons également supposer que $C > 1$. Afin de prouver la vitesse de convergence (33.133), nous allons encore redéfinir r en demandant $r < 1/C^2$. De cette manière nous avons

$$\|x_0 - a\| \leqslant \frac{1}{C^2} \tag{33.142}$$

et la récurrence sur n est :

$$\|x_{n+1} - a\| = \|F(x_n) - a\| \leqslant C\|x_n - a\|^2 \leqslant C\big(C^{-1-2^n}\big)^2 = C^{-1-2^{n+1}}. \tag{33.143}$$

Note : ce dernier calcul est le lemme 33.44 appliqué à $r = (1/C^2)$. □

Remarque 33.62.
La valeur de la constante C a été fixée par l'équation (33.141). Certes nous pouvons toujours choisir C plus grand affin d'augmenter la vitesse de convergence, mais le point de départ x_0 devant être dans une boule de taille $1/C^2$ autour de a, demander C plus grand revient à demander un point de départ plus précis.

9. Nous pouvons voir df comme l'application qui à x fait correspondre la matrice $df_x \in \mathbb{M}(n, \mathbb{R})$. Cette application étant continue et la non inversibilité d'une matrice étant donnée par l'annulation du déterminant, les matrices inversibles forment un ouvert dans l'ensemble des matrices.

10. Pour l'inversibilité de $\|h\|\tau(h)$, notons que df_a est inversible et que par hypothèse la somme $df_a + \|h\|\tau(h)$ est inversible.

33.7 Estimation de l'ordre de convergence

Définition 33.63 ([510]).
Nous disons que la suite (x_n) de limite x est ***convergente d'ordre*** *q pour $q > 1$ s'il existe $\mu > 0$ tel que*

$$\lim_{k\to\infty} \frac{|x_{k+1} - x|}{|x_k - x|^q} = \mu. \tag{33.144}$$

En particulier :

- *la convergence d'ordre 2 est dite quadratique,*
- *la convergence d'ordre 3 est dite cubique,*
- *la convergence d'ordre 4 est dite quartique.*

Comment estimer numériquement l'ordre p de convergence de la méthode ? Soit une suite (x_n) convergente vers α. Considérons les 4 termes x_{n-3}, x_{n-2}, x_{n-1}, x_n. Alors nous pouvons écrire l'approximation

$$\frac{|x_n - x_{n-1}|}{|x_{n-1} - x_{n-2}|} \simeq \left(\frac{|x_{n-1} - x_{n-2}|}{|x_{n-1} - x_{n-3}|}\right)^p. \tag{33.145}$$

Cette approximation ne serait pas trop mauvaise tant que n est assez grand pour que la convergence soit bien engagée. Passons au logarithme :

$$\ln \frac{|x_n - x_{n-1}|}{|x_{n-1} - x_{n-2}|} \simeq p \ln \left(\frac{|x_{n-1} - x_{n-2}|}{|x_{n-1} - x_{n-3}|}\right). \tag{33.146}$$

et donc

$$p \simeq \frac{\ln\left(\frac{|x_n - x_{n-1}|}{|x_{n-1} - x_{n-2}|}\right)}{\ln\left(\frac{|x_{n-1} - x_{n-2}|}{|x_{n-1} - x_{n-3}|}\right)}. \tag{33.147}$$

Avec cette approximation, en réalité nous calculons une suite (p_i) qui sont les approximations de p à partir des termes i à $i + 3$ de la suite (x_n). Il s'agit d'une suite d'estimations de p.

(1) Dans le cas de la bisection, nous obtenons toujours $p_i = 1$.

(2) Dans le cas de la méthode de Newton (33.6) nous avons $p = 2$. Mais les premières valeurs de p_i peuvent être aussi bien 0 que 7. Après quelques itérations pourtant les p_i se regroupent autour de 2.

En tout cas, le plus important est de savoir si $p > 1$ ou non. Rappel : nous voulons la superlinéarité parce que nous voulons utiliser le test d'arrêt de la différence entre deux termes, voir 33.46.

33.8 Autres méthodes

33.8.1 Méthode de Schröder

La formule est

$$x_{n+1} = x_n - \frac{f(x_n)f'(x_n)}{f'(x_n)^2 - f(x_n)f''(x_n)} \tag{33.148}$$

Cette méthode est d'ordre 2 pour toute racine et toute valeur de multiplicité. Le problème de cette méthode est qu'elle demande 3 évaluations de f. Son efficacité :

$$E = \sqrt[3]{2} \simeq 1.25 \tag{33.149}$$

Cela est donc moins efficace que Newton.

33.8.2 Halley

Il a $p = 3$ lorsque α est racine simple. Mais encore $p = 1$ pour les racines multiples. Plus efficace que Newton pour les racines simples, mais même problème pour les racines multiples.

$$x_{n+1} = x_n - \frac{2f(x_n)f'(x_n)}{2f'(x_n)^2 - f(x_n)f''(x_n)} \tag{33.150}$$

33.9 Méthode des sécantes variables

Si nous n'avons pas de formule analytique pour f, mais seulement la possibilité de calculer $f(x)$ pour tout x. Newton ne fonctionne pas, mais la bisection fonctionne.

Nous pouvons approximer

$$f'(x_n) = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}. \tag{33.151}$$

En substituant dans la formule de Newton, nous obtenons

$$x_{n+1} = x_n - \frac{f(x_n)(x_n - x_{n-1})}{f(x_n) - f(x_{n-1})}. \tag{33.152}$$

Il s'agit de prendre la droite qui passe par $(x_{n-1}, f(x_{n-1}))$ et par $(x_n, f(x_n))$ et de prendre l'intersection de cette droite avec l'axe $y = 0$. Cela donne le x_{n+1}.

Pour cette méthode, il ne faut pas seulement x_0 mais également x_1.

L'ordre de convergence est le nombre d'or

$$p = \frac{1 - \sqrt{5}}{2} \simeq 1.618. \tag{33.153}$$

Cela est donc superlinéaire.

Le nombre d'évaluations est $s = 1$ (il y a deux apparitions de f dans la formule, mais l'une des deux est récupérée dans l'itération suivante). Donc l'efficacité est

$$E = p. \tag{33.154}$$

Donc bien efficace.

Proposition 33.64.
Si α est racine simple, il existe un voisinage de α tel que pour tout choix de x_0, x_1 dans ce voisinage, la méthode converge.

Psychologiquement, on est tenté de prendre x_0 et x_1 de part et d'autre de α (pensant à la bisection), mais en réalité ce n'est pas obligatoire du tout et n'a aucune influence. Il faut seulement les prendre très proches de α.

Remarque 33.65.
La méthode de la sécante est souvent écrite sous la forme

$$x_{n+1} = \frac{x_{n-1}f(x_n) - x_n f(x_{n-1})}{f(x_n) - f(x_{n-1})}. \tag{33.155}$$

C'est évidemment algébriquement équivalent.

Les formules (33.153) et (33.155) ont toutes deux des erreurs de cancellation. Laquelle est la plus grave ?

Dans la première, si la fraction est mal calculée, elle ne fait que modifier x_n. C'est-à-dire qu'on peut espérer qu'à la prochaine itération, ça aille mieux. En tout cas, dans ce cas si la fraction est mal calculée, ça ne détruit pas tout.

Dans la seconde, c'est la valeur elle-même qui risque d'être mal calculée. Et si la fraction est mal calculée, alors on casse complètement l'éventuelle bonne approximation que nous avions déjà.

33.9.1 Aitken

La méthode du Δ^2 de Aitken est une méthode d'accélération de la convergence.

Soit (x_n) une suite qui converge. Nous voudrions une nouvelle suite (y_n) telle que

$$\lim_{n\to\infty} \frac{y_n - \alpha}{x_n - \alpha} \tag{33.156}$$

C'est la définition d'une convergence accélérée.

La façon de faire est :

$$y_n = \frac{x_{n+2}x_n - x_{n+1}^2}{x_{n+2} - 2x_{n+1} + x_n} = x_n - \frac{(x_{n+1} - x_n)}{x_{n+2} - 2x_{x+1} + x_n}. \tag{33.157}$$

La première expression a deux cancellations (la seconde une seule) et de plus la première est y_n elle-même alors que la seconde est une correction.

Donc la seconde expression est numériquement meilleure.

L'opérateur Δ appliqué à une suite est :

$$(\Delta x)_n = x_{n+1} - x_n \tag{33.158}$$

Donc

$$(\Delta^2 x)_n = (\Delta x)_{n+1} - (\Delta x)_n = x_{n+2} - x_{n+1} - x_{n+1} + x_n = x_{n+2} - 2x_{n+1} + x_n. \tag{33.159}$$

L'accélération a alors la formule

$$y_n = \frac{(\Delta x)_n^2}{(\Delta^2 x)_n}. \tag{33.160}$$

Le problème est que ça accélère tellement que l'on arrive vite à des erreurs de cancellations, et donc à une précision en pics oscillants.

33.10 Équations algébrique

C'est une équation du type $P(x) = 0$ où P est un polynôme. Soit un polynôme de degré n. Nous en savons des choses.

(1) L'équation a exactement n solutions dans $\mathbb{C}$ en comptant les multiplicités.

(2) Les racines complexes arrivent par paire complexes conjuguée. Elles sont donc toujours en nombre pair.

Si donc nous avons $n = 3$, nous ne pouvons pas avoir 2 racine réelles. Il y en a donc 1 ou 3 réelles. Pas zéro ni deux.

Quelques méthodes : Müller, matrice compagnon, Laguerre.

33.10.1 Résoudre un système linéaire

Pour résoudre un système linéaire d'équations, nous échelonnons la matrice du système. Soit à résoudre le système $Ax = b$ où

$$A = \begin{pmatrix} 2 & 4 & -6 \\ 1 & 5 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \text{et} \quad b = \begin{pmatrix} -4 \\ 10 \\ 5 \end{pmatrix}. \tag{33.161}$$

En termes de problèmes, on écrit $F\big(x, (A, b)\big) = Ax - b$. La donnée de ce problème est le couple (A, b).

En ce qui concerne l'algorithme, on pose comme premier problème

$$F_1\big(x_1, (A_1, b_1)\big) = A_1 x_1 - b_1 = 0 \tag{33.162}$$

avec $A_1 = A$ et $b_1 = b$.

Ensuite, on commence à échelonner et le second problème est

$$F_2\big(x_2, (A_2, b_2)\big) = A_2 x_2 - b_2 = 0 \tag{33.163}$$

avec

$$A = \begin{pmatrix} 2 & 4 & -6 \\ 0 & 3 & 6 \\ 0 & 1 & 5 \end{pmatrix}, \quad \text{et} \quad b = \begin{pmatrix} -4 \\ 12 \\ 13 \end{pmatrix}. \tag{33.164}$$

Le troisième problème sera

$$F_3\big(x_3, (A_3, b_3)\big) = A_3 x_3 - b_3 = 0 \tag{33.165}$$

avec

$$A = \begin{pmatrix} 2 & 4 & -6 \\ 0 & 3 & 6 \\ 0 & 0 & 3 \end{pmatrix}, \quad \text{et} \quad b = \begin{pmatrix} -4 \\ 12 \\ 3 \end{pmatrix}. \tag{33.166}$$

Ce problème est facile à résoudre « à la main ». Nous nous arrêtons donc ici avec l'algorithme, et nous trouvons le x_3 qui résous le problème F_3.

33.10.2 Caractéristiques

L'algorithme de résolution de systèmes linéaires d'équations a les propriétés suivantes, à mettre en contraste avec celles de Newton :

(1) Pour résoudre le problème numéro n, il n'a pas fallu résoudre le problème numéro $n-1$.

(2) Toutes les solutions x_n des problèmes intermédiaires sont solutions du problème de départ. Nous avons $F_n(x, d_n) = 0$ pour tout n (ici, $d_n = (A_n, b_n)$).

(3) D'un problème à l'autre, les données changent énormément : la matrice échelonnée peut être très différente de la matrice de départ.

33.10.3 Définitions

Nous allons maintenant formaliser en donnant quelques définitions pour nommer les propriétés que nous avons vues. D'abord, un algorithme est une suite de problèmes. Un **algorithme** pour résoudre un problème $F(x, d) = 0$ est une suite de problèmes $\{F_n(x_n, d_n) = 0\}_{n\in\mathbb{N}}$.

Définition 33.66.
*Un tel algorithme est dit **fortement consistant** si pour toutes données admissibles d_n, on a*

$$F_n(x, d_n) = 0 \quad \forall\, n, \tag{33.167}$$

où x est la solution de $F(x, d) = 0$.

L'algorithme des matrices est fortement consistant, mais pas l'algorithme de Newton.

Définition 33.67.
*Un algorithme est **consistant** si $\lim_{n\to\infty} F_n(x, d_n) = 0$.*

Dans le cas de l'algorithme de Newton, c'est plutôt une telle consistance qu'on attend.

L'algorithme est dit **stable** si pour tout n le problème correspondant est stable. Dans ce cas, on note K^{num} le **conditionnement relatif asymptotique** défini par

$$K^{\text{num}} = \limsup_n K_n \tag{33.168}$$

où K_n est le conditionnement relatif du problème $F_n(x_n, d_n) = 0$.

Définition 33.68.
*Un algorithme est dit **convergeant** (en d) si pour tout $\epsilon > 0$, il existe $N = N(\epsilon)$ et $\delta = \delta(N, \epsilon)$ tels que pour $n \geqslant 0$ et $|d - d_n| < \delta$, on ait $|x(d) - x_n(d_n)| < \epsilon$.*

Remarque 33.69.
Dans le cas de l'algorithme de Newton, nous avons vu que la donnée d_n du problème F_n était en fait la même que la donnée initiale d, donc nous avons $d_n = d$, et par conséquent nous avons toujours $|d - d_n| < \delta$. Dans ce cas, la définition de la convergence revient à demander que la suite numérique des x_n converge vers la solution x.

Remarque 33.70.
Dans le cas des matrices par contre, les données sont très différentes les unes des autres, nous avons donc en général que $|d-d_n| > \delta$. Mais en revanche nous savons que tous les problèmes intermédiaires F_n acceptent une solution unique[11] $x_n(d_n) = x(d)$. Par conséquent, $|x_n(d_n) - x(d)|$ est toujours plus petit que ϵ. L'algorithme des matrices est donc toujours un algorithme convergeant.

33.11 Équations non linéaires

Certaines équations non linéaires sont résolubles explicitement, par exemples les polynômes de degré jusqu'à 4 ou des choses comme

$$\sin^2(x) + 3\sin(x) + 5 = 0. \tag{33.169}$$

Mais ces exemples sont très rares.

Nous allons étudier des équations du type $f(x) = 0$, dans $\mathbb{R}$.

(1) Un problème écrit sous la forme $x = g(x)$ peut utiliser des théorèmes de points fixes.
(2) Un problème sous la forme $f(x) = 0$ peut utiliser des méthodes de bisection, Newton ou autres.

Il y a évidemment beaucoup de façons de transformer un problème pour passer d'une forme à l'autre.

Exemple 33.71.
Soit $f(x) = x^2 - a = 0$ avec $a > 0$. Nous pouvons l'écrire

$$x^2 + x - a = x \tag{33.170}$$

qui donne une forme $g(x) = x$ pour $g(x) = x^2 + x - a$.

Ou encore $x = \frac{a}{x}$ et donc $g(x) = a/x$ (si par ailleurs on sait que $x \neq 0$). Notons que $x \neq 0$ n'est pas une hypothèse très forte parce qu'on la vérifie directement sur a. △

Exemple 33.72.
Soit l'équation à résoudre

$$f(x) = x^2 - 2 - \ln(x) = 0 \tag{33.171}$$

Les solutions de cette équation peuvent être vues comme les intersections avec l'axe X du graphe $y = x^2 - 2 - \ln(x)$. Tracer peut donc aider. Par ailleurs, il faut noter que

$$\lim_{x\to\pm\infty} f(x) = \infty, \tag{33.172}$$

donc les solutions sont certainement contenues dans un compact de $\mathbb{R}$.

À part tracer nous pouvons écrire

$$x^2 - 2 = \ln(x). \tag{33.173}$$

Et là, ce sont deux fonctions dont nous pouvons tracer le graphe pour trouver graphiquement les points d'intersection. Une étude de fonction montre vite qu'il y a exactement deux solutions, qu'elles sont strictement positives. Pour trouver des bornes, il faut calculer par exemple pour $x = 2$ les valeurs de $\ln(x)$ et $x^2 - 2$ pour voir si le graphe de $x^2 - 2$ est déjà plus haut. △

11. Nous n'envisageons que le cas où le déterminant est non nul.

La majorité des méthodes numériques de résolution d'équations du type $f(x) = 0$ ou $x = g(x)$ seront sous la forme de suites. Avec questions à la clef :

(1) Quel point de départ choisir ?

(2) Convergence ?

(3) Est-ce que la limite est bien une solution ?

(4) Vu que la limite est unique, comment faire si l'équation a plusieurs solutions ? (souvent c'est le choix du point initial qui va jouer sur ce point)

33.73.
Si la fonction est très plate, il est possible d'avoir

$$|f(\tilde{\alpha})| \leqslant \epsilon \tag{33.174}$$

sans que $\tilde{\alpha}$ ne soit une bonne approximation.

Lorsqu'on fait tourner une méthode itérative résolvant $f(x) = 0$, il n'est pas suffisant de s'arrêter lorsque

$$f(x_n) \leqslant \epsilon_1. \tag{33.175}$$

Il faut aussi s'assurer que, si $\bar{x}$ est la solution exacte, $|x_n - \bar{x}| \leqslant \epsilon_2$. Ici ϵ_1 et ϵ_2 sont deux « précisions » que nous nous fixons au départ.

Évidemment, vérifier la condition $|x_n - \bar{x}| \leqslant \epsilon_2$, il faudrait savoir $\bar{x}$. Et savoir $\bar{x}$ c'est justement le problème. Nous sommes donc amenés à faire des estimations de $|x_n - \bar{x}|$.

33.74.
Lorsque nous effectuons une méthode itérative, il faut donc contrôler deux grandeurs :

$$|\bar{x} - x_n| \leqslant \epsilon_1 \tag{33.176a}$$

$$|x_{n+1} - x_n| \leqslant \epsilon_2. \tag{33.176b}$$

Proposition 33.75.
Soit p l'ordre de convergence de la suite (x_n) vers $\bar{x}$. Si $p > 1$ et $|x_{n+1} - x_n| \leqslant \epsilon_2$ alors $|\bar{x} - x_n| \leqslant \epsilon_2$.

33.11.1 Méthode de bisection

Il y a ce théorème des valeurs intermédiaires.

Théorème 33.76.
Soit f continue sur $[a, b]$ telle que $f(a)f(b) < 0$. Alors il existe au moins une solution à l'équation $f(x) = 0$ sur l'intervalle $]a, b[$.

Pour démarrer une bisection, il est toujours bon de prendre l'intervalle $[a, b]$ de façon à ne contenir qu'une seule solution.

Soit donc un premier intervalle $[a_0, b_O]$ tel que $f(a_0)f(b_0) < 0$ et ne contenant qu'une seule solution. À chaque itération nous considérons la moitié de l'intervalle précédent, mais la moitié contenant la solution.

Le test d'arrêt de la méthode de bisection se base uniquement sur la taille de l'intervalle qui reste. En effet si nous avons

$$|b_n - a_n| \leqslant \epsilon \tag{33.177}$$

nous avons certainement

$$|\bar{x} - x_n| \leqslant \frac{\epsilon}{2} \tag{33.178}$$

où x_n est le point du milieu de $[a_n, b_n]$.

33.77.
La fonction f n'intervient dans la méthode que via son signe, pas via ses valeurs exactes.

33.78.
Notons que le théorème des valeurs intermédiaires n'est pas très puissant pour choisir l'intervalle de départ ; penser à la fonction

$$f(x) = x^2 - 5 \tag{33.179}$$

sur l'intervalle $[-10, 10]$. Il y a bien deux solutions dans l'intervalle, mais elles sont invisibles du théorème des valeurs intermédiaires. La fonction $x \mapsto x^2$ a sa solution en $x = 0$, mais elle aussi n'est pas visible.

33.79.
Certes la méthode de bisection assure la convergence vers une solution, mais elle n'assure pas la convergence monotone. Il peut arriver que $|\bar{x} - x_n| < |\bar{x} - x_{n+1}|$. C'est le cas lorsque la solution est très proche du milieu de l'intervalle choisit. Le x_0 est alors proche de $\bar{x}$ alors que x_1 sera à une distance de $\bar{x}$ d'environ un quart de l'intervalle de départ.

Supposons déjà avoir trouvé un intervalle $[a, b]$ dans lequel se trouve une unique solution à $f(x) = 0$. Voici un algorithme possible.

```
from __future__ import division

def bissection(f, a, b, toll, mmax):
    """
    f : une fonction
    a,b : les limites de l'intervalle
    toll : la tolérance. C'est l'amplitude de l'intervalle à ↩
       partir duquel nous nous arrêtons.
    nmax : le nombre maximum d'itérations.

    Nous supposons que \( b>a\).

    Retourne un tuple (x,n) où 'x' est la solution approchée et 'n↩
       ' est le nombre d'itérations effectuées
    """
    n = -1
    amp = toll + 1 # Pour s'assurer que l'on entre dans le cycle
    while amp > toll and n < nmax :
        n = n + 1
        amp=abs(b - a)
        x = a + amp / 2
        if f(a) * f(x) < 0:
            b = x
        elif f(a)*f(x) > 0:
            a = x
        else :       # Problème ZERO
            amp = 0
    return (x, n)
```

tex/frido/codeSnip_1.py

Plusieurs remarques :

(1) Le fait de retourner le nombre d'itérations effectuées permet à l'utilisateur de savoir la précision et si le nombre maximum d'itérations est dépassé. Si ce `n` retourné est égal à `nmax`, l'utilisateur sait que le `x` retourné n'est pas fiable.

(2) La ligne `from __future__ import division` fait en sorte que l'opération `/` est bien la division usuelle. Sinon, le défaut en python 2 est que `/` soit la division *entière*, c'est-à-dire

que $1/2 = 0$ en python 2. En python 3, le symbole `/` désigne bien la division usuelle, mais Sage utilise Python 2.

(3) Même si l'intervalle $[a, b]$ contient plus d'une solution, la méthode fonctionne et donne une solution. Il est simplement éventuellement très compliqué de savoir laquelle.

(4) Nous faisons `amp=toll+1` parce que nous voulons absolument lancer le cycle au moins une fois. Sinon, le `x` à retourner ne serait pas définit au moment de sortir du cycle (si le cycle n'est pas exécuté).

(5) Calculer le point milieu d'un intervalle $[a, b]$ est par la formule $(a + b)/2$ sauf que cette opération est numériquement dangereuse parce qu'à cause de l'arithmétique en précision finie, il est possible que cela tombe *exactement* sur a ou b. D'où le fait de calculer le point milieu par

$$x = a + \frac{amp}{2}. \tag{33.180}$$

(6) Dans le cas `Problème ZERO` nous déduisons $f(x) = 0$. Attention que c'est pas que $f(x) = 0$ mais simplement que en mettant x dans f, la *machine* retourne son zéro.

Il peut cependant avoir une fonction telle que $f(1) = 10^{-50}$ et $f(2) = 0$. L'algorithme de bisection risque de s'arrêter si $x_n = 1$. Parce que la machine risque de calculer $f(x_n) = 0$.

Quoi qu'il en soit, nous y mettons `amp=0` pour être sûr de sortir de la boucle dès la prochaine vérification.

(7) Il y a moyen de sauver les valeurs de $f(a)$ et $f(x)$ pour ne pas les recalculer, et en particulier au moment de faire `b=x` nous pouvons poser `fa=fx`.

Si τ est la précision de la solution voulue, nous pouvons fixer à priori le nombre d'itérations à faire grâce à la formule

$$n \geqslant \left\lceil \log_2 \left(\frac{b-a}{\tau}\right) \right\rceil. \tag{33.181}$$

Il y a un "$\geqslant$" et non une égalité parce qu'en arithmétique numérique, le nombre obtenu à droite pourrait ne pas être le bon à 1 près.

Ici pour $\nu \in \mathbb{R}$ le nombre $\lceil \nu \rceil$ est le plus petit entier à être plus grand ou égal à ν.

33.80.
Notons l'importance de la continuité de f. Par exemple que ferait la bisection sur la fonction $f(x) = 1/x$ pour l'intervalle $[-3, 1]$?

Il y a changement de signe sans avoir de racine.

Vu que 2^{10} est déjà 1024. Donc si on veut de la précision de l'ordre de 1/1000, dix itérations suffisent. Si donc nous avons besoin de 200 itérations pour atteindre la précision voulue, c'est l'occasion de trouver un intervalle plus petit. Par exemple en traçant la fonction, en faisant un zoom et en trouvant des valeurs de a et b qui sont déjà proches.

33.81.
Dans le monde réel, il arrive souvent d'utiliser une méthode de bisection pour se donner un point de départ pour une autre méthode.

33.12 Efficacité

Définition 33.82.
*L'**efficacité** est le nombre*

$$E = \sqrt[s]{p} \tag{33.182}$$

où p est l'ordre de convergence de la méthode et s est le nombre de fois qu'il faut calculer une valeur de la fonction à chaque itération (nous ne comptons pas l'initialisation).

Que le nombre de d'évaluations de f intervienne est logique parce que chaque évaluation provoque une erreur possible.

Exemple 33.83 (Bisection)**.**
Pour la méthode de bisection, nous avons $s = 1$ parce que chercher x_{n+1}, il faut seulement calculer $f(x_n)$. △

Exemple 33.84 (Newton)**.**
Pour l'algorithme de Newton nous avons $p = 2$ et il y a deux évaluations à chaque itération (une fois f et une fois f'), donc $s = 2$ et $E = \sqrt{2}$. △

33.13 Exemples sous forme d'exercices

Exemple 33.85.
Nous supposons une machine acceptant 5 chiffres significatifs. Elle retient les nombres sous la forme $\pm 0.x \cdots \times 10^{\cdots}$.

$x = 1.403$, $y = 0.4112 \times 10^{-3}$, $z = -0.4111 \times 10^{-3}$.

Soient $a = (x \oplus y) \oplus z$ et $b = (y \oplus z) \oplus x$.

(1) D'abord la calculer à la main.

(2) Quel est le calcul préférable ?

(3) Donner l'erreur relative avec 3 chiffres significatifs.

Dans le cas du calcul à la main, il faut en faire un seul parce que, algébriquement, $a = b$.

Nous avons $x = 1.403$, $y = 0.0004112$ et $z = -0.0004111$. Et la somme donne :

$$a = b = 1.4030001 = 0.14030001 \times 10^1. \tag{33.183}$$

Faisons d'abord la normalisation de x, c'est-à-dire $\mathrm{fl}(x)$.

$$\mathrm{fl}(x) = 0.1403 \times 10^1. \tag{33.184}$$

et y est déjà normalisé :

$$\mathrm{fl}(y) = 0.4112 \times 10^{-3}. \tag{33.185}$$

Il n'y a pas d'erreurs d'assignation pour ces deux nombres.

Pour faire la somme, il faudra déjà un peu casser les nombres pour les écrire de façon à pouvoir les sommer. En effet, il faut écrire les deux nombres avec le même exposant de 10 (le plus grand), pour pouvoir les mettre en colonne :

$$0.1404 \times 10^1 \to 0.1404 \times 10^1 \tag{33.186a}$$

$$0.4112 \times 10^{-3} \to 0.00004 \times 10^1. \tag{33.186b}$$

La somme donne 0.14034×10^1. Et ça, c'est à nouveau arrondi. Le premier chiffre supprimé est un 4, donc

$$x \oplus y = 0.1403 \times 10^1. \tag{33.187}$$

Et là on remarque que nous avons la même chose que x. C'est un classique du calcul numérique.

Nous avons aussi

$$\mathrm{fl}(z) = 0.4111 \times 10^{-3}. \tag{33.188}$$

Et pour faire la somme de cela avec $x \oplus y$ nous devons le remettre sous la forme d'un 10^1 :

$$\mathrm{fl}(z) \to -0.00004 \times 10^1 \tag{33.189}$$

(erreur de conversion), et en sommant on trouve

$$(x \oplus y) \oplus z = 0.140216 \times 10^1, \tag{33.190}$$

qui est encore arrondi. Le premier chiffre supprimé est un 6, donc

$$\mathrm{fl}(a) = 0.1403 \times 10^1, \tag{33.191}$$

Le nom de l'erreur qui consiste à avoir $x \oplus y = x$ est "relation annirmale".

Calculons b.

Les nombres y et z ont même ordre de grandeur, donc pas d'erreur au moment de les mettre sous forme sommable.

$$\mathrm{fl}(x) + \mathrm{fl}(y) = 0.00010 \times 10^{-3}. \tag{33.192a}$$

Cela est renormalisé et arrondi : $\mathrm{fl}(x) \oplus \mathrm{fl}(y) = 0.1000 \times 10^{-6}$.

Notons que nous avons ici commis potentiellement une erreur de cancellation parce que entre y et z, il y a 3 chiffres sur 4 qui sont identiques. Seul le chiffre 1 est significatif en réalité.

Il faut maintenant ajouter x à cela. D'abord

$$\mathrm{fl}(x) = 0.1403 \times 10^{1}. \tag{33.193}$$

Pour cette somme, il faudra remettre notre 0.1000×10^{-6} avec une puissance 10^1. Et là, nous obtenons zéro parce que vraiment ce nombre est trop petit pour être écrit avec 10^1. Résultat des courses :

$$\mathrm{fl}(b) = 0.1403 \times 10^{1}. \tag{33.194}$$

Dans le premier calcul nous avons deux "relations anormales" et dans le second nous en avons une plus une cancellation.

Nous préférons avoir deux relations anormales, parce que l'erreur de cancellation est plus grave : elle consiste à une perte de chiffre significatifs. Le fait est que faisant la différence à l'ordinateur nous avons obtenu 0.1 qui est certes exact, mais qui est un coup de bol : la différence aurait aussi bien pu être 0.19 avec d'autres nombres, machinement égaux.

Note : avec les données ici, il n'y a en fait pas d'erreur de cancellation. Mais il y a une erreur potentielle de cancellation, potentiellement grave.

En ce qui concerne l'erreur relative. Dans la formule

$$\epsilon_r = \frac{|a - a^*|}{|a|}, \tag{33.195}$$

la différence ne peut pas être calculée à la calculatrice justement parce qu'elle est très potentiellement sujette à erreur de cancellation.

$$\epsilon_r = \frac{|0.1030001 \times 10^1 - 0.1403 \times 10^1|}{0.14030001 \times 10^1} = \frac{0.1 \times 10^{-6}}{0.14030001 \times 10^1} \simeq 0.712758 \times 10^{-7}. \tag{33.196}$$

En passant à 3 chiffres significatifs, 0.713×10^{-7} (le premier chiffre supprimé est un 7).

△

Exemple 33.86.
Soient $x = 0.1 \times 10^{21}$ et $y = 0.5 \times 10^20$ et les expressions

(1) $z_1 = \frac{x-y}{y} + \frac{x+y}{x}$

(2) $z_2 = \frac{x^2+y^2}{xy}$.

Ces deux expressions sont algébriquement équivalentes.

(1) Calculer les valeurs.

(2) On suppose une machine en précision simple. Laquelle des deux expressions est préférable ?

Pour z_1, en arithmétique exacte :

$$z_1 = \frac{0.5 \times 10^{20}}{0.5 \times 10^{20}} + \frac{1.5 \times 10^{20}}{1 \times 10^{20}} = 0.25 \times 10^{1}. \tag{33.197}$$

Le calcul exact de z_2 donne la même chose.

Calcul de z_1 Les deux valeurs sont mémorisables et la différence $x - y$ se fait sans erreurs de cancellation. Idem pour la somme $x + y$. Idem pour les divisions.

Calcul de z_2 Pour faire x^2, c'est pas possible parce que c'est de l'ordre de 10^{40} alors que nous sommes en précision simple. Idem pour le produit xy.

Morale : z_2 donne un `overflow` alors que z_1 fonctionne de façon exacte.

Remarque 33.87.
En réalité le z_1 n'est pas tout à fait calculable de façon exacte sur la machine parce qu'elle doit d'abord convertir en binaire, ce qui n'est pas toujours possible. Mais sur notre machine qui fonctionne en base 10, il n'y a pas de problèmes.

△

Exemple 33.88.
Soit la fonction

$$f(x) = 2x^2 - 4x + 2 - e^{-x}. \tag{33.198}$$

(1) Identifier la plus grande des solutions réelles de $f(x) = 0$.

(2) Effectuer une bisection pour la savoir.

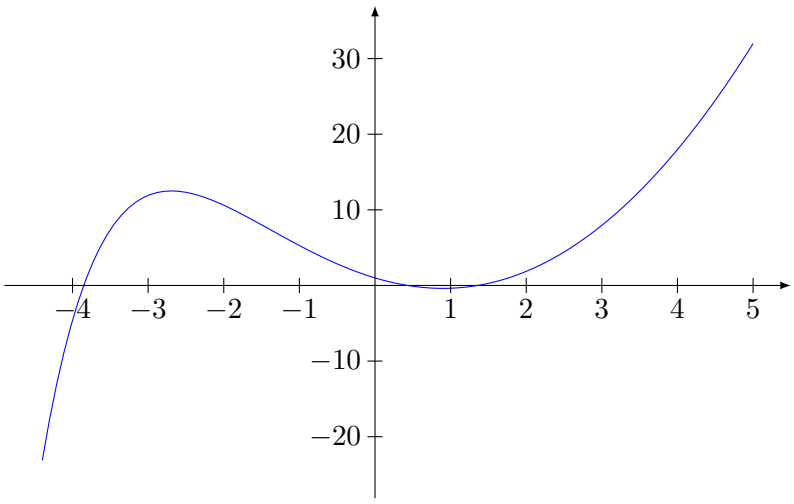
(3) Sachant que

$$\alpha \simeq 1.358500220734946, \tag{33.199}$$

quelle est l'erreur relative ?

Note : si c'est pour chercher à la main des approximations pour démarrer, il est évidemment préférable de dessiner $f_1(x) = 2x^2 - 4x + 2$ et $f_2(x) = -e^{-x}$ séparément.

Quoi qu'il en soit, voici un graphique :



Nous voyons trois racines : $\alpha_1 \in [0, 0.5,]$, $\alpha_2 \in [1, 1.5]$ et $\alpha_0 \in [-4, -3.5]$.

La plus grande solution est α_2. Nous pouvons déjà remplir le tableau des précisions :

n	a_n	b_n	x_n	$g(x_n)$	$\lvert b_n - a_n \rvert$
< ++ >	< ++ >	< ++ >	< ++ >	< ++ >	0.5
< ++ >	< ++ >	< ++ >	< ++ >	< ++ >	0.25
< ++ >	< ++ >	< ++ >	< ++ >	< ++ >	0.125

Et nous calculons les valeurs de f aux points d'extrémité de l'intervalle. Note que seul le signe nous importe :

$$f(1) \simeq -0.368 \tag{33.200a}$$

$$f(1.5) \simeq 0.278. \tag{33.200b}$$

Voici donc le tableau avec le signe de f indiqué :

n	a_n	b_n	x_n	$g(x_n)$	$\lvert b_n - a_n \rvert$
0	$1(+)$	$1.5(-)$	$< ++ >$	-0.162×10^0	0.5
1	$< ++ >$	$< ++ >$	$< ++ >$	$< ++ >$	0.25
2	$< ++ >$	$< ++ >$	$< ++ >$	$< ++ >$	0.125

Puis :

n	a_n	b_n	x_n	$g(x_n)$	$\lvert b_n - a_n \rvert$
0	$1(+)$	$1.5(-)$	$1.25(-)$	-0.162×10^0	0.5
1	$< ++ >$	$< ++ >$	$< ++ >$	$< ++ >$	0.25
2	$< ++ >$	$< ++ >$	$< ++ >$	$< ++ >$	0.125

Et enfin :

n	a_n	b_n	x_n	$g(x_n)$	$\lvert b_n - a_n \rvert$
0	$1(+)$	$1.5(-)$	$1.25(-)$	-0.162×10^0	0.5
1	$1.25(-)$	$1.5(+)$	$1.375(+)$	$+0.284 \times 10^{-1}$	0.25
2	$1.25(-)$	$1.375(+)$	$1.3125(-)$	$)0.738 \times 10^{-1}$	0.125

Note que les $f(x_n)$ restent toujours du même ordre de grandeur. Si un moment on voit un 1.6×10^7, c'est qu'une erreur a été commise.

En ce qui concerne le calcul de l'erreur relative, la première chose à faire est de vérifier que le α proposé est dans l'intervalle qui nous reste. Sinon c'est qu'une erreur a été commise.

De plus notre approximation est $x_n = 1.3125$, dont déjà deux chiffres sont corrects. En deux itérations de bisection en partant de 0.5, nous ne pouvons pas nous attendre à mieux.

△

33.14 Approximations de fonctions

(1) D'habitude on n'approxime pas une fonction sur tout son domaine, mais seulement sur une partie.

(2) Il y a le problème du choix de la classe des fonctions qui vont approximer. Nous allons travailler avec des polynômes.

(3) Il nous faut un critère disant si une approximation est bonne ou non.

33.14.1 Critère d'interpolation

À partir de $n+1$ abscisses points distinctes x_i, nous calculons $y_i = f(x_i)$. Il y a ce théorème qui dit qu'il existe un unique polynôme de degré (au plus) $n+1$ passant par les points (x_i, y_i).

Proposition-définition 33.89 (Base de Lagrange)**.**
Étant donnés $n+1$ valeurs distinctes x_i, l'espace des polynômes de degré n admet la base

$$L_i^{(n)}(x) = \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{x - x_j}{x_i - x_j} \tag{33.201}$$

pour $i = 0, \ldots, n$.

Soit par exemple les valeurs de f données dans

x_i	$y_i = f(x_i)$
$x_0 = 5$	1
$x_1 = -7$	-23
$x_2 = 6$	-54
$x_3 = 0$	-954

Les polynômes de Lagrange pour ces données dépendent seulement des x_i, pas des y_i. En particulier,

$$L_0^{(3)}(x) = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}, \tag{33.202}$$

etc.

La réponse est que

$$L_0^{(3)}(x) = \frac{x^3+13x^2+42x}{660} \tag{33.203a}$$

$$L_1^{(3)}(x) = \frac{-x^3-x^2+30x}{84} \tag{33.203b}$$

$$\textit{etc.} \tag{33.203c}$$

Ce qu'il y a de bien avec cette base est que en posant $a_i = f(x_i)$ alors le polynôme

$$\sum_{i=0}^{n} a_i L_i^{(n)}(x) \tag{33.204}$$

passe par les points $(x_i, f(x_i))$. Du coup il suffit d'écrire

$$P_3(x) = L_0^{(3)}(x) - 23L_1^{(n)}(x) - 54L_2^{(n)}(x) - 954L_3^{(n)}(x) = 4x^3 + 35x^2 - 84x - 954. \tag{33.205}$$

Un inconvénient de cette base est qu'elle est complètement dépendante des points choisis. Si on ajoute un point ou qu'on en prend un à peine différent, tous les coefficients changent. Mais en pratique, ajouter des points est quelque chose qui arrive souvent parce que souvent, après avoir vu le résultat d'un polynôme d'interpolation, on veut ajouter un point pour avoir un meilleur résultat.

33.90.
Une habitude : le premier et le dernier nœud se choisissent aux extrémités de l'intervalle sur lequel nous voulons une approximation.

Le but d'une approximation est d'avoir des approximations de $f(x^*)$ pour des valeurs de x^* qui ne soit pas une des abscisses données (parce que sur ces points, le polynôme et la fonction sont égaux). Nous considérons donc

$$f(x^*) \simeq P_n(x^*). \tag{33.206}$$

Si x^* est dans l'intervalle $I = [x_{min}, x_{max}]$ alors nous disons que nous calculons f par **interpolation**. Si au contraire x^* est en dehors de cet intervalle nous parlons d'**extrapolation**.

Si x^* est pris à l'extérieur de I, alors l'erreur risque d'être très grande, surtout parce que les polynômes tendent tous vers $\pm\infty$ lorsque $x \to \pm\infty$.

Autant l'interpolation via polynômes est le plus souvent valable, il faut garder à l'esprit que les extrapolations sont souvent mauvaises si x^* est trop loin des extrémités de I.

33.14.2 Base de Newton

Après la base canonique et la base de Lagrange, nous voyons la base de Newton. Soient encore $n+1$ points donnés du graphe de f.

Définition 33.91.
*La **base de Newton** pour les abscisses x_i est l'ensemble des polynômes suivants :*

$$1, (x-x_0), (x-x_0)(x-x_1), \dots, (x-x_0)(x-x_1)\dots(x-x_{n-1}) \tag{33.207}$$

Notons que ces polynômes n'utilisent pas le dernier point des x_i. Le polynôme passant par les points est

$$P_n(x) = x_0 + \sum_{i=1}^{n} c_i \prod_{j=0}^{n-1}(x-x_j). \tag{33.208}$$

Le calcul des c_i n'est pas absolument évident. Mais si nous ajoutons un point d'interpolation, les polynômes déjà calculés sont encore bons ; en particulier

$$P_{n+1}(x) = P_n(x) + c_{n+1}\prod_{j=0}^{n}(x - x_j). \tag{33.209}$$

Et cela est bien, parce que ça donne une façon de les calculer par récurrence.

Il y a plusieurs façons de calculer les c_i.

Les différences divisées sont des façons d'approximer les dérivées.

Définition 33.92.
*Soient $n+1$ nœuds x_i pour la fonction f. La **différence divisée** sont :*

Ordre 0

$$f[x_i] = f(x_i) \tag{33.210}$$

Ordre 1

$$f[x_i, x_j] = \frac{f[x_i] - f[x_j]}{x_i - x_j}. \tag{33.211}$$

Ordre 2

$$f[x_i, x_j, x_k] = \frac{f[x_i, x_j] - f[x_j, x_k]}{x_i - x_k}. \tag{33.212}$$

Ordre n

$$f[x_0, \dots, x_n] = \frac{f[x_0, \dots, x_{n-1}] - f[x_1, \dots, x_n]}{x_0 - x_n}. \tag{33.213}$$

Les ordres font référence à l'ordre de dérivation qui est approximé.

Nous avons alors

$$c_i = f[x_0, \dots, x_i]. \tag{33.214}$$

Cela donne effectivement une méthode de récurrence pour trouver les coefficients c_i.

Remarque 33.93.
Pour calculer c_0, il faut seulement calculer $f[x_0] = f(x_0)$. Mais pour calculer c_1 il faut $f[x_0]$ et $f[x_1]$. Et pour c_2 il faut $f[x_0, x_1, x_2]$ qui demande $f[x_0, x_1]$ et $f[x_0, x_1]$, qui demande etc.

Il faut donc calculer en réalité tous les $f[x_i]$ pour terminer le calcul. Par contre, pour ajouter un point, il ne faut pas tout recalculer, et même pas tout conserver en mémoire. Il faut seulement garder en mémoire la dernière diagonale.

Exemple 33.94.
Soit les nœuds

x	$f(x)$
3	1
1	-3
5	2
6	4

Trouver le polynôme d'interpolation via la base de Newton.

x_i	$f[x_i]$	$f[x_i, x_j]$	$f[x_i, x_j, x_k]$
3	1	$\frac{1+3}{2} = 2$	$\frac{f[x_0,x_1]-f[x_1,x_2]}{x_0-x_2} = \frac{2-\frac{5}{4}}{-2} = -\frac{3}{8}$
1	-3	$\frac{-3-2}{-4} = \frac{5}{4}$	$\frac{\frac{5}{4}-2}{-5} = \frac{3}{20}$
5	2	$\frac{2-4}{-1} = 2$	$\frac{f[x_0,x_1,x_2]-f[x_1,x_2,x_3]}{x_0-x_3} = \frac{4}{40}$
6	4	$< ++ >$	$< ++ >$

Le polynôme d'interpolation sera

$$P_3(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)(x - x_1) + c_3(x - x_0)(x - x_1)(x - x_2). \tag{33.215}$$

△

<++>

Un exercice typique serait de donner tout pour 3 points puis de demander le polynôme qui aurait un quatrième point.

33.14.3 Méthode des minimums quadratiques

Soient $m+1$ points connus sur le graphe de la fonction f que nous devons approximer. Au lieu d'exiger que notre approximation ne passe par tous les points, nous allons chercher une approximation qui minimise la somme des carrés des erreurs sur ces points.

Soit $\mathcal{F}$ une classe de fonctions dans laquelle nous allons chercher l'approximation. Nous cherchons $g \in \mathcal{F}$ qui minimise

$$E(g) = \sum_{i=0}^{m} \big(f(x_i) - g(x_i)\big)^2 \omega_i \tag{33.216}$$

où $\omega_i > 0$ est une pondération. Souvent on prend $\omega_i = 1$, mais pas toujours. La fonction E sur $\mathcal{F}$ est la **fonction d'erreur**.

33.95.
À part dans les exercices à la main, le nombre de points est grand, du type du milliard. Il est bien entendu pas envisageable de faire passer un polynôme *exactement* par un milliard de points, parce que cela demanderait un polynôme de degré un milliard.

Plus généralement, d'un point de vue scientifique, avoir n paramètres libres pour n données expérimentales, ça ne passe pas Popper.

Afin de faire de la science qui passe Popper nous nous restreignons à une classe de fonction $\mathcal{F}$ dont la dimension n'est pas grande : $\dim(\mathcal{F}) \ll m$. Et nous notons $\dim(\mathcal{F}) = n+1$.

Exemple 33.96.
La qualité d'une expérience peut être influencée par des paramètres extérieurs comme l'humidité, le vent, etc. Donc il est normal d'avoir des expériences moins précises que d'autres. On le pèse moins. △

Exemple 33.97.
Dans un questionnaire, il se met des questions volontairement contradictoires. Si quelqu'un répond « oui » aux deux questions, il y a une indication que la personne a répondu un peu n'importe comment, et il faut moins peser ses réponses. △

Soit $g \in \mathcal{F}$, et une base $\{g_i\}_{i=0,\dots,n}$ de $\mathcal{F}$. Nous écrivons

$$g = a_0 g_0 + a_1 g_1 + \cdots + a_n g_n \tag{33.217}$$

La fonction donnée E donnée en (33.216), est, à partir du moment où $\mathcal{F}$ et une base sont choisis, une fonction des paramètres a_i que nous nommons $F(a_0, \dots, a_i)$. Il faut minimiser F, c'est-à-dire poser

$$\frac{\partial F}{\partial a_j} = 0 \tag{33.218}$$

pour $j = 0, \dots, n$. Cela sont $n+1$ équations pour $n+1$ inconnues. Notons que ces équations sont linéaires parce que chacun des termes est du type

$$\left(f(x_i) - \sum_{j=0}^{m} a_j g_j(x_i)\right)^2, \tag{33.219}$$

et lors de la dérivation par rapport à a_j, nous obtenons du degré 1.

33.14.4 Notre espace de Hilbert

Nous allons maintenant formaliser un peu tout cela. Dans [511] il est expliqué que si ω est une fonction strictement positive, alors l'espace $L^2_\omega\big([a,b]\big)$ dérivé de la norme

$$\|f\|^2_{L^2_\omega} = \int_a^b |f(x)|^2\omega(x)dx \tag{33.220}$$

est un espace de Hilbert. Nous allons tenter le coup avec $\omega = \sum_{i=0}^m \omega_i\delta_{x_i}$ où δ_a est la distribution de Dirac [12] centrée en a.

Sur l'ensemble des fonctions $\mathbb{R} \to \mathbb{R}$ nous considérons la relation d'équivalence $f \sim g$ si $f(x_i) = g(x_i)$ pour tout $i = 0, \dots, m$. Nous notons L^2_ω cet ensemble.

Proposition 33.98.
La formule

$$\langle f, g\rangle = \sum_{i=0}^m f(x_i)g(x_i)\omega_i \tag{33.221}$$

définit un produit scalaire sur L^2_ω. Ce dernier devient un espace de Hilbert.

Démonstration. Pour être un produit scalaire (définition 7.186), la forme considérée doit être symétrique et strictement définie positive. La symétrie de la formule (33.221) ne fait pas de doute. Le fait que ce soit semi-défini positif non plus. Pour le strict,

$$\langle f, f\rangle = \sum_{i=0}^m |f(x_i)|^2\omega_i. \tag{33.222}$$

Étant donné que $\omega_i > 0$ pour tout i, l'annulation de $\langle f, f\rangle$ implique l'annulation de $f(x_i)$ pour tout i. Cela signifie que f est dans la classe de 0 et donc est nul dans L^2_ω.

En ce qui concerne la complétude, la proposition 7.298 répond à notre place, étant donné que L^2_ω est de dimension finie. Une base est donnée par exemple par $e_i(x) = \delta_{x,x_i}$. Ici le δ est celui de Kronecker, et non celui de Dirac. □

Lemme 33.99.
Si la classe de fonctions $\mathcal{F}$ est un sous-espace vectoriel de L^2_ω et si $f \in L^2_\omega$ il existe un unique élément g de $\mathcal{F}$ minimisant la distance à f.

Démonstration. Le théorème de projection (au choix 11.232 ou 24.5) nous assure l'existence et l'unicité d'un élément de $\mathcal{F}$ minimisant la distance à $f \in L^2_\omega$. □

33.100.
Ce lemme est gentil, mais ne nous donne pas de méthodes pour trouver ce minimum. Nous allons donc écrire explicitement un système d'équations permettant de le trouver. Si $\{g_\alpha\}$ est une base (finie) de $\mathcal{F}$ alors nous cherchons le minimisant sous la forme $f = \sum_\alpha a_\alpha g_\alpha$.

Nous devons minimiser

$$E(g) = \sum_{i=0}^m \big(f(x_i) - g(x_i)\big)^2\omega_i = \sum_{i=0}^m \big(f(x_i) - \sum_\alpha a_\alpha g_\alpha(x_i)\big)^2\omega_i. \tag{33.223}$$

Vu que cela est maintenant plutôt une fonction des coefficients a_α que de la fonction g nous la notons $F(a_0, \dots, a_n)$. Il s'agit d'étudier le système d'équations

$$\frac{\partial F}{\partial a_\alpha} = 0. \tag{33.224}$$

Un tout petit peu de calcul mène au système

$$\sum_i\sum_\beta a_\beta\omega_i g_\alpha(x_i)g_\beta(x_i) = \sum_i \omega_i g_\alpha(x_i)f(x_i). \tag{33.225}$$

12. Définition 29.34.

À droite nous reconnaissons $\langle f, g_\alpha \rangle$. et à gauche, $\sum_\beta a_\beta \langle g_\alpha, g_\beta \rangle$. Donc le système s'écrit

$$\sum_\beta a_\beta \langle g_\alpha, g_\beta \rangle = \langle f, g_\alpha \rangle. \tag{33.226}$$

Il y a une équation pour chaque valeur de α.

La matrice $A \in \mathbb{M}(n+1), \mathbb{R}$ donnée par $\langle g_\alpha, g_\beta \rangle$ étant strictement définie positive (c'est un produit scalaire), le système a une unique solution. Et comme cette matrice est de plus symétrique, elle est diagonalisable par le théorème spectral 9.150. Toutes ses valeurs propres sont strictement positives.

Notons pour la curiosité que si l'on considère la matrice $B \in \mathbb{M}(m \times n)$ donnée par

$$B_{ij} = \sqrt{\omega_i} g_j(x_l), \tag{33.227}$$

alors nous avons $A = B^t B$.

33.14.5 Droite de régression

La droite de régression est la cas particulier $n = 1$, c'est-à-dire un système 2×2. Nous cherchons $P = a_0 + a_1 x$. Et la base choisie est $g_0(x) = 1$, $g_1(x) = x$. Nous avons

$$\langle g_0, g_0 \rangle = \sum_i \omega_i g_0(x_i) g_0(x_i) = \sum_i \omega_i \tag{33.228a}$$

$$\langle g_0, g_1 \rangle = \sum_i \omega_i x_i \tag{33.228b}$$

$$\langle g_1, g_1 \rangle = \sum_i \omega_i x_i^2. \tag{33.228c}$$

Donc pour approximer une fonction f il faut résoudre le système

$$\begin{pmatrix} \sum_i \omega_i & \sum_i \omega_i x_i \\ \sum_i \omega_i x_i & \sum_i \omega_i x_i^2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} \langle f, g_0 \rangle \\ \langle f, g_1 \rangle \end{pmatrix}. \tag{33.229}$$

Pour calculer les produits $\langle f, g_\alpha \rangle$ il suffit de savoir f sur les points x_i. Et encore heureux, parce que toute la méthode est basée sur le fait que nous ne connaissons pas f ailleurs. C'est pour cela que nous avons défini L^2_ω comme un ensemble quotient.

Exemple 33.101.
Faisons la droite de régression pour les données avec tous les poids $\omega_i = 1$.

x_i	$f(x_i)$
-5	18
-3	7
1	0
3	7
4	16
6	50
8	67

Nous avons

$$\langle g, g_0 \rangle = \sum_i f(x_i) = 165 \tag{33.230a}$$

$$\langle f, f_1 \rangle = \sum_i f(x_i) x_i = 810. \tag{33.230b}$$

et donc le système

$$\begin{pmatrix} 7 & 14 \\ 14 & 160 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 165 \\ 810 \end{pmatrix} \tag{33.231}$$

dont résolution donne la droite de régression. △

Proposition 33.102.
Si tous les poids sont identiques, alors la droite de régression passe par le barycentre des points donnés :

$$\begin{cases} x_M = \dfrac{1}{m+1}\displaystyle\sum_{i=0}^{m} x_i & (33.232a) \\ y_M = \dfrac{1}{m+1}\displaystyle\sum_{i=0}^{m} y_i. & (33.232b) \end{cases}$$

Cela donne une vérification possible de la réponse trouvée.

Définition 33.103.
*L'**erreur quadratique** est la fonction $F(a_0, \dots, a_n)$ dont il est question plus haut. Et si une solution est connue, son erreur quadratique est la valeur de F pour cette solution.*

33.15 Conditionnement d'une matrice

Soit le système d'équations linéaires $Au = b$ avec la matrice inversible A ainsi que le système perturbé $(A + \Delta A)u' = (b + \Delta b)$. Nous notons $\Delta u = u' - u$ et nous voudrions pouvoir dire des choses de l'erreur relative $\frac{\|\Delta u\|}{\|u\|}$.

Exemple 33.104 ([186])**.**
Soit la matrice

$$A = \begin{pmatrix} 10 & 7 \\ 7 & 5 \end{pmatrix} \tag{33.233}$$

et $b = \begin{pmatrix} 32 \\ 23 \end{pmatrix}$. La solution de $Au = b$ est $u = \begin{pmatrix} -1 \\ 6 \end{pmatrix}$. Si nous conservons la même matrice mais nous considérons $b = \begin{pmatrix} 32.1 \\ 22.9 \end{pmatrix}$. La solution devient $u' = \begin{pmatrix} 0.2 \\ 4.3 \end{pmatrix}$

En norme $\|.\|_\infty$ nous avons[13]

$$\frac{\|\Delta b\|}{\|b\|} = \frac{0.1}{32} = 0.003125 \tag{33.234}$$

et

$$\frac{\|\Delta u\|}{\|u\|} = \frac{1.7}{6} = 0.28. \tag{33.235}$$

Cela montre environ amplification d'un facteur 100 entre l'erreur sur b et l'erreur sur la solution. △

Définition 33.105.
*Le **conditionnement** de la matrice inversible $A \in \mathrm{GL}(n, \mathbb{C})$ est le nombre positif*

$$\mathrm{Cond}(A) = \|A\|\|A^{-1}\|. \tag{33.236}$$

Cette dénomination sera justifié par le corolaire 33.110 parce qu'il est évident que le conditionnement d'une matrice est lié au conditionnement du problème de résolution d'un système linéaire.

Remarque 33.106.
Le conditionnement dépend de la norme choisie, mais cette dependence est contrôlée par la proposition 7.223 qui nous indique que si le conditionnement d'une matrice est grand dans une norme, il sera grand dans une autre norme.

13. La proposition 7.223(3) montre que si nous voulions des estimations en norme $\|.\|_2$, il y aurait au maximum un facteur $\sqrt{2}$ par-ci par là.

D'autre part, lorsque nous écrirons $\|A\|$ nous supposerons toujours que $\|.\|$ est une norme d'algèbre[14] et donc que nous avons toujours

$$\|AB\| \leqslant \|A\|\|B\|. \tag{33.237}$$

De plus nous supposerons toujours avoir une norme subordonnée à une norme sur l'espace $\mathbb{C}^n$, de telle sorte à avoir

$$\|Au\| \leqslant \|A\|\|u\| \tag{33.238}$$

pour tout $u \in \mathbb{C}^n$. Voir aussi le lemme 10.10.

Proposition 33.107 ([186]).
Si A est une matrice inversible et si $\alpha \in \mathbb{C}$ nous avons :

(1) $\mathrm{Cond}(A) \geqslant 1$
(2) $\mathrm{Cond}(A) = \mathrm{Cond}(A^{-1})$
(3) $\mathrm{Cond}(\alpha A) = \mathrm{Cond}(A)$.

Si $Q \in \mathrm{O}(n)$ alors

(1) Nous avons $\mathrm{Cond}_2(Q) = 1$ où Cond_2 est le conditionnement pour la norme $\|.\|_2$.
(2) Nous avons aussi

$$\mathrm{Cond}_2(A) = \mathrm{Cond}_2(AQ) = \mathrm{Cond}_2(QA). \tag{33.239}$$

Démonstration. Nous savons que $\mathrm{Cond}(\mathbb{1}) = 1$ et donc

$$1 = \|\mathbb{1}\| \leqslant \|A\|\|A^{-1}\| \tag{33.240}$$

parce que la norme utilisée est une norme matricielle.

Les deux autres formules sont évidentes à partit du fait que la définition du conditionnement de A est symétrique entre A et A^{-1}.

En ce qui concerne les formules relatives à la matrice orthogonale Q nous savons par la proposition 9.42(3) qu'une matrice orthogonale est une bijection de l'ensemble $\{x \in \mathbb{R}^n \text{ tel que } \|x\| = 1\}$. Par conséquent

$$\|AQ\| = \sup_{x \text{ tel que } \|x\|=1} \|AQx\| = \sup_{Q^{-1}x \text{ tel que } \|x\|=1} \|AQQ^{-1}x\| = \|A\|. \tag{33.241}$$

Donc $\|AQ\| = \|A\|$. Les assertions s'ensuivent immédiatement en remarquant que Q^{-1} est également orthogonale. □

Soit une matrice inversible $A \in \mathrm{GL}(n, \mathbb{C})$. La matrice A^*A est hermitienne[15] et le théorème 9.144 nous assure que ses valeurs propres sont réelles. Par la remarque 9.145, ses valeurs propres sont même positives.

Proposition 33.108 ([186]).
Soit une matrice inversible $A \in \mathrm{GL}(n, \mathbb{C})$, et $\mu_1 \leqslant \ldots \leqslant \mu_n$ les valeurs propres de A^*A. Alors nous avons la formule

$$\mathrm{Cond}_2(A) = \sqrt{\frac{\mu_n}{\mu_1}}. \tag{33.242}$$

Démonstration. Par le théorème 10.21, la norme de A est liée au au rayon spectral de A^*A par

$$\|A\|_2 = \sqrt{\rho(A^*A)} = \sqrt{\mu_n}. \tag{33.243}$$

Vu que le spectre de AA^* est le même que celui de A^*A (lemme 9.148) nous avons aussi

$$\|A^{-1}\|_2 = \sqrt{\rho\big((A^{-1})^*A^{-1}\big)} = \sqrt{\rho\big((A^*A)^{-1}\big)} = \frac{1}{\sqrt{\mu_1}} \tag{33.244}$$

14. Définition 10.7.
15. Définition 9.35.

parce que la plus grande valeur propre de $(A^*A)^{-1}$ est l'inverse de la plus petite de A^*A.

Ces deux calculs étant,

$$\mathrm{Cond}_2(A) = \|A\|_2\|A^{-1}\|_2 = \sqrt{\frac{\mu_n}{\mu_1}}. \tag{33.245}$$

□

Problèmes et choses à faire

À mon avis ce qui est dans la proposition 16.109 est le conditionnement de la matrice ou sa racine carrée ou un truc du genre. Il faut voir le lien entre les valeurs propres de A et celles de AA^*.

33.15.1 Perturbation du vecteur

Proposition 33.109 (Système linéaire : perturbation du vecteur[186])**.**
Soit une matrice inversible A et les systèmes d'équations linéaires

$$Au = b \tag{33.246a}$$
$$Au' = b'. \tag{33.246b}$$

En notant $\Delta u = u' - u$ et $\Delta b = b' - b$ nous avons

$$\frac{\|\Delta u\|}{\|u\|} \leqslant \mathrm{Cond}(A)\frac{\|\Delta b\|}{\|b\|}. \tag{33.247}$$

Démonstration. En soustrayant les équations (33.246) nous avons $\Delta b = A\Delta u$, et donc $\Delta u = A^{-1}\Delta b$. D'une part nous avons alors

$$\|\Delta u\| \leqslant \|A^{-1}\|\|\Delta b\|. \tag{33.248}$$

Et d'autre part, $\|b\| \leqslant \|A\|\|u\|$, ce qui donne

$$\frac{\|b\|}{\|A\|} \leqslant \|u\|. \tag{33.249}$$

En mettant les deux ensemble,

$$\frac{\|\Delta u\|}{\|u\|} \leqslant \frac{\|A^{-1}\|\|\Delta b\|}{\|b\|}\|A\| = \mathrm{Cond}(A)\frac{\|\Delta b\|}{\|b\|}. \tag{33.250}$$

□

Le corolaire suivant justifie le nom « conditionnement » au conditionnement d'une matrice.

Corolaire 33.110.
Soit $A \in \mathrm{GL}(n, \mathbb{C})$ fixée et le problème de résoudre $Au = b$, c'est-à-dire la fonction

$$F(u, b) = Au - b. \tag{33.251}$$

(1) Ce problème est stable pour toute valeur de b.

(2) Nous avons une majoration pour le conditionnement relatif[16] *:*

$$K_{rel}(\eta, b_0) \leqslant \mathrm{Cond}(A). \tag{33.252}$$

Démonstration. **Stabilité** Vu que A est inversible, il existe une solution unique à tout système de la forme $Au = b'$. De plus $u(b) = A^{-1}b$, donc

$$\|u(b) - u(b_0)\| = \|A^{-1}(b - b_0)\| \leqslant \|A^{-1}\|\|b - b_0\|, \tag{33.253}$$

de telle sorte que la condition 33.25(2) fonctionne avec $K = \|A^{-1}\|$.

16. Si vous doutez de la norme à prendre, lisez la remarque 33.111

Conditionnement En partant de la définition 33.49, et en utilisant la majoration de la proposition 33.109 sous la forme

$$\|u(b)-u(b_0)\| \leqslant \mathrm{Cond}(A)\|u(b_0)\|\frac{\|\Delta b\|}{\|b_0\|}, \tag{33.254}$$

nous obtenons :

$$\begin{aligned}
K_{rel}(b_0,\eta) &= K_{abs}(b_0,\eta)\frac{\|b_0\|}{\|u(b_0)\|} && (33.255a)\\
&= \sup_{\|b-b_0\|\leqslant\eta}\frac{\|u(b)-u(b_0)\|}{\|b-b_0\|}\frac{\|b_0\|}{\|u(b_0)\|} && (33.255b)\\
&\leqslant \sup_b \mathrm{Cond}(A)\frac{\|u(b_0)\|}{\|b_0\|}\|\Delta b\|\frac{1}{\|b-b_0\|}\frac{\|b_0\|}{\|u(b_0)\|} && (33.255c)\\
&= \mathrm{Cond}(A). && (33.255d)
\end{aligned}$$

□

Remarque 33.111.
La notion de conditionnement relatif dépend aussi de la norme choisie. Dans la formule (33.252) il faut prendre le conditionnement $\mathrm{Cond}(A)$ pour la norme dans laquelle le K_{rel} est écrit. Encore une fois, toutes les normes étant équivalentes, cette majoration est à constante près bonne pour toutes les normes. Si la dimension est très grande, cette constante peut par contre être grande.

33.15.2 Perturbation de la matrice

Proposition 33.112 (Système linéaire : perturbation de la matrice[186])**.**
Soient les systèmes linéaires

$$Au = b \tag{33.256a}$$
$$A'u' = b \tag{33.256b}$$

avec A et A' inversibles. Nous notons $\Delta A = A' - A$. Alors

(1)

$$\frac{\|\Delta u\|}{\|u'\|} \leqslant \mathrm{Cond}(A)\frac{\|\Delta A\|}{\|A\|} \tag{33.257}$$

(2)

$$\frac{\|\Delta u\|}{\|u\|} \leqslant \mathrm{Cond}(A)\frac{\|\Delta A\|}{\|A\|}\big(1+\alpha(\|\Delta A\|)\big) \tag{33.258}$$

où $\lim_{x\to 0}\alpha(x) = 0$.

Démonstration. D'abord nous avons

$$\begin{aligned}
0 &= Au' - Au && (33.259a)\\
&= (A'-A)u' - Au' - Au && (33.259b)\\
&= \Delta A u' + A\Delta u. && (33.259c)
\end{aligned}$$

Par conséquent, $\Delta u = -A^{-1}(\Delta A)u'$ et

$$\|\Delta u\| \leqslant \|A^{-1}\|\|\Delta A\|\|u'\|. \tag{33.260}$$

Donc

$$\frac{\|\Delta u\|}{\|u'\|} \leqslant \|A^{-1}\|\|A\|\frac{\|\Delta A\|}{\|A\|} = \mathrm{Cond}(A)\frac{\|\Delta A\|}{\|A\|}. \tag{33.261}$$

Cela est (1).

Pour l'autre inégalité, nous avons $A' = A + \Delta A$ et donc

$$\|A'^{-1}\| = \|(A + \Delta A)^{-1}\| \tag{33.262}$$

Nous repartons alors de (33.260) en changeant le rôle de A et A' (et donc aussi de u et u'). Ce changement étant, $\|\Delta u\|$ et $\|\Delta A\|$ ne changent pas. Nous avons :

$$\frac{\|\Delta u\|}{\|u\|} \leqslant \|A'^{-1}\|\|\Delta A\| \tag{33.263a}$$

$$= \|(A + \Delta A)^{-1}\|\|\Delta A\| \frac{\mathrm{Cond}(A)}{\|A\|\|A^{-1}\|} \tag{33.263b}$$

$$= \frac{\|(A + \Delta A)^{-1}\|}{\|A^{-1}\|} \frac{\|\Delta A\|}{\|A\|} \mathrm{Cond}(A). \tag{33.263c}$$

Il reste à voir que

$$\lim_{\|\Delta A\| \to 0} \frac{\|(A + \Delta A)^{-1}\|}{\|A^{-1}\|} = 1, \tag{33.264}$$

ou autrement dit que

$$\lim_{A \to A'} \frac{\|A'^{-1}\|}{\|A^{-1}\|} = 1 \tag{33.265}$$

où la limite est celle dans $\mathrm{GL}(n, \mathbb{C})$. Par définition de la topologie, la norme est continue (quelle qu'elle soit par l'équivalence de norme 7.224). Par le théorème 10.153, l'application $A \mapsto A^{-1}$ est également continue et commute donc avec la limite. Nous avons donc

$$\lim_{A' \to A} \|A'^{-1}\| = \|(\lim_{A' \to A} A')^{-1}\| = \|A^{-1}\|. \tag{33.266}$$

Donc la limite du quotient (33.265) est bien 1. □

33.16 Système linéaires (généralités)

Soit un système d'équations linéaires $Ax = b$ avec $A \in \mathbb{M}(n, \mathbb{R})$. Le problème est évidemment de savoir s'il existe une unique solution x et de la déterminer. Nous supposons l'existence et l'unicité. C'est-à-dire que les conditions équivalentes [17] sont vérifiées :

(1) A est inversible, c'est-à-dire qu'il existe une matrice notée A^{-1} telle que $AA^{-1} = A^{-1}A = \mathbb{1}$.

(2) $\det(A) \neq 0$.

Note : si nous avons un système pas carré du type $Bx = v$ avec $B \in \mathbb{M}(n \times m)$ alors nous pouvons nous ramener à un système carré en écrivant

$$B^t Bx = B^t v. \tag{33.267}$$

Mais attention : bien que $B^t B$ soit symétrique et semi-définie positive, certaines valeurs propres peuvent être nulles.

33.113.
Deux choses générales en calcul numérique :

(1) On ne calcule pas l'inverse d'une matrice.

(2) On ne calcule même pas son déterminant.

Par conséquent nous ne faisons pas $x = A^{-1}v$.

Il faut garder en tête le fait que dans la pratique, la matrice A possède des millions de lignes et colonnes, si pas pire. Pour une matrice de taille de l'ordre du million, il y a 1000 milliards d'entrées. Si on compte 32 bits par nombre (précision simple, définition 33.8), c'est-à-dire 4 octets, il faut

17. L'équivalence est la proposition 9.10(2).

4000 giga-octets pour enregistrer la matrice. Même pour la mémoire actuellement disponible, ce n'est pas rien. Surtout que souvent, la précision simple n'est pas utilisée, mais la précision double, ce qui donne 8000 giga pour enregistrer la matrice.

Heureusement, dans la majorité des cas pratiques, les matrices géantes qui apparaissent sont pleines de zéros.

Définition 33.114.
Une matrice est ***creuse*** *si elle possède beaucoup de zéros. Une matrice non creuse est dite* ***dense****.*

Notons que lorsqu'on parle de matrice comprenant beaucoup de « zéros », nous pensons à des éléments très petits, et non de vrai zéros.

Les matrices creuses ne sont pas mémorisées entièrement, mais plutôt comme un dictionnaire (i, j, v) qui donne la valeur v de A_{ij}.

Définition 33.115.
Une matrice est de « grande dimension » si elle ne peut pas être mise en mémoire sur un ordinateur donné. Sur certains ordinateurs, ça commence à 5000 *inconnues. Mais sur des plus forts, on peut aller jusqu'au million ou le milliard.*

Si la matrice est de petite dimension, il est possible d'utiliser des méthodes dites « directes ». Sinon, il faudra utiliser des méthodes itératives.

33.16.1 Les méthodes directes

Une méthode directe consiste à successivement transformer un système $A^{(0)}x = b^{(0)}$ en de nouveaux systèmes $A^{(i)}x = b^{(i)}$ dont la solution est identique jusqu'à obtenir un système $A^{(n-1)}x = b^{(n-1)}$ qui est à résolution immédiate.

L'avantage d'une méthode directe est qu'elle fournit une réponse exacte, pour autant que les calculs intermédiaires soient bien faits (ce qui n'est pas le cas sur un ordinateur).

Une méthode directe fonctionne en général avec un nombre de pas fixés par la taille du système. Par exemple pour un système $n \times n$, la méthode de Gauss demande exactement n pas, et il n'y a pas moyen de faire mieux. Or chaque pas demande de recalculer tous les éléments de la matrice. Encore une fois, si la matrice a une taille de l'ordre du milliard, cela fait 10^{18} éléments à recalculer un milliard de fois (sans compter les éléments du vecteur b). Infaisable.

Souvent une méthode directe passe par une factorisation $A = BC$ avec $B, C \in \mathbb{M}(n \times n)$.

Quelques types de matrices dont la résolution est immédiate :

— Matrice diagonale.

— Matrice orthogonale parce que si A est orthogonale alors $Ax = v$ se résout par $x = A^t v$ qui n'est pas particulièrement lourd à faire numériquement.

— Matrice triangulaire.

Remarque 33.116.
Pour une matrice diagonale, le déterminant et l'inverse sont faciles. Mais également pour la triangulaire. Pour une matrice triangulaire, le déterminant est le produit des éléments diagonaux, et il se fait qu'il y a une algorithme facile pour calculer l'inverse.

Donc en fait les matrices à résolution immédiates sont des matrices pour lesquelles l'inverse et le déterminant sont facile à calculer.

33.16.2 Méthodes itératives

Si la matrice est trop grande, il n'est pas possible de faire des manipulations de matrices à chaque itération.

En général, les méthodes itératives ne convergent pas toujours. Mais lorsqu'une méthode converge, c'est une propriété de la matrice, et donc la convergence aura lieu pour tout vecteur de départ x_0. Cela est très différent du cas des équations non linéaires type Newton pour lesquelles la convergence peut fortement dépendre du point de départ.

33.17 Système linéaires (méthodes directes)

Les matrices que nous sommes autorisés à inverser sont les matrices

— orthogonales : l'inverse est la transposée

— diagonales : l'inverse est diagonale avec les inverses sur la diagonale

— triangulaires : nous en parlons maintenant.

33.17.1 Inversion de matrice triangulaire

Si T est une matrice triangulaire (mettons supérieure pour fixer les idées), il est possible d'en calculer l'inverse sans trop d'efforts. Notons B la matrice inverse que nous allons construire ligne par ligne. Vu que $BT = \mathbb{1}$ nous avons

$$\delta_{1j} = \sum_{k=1}^{n} B_{1k}T_{kj} = \sum_{k=1}^{j} B_{1k}T_{kl} \tag{33.268}$$

parce que $T_{kj} = 0$ pour $k > j$. Donc nous pouvons calculer les éléments B_{1j} un par un parce que chacun ne dépend que des précédents. Le même procédé fonctionne pour les autres lignes :

$$\delta_{ij} = \sum_{k=1}^{j} B_{ij}T_{kj}. \tag{33.269}$$

Et tu notes que le calcul peut être parallélisé : le calcul de la ligne numéro j ne dépend pas du résultat des autres lignes.

33.17.2 Transformation gaussienne

Définition 33.117 (Transformation gaussienne[512])**.**
*Soit $x \in \mathbb{R}^n$ avec $x_k \neq 0$. La k^e**transformation gaussienne** pour x est la matrice*

$$M_k(x) = \mathbb{1} - T_k(x) \tag{33.270}$$

où $T_k(x)$ est la matrice unité à qui on a ajouté le vecteur

$$\tau_k(x) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ x_{k+1}/x_k \\ \vdots \\ x_n/x_k \end{pmatrix} \tag{33.271}$$

à la k^ecolonne.

Autrement dit, la matrice $M_k(x)$ est la matrice

$$M_k(x) = \begin{pmatrix} 1 & & & & & \\ 0 & \ddots & & & & \\ \vdots & 0 & 1 & & & \\ \vdots & \vdots & -x_{k+1}/x_k & 1 & & \\ \vdots & \vdots & \vdots & 0 & \ddots & \\ 0 & 0 & -x_n/x_k & 0 & & 1 \end{pmatrix} \tag{33.272}$$

En coordonnées nous avons

$$\big(M_k(x)\big)_{ij} = \delta_{ij} - \tau_k(x)_i\delta_{kj}. \tag{33.273}$$

33.118.
Les matrices de transformation gaussienne sont des matrices triangulaires de diagonale unitaire (c'est-à-dire avec des 1 sur la diagonale).

Lemme 33.119.
Si $x \in \mathbb{R}^n$ alors nous avons

$$M_k(x)x = \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{33.274}$$

Démonstration. Nous avons

$$\big(M_k(x)x\big)_i = \sum_l M_k(x)_{il}x_l = \sum_l \big(\delta_{il} - \tau_k(x)_i\delta_{kl}\big)x_l = x_i - \tau_k(x)_i x_k. \tag{33.275}$$

Si $i \leqslant k$ nous avons $\tau_k(x)_i = 0$ et donc $\big(M_k(x)x\big)_i = x_i$. Si par contre $i \geqslant k+1$ alors $\tau_k(x)_i = \frac{x_i}{x_k}$ et alors $\big(M_k(x)x\big)_i = 0$. □

Lemme 33.120.
Si $y \in \mathbb{R}^n$ vérifie $y_i = 0$ pour $i > k$ alors $M_{k+1}(x)y = y$.

Démonstration. C'est une simple vérification :

$$\big(M_{k+1}(x)y\big)_i = \sum_l \big(\delta_{il} - \tau_{k+1}(x)_i\delta_{k+1,l}\big)y_l = y_i - \tau_{k+1}(x)_i y_{k+1}. \tag{33.276}$$

Mais comme $y_{k+1} = 0$ il nous reste automatiquement y_i. □

Le sens de ce lemme est si un vecteur est déjà « gaussiannisé » au niveau k, alors en lui appliquant une transformation gaussienne de niveau plus élevé que k, il ne change pas. Ce fait est important parce qu'il assure que lorsque l'on avance dans le processus de Gauss, chaque étape ne détruit pas les précédentes.

Le lemme suivant nous indique que l'inverse d'une matrice de transformation gaussienne est facile à calculer[18].

Lemme 33.121.
L'inverse de la transformation gaussienne

$$M_k(x)_{ij} = \delta_{ij} - \tau_k(x)_i\delta_{kj}. \tag{33.277}$$

est la matrice donnée par

$$M_k(x)^{-1}_{ij} = \delta_{ij} + \tau_k(x)_i\delta_{kj}. \tag{33.278}$$

Autrement dit, il suffit de changer le signe de la partie non diagonale.

Démonstration. Il s'agit d'une simple vérification, utilisant le produit matriciel explicite, et en remarquant que $\tau_k(x)_k = 0$ pour tout k. □

33.17.3 Méthode de Gauss pour résoudre des systèmes d'équations linéaires

Pour résoudre un système d'équations linéaires, on procède comme suit :

(1) Écrire le système sous forme matricielle.

$$\text{p.ex.} \begin{cases} 2x + 3y &= 5 \\ x + 2y &= 4 \end{cases} \Leftrightarrow \left(\begin{array}{cc|c} 2 & 3 & 5 \\ 1 & 2 & 4 \end{array}\right)$$

18. Elle rentre d'ailleurs dans la catégorie des matrices triangulaires dont nous avons déjà discuté l'inverse.

(2) Se ramener à une matrice avec un maximum de 0 dans la partie de gauche en utilisant les transformations admissibles :

(2a) Remplacer une ligne par elle-même + un multiple d'une autre ;

$$\text{p.ex.} \left(\begin{array}{cc|c} 2 & 3 & 5 \\ 1 & 2 & 4 \end{array}\right) \overset{L_1 - 2.L_2 \mapsto L_1'}{\Longrightarrow} \left(\begin{array}{cc|c} 0 & -1 & -3 \\ 1 & 2 & 4 \end{array}\right)$$

(2b) Remplacer une ligne par un multiple d'elle-même ;

$$\text{p.ex.} \left(\begin{array}{cc|c} 0 & -1 & -3 \\ 1 & 2 & 4 \end{array}\right) \overset{-L_1 \mapsto L_1'}{\Longrightarrow} \left(\begin{array}{cc|c} 0 & 1 & 3 \\ 1 & 2 & 4 \end{array}\right)$$

(2c) Permuter des lignes.

$$\text{p.ex.} \left(\begin{array}{cc|c} 0 & 1 & 3 \\ 1 & 0 & -2 \end{array}\right) \overset{L_1 \mapsto L_2' \text{ et } L_2 \mapsto L_1'}{\Longrightarrow} \left(\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 3 \end{array}\right)$$

(3) Retransformer la matrice obtenue en système d'équations.

$$\text{p.ex.} \left(\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 3 \end{array}\right) \Leftrightarrow \begin{cases} x &= -2 \\ y &= 3 \end{cases}$$

Remarques :

— Si on obtient une ligne de zéros, on peut l'enlever :

$$\text{p.ex.} \left(\begin{array}{ccc|c} 3 & 4 & -2 & 2 \\ 4 & -1 & 3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \Leftrightarrow \left(\begin{array}{ccc|c} 3 & 4 & -2 & 2 \\ 4 & -1 & 3 & 0 \end{array}\right)$$

— Si on obtient une ligne de zéros suivie d'un nombre non-nul, le système d'équations n'a pas de solution :

$$\text{p.ex.} \left(\begin{array}{ccc|c} 3 & 4 & -2 & 2 \\ 4 & -1 & 3 & 0 \\ 0 & 0 & 0 & 7 \end{array}\right) \Leftrightarrow \begin{cases} \cdots \\ \cdots \\ 0x + 0y + 0z = 7 \end{cases} \Rightarrow \textbf{Impossible}$$

— Si on moins d'équations que d'inconnues, alors il y a une infinité de solutions qui dépendent d'un ou plusieurs paramètres :

$$\text{p.ex.} \left(\begin{array}{ccc|c} 1 & 0 & -2 & 2 \\ 0 & 1 & 3 & 0 \end{array}\right) \Leftrightarrow \begin{cases} x - 2z = 2 \\ y + 3z = 0 \end{cases} \Leftrightarrow \begin{cases} x = 2 + 2\lambda \\ y = -3\lambda \\ z = \lambda \end{cases}$$

33.17.4 Méthode de Gauss sans pivot (décomposition LU)

La méthode de Gauss est encore utilisée aujourd'hui dans les vrais problèmes.

La méthode de Gauss est souvent aussi appelée méthode « LU » qui va décomposer $A = LU$ où L est triangulaire inférieure et U est triangulaire supérieure. La décomposition est même plus précise que cela : on demande que L ait seulement des 1 sur la diagonale.

Si A est une matrice nous notons

$$\Delta_k(A) = (A_{ij})_{1 \leqslant i,j \leqslant k} \tag{33.279}$$

la matrice tronquée dont nous ne gardons que le carré $k \times k$ en haut à gauche.

Lemme 33.122.
Soit S une matrice triangulaire inférieure. Soient également A et B telles que $B = SA$. Alors

$$\Delta_k(B) = \Delta_k(S)\Delta_k(A). \tag{33.280}$$

Démonstration. En effet nous avons

$$\Delta_k(B)_{ij} = \sum_{l=1}^{n} S_{il}A_{lj}. \tag{33.281}$$

Dans la somme sur l il ne reste que les termes $l \leqslant i$. Mais dans le calcul des éléments de matrice $\Delta_k(B)_{ij}$, nous avons évidemment $i, j \leqslant k$. Donc $l \leqslant i \leqslant k$. Les seuls éléments de matrice de A qui sont utilisés dans la somme (33.281) sont les éléments A_{lj} avec $l, j \leqslant k$.

Nous pouvons donc limiter la somme à $l = k$ au lieu de $l = n$ et écrire $\Delta_k(A)_{lj}$ au lieu de A_{lj}.

Même chose en ce qui concerne S. À partir du moment où l est limité à k, les éléments S_{il} et $\Delta_k(S)_{il}$ sont les mêmes. □

Théorème 33.123 (Décomposition LU[513, 1])**.**
Soit une matrice A inversible telles que $\det(\Delta_k(A)) \neq 0$ pour tout k. Alors il existe un unique couple de matrices (L, U) telles que

- *U soit triangulaire supérieure*
- *L soit triangulaire inférieure, de diagonale unité*
- *$A = LU$.*

De plus pour tout $k \leqslant n$ nous avons

$$\Delta_k(A) = \Delta_k(L)\Delta_k(U). \tag{33.282}$$

Démonstration. Nous allons prouver par récurrence le fait suivant : pour tout $1 \leqslant k \leqslant n-1$ il existe des matrices E_i $(i = 1, \dots, k)$ telles que en posant

$$A_k = E_k \dots E_1 A, \tag{33.283}$$

- E_j est une transformation gaussienne pour la j^{e}colonne,
- pour tout $j \leqslant k$, $A_{ij} = 0$ dès que $i > j$. Autrement dit la matrice A_k est triangulaire supérieure jusqu'à y compris la $(k+1)^{\text{e}}$colonne (laquelle est quelconque). Exemple pour fixer les idées : pour une matrice $A \in \mathbb{M}(4 \times 4)$, la matrice A_2 doit avoir la forme

$$A_2 = E_2E_1A = \begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & \circledast & * \\ 0 & 0 & * & * \end{pmatrix} \tag{33.284}$$

  où les éléments notés $*$ sont à priori non nuls,
- l'élément de matrice $(A_k)_{k+1,k+1}$ est non nul (celui entouré dans l'exemple).

La chose un peu triste dans cette démonstration est que l'initialisation va être très ressemblante au pas de récurrence.

<u>Initialisation : $k = 1$</u> Vu que $\Delta_1(A)$ est inversible, l'élément A_{11} est non nul. Il existe donc une transformation gaussienne E_1 telle que la première colonne de la matrice $A_1 = E_1A$ soit nul sauf la première composante. En particulier $(A_1)_{21} = 0$.

Par le lemme 33.122, nous avons $\Delta_2(A_1) = \Delta_2(E_1)\Delta_2(A)$, donc[19]

$$\det\big(\Delta_2(A_1)\big) = \det\big(\Delta_2(E_1)\big)\det\big(\Delta_2(A)\big). \tag{33.285}$$

Étant donnée la forme (33.272), toutes les matrices du type $\Delta_k(E_i)$ ont un déterminant unité, et par hypothèse $\Delta_2(A)$ est inversible, donc de déterminant non nul. Par conséquent $\det\big(\Delta_2(A_1)\big) \neq 0$. Mais comme ce déterminant est le produit des éléments diagonaux (c'est une matrice triangulaire), ces derniers ne sont pas nuls. Finalement, $(A_1)_22 \neq 0$.

19. Le déterminant est multiplicatif, proposition 9.10(1).

Le pas de récurrence Nous supposons avoir $A_k = E_k \dots E_1 A$ avec $(A_k)_{k+1,k+1} \neq 0$. Alors il existe une transformation gaussienne E_{k+1} de la $(k+1)^{\text{e}}$colonne telle que $A_{k+1} = E_{k+1}A_k$ soit une matrice dont la $(k+1)^{\text{e}}$colonne n'ait que des zéros en dessous de la $(k+1)^{\text{e}}$position. Vu le lemme 33.120, cette transformation n'affecte pas les colonnes précédentes.

La matrice A_{k+1} est donc triangulaire supérieure jusqu'à la $(k+2)^{\text{e}}$colonne.

Vu que le produit $E_k \dots E_1$ est une matrice triangulaire inférieure, le lemme 33.122 fonctionne encore et nous avons

$$\Delta_{k+1}(A_k) = \Delta_{k+1}(E_k \dots E_1)\Delta_{k+1}(A). \tag{33.286}$$

En ce qui concerne les déterminants, par hypothèse, nous avons $\det\big(\Delta_{k+1}(A)\big) \neq 0$ ainsi que $\det\big(\Delta_{k+1}(E_k \dots E_1)\big) = 1$. Donc

$$\det\big(\Delta_{k+1}(A_k)\big) \neq 0. \tag{33.287}$$

Cette matrice étant triangulaire de déterminant non nul, ses éléments diagonaux sont tous non nuls ; en particulier $(A_k)_{k+1,k+1} \neq 0$.

En poussant la récurrence jusqu'au bout, la matrice

$$A_{n-1} = E_{n-1} \dots E_n A \tag{33.288}$$

est triangulaire supérieure.

Nous posons alors $L = (E_{n-1} \dots E_n)^{-1}$ et $U = A_{n-1}$. Cela prouve l'existence parce que

$$A = (E_{n-1} \dots E_1)^{-1} A_{n_1}. \tag{33.289}$$

Encore une fois, le lemme 33.122 nous donne

$$\Delta_k(A) = \Delta_k\Big((E_{n_1} \dots E_1)^{-1}\Big)\Delta_k(A_{n-1}), \tag{33.290}$$

ou encore $\Delta_k(A) = \Delta_k(L)\Delta_k(U)$.

En ce qui concerne l'unicité, si $A = L_1U_1 = L_2U_2$ alors $L_2^{-1}L_1 = U_2U_1^{-1}$. Vu qu'à gauche nous avons une matrice triangulaire inférieure et que à droite nous avons une triangulaire inférieure, nous savons que les deux membres représentent une matrice diagonale. Mais à gauche, la diagonale est unitaire. Donc les deux membres représentent la matrice unité. □

33.124.
En pratique, pour résoudre $Ax = b$, il faut seulement appliquer les transformations gaussiennes à la matrice élargie $(A|b)$ pour finir sur un système du type

$$Ux = b' \tag{33.291}$$

qui est immédiatement soluble. Autrement dit, en effectuant les annulations de colonnes, la matrice U est « gratuite ».

Il n'est pas indispensable de calculer la matrice L qui, elle, demande à chaque étape de se souvenir de la matrice E_i utilisée. S'il faut résoudre plusieurs systèmes $Ax_i = b_i$, nous pouvons encore travailler avec la matrice encore plus élargie $(A|b_1 \dots b_m)$.

Si par contre nous ne connaissons pas à l'avance l'ensemble des vecteurs b avec lesquels il faudra résoudre le système, il est bon de calculer la décomposition $A = LU$ in extenso, c'est-à-dire de garder une trace des matrices L et U séparément. Dans ce cas, résoudre $Ax = b$ revient à résoudre $Ly = b$, et ensuite $Ux = y$. Ce sont deux systèmes de résolution directe parce que les matrices sont triangulaires.

33.125.
Le fait que

$$\Delta_k(A) = \Delta_k(L)\Delta_k(U) \tag{33.292}$$

nous dit que si après avoir calculé L et U nous remarquons que le système est un peu plus petit ou un peu plus grand que prévu, tout le travail n'est pas perdu. En particulier si le système est plus petit que prévu, l'adaptation de L et U est immédiate.

Notons que U et L sont inversibles, et que $\det(L) = 1$. Donc $\det(U) = \det(A)$.

Exemple 33.126.
Pour travailler la méthode de Gauss pour le système $Ax = b$, nous introduisons la matrice un peu augmentée $(A|b)$. Nous faisons un exemple. Soit à résoudre

$$\begin{pmatrix} 2 & 1 & 3 \\ 4 & 3 & 10 \\ -2 & 1 & 73 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 11 \\ 28 \\ 3 \end{pmatrix}. \tag{33.293}$$

Nous introduisons la matrice augmentée

$$(A|b)^{(0)} = \begin{pmatrix} 2 & 1 & 3 & 11 \\ 4 & 3 & 10 & 28 \\ -2 & 1 & 7 & 3 \end{pmatrix}. \tag{33.294}$$

Le premier pas consiste à annuler tous les éléments sous la diagonale de la première colonne. Autrement dit, nous prenons le 2 comme pivot. Nous introduisons les multiplicateurs $l_{ij} = \frac{A_{ij}}{A_{i1}}$. La nouvelle matrice est :

$$(A|b)^{(1)} = \begin{pmatrix} 2 & 1 & 3 & 11 \\ 0 & 1 & 4 & 6 \\ 0 & 2 & 10 & 14 \end{pmatrix} \tag{33.295}$$

où nous avons utilisé les multiplicateurs $l_{21} = 2$, $l_{31} = -1$.

Et la matrice suivante est :

$$(A|b)^{(2)} = \begin{pmatrix} 2 & 1 & 3 & 11 \\ 0 & 1 & 4 & 6 \\ 0 & 0 & 2 & 2 \end{pmatrix} \tag{33.296}$$

où nous avons utilisé le multiplicateur $l_{32} = 2$.

Cela est un système de résolution immédiate :

$$\begin{cases} 2x + y + 3z = 11 & \text{(33.297a)} \\ y + 4z = 6 & \text{(33.297b)} \\ 2z = 2. & \text{(33.297c)} \end{cases}$$

La troisième donne $z = 1$. Ensuite $y + 4 = 6$, donc $y = 2$. Et la première donne : $2x + 2 + 3 = 11$, c'est-à-dire $2x = 6$, enfin : $x = 3$.

Solution : $(x, y, z) = (3, 2, 1)$.

Nous notons surtout que dans $(A|b)^{(2)}$ nous avons une matrice triangulaire supérieure. Où est la matrice triangulaire inférieure ? En réalité la matrice L est la matrice des multiplicateurs :

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix}. \tag{33.298}$$

△

Le problème de cette méthode est que faisant ainsi nous risquons d'avoir un zéro sur un des pivots. Par exemple tomber sur

$$(A|b) = \begin{pmatrix} 2 & 1 & 3 & 11 \\ 0 & 0 & 4 & 6 \\ 0 & 2 & 10 & 14 \end{pmatrix}. \tag{33.299}$$

Le zéro sur la deuxième ligne nous ennuie si nous voulons tout faire dans l'ordre. Mais notons qu'en échangeant les deux dernières lignes, tout va bien : le système donné par

$$(A|b) = \begin{pmatrix} 2 & 1 & 3 & 11 \\ 0 & 2 & 10 & 14 \\ 0 & 0 & 4 & 6 \end{pmatrix} \tag{33.300}$$

fonctionne très bien. Et même tellement bien qu'il est de résolution immédiate, dans ce cas.

Un autre problème est que si un des pivots est 10^{-14}, le multiplicateur sera de l'ordre 10^{14}, qui est mal représenté en mémoire. Il est donc bon de prendre les pivots le plus grand possible. Si le pivot est le plus grand nombre en valeur absolue d'une colonne, alors les nombres x_{k+i}/x_k qui entrent dans la matrice de transformation gaussienne sont des nombres dans $[-1, 1]$ qui sont bien représentés en mémoire.

Tout cela nous incite à développer une méthode de Gauss qui permet de tenir une trace des permutations.

33.17.5 Matrice de permutation élémentaire

Définition 33.127.
*Une **matrice de permutation élémentaire** est une matrice obtenue en permutant deux lignes de la matrice identité. Nous notons P_{ij} la matrice obtenue en inversant les lignes i et j de la matrice identité.*

Exemple 33.128.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = P_{12}. \tag{33.301}$$

△

Lemme 33.129.
La matrice $P_{ij}A$ est la matrice A avec ses lignes i et j inversées.

Démonstration. Il suffit d'écrire

$$(P_{ij}A)_{kl} = \sum_m (P_{ij})_{km} A_{ml} \tag{33.302}$$

et de faire trois cas selon que $k = i$, $k = j$ ou k différent de i et j. Si $k = i$ alors $(P_{ij})_{im} = \delta_{mj}$ et si k est différent de i et j alors $(P_{ij})_{mk} = \delta_{km}$ (troisième cas similaire au premier). □

Et la matrice AP_{12} est la A avec ses deux premières *colonnes* échangées.

Avec ces notations, notre matrice $(A|b)^{0'}$ est

$$P_{12}(A|b)^{(0)}. \tag{33.303}$$

Puis la matrice $(A|b)^{(1')}$ est

$$P_{23}(A|b)^{(1)}. \tag{33.304}$$

Et la matrice P qui arrive dans $PA = LU$ est la matrice $P = P_{23}P_{21}$, qui est une matrice de permutations non élémentaire. Elle vaut :

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}. \tag{33.305}$$

Lemme 33.130 ([514])**.**
Si $i, j > k$ alors les matrices de permutation élémentaires ont la relation de « commutation » suivante avec les transformations gaussiennes :

$$M_k(x)P_{ij} = P_{ij}M_k\big(P_{ij}(x)\big). \tag{33.306}$$

Démonstration. Il suffit de calculer les éléments de matrice :

$$\big(P_{ij}M_k(x)\big)_{st} = (P_{ij})_{st} - \sum_m (P_{ij})_{sm}\tau_k(x)_m\delta_{kt}, \tag{33.307}$$

mais

$$\sum_m (P_{ij})_{sm}\tau_k(x)_m = \big(P_{ij}\tau_k(x)\big)_s = \tau_k\big(P_{ij}(x)\big)_s \tag{33.308}$$

parce que $i, j > k$ implique que dans $P_{ij}\tau_k(x)$ nous inversons deux élément non nuls de $\tau_k(x)$, tout en laissant le k^{e}élément. Le dénominateur ne change pas et il s'agit réellement d'une inversion de ligne. Donc

$$\big(P_{ij}M_k(x)\big)_{st} = (P_{ij})_{st} - \tau_k\big(P_{ij}x\big)_s\delta_{kt}. \tag{33.309}$$

De l'autre côté,

$$\big(M_k(y)P_{ij}\big)_{st} = (P_{ij})_{st} - \tau_k(y)_s(P_{ij})_{kt}. \tag{33.310}$$

Mais comme $i, j > k$ la k^{e}ligne de P_{ij} est la même que celle de la matrice unité, donc $(P_{ij})_{kt} = \delta_{kt}$.

$$\big(M_k(y)P_{ij}\big)_{st} = (P_{ij})_{st} - \tau_k(y)_s\delta_{kt}. \tag{33.311}$$

Cela correspond bien à (33.309). □

33.18 Méthode de Gauss avec pivot partiel (décomposition PLU)

33.18.1 L'idée

À chaque pas, nous faisons une permutation de ligne. Nous permutons à chaque pas la première ligne avec celle qui a le pivot le plus grand (en valeur absolue). Donc :

$$(A|b)^{(0)} = \begin{pmatrix} 2 & 1 & 3 & 11 \\ 4 & 3 & 10 & 28 \\ -2 & 1 & 7 & 3 \end{pmatrix} \tag{33.312}$$

Nous commençons par déplacer des lignes :

$$(A|b)^{(0')} = \begin{pmatrix} 4 & 3 & 10 & 28 \\ 2 & 1 & 3 & 11 \\ -2 & 1 & 7 & 3 \end{pmatrix}. \tag{33.313}$$

Les multiplicateurs sont $l_{21} = 1/2$ et $l_{31} = -1/2$. Le fait est que les multiplicateurs ont toujours le plus grand dénominateur possible et nous avons alors toujours $0 \leqslant |l_{ij}| \leqslant 1$, qui sont des nombres relativement petits, et bien représentés en mémoire.

Nous avons la nouvelle matrice

$$(A|b)^{(1)} = \begin{pmatrix} 4 & 3 & 10 & 28 \\ 0 & -1/2 & -2 & -3 \\ 0 & 5/2 & 12 & 17 \end{pmatrix}. \tag{33.314}$$

Le pivot serait $-1/2$. Nous cherchons un pivot plus grand en dessous de ce $-1/2$ (et pas au dessus, sinon on casserait les zéros déjà trouvés). Nous trouvons le $5/2$ qui est plus grand. Nous permutons donc les deux dernières lignes :

$$(A|b)^{(1')} = \begin{pmatrix} 4 & 3 & 10 & 28 \\ 0 & 5/2 & 12 & 17 \\ 0 & -1/2 & -2 & -3 \end{pmatrix} \tag{33.315}$$

où le pivot est maintenant $l_{32} = -1/5$. La matrice suivante :

$$(A|b)^{(1)} = \begin{pmatrix} 4 & 3 & 10 & 28 \\ 0 & 5/2 & 12 & 17 \\ 0 & 0 & 2/5 & 2/5 \end{pmatrix} \tag{33.316}$$

Dans ce cas, la matrice L n'est pas aussi simple à construire parce que nous avons permuté des choses. Dans ce cas, la matrice L est encore de la forme

$$L = \begin{pmatrix} 1 & 0 & 0 \\ . & 1 & 0 \\ . & . & 1 \end{pmatrix}. \tag{33.317}$$

Mais vu que nous avons permuté les lignes 2 et 3 au deuxième pas, nous devons permuter l_{21} et l_{31} avant de remplir la matrice L avec les multiplicateurs :

$$L = \begin{pmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 1/2 & -1/5 & 1 \end{pmatrix}. \tag{33.318}$$

Notons que ces L et U ne sont pas les mêmes que le LU obtenu sans pivot. Où est l'unicité ? Elle est que en fait maintenant nous n'avons pas $A = LU$, mais

$$PA = LU \tag{33.319}$$

où P est une matrice de permutation.

33.18.2 Le théorème

Proposition 33.131 (Méthode de Gauss avec pivot partiel[514])**.**
Soit une matrice inversible $A \in \mathbb{M}(n, \mathbb{C})$. Il existe

- *une matrice de permutations P*
- *une matrice triangulaire inférieure de diagonale unitaire L,*
- *une matrice triangulaire supérieure inversible U*

telles que

$$PA = LU. \tag{33.320}$$

Notons que cette proposition ne demande que l'hypothèse d'inversibilité pour A. Il n'y a pas d'hypothèses sur tous les mineurs comme c'était le cas avec Gauss sans pivot.

Démonstration. Nous prouvons par récurrence qu'il existe des matrices Q_k, $E_1, \dots, E_k$ et A_k telles que

$$Q_k A = E_1 \dots E_k A_k \tag{33.321}$$

avec

(1) Q_k est une matrice de permutation
(2) E_i est une transformation gaussienne sur la i^{e}colonne
(3) A_k est triangulaire supérieure jusqu'à la k^{e}colonne.

Sachant que $\det(Q_k) = \pm 1$, et que $\det(E_i) = 1$, le passage au déterminant dans (33.321) nous donne $\det(A_k) \neq 0$ et si nous notons $\Omega_k(A)$ la matrice tronquée de A, ne gardant que les entrées plus grandes que k, nous avons

$$\det(A_k) = \prod_{i=1}^{k} (A_k)_i i \det\big(\Omega_{k+1}(A_k)\big). \tag{33.322}$$

Donc : $(A_k)_{ii} \neq 0$ pour $i \leqslant k$ et $\det\big(\Omega_{k+1}(A_k)\big) \neq 0$.

Pour fixer les idées, voici une image de $k = 2$:

$$\begin{array}{c} \Delta_k(A_2) \\ \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & * & * & * \end{pmatrix} \\ \Omega_{k+1}(A_2) \end{array} \tag{33.323}$$

Étant donné que $\det\big(\Omega_{k+1}(A_k)\big) \neq 0$, parmi les nombres $(A_k)_{i,k+1}$ $(i \geqslant k+1)$, au moins un est non nul et nous posons r_{k+1} tel que $|(A_k)_{r_{k+1},k+1}|$ soit maximum parmi ces éléments.

Le nombre r_{k+1} est enregistré parce qu'il servira à écrire la matrice P plus tard. Les matrices E_i ne sont pas enregistrées, parce que nous verrons qu'elles vont encore changer. Seule la dernière sera enregistrée.

La composante $(k+1, k+1)$ de la matrice

$$P_{r_{k+1},k+1}A_k \tag{33.324}$$

est non nulle et peut donc servir de pivot. Soit M_{k+1} la transformation gaussienne pour la $(k+1)^{\text{e}}$colonne de la matrice $P_{r_{k+1},k+1}A_k$. La matrice

$$A_{k+1} = M_{k+1}P_{r_{k+1},k+1}A_k \tag{33.325}$$

est alors une matrice triangulaire supérieure jusqu'à la $(k+1)^{\text{e}}$colonne. En posant $E_{k+1} = M_{k+1}^{-1}$ nous avons

$$P_{r_{k+1},k+1}E_{k+1}A_{k+1} = A_k, \tag{33.326}$$

et nous nous sentons en droit de récrire l'équation de départ (33.321) :

$$Q_kA = E_1 \dots E_kA_k = E_1 \dots E_kP_{r_{k+1},k+1}E_{k+1}A_{k+1}. \tag{33.327}$$

Le lemme 33.130 nous permet de ramener la matrice $P_{r_{k+1},k+1}$ en première position, quitte à modifier un peu (pas beaucoup) chacune des matrices E_i $(i = 1, \dots, k)$. C'est pour cela que nous n'enregistrons pas les matrices E_i. Nous avons donc

$$P_{r_{k+1},k+1}Q_kA = E'_1 \dots E'_kE_{k+1}A_{k+1} \tag{33.328}$$

où

— Le produit $P_{r_{k+1},k+1}Q_k$ est encore une matrice de permutation, et mieux : elle vaut

$$\prod_{i=1}^{k+1} P_{r_i,i}. \tag{33.329}$$

Cela montre qu'il est suffisant d'enregistrer les nombres r_i pour reconstituer cette partie.

— La matrice E'_i est une transformation gaussienne pour la i^{e}colonne.

— La matrice A_{k+1} est triangulaire supérieure jusqu'à la $k+1^{\text{e}}$colonne.

La récurrence est maintenant finie et nous pouvons écrire avec $k = n$:

$$Q_nA = E_1 \dots E_nA_n \tag{33.330}$$

où le produit $E_1 \dots E_n$ est triangulaire inférieure et A_n est triangulaire supérieur.

Maintenant nous enregistrons la matrice $U = A_n$, le produit $L = \prod_{i=1}^n E_n$ et les nombres r_i qui permettent de retrouver P. □

Note : dans l'équation (33.330) nous avons bien entendu massivement renommé les E'_i en E_i. En réalité la matrice E_1 vient avec n primes sur la tête.

Dans les exemples 33.134, 33.135 et 33.136, nous allons résoudre le système

$$\begin{pmatrix} 10^{-9} & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{33.331}$$

d'abord de façon exacte, et ensuite en supposant une machin ne tenant que 8 chiffres significatifs en utilisant la méthode de Gauss avec ou sans pivot.

Commençons par voir comment se passe en pratique la décomposition $PA = LU$ de Gauss avec pivot partiel.

Exemple 33.132.
Décomposons la matrice

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 0 \\ 3 & 8 & 0 \end{pmatrix}. \tag{33.332}$$

Sur la première colonne, le plus grand nombre est 3. Nous commençons par permuter la première et la troisième ligne en utilisant la matrice de permutation $P_1 = P_{3,1}$ et nous enregistrons $r_1 = 3$. Nous avons alors la matrice

$$A'_0 = \begin{pmatrix} 3 & 8 & 0 \\ 2 & 5 & 0 \\ 1 & 3 & 3 \end{pmatrix}. \tag{33.333}$$

Pour trouver la matrice A_1 nous suivons l'équation (33.325). Bien que le résultat net soit des combinaisons de lignes : $L_2 \to L_2 - 2L_1/3$ et $L_3 \to L_3 - L_1/3$ (que nous pourrions savoir dès à présent), il est important de passer par la matrice gaussienne pour obtenir la matrice L_1.

La matrice de transformation gaussienne pour la première colonne de (33.333) est :

$$M_1 = \begin{pmatrix} 1 & 0 & 0 \\ -2/3 & 1 & 0 \\ -1/3 & 0 & 1 \end{pmatrix} \tag{33.334}$$

et $L_1 = M_1^{-1}$. Le lemme 33.121 nous dit comment calculer facilement cet inverse :

$$L_1 = \begin{pmatrix} 1 & 0 & 0 \\ 2/3 & 1 & 0 \\ 1/3 & 0 & 1 \end{pmatrix} \tag{33.335}$$

En suivant l'équation (33.325) nous posons $A_1 = M_1 A'_0$:

$$A_1 = \begin{pmatrix} 1 & 0 & 0 \\ -2/3 & 1 & 0 \\ -1/3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 8 & 0 \\ 2 & 5 & 0 \\ 1 & 3 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 8 & 0 \\ 0 & -1/3 & 0 \\ 0 & -2/3 & 3 \end{pmatrix} \tag{33.336}$$

et nous avons

$$Q_1 A = L_1 A_1 \tag{33.337}$$

où L_1, A_1 et $r_1 = 3$ sont enregistrés. La matrice Q_1 peut être retrouvée en sachant r_1 parce que P est la matrice de permutation $P_{r_1,1}$.

Nous travaillons maintenant sur la deuxième colonne de A_1. Le plus grand élément en valeur absolue (sur ou sous la diagonale) est $-2/3$. Nous posons $r_2 = 3$ et

$$A'_1 = \begin{pmatrix} 3 & 8 & 0 \\ 0 & -2/3 & 3 \\ 0 & -1/3 & 0 \end{pmatrix} \tag{33.338}$$

et la matrice gaussienne pour la deuxième colonne est

$$M_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/2 & 1 \end{pmatrix} \tag{33.339}$$

Le $-1/2$ provient du calcul $-\big((-1/3)/(-2/3)\big)$. L'inverse de cette matrice est facile :

$$L_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/2 & 1 \end{pmatrix} \tag{33.340}$$

et la matrice suivante à enregistrer est

$$A_2 = M_2 P_{3,2} A_1 = M_2 A_1' = \begin{pmatrix} 3 & 8 & 0 \\ 0 & -2/3 & 3 \\ 0 & 0 & -3/2 \end{pmatrix}. \tag{33.341}$$

Notons toutefois que pour calculer cette matrice, seul le dernier élément demande un calcul. La première colonne ne change pas (par construction), la seconde gagne un zéro en dernière ligne (la matrice M_2 sert à ça) et sur la dernière colonne, seule la dernière ligne est sujette à changement.

Avec la matrice A_2, la trigonalisation supérieure est faite. La décomposition n'est cependant pas terminée. Nous devons encore trouver la partie triangulaire inférieure. Nous en sommes à

$$Q_1 A = L_1 A_1 = L_1 P_{3,2} L_2 A \tag{33.342}$$

où Q_1 est la première matrice de permutation.

Utilisant le lemme 33.130, il est facile de permuter L_1 avec $P_{3,2}$:

$$L_1 P_{3,2} = P_{3,2} \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 2/3 & 0 & 1 \end{pmatrix}}_{L_1'} \tag{33.343}$$

Nous avons donc

$$P_{3,2} P_{3,1} A = L_1' L_2 A \tag{33.344}$$

Deux multiplications matricielles plus tard nous terminons :

$$PA = LU \tag{33.345}$$

avec

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 0 & 0 \\ 1/3 & 1 & 0 \\ 2/3 & 1/2 & 1 \end{pmatrix}, \quad U = A_2 = \begin{pmatrix} 3 & 8 & 0 \\ 0 & -2/3 & 3 \\ 0 & 0 & -3/2 \end{pmatrix}. \tag{33.346}$$

△

Notons que Sage utilise la méthode de Gauss avec pivots :

```
sage: A=matrix([ [1,2,3],[2,5,0],[3,8,0] ])
sage: A
[1 2 3]
[2 5 0]
[3 8 0]
sage: A.LU()
(
[0 1 0]   [   1    0    0]   [    3     8     0]
[0 0 1]   [1/3    1    0]   [    0  -2/3     3]
[1 0 0],  [2/3  1/2    1],  [    0     0  -3/2]
)
```

tex/sage/sageSnip006.sage

Mais attention : Sage crée une décomposition $A = PLU$ et non $PA = LU$. D'où le fait que la matrice de permutations de Sage est l'inverse de celle donnée ici.

33.18.3 D'un point de vue algorithmique

Problèmes et choses à faire

Je ne suis pas certain de l'optimalité de ce que je raconte ici. Je décris simplement ce que j'ai fait pour écrire mon programme finitediff.

Si vous êtes expert en calcul numérique, n'hésitez pas à donner votre avis.

Nous décrivons à présent la décomposition $A = PLU$ (du théorème 33.131, avec le P à droite). En suivant l'exemple 33.132 nous voyons assez bien comment créer les matrices U et P au fur et à mesure. La construction de L est peut-être moins évidente.

Écrivons un exemple très explicite pour

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 3 & 10 \\ -2 & 1 & 7 \end{pmatrix}. \tag{33.347}$$

Nous commençons par permuter des lignes pour avoir un grand pivot :

$$P_{12}A = \begin{pmatrix} 4 & 3 & 10 \\ 2 & 1 & 3 \\ -2 & 1 & 7 \end{pmatrix}. \tag{33.348}$$

Et nous effectuons l'élimination avec la matrice

$$M_1 = \begin{pmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 1/2 & 0 & 1 \end{pmatrix}. \tag{33.349}$$

Cela donne le premier résultat :

$$M_1P_{12}A = \begin{pmatrix} 4 & 3 & 10 \\ 0 & -1/2 & -2 \\ 0 & 5/2 & 12 \end{pmatrix} \tag{33.350}$$

Nous continuons avec P_{23} pour avoir un nouveau grand pivot :

$$P_{23}M_1P_{12}A = \begin{pmatrix} 4 & 3 & 10 \\ 0 & 5/2 & 12 \\ 0 & -1/2 & -2 \end{pmatrix}. \tag{33.351}$$

Nous utilisons la matrice

$$M_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/5 & 1 \end{pmatrix} \tag{33.352}$$

et au final :

$$M_2P_{23}M_1P_{12}A = \begin{pmatrix} 4 & 3 & 10 \\ 0 & 5/2 & 12 \\ 0 & 0 & 2/5 \end{pmatrix} = U. \tag{33.353}$$

L'égalité obtenue est

$$M_2P_{23}M_1P_{12}A = U. \tag{33.354}$$

Pour avoir la décomposition PLU il faut écrire

$$A = P_{12}M_1^{-1}P_{23}M_2^{-1}U, \tag{33.355}$$

et permuter P_{23} avec M_1^{-1}, ce qui est facile par le lemme 33.130.

Remarque 33.133.
Nous ne devons permuter la matrice M_k avec une matrice de permutations qu'à partir de la deuxième étape. En effet l'équation (33.350) revient à

$$A = M_1^{-1}P_{12}m_U \tag{33.356}$$

qui est dans le bon ordre. Ce n'est qu'à partir de la seconde étape que des matrices de permutations apparaissent à droite des matrices gaussiennes.

Cependant dans un cas 4×4, cette méthode deviendrait fastidieuse parce que nous aurions encore des étapes à faire. En repartant de (33.355), mais avec m_U (la matrice pas encore tout à fait triangularisée) au lieu de U, nous aurons, pour un certain $k > 3$:

$$M_3 P_{3k} M_2 P_{23} M_1 P_{12} A = U, \tag{33.357}$$

ce qui fait :

$$A = P_{12} M_1^{-1} P_{23} M_2^{-1} P_{3k} M_3^{-1} U. \tag{33.358}$$

Tous les P_{ij} peuvent être mis à gauche parce que leurs indices sont toujours strictement supérieurs à ceux des M_l placés devant eux. Mais c'est fastidieux.

Nous allons donc permuter à chaque étape pour ne retenir que l'important. Si à une certaine étape nous avons

$$A = P_{1,r_1} \dots P_{k,r_k} M_1^{-1} \dots M_k^{-1} m_U \tag{33.359}$$

avec

$$m_U = P_{k+1,r_{k+1}} M_{k+1}^{-1} U \tag{33.360}$$

alors nous allons directement permuter $P_{k+1,r_{k+1}}$ avec tous les M_i^{-1}. Si nous notons P_k la permutation (pas élémentaire) à l'étape k et L_k la matrice triangulaire inférieure à de l'étape k,

$$A = P_k L_k m_U = P_k L_k P_{k+1,r_{k+1}} M_{k+1}^{-1} m'_U. \tag{33.361}$$

Nous enregistrons alors $P_{k+1} = P_k P_{k+1,r_{k+1}}$ et pour L_{k+1} nous partons de L_k et nous faisons deux opérations suivantes :

— nous permutons, sur ses colonnes non triviales, les indices $k+1$ et r_{k+1},

— nous multiplions par M_{k+1}^{-1}, ce qui revient à simplement lui ajouter une colonne non triviale.

Notons que $r_{k+1} \geqslant k+1$, de telle sorte que sur les colonnes non triviales (qui sont jusqu'au numéro k), la permutation des lignes $k+1$ et r_{k+1} ne change pas l'aspect de la matrice : elle reste multi-gaussienne de dernière colonne k.

33.18.4 Exemples

Nous nous lançons dans la résolution du système

$$\begin{pmatrix} 10^{-9} & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}. \tag{33.362}$$

Exemple 33.134.
Nous commençons de façon exacte, par la méthode de Gauss sans pivot. La première transformation gaussienne est

$$E_1 = \begin{pmatrix} 1 & 0 \\ -10^9 & 1 \end{pmatrix} \tag{33.363}$$

et nous calculons

$$E_1 A = \begin{pmatrix} 1 & 0 \\ -10^9 & 1 \end{pmatrix} \begin{pmatrix} 10^{-9} & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 10^{-9} & 1 \\ 0 & 1 - 10^{-9} \end{pmatrix}. \tag{33.364}$$

Vu que cette dernière est triangulaire supérieure, nous avons fini la méthode de Gauss et $U = E_1 A$. En ce qui concerne la matrice L, elle est donnée par $L = E_1^{-1}$, c'est-à-dire

$$L = \begin{pmatrix} 1 & 0 \\ -10^9 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 10^9 & 1 \end{pmatrix}. \tag{33.365}$$

Au final nous avons la décomposition $A = LU$ exacte suivante :

$$L = \begin{pmatrix} 1 & 0 \\ 10^9 & 1 \end{pmatrix} \quad U = \begin{pmatrix} 10^{-9} & 1 \\ 0 & 1 - 10^9 \end{pmatrix}. \tag{33.366}$$

Résoudre le système $Ax = b$ revient à résoudre $LUx = b$ et donc résoudre successivement les systèmes

$$\begin{cases} Ly = b & (33.367a) \\ Ux = y. & (33.367b) \end{cases}$$

D'abord le système

$$\begin{pmatrix} 1 & 0 \\ 10^9 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{33.368}$$

donne $y_1 = 1$ et $y_2 = 2 - 10^9$.

Ensuite nous résolvons

$$\begin{pmatrix} 10^{-9} & 1 \\ 0 & 1 - 10^9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 - 10^9 \end{pmatrix}. \tag{33.369}$$

Cela donne

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -\frac{10^9}{1-10^9} \\ \frac{2-10^9}{1-10^9} \end{pmatrix} \simeq \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{33.370}$$

C'est également le résultat que trouve Sage :

```
sage: var('y')
y
sage: solve(    [10**(-9)*x+y==1,x+y==2],[x,y]    )
[[x == (1000000000/999999999), y == (999999998/999999999)]]
```

tex/sage/sageSnip007.sage

△

Exemple 33.135 ([514]).
Nous recommençons tout le calcul avec une précision limitée à 8 chiffres significatifs, sans pivot.

Nous avons à nouveau la transformation gaussienne

$$E_1 = \begin{pmatrix} 1 & 0 \\ -10^9 & 1 \end{pmatrix}, \tag{33.371}$$

mais pour calculer U nous effectuons le produit matriciel

$$U = E_1 A = \begin{pmatrix} 1 & 0 \\ -10^9 & 1 \end{pmatrix} \begin{pmatrix} 10^{-9} & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 10^{-9} & 1 \\ 0 & * \end{pmatrix}. \tag{33.372}$$

Nous détaillons à présent le calcul de l'élément noté $*$. Le calcul de $10^9 \ominus 1)$ donne

$$999999999 = 9.99999999 \times 10^8, \tag{33.373}$$

mais la précision étant limitée à 8 chiffres, un arrondit arrive. Étant donné que le premier chiffres supprimé est un 9 nous retombons sur 10^9, et donc notre machine à précision limitée donnera

$$U = \begin{pmatrix} 10^{-9} & 1 \\ 0 & -10^9 \end{pmatrix}. \tag{33.374}$$

Ensuite le calcul de $L = E_1^{-1}$ ne cause pas de problèmes :

$$L = \begin{pmatrix} 1 & 0 \\ -10^9 & 1 \end{pmatrix}. \tag{33.375}$$

Maintenant il s'agit de résoudre les systèmes $Ly = b$ et $Ux = y$. Du système

$$\begin{pmatrix} 1 & 0 \\ 10^9 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \tag{33.376}$$

nous tirons tout de suite $y_1 = 1$ et ensuite $10^9 + y_2 = 2$, c'est-à-dire $y_2 = 2 - 10^9$, qui en précision limitée donne encore $y_2 = -10^9$. À résoudre maintenant :

$$\begin{pmatrix} 10^{-9} & 1 \\ 0 & -10^9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -10^9 \end{pmatrix}. \tag{33.377}$$

Cela donne immédiatement $x_2 = 1$ et ensuite

$$10^{-9} x_1 + 1 = 1, \tag{33.378}$$

donc $x_1 = 0$. La solution trouvée est

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{33.379}$$

qui est complètement faux au niveau de la première variable. △

Exemple 33.136 ([514]).
Nous résolvons encore le même système en précision limitée, mais en utilisant cette fois la méthode de Gauss avec pivot partiel.

Le plus grand élément de la première colonne est 1 ; nous utilisons donc la permutation $P_{1,2}$:

$$P_{1,2} A = \begin{pmatrix} 1 & 1 \\ 10^{-9} & 1 \end{pmatrix}. \tag{33.380}$$

La matrice de transformation gaussienne pour la première colonne de cette matrice est

$$M_1 = \begin{pmatrix} 1 & 0 \\ -10^{-9} & 1 \end{pmatrix} \tag{33.381}$$

et nous posons

$$A_1 = M_1 P_{1,2} A = \begin{pmatrix} 1 & 0 \\ -10^{-9} & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 10^{-9} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & -10^{-9} + 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = U \tag{33.382}$$

où un arrondi a eu lieu pour $-10^{-9} + 1 = 1$. En inversant M_1 nous avons

$$L_1 = M_1^{-1} = \begin{pmatrix} 1 & 0 \\ 10^{-9} & 1 \end{pmatrix}. \tag{33.383}$$

La décomposition est

$$A = \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}}_{P} \underbrace{\begin{pmatrix} 1 & 0 \\ 10^{-9} & 1 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}}_{U} \tag{33.384}$$

Le moment de résoudre est venu. Vu que $PLUx = b$ nous devons résoudre les systèmes

$$\begin{cases} Pz = b & \text{(33.385a)} \\ Ly = z & \text{(33.385b)} \\ Ux = y. & \text{(33.385c)} \end{cases}$$

Pour z c'est facile :

$$z = \begin{pmatrix} 2 \\ 1 \end{pmatrix}. \tag{33.386}$$

Pour y il y a un arrondi :

$$\begin{pmatrix} 1 & 0 \\ 10^{-9} & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}. \tag{33.387}$$

Tout de suite : $y_1 = 2$ et ensuite $2 \times 10^{-9} + y_2 = 1$, ce qui donne $y_2 = 1 \ominus 2 \times 10^{-9} = 1$. Donc

$$y = \begin{pmatrix} 2 \\ 1 \end{pmatrix}. \tag{33.388}$$

Et enfin pour x c'est le système

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}. \tag{33.389}$$

Nous avons $x_2 = 1$ et ensuite $x_1 + 1 = 2$ c'est-à-dire $x_1 = 1$. Au final la solution trouvée est

$$x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{33.390}$$

Cette solution est considérablement meilleure que (33.379). △

33.137.
L'utilisation du pivot non seulement assure le fait que la trigonalisation va bien se passer (on évite les zéros en pivot), mais aussi et surtout, en choisissant de prendre le plus grand pivot possible, nous obtenons une meilleur stabilité numérique.

33.19 Résolution de systèmes linéaires (suite)

33.19.1 Déterminant

Pour calculer un déterminant lorsque nous avons la décomposition $A = LU$ nous pouvons faire

$$\det(A) = \det(LU) = \det(L)\det(U) = \det(U) \tag{33.391}$$

parce que L est triangulaire avec des 1 sur la diagonale.

Si par contre nous avons fait des pivots, nous avons $PA = LU$. Il nous faut le déterminant de P, qui n'est autre que ± 1. Nous avons

$$\det(P) = (-1)^s \tag{33.392}$$

où s est le nombre de permutations effectives effectuées. Nous précisons « effectives » parce qu'il ne faut pas compter le pas où nous n'avons pas permuté (les cas où le bon pivot était présent du premier coup). Nous avons alors

$$\det(A) = (-1)^s \det(U). \tag{33.393}$$

33.19.2 Plusieurs termes indépendants

Mettons un système $Ax = b$ qu'il faut résoudre pour plusieurs b différents. C'est typiquement le cas où l'on voudrait calculer l'inverse de A. Mais on va directement se calmer. Soient donc à résoudre $Ax_1 = b_1$, ..., $Ax_n = b_n$.

Les opérations (avec ou sans pivot) que nous faisons ne dépendent que de la matrice A, mais aucune décisions concernant les pivots ou la matrice des multiplicateurs ne dépend de b. Autre façon de dire : si le système $(A|b_1)$ devient $(U|y_1)$, le système $(A|b_i)$ devient $(U|y_i)$ avec le même U.

Nous ne sommes donc pas obligés de faire tout le travail autant de fois qu'il n'y a de systèmes à résoudre. Donc si on a plusieurs systèmes à résoudre avec la même matrice, on fait mieux de retenir une fois pour toute la décomposition LU (avec ou sans pivots), avant de vraiment résoudre.

Ou alors on peut aussi faire que, au lieu de faire $(A|b_i)$ plein de fois, faire une seule fois

$$(A|b_1 \dots b_n). \tag{33.394}$$

Et on fait tout le travail sur tous les vecteurs d'un en même temps.

Soit e_i la base canonique. Si nous notons x_n les solutions des problèmes $Ax_i = e_i$, tous les problèmes $Ax_i = e_i$ s'écrivent d'un seul coup

$$AX = Y \tag{33.395}$$

où X est la matrice des x_i en colonnes, et Y est celle des e_i en colonnes. Oh, mais $Y = \mathbb{1}$ évidemment. Donc

$$AX = \mathbb{1}. \tag{33.396}$$

Si nous supposons que A est inversible, alors ce X est l'inverse.

Donc pour calculer l'inverse d'une matrice de dimension non trop grande, il suffit d'utiliser la méthode de Gauss sur les vecteurs de la base canonique. Cette idée est la base du calcul de l'inverse par matrice companion. En effet, si nous partons du problème

$$(A|\mathbb{1}) \tag{33.397}$$

et nous appliquons la méthode de Gauss avec pivot, nous arrivons à

$$(U|L^{-1}P). \tag{33.398}$$

Attention : le produit $L^{-1}P$ est une permutation des *colonnes* de L^{-1}. Vu que L est triangulaire inférieure avec des 1 sur la diagonale, L^{-1} est triangulaire inférieure avec des 1 sur la diagonale. Donc si la matrice n'est pas trop grande, on peut assez facilement remettre les colonnes de $L^{-1}P$ dans l'ordre pour recomposer une matrice triangulaire inférieure avec 1 sur la diagonale.

Une autre façon de calculer l'inverse, si $A = LU$ est connue, il suffit de faire

$$A^{-1} = U^{-1}L^{-1}. \tag{33.399}$$

Et il existe un algorithme facile pour l'inverse d'une matrice triangulaire.

33.19.3 Cholesky

Le commandant Cholesky travaillait sur le tir de canon (chose éminemment liée à de nombreuses mathématiques ingénieuses). La méthode de Cholesky est encore utilisée aujourd'hui dans les vrais problèmes.

La méthode de Gauss s'applique sans hypothèses sur la matrice A, à part qu'elle doit être de petite dimension, comme pour toute méthode directe. Souvent nous savons des choses sur la matrice. Ici nous allons supposer que A est symétrique et définie positive.

Comment numériquement vérifier ces hypothèses ? En ce qui concerne la symétrique, il suffit de faire le test complet :

$$A^t = A. \tag{33.400}$$

La vérification de cela coûte au maximum n^2 comparaisons (et en fait la moitié de ça moins la diagonale).

Le fait que A soit définie positive est facile à vérifier pour utiliser Cholesky parce que il suffit de le faire, et s'il n'y a pas de nombres complexes qui arrivent, c'est que la matrice était définie positive.

Un lemme très simple à mettre en oeuvre numériquement nous permet de traiter certains cas.

Lemme 33.138.
Une matrice symétrique possédant un élément négatif sur la diagonale n'est pas définie positive.

Démonstration. Un simple calcul ou effort d'imagination montre que $\langle Me_k, e_k\rangle = M_{kk}$. Donc si M doit être définie positive, M_{kk} doit être positive par le lemme 9.157. □

Ce lemme est un moyen déjà de faire quelques vérifications. Et si les éléments diagonaux de A sont tous négatifs, on peut prendre $-A$.

Lemme 33.139.
Si A est une matrice symétrique strictement définie positive, alors pour tout k, la matrice tronquée $\Delta_k(A)$ l'est également.

Démonstration. Le fait que $\Delta_k(A)$ soit symétrique est évidemment. Le fait qu'elle soit définie positive l'est moins. Soit $y \in \mathbb{R}^k$ et le vecteur $\tau y \in \mathbb{R}^n$, qui est « complété » avec des zéros.

Nous avons $\langle \Delta_k(A)y, y\rangle_k = \langle A\tau y, \tau\rangle_n$. En effet

$$\langle \Delta_k(A)y, y\rangle = \sum_{i=1}^{k}\sum_{l=1}^{k} A_{il}y_l y_i. \tag{33.401}$$

Et à droite :

$$\langle A\tau y, \tau y\rangle = \sum_{i=1}^{n}(A\tau y)_i(\tau y)_i = \sum_{i=1}^{k}(A\tau y)_i y_i = \sum_{i=1}^{k}\sum_{l=1}^{n} A_{il}(\tau y)_l y_i = \sum_{i=1}^{k}\sum_{l=1}^{k} A_{il}y_l y_i \tag{33.402}$$

où nous avons utilisé le fait que $(\tau y)_i = 0$ dès que $i > k$ et que $(\tau y)_i = y_i$ sinon.

En conséquence de quoi $\langle \Delta_k(A)y, u\rangle > 0$ pour tout $y \in \mathbb{R}^k$ et la matrice $\Delta_k(A)$ est strictement définie positive. □

Lemme 33.140.
Si T est une matrice triangulaire, alors $(T_{ii})^{-1} = (T^{-1})_{ii}$.

Démonstration. Il suffit de se rendre compte que le coefficient ii de l'égalité $\mathbb{1} = TT^{-1}$ donne

$$1 = \sum_l T_{il}(T^{-1})li. \tag{33.403}$$

Dans la somme il ne reste que le terme $l = i$. □

Nous allons chercher une décomposition de type LU sous la forme $A = LL^t$, c'est-à-dire $U = L^t$. Attention : maintenant nous n'avons plus des 1 sur la diagonale. Ce n'est donc pas exactement la décomposition LU dont nous parlions plus haut. C'est pour cela que nous n'allons pas la noter LL^t mais BB^t.

Théorème 33.141 (Cholesky[513]).
Soit une matrice réelle symétrique strictement définie positive. Il existe une unique matrice réelle B telle que

- *B est triangulaire inférieure,*
- *la diagonale de B est positive,*
- *$A = BB^t$.*

Démonstration. Par la décomposition LU du théorème 33.123 nous avons des matrices L et U telles que $A = LU$. Soit D la matrice diagonale donnée par

$$D_{ii} = \sqrt{U_{ii}}. \tag{33.404}$$

Cette définition fonctionne parce que $U_{ii} > 0$. En effet nous savons que $\Delta_k(A) = \Delta_k(L)\Delta_k(U)$, et en passant au déterminant,

$$\det\big(\Delta_k(A)\big) = \det\big(\Delta_k(U)\big). \tag{33.405}$$

Vu que $\Delta_k(A)$ est strictement définie positive par le lemme 33.139, son determinant est strictement positif[20] et nous avons

$$\det\big(\Delta_k(U)\big) > 0. \tag{33.406}$$

En appliquant cela à $k = 1$ nous avons $U_{11} > 0$ puis de proche en proche, $U_i i > 0$ pour tout i.

Nous posons :

$$B = LD \qquad \text{qui est triangulaire inférieure} \tag{33.407a}$$

$$C = D^{-1}U \qquad \text{qui est triangulaire supérieure.} \tag{33.407b}$$

20. Le théorème 9.150 donne une diagonalisation par des matrices de déterminant 1. Vu que les valeurs propres forment sur la diagonale, et qu'elles sont toutes positives, el déterminant est positif.

Nous avons bien entendu $A = BC$ et nous allons prouver que $C = B^t$. Vu que $A = A^t$ nous pouvons identifier BC et C^tB^t :

$$BC = C^tB^t. \tag{33.408}$$

En mettant les matrices triangulaires supérieures à gauche et inférieures à droite :

$$C(B^t)^{-1} = B^{-1}C^t, \tag{33.409}$$

qui sont donc deux matrices diagonales. Nous montrons que cette diagonale est en réalité l'identité.

D'abord

$$B_{ii} = \sum_{l=1}^{n} L_{il}D_{li} = L_{ii}\sqrt{U_{ii}} = \sqrt{U_{ii}} \tag{33.410}$$

parce que $L_{ii} = 1$. Notons en passant que la diagonale de B est positive. Ensuite

$$C_{ii} = \sum_{l=1}^{n}(D^{-1})_{il}U_{li} = (D^{-1})_{ii}U_{ii} = \frac{1}{\sqrt{U_i i}}U_{ii} = \sqrt{U_{ii}}. \tag{33.411}$$

Donc B et C ont des diagonales égales. Calculons alors la diagonale de $B^{-1}C^t$:

$$\left(B^{-1}C^t\right)_{ii} = \sum_{l}(B^{-1})_{il}(C^t)_{li} = (B^{-1})_{ii}C_{ii} \tag{33.412}$$

parce que encore une fois, de la somme il ne reste que le terme $l = i$.

Mais B est une matrice triangulaire qui tombe sous le coup du lemme 33.140. Donc $(B^{-1})_{ii} = (B_{ii})^{-1} = (C_{ii})^{-1}$. Nous avons alors

$$(B^{-1}C^t)_{ii} = 1. \tag{33.413}$$

Cela conclu l'existence de la décomposition de Cholesky.

En ce qui concerne l'unicité, soient $A = BB^t = CC^t$. Nous regroupons les supérieures et les inférieures :

$$B^t(C^t)^{-1} = B^{-1}C. \tag{33.414}$$

Ces deux matrices sont donc diagonales et nous posons $D = B^{-1}C$, c'est-à-dire $C = BD$. Nous remplaçons donc C par BD dans (33.414) :

$$A = BB^t = BD(BD)^t = BDD^tB^t. \tag{33.415}$$

Donc $DD^t = \mathbb{1}$, ce qui signifie que les éléments diagonaux de D sont ± 1. Nous montrons qu'ils sont positifs : à partir de $C = BD$ nous déballons

$$C_{ii} = \sum_{l} B_{il}D_{li}, \tag{33.416}$$

et donc

$$B_{ii}D_{ii} = C_{ii}. \tag{33.417}$$

En sachant que les conditions de la décomposition de Cholesky demandent les éléments diagonaux positifs nous en déduisons que D_{ii} est positif et donc égal à 1. Finalement $D = \mathbb{1}$ et $B = C$. □

Prenons la matrice

$$A = \begin{pmatrix} 4 & 2 & -2 \\ 2 & 10 & -7 \\ -2 & -7 & 9 \end{pmatrix} \tag{33.418}$$

Elle est symétrique et définie positive. Nous posons

$$\{ \quad l_{11} = \sqrt{a_{11}} l_{i1} = a_{i1}/l_{11} \tag{33.419a}$$

pour $i = 2, \dots, n$. Et aussi

$$\begin{cases} l_{jj} = (a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2)^{1/2} & (33.420a) \\ l_{ij} = (a_{ij} - \sum_{k=1}^{j-1} l_{ik} l_{jk})/a_{jj} & (33.420b) \end{cases}$$

pour $i = j+1, \dots, n$.

Les formules (33.419) nous disent comment remplir la première colonne. Cela donne la matrice

$$L = \begin{pmatrix} \sqrt{4} = 2 & . & . \\ 2/2 = 1 & . & . \\ -2/2 = -1 & . & . \end{pmatrix} \tag{33.421}$$

Les formules (33.420) donnent les autres colonnes en fonction des précédentes.

Dans Sage :

```
sage: A=matrix(  [  [4,2,-2],[2,10,-7],[-2,-7,9]  ]   )
sage: A
[ 4  2 -2]
[ 2 10 -7]
[-2 -7  9]
sage: A.cholesky()
[ 2  0  0]
[ 1  3  0]
[-1 -2  2]
```

tex/sage/sageSnip005.sage

33.20 Système linéaire (méthodes itératives)

Nous trouvons des méthodes itératives lorsque les matrices sont grandes, ce qui arrive lorsque l'on discrétise une équation différentielle.

Nous allons chercher des méthodes de la forme $x_{n+1} = Bx_n + q$; ce sont des méthodes stationnaires. La convergence d'une méthode est toujours liée à la matrice B et en général, la convergence ne dépend pas du choix du vecteur initial. Nous faisons donc souvent $x_0 = 0$ et donc $x_1 = q$. Voilà donc une itération de faite gratuitement.

Nous notons e_k le **vecteur d'erreur** qui est définit par $e_k = x - x_k$. Et le **vecteur résidu** $r_k = b - Ax_k$. Attention : ici k n'est pas un indice mais un numéro de vecteur.

Notons que si x est solution, alors $b - Ax = 0$, ce qui motive le vecteur résidu.

Les conditions d'arrêt d'un algorithme seraient

$$\begin{cases} \|e_k\|_\infty \ll \epsilon_1 & (33.422a) \\ \|r_k\|_\infty < \epsilon_2 & (33.422b) \end{cases}$$

où ϵ_1 et ϵ_2 sont des précisions décidées à l'avance par l'utilisateur.

Proposition 33.142.
Si A est une matrice inversible, alors

$$\lim_{k\to\infty} e_k = \lim_{k\to\infty} r_k = 0. \tag{33.423}$$

Vu que $r_k = Ae_k$, si la matrice A est mal conditionnée, il peut arriver que r_k reste grand alors que e_k est déjà petit.

Remarque 33.143.
Dans les méthode stationnaires, nous avons $x_{n+1} = Bx_n + q$ avec B et q fixés au départ de l'algorithme. Il existe des méthodes non stationnaires pour lesquelles l'itération prend la forme $x_{n+1} = B_n x_n + q_n$ avec B_n et q_n qui changement avec les étapes.

Proposition 33.144.
Pour la méthode $x_{n+1} = Bx_n + q$ nous avons équivalence de

(1) La méthode converge pour tout x_0

(2) B est une matrice convergente[21]

(3) $\rho(B) < 1$ (rayon spectral).

De plus si $\|B\| < 1$ alors la méthode converge (quelle que soit la norme algébrique).

La norme d'une matrice (en tout cas, certaines normes) est quelque chose de facile à calculer à l'ordinateur. Typiquement $\|.\|_\infty$ est un simple maximum. Cependant si après avoir calculé $\|B\|_i$ pour des dizaines de normes i différentes, nous avons toujours $\|B\|_i \geqslant 1$, alors nous ne pouvons rien conclure.

33.20.1 La méthode générale

Nous décomposons la matrice A sous la forme $A = M - N$ avec M inversible. Le système $Ax = b$ devient

$$Mx - Nx = b \tag{33.424}$$

puis $Mx = Nx + b$ et finalement

$$x = M^{-1}Nx + M^{-1}b, \tag{33.425}$$

et voilà une méthode stationnaire avec $B = M^{-1}N$ et $q = M^{-1}b$.

Mais ici nous voyons que M doit être non seulement inversible, mais en plus doit être facilement calculable. En sachant que nous travaillons avec des grandes matrices, il n'est pas question d'inverser M avec une méthode de Gauss.

En bref, il faut choisir M triangulaire parce que c'est en gros la seule que nous pouvons inverser facilement[22].

Remarque 33.145.
La matrice B ne doit pas spécialement être inversible. Si elle ne l'est pas, ce n'est pas un problème.

33.20.2 Jacobi

Nous décomposons

$$A = D - E - F \tag{33.426}$$

où D est la diagonale de A, $-F$ est la partie triangulaire supérieure (sans la diagonale) et $-E$ la triangulaire inférieure (sans la diagonale). Donc D, E et F sont simplement des extractions de parties de la matrice A (et quelques changements de signes).

La méthode de Jacobi prend $M = D$ et $N = (E + F)$. L'inverse de M est facile à calculer parce que M est diagonale. Nous notons B_J la matrice B de la méthode de Jacobi.

Remarque 33.146.
Il se peut que la matrice A ait des zéros sur la diagonale, même si elle est inversible. Et cela est un problème parce qu'alors la matrice D ici construite n'est pas inversible. Dans ce cas, avant de nous lancer dans la méthode de Jacobi, il faut permuter deux lignes de A et donc de b.

Attention cependant que l'on pourrait vouloir effectuer ces permutations en mettant sur la diagonale des nombres les plus grands possibles (parce qu'ensuite, ce qui rentre dans les calculs,

21. C'est-à-dire $\lim_{k\to\infty} B^k = 0$.

22. Les matrices orthogonales sont aussi facilement inversibles, mais ne se prêtent pas bien à une décomposition de type somme.

c'est D^{-1} qui aura alors des petits nombres). Mais il faut toutefois faire en sorte que le rayon spectral de la matrice B résultante reste plus petit que 1.

Chaque changement dans A induit des changements dans B et donc sur la convergence de la méthode.

33.20.3 Gauss-Seidel

Nous partons de la même décomposition $A = D - E - F$ que dans (33.426). La méthode de Gauss-Seidel prend $M = (D - E)$ et $N = F$.

33.20.4 Autres

Voir la méthode des gradients, et des gradients conjugués.

33.21 Indices connectés, matrice irréductible

Définition 33.147.
*Soit $A \in \mathbb{M}(n, \mathbb{R})$ et $i, j \in \{1, \dots, n\}$. Nous disons que les indices i et j sont **directement connectés** si $A_{ij} \neq 0$ ou $A_{ji} \neq 0$.*

Définition 33.148.
*Soit $A \in \mathbb{M}(n, \mathbb{R})$ et $i, j \in \{1, \dots, n\}$. Nous disons que les indices i et j sont **connectés** si il existe un ensemble d'indices $i_0 = i, i_1, \dots, i_{r-1}, i_r = j$ tels que $A_{i_k, i_{k+1}} \neq 0$ pour tout $0 \leqslant k \leqslant r$.*

Par exemple pour que les indices 1 et 4 soient connectés, on peut avoir les éléments A_{13}, A_{32} A_{24} non nuls.

Définition 33.149 ([515])**.**
*Une matrice carrée A est **réductible** si il existe une permutation σ telle que*

$$\sigma^t A \sigma = \begin{pmatrix} K & L \\ 0 & M \end{pmatrix} \tag{33.427}$$

où K et M sont carrées.

Notons que par définition de la matrice d'une application linéaire,

$$B_{ij} = \langle e_i, Be_j \rangle = \langle e_i, \sigma^t A \sigma e_j \rangle = \langle \sigma e_i, A \sigma e_j \rangle = A_{\sigma(i), \sigma(j)}. \tag{33.428}$$

Proposition 33.150 ([515])**.**
Soit une matrice carrée A. Les faits suivants sont équivalents :

(1) A est réductible.

(2) Il existe une partition non triviale I, J de $\{1, \dots, n\}$ telle que $I \cup J = \{1, \dots, n\}$, $I \cap J = \varnothing$ et pour tout $i \in I$, et pour tout $j \in J$, $A_{ij} = 0$.

(3) La matrice A admet des indices non connectés (définition 33.148).

Démonstration. Dans plusieurs sens...

(1) implique (2) Nous notons j^* la taille de la matrice K dans (33.427). Nous avons $B_{ij} = 0$ si

$$\begin{cases} J^* + 1 \leqslant i \leqslant n & \text{(33.429a)} \\ 1 \leqslant j \leqslant j^*. & \text{(33.429b)} \end{cases}$$

Donc en posant $I = \sigma\{j^* + 1, \dots, n\}$ et $J = \sigma\{1, \dots, j^*\}$ nous avons une partition non triviale de $\{1, \dots, n\}$ telle que si $i \in I$ et $j \in J$ alors $i = \sigma(i_0)$, $j = \sigma(j_0)$ et

$$A_{ij} = A_{\sigma(i_0), \sigma(j_0)} = B_{i_0, j_0} = 0. \tag{33.430}$$

(2) implique (1) Soit une partition I, J comme indiquée dans l'hypothèse. Soit j^* le nombre d'éléments dans J. Soit σ une permutation de $\{1, \dots, n\}$ telle que $\sigma\{j^* + 1, \dots, n\} = I$ et $\sigma\{1, \dots, j^*\} = J$. Nous posons ensuite $B = \sigma^t A \sigma$. Par construction si $i \in I$ et $j \in J$ alors $A_{ij} = 0$.

Mais si

$$\begin{cases} J^* + 1 \leqslant i \leqslant n & (33.431a) \\ 1 \leqslant j \leqslant j^*. & (33.431b) \end{cases}$$

alors $B_{ij} = A_{\sigma(i)\sigma(j)} = 0$. Donc B a la bonne forme.

(3) implique (2) Soient i et j deux indices non connectés : il n'existe pas de chaines partant de i et arrivant à j. Nous notons I l'ensemble des indices connectés à i, et J les autres. Pat hypothèses ces ensembles sont non vides.

Si $k \in i$ et $l \in J$ alors $A_{kl} = 0$ parce que sinon on aurait une chaine de i à k puis de k à l et donc de i à l, ce qui signifierait que l est connecté à i.

(2) implique (3) Soit une partition I, J comme dans l'hypothèse. Si $j \in J$ est connecté à $i \in I$ alors il existe une chaine

$$i = i_0, i_1, \dots, i_r = j. \tag{33.432}$$

Si i_s est le premier dans J alors $i_{s-1} \in I$ et $A_{i_{s-1}, i_s} = 0$, ce qui empêche la chaine de connecter j à i.

□

33.22 Localisation des valeurs propres

Sur l'ensemble $\mathbb{M}(n, \mathbb{R})$ des matrices $n \times n$ à coefficients réels nous introduisons l'ordre partiel [23] donné par $A \geqslant B$ lorsque $A_{ij} \geqslant B_{ij}$ pour tout i et j. Nous définissons de façon similaire les relations $A \leqslant B$, $A < B$ et $A > B$.

Si $x \in \mathbb{R}^n$ nous notons $|x| = \big(|x_1|, \dots, |x_n|\big)$ et $x \leqslant y$ lorsque $x_i \leqslant y_i$ pour tout i.

Proposition 33.151.
Soit $A \in \mathbb{M}(n, \mathbb{R})$ et $x, y \in \mathbb{R}^n$.

(1) Si $A \geqslant 0$ et si $x \leqslant y$ alors $Ax \leqslant Ay$.

(2) Si $A \geqslant 0$ alors $Ax \leqslant |Ax| \leqslant A|x|$.

Démonstration. Pour la première inégalité, pour tout i et k nous avons $A_{ik}x_k \leqslant A_{ik}y_k$ et donc

$$(Ax)_i = \sum_k A_{ik}x_k \leqslant \sum_k A_{ik}y_k = (Ay)_k. \tag{33.433}$$

Pour la seconde, d'abord l'inégalité $Ax \leqslant |Ax|$ est évidente. Ensuite vu que $A_{ik} \geqslant 0$ nous avons

$$|Ax|_i = |\sum_k A_{ik}x_k| \leqslant \sum_k A_{ik}|x_k| = \big(A|x|\big)_i. \tag{33.434}$$

□

Soit une matrice $A \in \mathbb{M}(n, \mathbb{R})$. Nous notons

$$r_i = \sum_{j \neq i} |A_{ij}|. \tag{33.435}$$

Notons la somme sur la ligne i, pas sur la colonne : la somme est horizontale.

23. Définition 1.8.

Définition 33.152.
Les ensembles

$$D_i = \{z \in \mathbb{C} \text{ tel que } |z - A_{ii}| \leqslant r_i\} \tag{33.436}$$

sont les ***disques de Gershgorin****. Nous allons également noter $B_i = \mathrm{Int}(D_i)$ les boules ouvertes correspondantes.*

Théorème 33.153 (Gershgorin)**.**
Soit $A \in \mathbb{M}(n, \mathbb{R})$. Si $\lambda \in \mathbb{C}$ est valeur propre de A alors $\lambda \in D_i$ pour un certain i.

Démonstration. Soit une valeur propre λ et un de ses vecteurs propres $u \in \mathbb{R}^n$: $Au = \lambda u$ avec $u \neq 0$. Soit i un indice réalisant le maximum $|u_i| = \max\{|u_k|\}_k$. Nous écrivons la i^{e}ligne de $Au = \lambda u$:

$$\sum_k A_{ik} u_k = \lambda u_i, \tag{33.437}$$

c'est-à-dire $A_{ii} u_i + \sum_{k \neq i} A_{ik} u_k = \lambda u_i$, ou encore

$$A_{ii} + \sum_{k \neq i} A_{ik} \frac{u_k}{u_i} = \lambda, \tag{33.438}$$

qui donne

$$|A_{ii} - \lambda| \leqslant \sum_{k \neq i} |A_{ik}| \frac{|u_k|}{|u_i|} \leqslant \sum_{k \neq i} |A_{ik}| \tag{33.439}$$

pare que $|u_i| \geqslant |u_k|$. Notons que sur la ligne précédente, $|.|$ est le module dans $\mathbb{C}$, pas la valeur absolue dans $\mathbb{R}$. □

Théorème 33.154 (Gershgorin 2[515])**.**
Soit une matrice irréductible $A \in \mathbb{M}(n, \mathbb{R})$ et une valeur propre λ de A. Si elle est sur la frontière de l'union des disques de Gershgorin, alors elle est sur le bord de tous les disques.

Démonstration. Soit une valeur propre λ de A telle que $\lambda \in \partial\big(\bigcup_i D_i\big)$. Alors λ n'est dans l'intérieur d'aucune boule et nous avons $|\lambda - A_{ii}| \geqslant r_i$ pour tout i.

Soit un vecteur propre u de A tel que $\|u\|_\infty = 1$. Nous posons $I = \{1 \leqslant i \leqslant n \text{ tel que } |u_i| = 1\}$ et $J = \{1 \leqslant j \leqslant n \text{ tel que } |i_j| < 1\}$. Par hypothèse I n'est pas vide, et de plus $I \cap J = \varnothing$ et $I \cup J = \{1, \dots, n\}$ parce qu'aucune composante de u n'a un module[24] plus grand que 1.

La i^{e}composante de la relation $Au = \lambda u$ peut s'écrire

$$(A_{ii} - \lambda) u_i + \sum_{k \neq i} A_{ik} u_k = 0. \tag{33.440}$$

Forts de cela nous écrivons les inégalités suivantes :

$$r_i \leqslant |\lambda - A_{ii}| = \big|(\lambda - A_{ii}) u_i\big| = |\sum_{k \neq i} A_{ik} u_k| \leqslant \sum_{k \neq i} |A_{ik}||u_k| \leqslant \sum_{k \neq i} |A_{ik}| = r_i. \tag{33.441}$$

Donc les inégalités sont des égalités :

$$r_i = |\lambda - A_{ii}| = \big|(\lambda - A_{ii}) u_i\big| = |\sum_{k \neq i} A_{ik} u_k| = \sum_{k \neq i} |A_{ik}||u_k| = \sum_{k \neq i} |A_{ik}|. \tag{33.442}$$

En particulier l'égalité $\sum_{k \neq i} |A_{ik}||u_k| = \sum_{k \neq i} |A_{ik}|$ donne

$$\sum_{k \neq i} |A_{ik}|\big(|u_k| - 1\big) = 0. \tag{33.443}$$

Donc pour tout $k \in J$ nous avons $A_{ik} = 0$. Vu que A est irréductible, cela donnerait une partition impossible $\{1, \dots, n\} = I \cup J$. Nous en déduisons que J est vide et donc que $|u_j| = 1$ pour tout j. En repartant de (33.442) nous avons alors

$$r_i = \big|(\lambda - A_{ii}) u_i\big| = |\lambda - A_{ii}||u_i| = |\lambda - A_{ii}|. \tag{33.444}$$

Cela prouve que $\lambda \in \partial D_i$ pour tout i. □

24. Les composantes de u sont à priori dans $\mathbb{C}$, et non spécialement dans $\mathbb{R}$, même si A est une matrice réelle.

Exemple 33.155.
Soit la matrice

$$B = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & -1/2 \\ -1 & 0 & 3 \end{pmatrix}. \tag{33.445}$$

D'abord nous rappelons que si vous voulez entrer cette matrice dans Sage (ou plus généralement dans Python2 [25]), vous devez faire attention au 1/2 qui, tel quel, est évalué à 0. Nous vous rappelons donc que tous vos codes Sage doivent commencer par ceci :

```
# -*- coding: utf8 -*-
from __future__ import unicode_literals
from __future__ division
```

tex/sage/sageSnip015.sage

Les éléments non nuls hors diagonale sont B_{13}, B_{31} et B_{23}. Elle n'est donc pas irréductible ; nous avons par exemple la partition $I = \{1, 3\}$, $J = \{2\}$ pour le critère de la proposition 33.150(2).

Les disques de Gershgorin sont

$$D_1 = \{z \in \mathbb{C} \text{ tel que } |z - 2| \leqslant 1\} \tag{33.446a}$$
$$D_2 = \{z \in \mathbb{C} \text{ tel que } |z - 1| \leqslant 1/2\} \tag{33.446b}$$
$$D_3 = \{z \in \mathbb{C} \text{ tel que } |z - 3| \leqslant 1\} \tag{33.446c}$$

Les valeurs propres de la matrice sont sur des bords de disques de Gershgorin, sans être sur tous les bords, comme ça aurait été le cas par le théorème 33.154 si la matrice avait été irréductible. Elles sont sur la figure 33.1 ; notez en particulier les valeurs propres λ_2 et λ_3 qui sont sur le bord de deux disques mais pas sur le bord des trois disques en même temps.

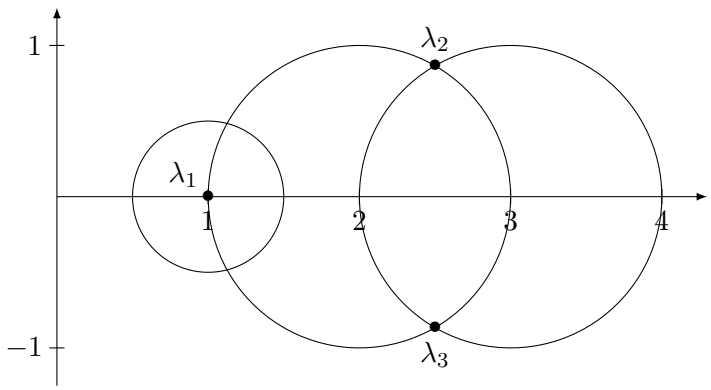


FIGURE 33.1 – Les disques de Gershgorin et les valeurs propres pour l'exemple 33.155.

△

Exemple 33.156.
Soit la matrice

$$A = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 1 & 2 \\ 3 & 0 & 2 \end{pmatrix}. \tag{33.447}$$

Nous avons

$$D_1 = \{z \in \mathbb{C} \text{ tel que } |z| \leqslant 1\}, \tag{33.448a}$$
$$D_2 = \{z \in \mathbb{C} \text{ tel que } |z - 1| \leqslant 2\}, \tag{33.448b}$$
$$D_3 = \{z \in \mathbb{C} \text{ tel que } |z - 2| \leqslant 3\}, \tag{33.448c}$$

25. Que vous n'avez aucune raison d'utiliser autre que Sage.

Le polynôme caractéristique est

$$\chi(\lambda) = -\lambda^3 + 3\lambda^2 - 2\lambda - 6. \tag{33.449}$$

Une fois remarqué que $\lambda_1 = -1$ est une racine, les autres sont faciles à trouver (division euclidienne de $\chi(\lambda)$ par $\lambda + 1$) : $\lambda_2 = 2 + i\sqrt{2}$ et $\lambda_3 = 2 - i\sqrt{2}$.

La matrice A est irréductible. En effet les éléments non diagonaux non nuls sont A_{12}, A_{23} et A_{31}. Ils peuvent former une chaine reliant tous les indices entre eux.

Les contraintes sur la localisation des valeurs propres est donc qu'elles doivent être dans ou sur les disques de Gershgorin, mais que celles qui sont sur le bord d'un disque doivent être sur le bord de tous les disques en même temps. C'est cela que nous observons sur la figure 33.2. Notez en particulier la position de la valeur propre λ_1.

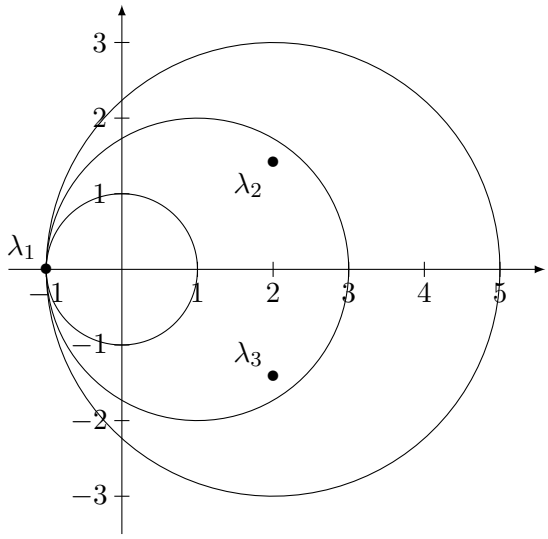


FIGURE 33.2 – Les disques de Gershgorin et les valeurs propres pour l'exemple 33.156.

△

33.22.1 Matrices à diagonale dominante

Définition 33.157 ([516, 515]).
*La matrice $A \in \mathbb{M}(n, \mathbb{K})$ est **à diagonale dominante** si pour tout i,*

$$\sum_{j \neq i} |A_{ij}| \leqslant |A_{ii}| \tag{33.450}$$

où $|.|$ est la module dans $\mathbb{C}$ ou la valeur absolue dans $\mathbb{R}$.

*Elle est **à diagonale fortement dominante** si elle est à diagonale dominante et si il existe un i tel que*

$$\sum_{j \neq i} |A_{ij}| < |A_{ii}|. \tag{33.451}$$

*Elle est **à diagonale strictement dominante** si*

$$\sum_{j \neq i} |A_{ij}| < |A_{ii}| \tag{33.452}$$

pour tout i (entier entre 1 et n).

Nous avons les inclusions suivantes :

$$\text{strictement dominante} \subset \text{fortement dominante} \subset \text{dominante}. \tag{33.453}$$

Lemme 33.158.
Si A est dans un des deux cas suivant :

— *diagonale strictement dominante,*

— *diagonale dominante et irréductible*[26]

alors $A_{ii} \neq 0$ pour tout i.

Démonstration. Si A est à diagonale strictement dominante, alors l'inégalité stricte (33.452) n'est pas possible.

Si A est à diagonale dominante, alors si $A_{ii} = 0$, toute la ligne est nulle. Dans ce cas, la matrice ne peut pas être irréductible. □

Proposition 33.159 ([516])**.**
Une matrice à diagonale strictement dominante est inversible.

Démonstration. Soit une matrice A à diagonale strictement dominante. Soit x tel que $Ax = 0$. Le but est de montrer que $x = 0$. Soit un indice i_0 réalisant la norme maximum :

$$|x_{i_0}| = \|x\|_\infty. \tag{33.454}$$

Nous écrivons la composante i_0 de l'égalité $Ax = 0$:

$$\sum_k A_{i_0k}x_k = 0, \tag{33.455}$$

et nous séparons le terme $k = i_0$ des autres :

$$\sum_{k\neq i_0} A_{i_0k}x_K + A_{i_0i_0}x_{i_0} = 0. \tag{33.456}$$

Nous prenons le module et majorons les sommes :

$$|A_{i_0i_0}||x_{i_0}| \leqslant \sum_{k\neq i_0} |A_{i_0k}||x_k| \leqslant \sum_{k\neq i_0} |A_{i_0k}||x_{i_0}|. \tag{33.457}$$

Si $|x_{i_0}|$ est non nul nous pouvons simplifier :

$$|A_{i_0i_0}| \leqslant \sum_{k\neq i_0} |A_{i_0k}|. \tag{33.458}$$

Hélas, l'hypothèse de diagonale strictement dominante implique l'inégalité stricte dans le sens inverse. Impossible. Nous en déduisons que $|x_{i_0}| = 0$. Donc $\|x\|_\infty = 0$, ce qui signifie que $x = 0$.

Le fait que le noyau de A se réduise à $\{0\}$ implique l'inversibilité de A. □

Proposition 33.160.
Soit $A \in \mathbb{M}(n, \mathbb{C})$ une matrice qui est dans un des deux cas suivants :

— *à diagonale strictement dominante*

— *à diagonale dominante et irréductible*

Si $A = D - M$ où D est la diagonale de A (et M est « le reste ») alors D est inversible et

$$\rho(D^{-1}M) < 1 \tag{33.459}$$

où ρ est le rayon spectral (thème 43).

26. Définition 33.149.

Démonstration. Le lemme 33.158 nous dit que les éléments diagonaux de A sont non nuls. Cela donne déjà le fait que la matrice D est inversible et que la produit $D^{-1}M$ ait un sens. Nous posons $T = D^{-1}M$. Nous avons alors

$$T_{ii} = \sum_k (D^{-1})_{ik} M_{ki}. \tag{33.460}$$

Si $k = i$ alors $M_{ki} = 0$ et si $k \neq i$ alors $D_{ik} = 0$. Donc $T_{ii} = 0$ pour tout i.

En ce qui concerne les autres éléments de T,

$$T_{ij} = \sum_k (D^{-1})_{ik} M_{kj} = \sum_k \frac{1}{A_{ik}} \delta_{ik} M_{kj} = -\frac{A_{ij}}{A_{ii}}. \tag{33.461}$$

Notes :

— Les hypothèses sur A jouent pour dire que $A_{ii} \neq 0$.

— Le signe moins est dû au fait que $M_{ij} = -A_{ij}$ lorsque $i \neq j$.

En faisant la somme des modules :

$$\sum_{j \neq i} |T_{ij}| = \sum_{j \neq i} \frac{|A_{ij}|}{|A_{ii}|} = \frac{1}{|A_{ii}|} \sum_{j \neq i} |A_{ij}| \leqslant 1. \tag{33.462}$$

La dernière inégalité est le fait que A soit à diagonale dominante.

Si A est à diagonale strictement dominante Alors nous avons l'inégalité stricte

$$\sum_{j \neq i} |T_{ij}| < 1. \tag{33.463}$$

Et le théorème de Gershgorin 33.153 dit que le spectre de T est contenu dans l'union des disques

$$D_i = \{z \in \mathbb{C} \text{ tel que } |z - T_{ii}| \leqslant r_i\} \tag{33.464}$$

où

$$r_i = \sum_{j \neq i} |T_{ij}|. \tag{33.465}$$

Mais nous avons prouvé que pour tout i, $T_{ii} = 0$ et $\sum_{j \neq i} |T_{ij}| < 1$. Donc toutes ces boules sont contenues dans $B(0,1)$. Cela prouve que $\rho(T) < 1$.

Diagonale dominante, irréductible La matrice T est alors également irréductible parce que les éléments non nuls de A et de T sont les mêmes : $T_{ij} = -A_{ij}/A_{ii}$. Nous utilisons alors le second théorème de Gershgorin 33.154. Si λ est une valeur propre de T, alors soit

$$\lambda \in \bigcup_i B\big(0, r_i\big) \tag{33.466}$$

soit

$$\lambda \in \bigcap_i \partial B(0, r_i). \tag{33.467}$$

Vu que $r_i \leqslant 1$ pour tout i, dans le premier cas λ est dans l'union des boules *ouvertes* de rayon 1. Le nombre λ est donc une la boule ouverte de rayon 1. Bref, $|\lambda| < 1$.

Dans le second cas, l'intersection de deux cercles de même centre sont soit vide soit tout le cercle (auquel cas les rayons sont égaux). Dans le second cas, ledit rayon est certainement strictement plus petit que 1 parce que

$$r_i = \sum_{j \neq i} |T_{ij}| = \sum_{j \neq i} \frac{|A_{ij}|}{|A_{ii}|} < 1. \tag{33.468}$$

□

33.22.2 M-matrice

Définition 33.161.
*Une matrice $A \in \mathbb{M}(n, \mathbb{R})$ est une **M-matrice** si*

(1) $A_{ii} > 0$ pour tout i,

(2) $A_{ij} \leqslant 0$ si $i \neq j$

(3) A est inversible et $A^{-1} \geqslant 0$.

Proposition 33.162.
Soit $A \in \mathbb{M}(n, \mathbb{R})$ telle que $A_{ii} > 0$ pour tout i et $A_{ij} \leqslant 0$ pour tout $i \neq j$. Nous posons $A = D - M$ où D est la diagonale de A.

La matrice A est une M-matrice si et seulement si $\rho(D^{-1}M) < 1$.

Démonstration. En deux morceaux.

Si $\rho(D^{-1}M) < 1$ Nous posons encore $T = D^{-1}M$. Par le théorème 14.115, la matrice $\mathbb{1} - T$ est inversible et

$$(\mathbb{1} - T)^{-1} = \sum_{k=0}^{\infty} T^k. \tag{33.469}$$

D'autre part, via des calculs déjà faits, et les hypothèses sur les signes des éléments de A,

$$T_{ij} = -\frac{A_{ij}}{A_{ii}} \geqslant 0. \tag{33.470}$$

Donc tous les éléments de T sont positifs (ou nuls). Par conséquent $T^k \geqslant 0$ pour tout k et $(\mathbb{1} - T)^{-1}$ est positive.

Mais $A = D - M = D(\mathbb{1} - D^{-1}M) = D(\mathbb{1} - T)$. Vu que D et $\mathbb{1} - T$ sont inversibles, nous savons que A est inversible et

$$A^{-1} = (\mathbb{1} - A)^{-1}D^{-1}, \tag{33.471}$$

qui est un produit de matrices positives. Donc $A^{-1} \geqslant 0$.

Au final, A est une M-matrice.

Si A est une M-matrice Soit une valeur propre λ de $T = D^{-1}M$ est un vecteur propre u : $Tu = \lambda u$. Vu que $T \geqslant 0$ nous avons d'une part $|\lambda u| = |\lambda|u|$ et d'autre part $|\lambda u| = |Tu| \leqslant T|u|$, ce qui donne

$$|\lambda||u| \leqslant T|u|. \tag{33.472}$$

Dans cette inégalité nous substituons T par $\mathbb{1} - (\mathbb{1} - T)$ pour avoir

$$|\mu||u| \leqslant |u| - (\mathbb{1} - T)|u| \tag{33.473}$$

ou encore

$$(\mathbb{1} - T)|u| \leqslant \big(1 - |\lambda|\big)|u|. \tag{33.474}$$

Mais $(\mathbb{1} - T)^{-1} = A^{-1}D \geqslant 0$ parce que A et D sont positives. Donc en appliquant $(\mathbb{1} - T)^{-1}$ à l'inégalité (33.474), elle est conservée (proposition 33.151(2)) :

$$|u| \leqslant (\mathbb{1} - T)^{-1}\big(1 - |\lambda|\big)|u|. \tag{33.475}$$

Si $|\lambda| \geqslant 1$ alors toutes les composantes de $\big(1 - |\mathbb{1}|\big)|u|$ sont négatives et l'inégalité n'est possible qu'avec $|u| = 0$. Dans ce cas, λ n'est pas une valeur propre (le vecteur propre soit être non nul).

Nous en déduisons que $|\lambda| < 1$ et donc que $\rho(T) = \rho(D^{-1}M) < 1$.

□

Le théorème suivant résume ce que nous avons vu en donnant une condition suffisante facile à vérifier pour être une M-matrice.

Théorème 33.163.
Soit $A \in \mathbb{M}(n, \mathbb{R})$ telle que

(1) $A_{ii} > 0$

(2) $A_{ij} \leqslant 0$ *pour* $i \neq j$

(3) vérifiant une des deux conditions suivantes :

— *à diagonale strictement dominante*

— *à diagonale dominante et irréductible.*

Alors A est une M-matrice.

Démonstration. Au vu de la proposition 33.162, il suffira de montrer que $\rho(D^{-1}M) < 1$ où D et M sont la décomposition $A = D - M$ habituelle. C'est le cas grâce à la proposition 33.160. □

Proposition 33.164.
Soit $A \in \mathbb{M}(n, \mathbb{R})$, une M-matrice irréductible. Alors $A^{-1} > 0$.

Démonstration. Nous posons $T = D^{-1}M$. En comparant la définition 33.161 de M-matrice et la caractérisation de la proposition 33.162, nous avons $\rho(D^{-1}M) < 1$. Par conséquent

$$(\mathbb{1} - T)^{-1} = \sum_{k=0}^{\infty} T^k \tag{33.476}$$

par la proposition 33.162. D'autre part, $A^{-1} = (\mathbb{1} - A)^{-1}D$ où les éléments D sont strictement positifs. Donc nous devons encore prouver que $(\mathbb{1} - T)^{-1} > 0$. Nous savons que $T \geqslant 0$, et vu que

$$(\sum_k T^k)_{ij} = \sum_k (T^k)_{ij} \tag{33.477}$$

il nous suffit de prouver que pour chaque (ij), un des $(T^k)_{ij}$ est strictement positif. Soient donc deux indices i et j. Vu que A est irréductible, ils sont connectés par une suite d'indice $i = i_0, i_1, \dots, , i_r = j$ tels que

$$T_{i_k, i_{k+1}} = -\frac{A_{i_k, i_{k+1}}}{A_{i_k, i_k}} > 0. \tag{33.478}$$

Or les indices i_k sont choisis de telle sorte que les numérateurs soient non nuls et donc strictement négatifs. Nous avons, en général :

$$(T^k)_{ij} = \sum_{l_1, \dots, l_{r-1}} T_{i,l_1} T_{l_1,l_2} \dots T_{l_{r-1},j}. \tag{33.479}$$

Chacun des termes est positif ou nul, mais pour $k = r$, il y a entre autres le terme

$$T_{i,i_1} T_{i_1,i_2} \cdots T_{i_r,j} \neq 0. \tag{33.480}$$

Donc $(T^r)_{ij} > 0$ et $\sum_{k=0}^{\infty} (T^k)_{ij} > 0$. Et par conséquent

$$A^{-1} = (\mathbb{1} - T)^{-1} D > 0. \tag{33.481}$$

□

Théorème 33.165.
Soit une M-matrice $A \in \mathbb{M}(n, \mathbb{R})$ et $g \in \mathbb{R}^n$ tel que $(Ag)_i \geqslant 1$ pour tout i. Alors $\|A^{-1}\|_\infty \leqslant \|g\|_\infty$.

Démonstration. Nous posons $u = (1, \dots, 1)$ et considérons $x \in \mathbb{R}^n$. Vu que A est une M-matrice, nous avons $A^{-1} \geqslant 0$, donc

$$|A^{-1}x| \leqslant A^{-1}|x| \leqslant \|x\|_\infty A^{-1}u \leqslant \|x\|_\infty g. \tag{33.482}$$

Justifications :

— La première inégalité est la proposition 33.151(2).

— La seconde provient de

$$\big(B|x|\big)_i = \sum_k B_{ik}|x_k| \leqslant \sum_k B_{ik}\|x\|_\infty = \|x\|_\infty \sum_k B_{ik}u_k = \|x\|_\infty Bu. \tag{33.483}$$

— Étant donné que $A^{-1} \geqslant 0$ nous conservons l'inégalité et $Ag \geqslant u$ implique $g \geqslant A^{-1}u$ (c'est la proposition 33.151(1)).

En ce qui concerne la norme de A^{-1} nous avons donc

$$\|A^{-1}\|_\infty = \sup_{|x|_\infty=1} \|A^{-1}x\|_\infty \leqslant \sup_{\|x\|_\infty=1} \|x\|_\infty \|g\|_\infty = \|g\|_\infty. \tag{33.484}$$

□

Proposition 33.166.
Une matrice de $\mathbb{M}(n,\mathbb{R})$ qui

(1) est symétrique,

(2) Vérifie une des deux conditions suivantes

— *est irréductible à diagonale fortement dominante*

— *est à diagonale strictement dominante,*

(3) vérifie $A_{ii} > 0$ pour tout i

est strictement définie positive.

Démonstration. D'après le théorème de Gershgorin 33.153, chaque valeur propre de A est dans un des disques fermés

$$D_i = \{z \in \mathbb{C} \text{ tel que } |z - A_{ii}| \leqslant r_i\}. \tag{33.485}$$

Par hypothèse, les centres de ces disques sont réels et strictement positifs. Mais le fait que A soit à diagonale dominante donne que le rayon de ces cercles sont plus petits que A_{ii}. Donc D_i n'intersecte pas $]-\infty,0[$. Mais le fait que A soit symétrique implique que les valeurs propres soient réelles (théorème 9.150(1)). Cela montre que les valeurs propres de A sont toutes dans $[0,\infty[$.

Si la matrice A est à diagonale strictement dominante, alors les inégalités sont strictes et le théorème est prouvé.

Sinon nous somme dans le cas irréductible à diagonale fortement dominante et nous avons le théorème de Gershgorin numéro 2 33.154. Soit une valeur propre λ. Soit elle est dans un des disques ouvert (qui est inclus dans $]0,\infty[$), soit elle est dans l'intersection des bords des disques. Mais au moins un des disques n'intersecte pas 0 (parce que la diagonale est strictement dominante). Dans ce cas non plus λ ne peut pas être nul.

Nous en déduisons que dans tous les cas, les valeurs propres sont toutes réelles strictement positives. □

Chapitre 34

Méthode des différences finies

34.1 Problèmes de dimension un

Soit $u\colon \mathbb{R} \to \mathbb{R}$ une fonction et $h > 0$. Nous définissons les opérations suivantes (qui sont supposées approximer la dérivée $u'(x)$ lorsqu'elle existe).

Définition 34.1.
*La **différence progressive** est*

$$(D_h^+ u)(x) = \frac{u(x+h) - u(x)}{h}, \tag{34.1}$$

*la **différence régressive** est*

$$(D_h^- u)(x) = \frac{u(x) - u(x-h)}{h}, \tag{34.2}$$

*la **différence centrée** est*

$$(D_h^0 u)(x) = \frac{u(x+h) - u(x-h)}{2h}. \tag{34.3}$$

Nous ne noterons pas toujours la dépendance en h, c'est-à-dire que nous noterons D^+u au lieu de $D_h^+ u$ lorsque cela ne pose pas de problèmes.

Notons que u'' peut être approximé par D^+D^+u, D^0D^+, D^+D^-, et encore de nombreuses autres possibilités.

Voici un lemme qui dit que tout cela n'est pas si mal, pourvu que u soit assez régulière.

Lemme 34.2.
Soit un ouvert connexe Ω de $\mathbb{R}$, soit $x \in \Omega$ et $h > 0$ tel que $\overline{B(x,h)} \subset \Omega$.

(1) Si $u \in C^2(\Omega)$ alors

$$|u'(x) - D^+u(x)| \leqslant \frac{h}{2}\|u''\|_{\bar{\Omega}} \tag{34.4}$$

et

$$|u'(x) - D^-u(x)| \leqslant \frac{h}{2}\|u''\|_{\bar{\Omega}}. \tag{34.5}$$

(2) Si $u \in C^3(\bar{\Omega})$ alors

$$|u'(x) - D^0(x)| \leqslant \frac{h^2}{2}\|u^{(3)}\|_{\bar{\Omega}} \tag{34.6}$$

(3) Si $u \in C^4(\bar{\Omega})$ alors

$$|u''(x) - D^-D^+u(x)| \leqslant \frac{h^2}{12}\|u^{(4)}\|_{\bar{\Omega}}. \tag{34.7}$$

Démonstration. Nous prouvons le point (3). D'abord nous regardons de quoi nous avons besoin :

$$D^-D^+u(x) = \frac{(D^+u)(x) - (D^+u)(x-h)}{h} = \frac{u(x+h) - 2u(x) + u(x-h)}{h^2} \tag{34.8}$$

Nous allons y mettre les approximations de $u(x+h)$ et $u(x-h)$ par Taylor, proposition 11.386 :

$$u(x+h) = u(x) + hu'(x) + \frac{h^2}{2}u''(x) + \frac{h^3}{6}u'''(x) + \frac{h^4}{24}u^{(4)}(x+\theta_1 h) \tag{34.9}$$

avec $\theta_1 \in [0,1]$. De même,

$$u(x-h) = u(x) - hu'(x) + \frac{h^2}{2}u''(x) - \frac{h^3}{6}u'''(x) + \frac{h^4}{24}u^{(4)}(x-\theta_2 h) \tag{34.10}$$

avec $\theta_2 \in [0,1]$.

Donc

$$u(x+h) + u(x-h) - 2u(x) = h^2u''(x) + \frac{h^4}{4!}\Big(u^{(4)}(x+\theta_1 h) + u^{(4)}(x-\theta_2 h)\Big), \tag{34.11}$$

ce qui donne

$$(D^-D^+u)(x) = u''(x) + \frac{h^2}{4!}\Big(u^{(4)}(x+\theta_1 h) + u^{(4)}(x-\theta_2 h)\Big). \tag{34.12}$$

Chacun des deux termes dans la parenthèse peut être majoré par $\|u^{(4)}\|_{\bar\Omega}$. Notons que c'est une majoration très sauvage parce que $x+\theta_1 h$ ne prend ses valeurs que dans $[x, x+h]$ avec h supposé petit. Quoi qu'il en soit nous ne pouvons pas dire mieux que

$$|u''(x) - D^-D^+u(x)| \leqslant \frac{h^2}{12}\|u^{(4)}\|_{\bar\Omega}. \tag{34.13}$$

□

Remarque 34.3 ([1]).
Lorsque nous écrivons

$$|u'(x) - D^+u(x)| \leqslant \delta \tag{34.14}$$

pour tout x, nous ne pouvons pas écrire

$$\|u' - D^+u\|_\infty \leqslant \delta \tag{34.15}$$

parce que l'inégalité (34.14) n'est valable que pour les x tels que $[x-h, x+h] \subset \Omega$, de telle sorte que l'inégalité n'est pas spécialement correcte sur $\bar\Omega$.

34.1.1 Un schéma à cinq points

34.1.1.1 Poser le système

Soit $\Omega =]0,1[$ et l'équation différentielle

$$\begin{cases} -u''(x) + c(x)u(x) = f \quad \text{sur } \Omega \\ u(0) = \alpha \\ u(1) = \beta \end{cases} \tag{34.16}$$

où c est une fonction positive et $\alpha, \beta \in \mathbb{R}$. Nous considérons $h > 0$ assez petit pour que le reste ait un sens. Si nous cherchons des solutions dans $C^4(\bar\Omega)$, le lemme 34.2 nous dit que

$$|u''(x) - D^-D^+u(x)| = \eta(h^2) \tag{34.17}$$

où η est une fonction telle que $\lim_{t\to 0}\eta(t) = 0$. Nous pouvons récrire l'équation différentielle sous la forme

$$-D^-D^+u(x) + c(x)u(x) = f(x) + \eta(h^2). \tag{34.18}$$

Si nous négligeons le terme $\eta(h^2)$ qui est supposé être petit nous pouvons tenter de résoudre pour la fonction u_h

$$-D^-D^+u_h(x) + c(x)u_h(x) = f(x). \tag{34.19}$$

Notons ici l'importance de la notion de problème bien posé parce que en remplaçant le paramètre (fonctionnel) f par $f + \eta(h^2)$, nous modifions les solutions. Dans la mesure où le problème est bien posé, cette petite modification ne modifiera pas trop la solution et nous pouvons espérer que $\|u - u_h\|$ soit petit pour une norme ou une autre.

Utilisant l'expression (34.8) pour D^-D^+ nous avons l'équation suivante pour u_h :

$$\frac{1}{h^2}\Big(2u_h(x) - u_h(x+h) - u_h(x-h)\Big) + c(x)u_h(x) = f(x). \tag{34.20}$$

Avons-nous gagné quelque chose ? Pas encore. L'idée de la discrétisation est de ne considérer u_h qu'en certains points, et de prendre ces points à intervalles réguliers de taille h. Soient donc N un nombre entier et $h = 1/(N+1)$. Nous posons

$$x_k = kh \tag{34.21}$$

pour $i = 0, \dots, N+1$. Avec cela nous avons

$$\overline{\Omega} = \bigcup_{k=0}^{N-1} [x_k, x_{k+1}] \tag{34.22a}$$

$$x_0 = 0 \tag{34.22b}$$

$$x_{N+1} = 1. \tag{34.22c}$$

Nous posons surtout

$$\Omega_h = \{x_i\}_{i=1,\dots,N} \tag{34.23}$$

et

$$\bar{\Omega}_h = \{x_i\}_{i=0,\dots,N+1}. \tag{34.24}$$

Enfin, nous ne considérons plus u_h que comme une fonction $u_h \colon \bar{\Omega}_h \to \mathbb{R}$. C'est-à-dire que u_h est un vecteur à $N+2$ composantes.

L'équation (34.20) devient

$$\frac{1}{h^2}\big(2u_h(x_i) - u_h(x_{i+1}) - u_h(x_{i-1}) + c(x_i)u_h(x_i)\big) = f(x_i) \tag{34.25}$$

pour $i = 1, \dots, N$. Sur les bords, cette équation n'est pas possible parce que x_{i-1} ou x_{i+1} n'existerait pas. Au contraire, sur les bords nous avons les conditions aux bords

$$u_h(x_0) = \alpha \tag{34.26}$$

et

$$u_h(x_{N+1}) = \beta. \tag{34.27}$$

En posant $c_i = c(x_i)$ et $u_i = u_h(x_i)$, les inconnues du problème sont les nombres u_i ($i = 1, \dots, N+1$). Elles sont immédiatement résolues pour u_0 et u_{N+1}. Pour les autres, il faut écrire et résoudre un système d'équation linéaire.

L'écriture du système linéaire à résoudre consiste essentiellement à écrire (34.25) en séparant les cas $i = 0$ et $i = N$ parce que nous savons déjà les valeurs de u_0 et u_N. Le système que nous avons est :

$$\begin{cases} \left(\dfrac{2}{h^2} + c_1\right)u_1 - \dfrac{1}{h^2}u_2 = f_1 + \dfrac{\alpha}{h^2} & i = 1 \quad (34.28a) \\ \left(\dfrac{2}{h^2} + c_N\right)u_N - \dfrac{1}{h^2}u_{N-1} = f_N + \dfrac{\beta}{h^2} & i = N \quad (34.28b) \\ \left(\dfrac{2}{h^2} + c_i\right)u_i - \dfrac{1}{h^2}u_{i+1} - \dfrac{1}{h^2}u_i = f_i. & \text{autres} \quad (34.28c) \end{cases}$$

Cela se met sous la forme matricielle

$$L_h U_h = F_h \tag{34.29}$$

pour

$$F_h = \left(f_1 + \frac{\alpha}{h^2}, f_2, \dots, f_{N-1}, f_N + \frac{\beta}{h^2}\right) \tag{34.30}$$

et les éléments non nuls de L_h sont :

$$(L_h)_{i,i-1} = -\frac{1}{h^2} \qquad \text{pour } i = 2, \dots, N \tag{34.31a}$$

$$(L_h)_{i,i+1} = -\frac{1}{h^2} \qquad \text{pour } i = 1, \dots, N-1 \tag{34.31b}$$

$$(L_h)_{i,i} = \frac{2}{h^2} + c_i \qquad \text{pour } i = 1, \dots, N. \tag{34.31c}$$

Cette matrice est pleine de zéros, à part les trois diagonales centrales, et il existe des méthodes efficaces pour résoudre le système d'équation correspondant.

34.1.1.2 Propriétés du système

La matrice est la suivante :

$$L_h = \begin{pmatrix} \frac{2}{h^2} + c_1 & -1/h^2 & 0 & 0 & \cdots & 0 \\ -1/h^2 & \frac{2}{h^2} + c_2 & -1/h^2 & 0 & \cdots & 0 \\ 0 & -1/h^2 & \frac{2}{h^2} + c_3 & -1/h^2 & \ddots & \vdots \\ 0 & 0 & -1/h^2 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & -1/h^2 \\ 0 & 0 & \cdots & 0 & -1/h^2 & \frac{2}{h^2} + c_N \end{pmatrix} \tag{34.32}$$

où nous avions déjà posé l'hypothèse $c_i \geqslant 0$ pour tout i.

Lemme 34.4.
La matrice L_h est irréductible[1] *à diagonale fortement dominante*[2].

Démonstration. Nous décomposons la preuve en plusieurs parties.

La première ligne Sur la première ligne, seuls deux éléments sont non nuls et nous avons

$$L_{11} = \frac{2}{h^2} + c_1 \geqslant \frac{2}{h^2} > \frac{1}{h^2} = L_{12}. \tag{34.33}$$

La dernière ligne Elle est semblable à la première.

Les autres lignes Sur les autres lignes nous avons trois éléments non nuls et

$$\sum_{j \neq i} |A_{ij}| = \frac{2}{h^2} \leqslant \frac{2}{h^2} + c_i = L_{ii}. \tag{34.34}$$

Diagonale fortement dominante Nous avons prouvé jusqu'à présent que L_h était une matrice à diagonale fortement dominante.

Irréductible Nous allons utiliser la caractérisation de la proposition 33.150(3). Pour cela nous considérons le chaine d'éléments non nuls

$$A_{12}, A_{23}, \dots, A_{N-1,N} = -\frac{1}{h^2}. \tag{34.35}$$

Soient deux indices i et j avec $i < j$. Cette suite d'indice (ou un sous-suite) rend i et j connectés.

Si par contre $i > j$, il faut considérer la suite inversée grâce au fait que L_h est symétrique :

$$A_{N,N-1}, A_{N-1,N-2}, \dots, A_{32}, A_{21} = -\frac{1}{h^2}. \tag{34.36}$$

1. Caractérisation 33.150.
2. Définition 33.157.

□

Proposition 34.5.
Soit le problème

$$\begin{cases} -u''(x) + c(x)u(x) = f \quad \text{sur } \Omega \\ u(0) = \alpha \\ u(1) = \beta \end{cases} \tag{34.37}$$

où c est une fonction positive et $\alpha, \beta \in \mathbb{R}$. Nous considérons $h > 0$ assez petit pour que le reste ait un sens. Et nous approximons u'' par D^-D^+u.

La matrice L_h des différences finies associée à ce problème est

(1) une M-matrice,

(2) strictement définie positive,

(3) d'inverse $L_h^{-1} > 0$.

Démonstration. Le théorème 33.163 dit que L_h est une M-matrice. La Proposition 33.166 nous donne aussi que L_h est strictement définie positive.

Le lemme 34.4 dit que L_h est irréductible, ce qui permet à la proposition 33.164 de conclure que $L_h^{-1} > 0$. □

Cela étant rappelé, nous pouvons continuer.

Lemme 34.6.
Soit $\Omega =]0, 1[$, soit $N \in \mathbb{N}$ et $h = 1/(N+1)$. La solution $w_h \colon \Omega_h \to \mathbb{R}$ du problème discrétisé

$$\begin{cases} -(D^-D^+w_h)(x_i) = 1 & \text{(34.38a)} \\ w_h(0) = 0 & \text{(34.38b)} \\ w_h(1) = 0 & \text{(34.38c)} \end{cases}$$

pour tout $x_i = ih$ $(i = 1, \dots, N)$ donne les valeurs exactes des $w(x_i)$ lorsque w est la solution de

$$\begin{cases} -w''(x) = 1 & \text{(34.39a)} \\ w(0) = 0 & \text{(34.39b)} \\ w(1) = 0. & \text{(34.39c)} \end{cases}$$

Démonstration. Un enseignement de la proposition 34.5 est que le système (34.38) peut être écrit sous la forme d'un système linéaire $L_h^0 w_h = F_h$ où L_h^0 est inversible. Il y a donc unicité de la solution.

D'autre part, la solution du système (34.39) est $w(x) = \frac{1}{2}(x - x^2)$, qui est de classe C^∞. Le lemme 34.2(3) dit que $D^-D^+w = w''$. Donc les valeurs $w(x_i)$ résolvent aussi le système (34.38). □

Lemme 34.7 (Quelques estimations)**.**
La matrice L_h du problème sus-mentionné en (34.37) vérifie[3] *:*

(1) $\|L_h\|_\infty \leqslant \frac{4}{h^2} + \|c\|_\infty$

(2) $\|L_h^{-1}\|_\infty \leqslant \frac{1}{8}$.

Démonstration. Nous nous souvenons de la formule (10.8) :

$$\|A\|_\infty = \max_{i=1,\dots,n} \sum_{j=1}^{n} |A_{ij}|. \tag{34.40}$$

La première ligne a pour somme : $\frac{3}{h^2} + c_1$, la dernière a pour somme $\frac{3}{h^2} + c_n$ et les autres sont pour somme $\frac{4}{h^2} + c_i$. Elles sont donc toutes majorées par $\frac{4}{h^2} + \|c\|_\infty$.

3. Dans le CTES d'analyse numérique de Marseille, l'estimation donnée est $\|L_h^{-1}\|_\infty \leqslant \frac{1}{4}$.

Pour l'estimation de $\|L_h^{-1}\|_\infty$ nous allons nous appuyer sur le théorème 33.165.

Commençons par considérer le problème

$$\begin{cases} -w'' = 1 & (34.41a) \\ w(0) = w(1) = 0. & (34.41b) \end{cases}$$

La première équation dit que w est un polynôme de degré 2. En écrivant $w(x) = ax^2 + bx + c$ et en imposante toutes les contraintes, nous trouvons l'unique solution

$$w(x) = -\frac{1}{2}(x^2 - x). \tag{34.42}$$

Le lemme 34.6 nous dit que la fonction w prise aux points $x_i = ih$ donne les valeurs de w_h.

La matrice L_h^0 est une M-matrice et le vecteur w_h vérifie $L_h^0 w_h = \mathbb{1}$. Donc le théorème 33.165 s'applique et

$$\|(L_h^0)^{-1}\| \leqslant \|w_h\|_\infty = \frac{1}{8}. \tag{34.43}$$

L'obtention de 1/8 n'est rien d'autre que la recherche du maximum (en valeur absolue) de la parabole $x \mapsto (x - x^2)/2$ pour $x \in [0, 1]$. Le maximum est atteint pour $x = 1/2$; calcul de dérivée et tout ça ...

Nous retournons maintenant à notre matrice originale L_h. Nous avons

$$L_h - L_h^0 = \operatorname{diag}(c_1, \dots, c_n) \geqslant 0, \tag{34.44}$$

et aussi

$$L_h^{-1} - (L_h^0)^{-1} = \underbrace{L_h^{-1}}_{\geqslant 0} \underbrace{(L_h^0 - L_h)}_{\leqslant 0} \underbrace{(L_h^0)^{-1}}_{\geqslant 0} \tag{34.45}$$

parce que L_h est une M-matrice. Donc tous les coefficients de $L_h^{-1} - (L_h^0)^{-1}$ sont négatifs. Cela implique

$$L_h^{-1} \leqslant (L_h^0)^{-1}. \tag{34.46}$$

Mais nous savons que les coefficients de L_h^{-1} sont positifs, donc le maximum de ses coefficients en valeur absolue est plus petit que ceux de $(L_h^0)^{-1}$, c'est-à-dire

$$\|L_h^{-1}\|_\infty \leqslant \|(L_h^0)^{-1}\|_\infty \leqslant \frac{1}{8}. \tag{34.47}$$

□

34.1.2 Exemple

Soit $\Omega =]0, 1[$ et une fonction $u \colon \bar{\Omega} \to \mathbb{R}$ de classe C^4 vérifiant

$$\begin{cases} -u''(x) + u(x) = \sin(x) & (34.48a) \\ u(0) = 0 & (34.48b) \\ u(1) = 0. & (34.48c) \end{cases}$$

Nous allons écrire la méthode des différences finies pour $h = 1/4$. Nous posons donc les points

$$\begin{cases} x_0 = 0 & (34.49a) \\ x_1 = 1/4 & (34.49b) \\ x_2 = 1/2 & (34.49c) \\ x_3 = 3/4 & (34.49d) \\ x_4 = 1. & (34.49e) \end{cases}$$

Vu que nous avons supposé u de classe C^4, le lemme 34.2(3) nous donne [4]

$$u''(x) = (D^-D^+u)(x) + \alpha(h) \tag{34.50}$$

avec $\lim_{h\to 0} \alpha(h)/h = 0$. L'équation discrétisée serait alors

$$\begin{cases} -(D^-D^+u)(x) + u(x) = \sin(x) & \text{(34.51a)} \\ u(0) = u(1) = 0. & \text{(34.51b)} \end{cases}$$

où nous n'avons pas précisé l'indice h au bas des opérateurs D^+ et D^-. Les équations (34.51) ne doivent être posées que pour x_1, x_2 et x_3 parce que les valeurs en x_0 et x_4 sont déjà connues.

Pour x_1

$$\frac{u_2 - 2u_1 + u_0}{h^2} + u_1 = \sin(x_1) \tag{34.52}$$

Pour x_2

$$\frac{u_3 - 2u_2 + u_1}{h^2} + u_2 = \sin(x_2) \tag{34.53}$$

Pour x_3

$$\frac{u_4 - 2u_3 + u_2}{h^2} + u_3 = \sin(x_3). \tag{34.54}$$

Nous tenons compte du fait que $u_0 = u_4 = 0$ et que $h = 1/4$ pour écrire le système

$$\begin{pmatrix} -31 & 16 & 0 \\ 16 & -31 & 16 \\ 0 & 16 & -31 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} \tag{34.55}$$

A où les s_i sont des nombres parfaitement connus : par exemple $s_1 = \sin(x_1) = \sin(1/4) \simeq 0.247403959254523$.

34.2 Problèmes de dimension deux

Nous allons considérer le système

$$\begin{cases} -\Delta u = f & \text{sur } \Omega \\ u = g & \text{sur } \partial\Omega \end{cases} \tag{34.56}$$

où $\Omega =]0, a[\times]0, b[$.

Remarque 34.8.
Pourquoi un signe moins devant le laplacien ? Pour avoir la proposition 34.5 qui dira que la matrice correspondant aux différences finies appliquées à ce système est une M-matrice. Sinon, c'est la matrice $-L_h$ qui en serait une.

34.2.1 Discrétisation en croix

Nous allons maintenant déduire une discrétisation du laplacien en discrétisant les opérations ∂_x^2 et ∂_y^2. Nous discrétisons Ω en mailles carrés de côté h : $x_k = kkh$ et $y_k = kh$. L'opération de dérivée partielle ∂_x est discrétisée par

$$(D_x^+u)(x, y) = \frac{u(x + h) - u(x, y)}{h} \tag{34.57}$$

ou

$$(D_x^-u)(x, y) = \frac{u(x, y) - u(x - h, y)}{h} \tag{34.58}$$

4. Nous ferions n'importe quoi pour ne pas écrire $u''(x) = (D^-D^+u)(x) + o(h^2)$. Notez que vous faites ce que vous voulez : écrivez avec la notation « petit o » si cela vous chante.

ou

$$(D_x^0 u)(x,y) = \frac{u(x+h,y) - u(x-h,y)}{2h} \tag{34.59}$$

où le h est sous-entendu dans les opérateurs D^0, D^+ et D^-.

La dérivée partielle seconde $\partial_x^2 u$ peut être approximée par toutes les combinaisons imaginable, par exemple

$$(D_x^- D_x^+ u)(x,y) = \frac{u(x+h,y) - 2u(x,y) + u(x-h,y)}{h^2}. \tag{34.60}$$

Pour évaluer la différence entre $(\partial_x^2 u)(x,y)$ et $(D^- D^+ u)(x,y)$, il est possible de faire du Taylor en deux dimensions, mais nous pouvons également recycler ce qui a été fait. Nous posons $u_y(x) = u(x,y)$ et alors $(\partial_x^2 u)(x,y) = u_y''(x)$ et le lemme 34.2(3) donne, si u_y est de classe C^4,

$$|u_y''(x) - D^- D^+ u_y(x)| \leqslant \frac{1}{12} h^2 \|u_y^{(4)}\|_\infty. \tag{34.61}$$

Là, les opérateurs D^+ et D^- sont ceux à une dimension. Mais nous avons $(D^- D^+ u_y)(x) = (D^- D^+ u)(x,y)$ (à droite ce sont les opérateurs à deux dimensions), donc

$$\big|(\partial_x^2 u)(x,y) - (D^- D^+ u)(x,y)\big| \leqslant \frac{1}{12} h^2 \|\partial_x^4 u\|_\infty \tag{34.62}$$

et nous pouvons écrire

$$(\partial_x^2 u)(x,y) = (D^- D^+ u)(x,y) + h^2 R(x,y,h) \tag{34.63}$$

où R est une fonction qui dépend de x, y et h, mais aussi de u. Le point important est que R soit majoré par une quantité indépendante de h, de telle sorte que nous ayons quelques garanties que négliger ce terme soit une bonne approximation lorsque $h \to 0$.

Au niveau de la discrétisation, nous considérons x_i avec $i = 0, \dots, N_x$ et y_j avec $j = 0, \dots, N_y$. La discrétisation de $-(\Delta u)(x,y) = f(x,y)$ donne, pour $i = 1, \dots, N_x - 1$ et $j = 1, \dots, N_y - 1$,

$$\frac{1}{h^2}(-u_{i+1,j} + 4u_{ij} - u_{i-1,j} - u_{i,j+1} + u_{i,j-1}) = f_{ij}. \tag{34.64}$$

Les équations avec i ou j valant 0 ou N_x, N_y sont les valeurs au bords.

34.9.
Nous notons pour référence ultérieure la discrétisation suivante du laplacien :

$$(\Delta_h u)(x_i, y_j) = \frac{1}{h^2}\big(-4u_{i,j} + u_{i+1,j} + u_{i-1,j} + u_{i,j-1} + u_{i,j+1}\big). \tag{34.65}$$

Elle vérifie

$$\Delta_h f = \Delta f + h^2 \alpha(h). \tag{34.66}$$

Cette discrétisation est dite « en croix » parce que les points exploités forment une croix.

34.2.2 Discrétisation en carré

L'opérateur laplacien est défini par $\Delta = \partial_x^2 + \partial_y^2$, mais il existe de nombreuses autres façons de l'écrire.

Lemme 34.10.
Le laplacien est invariant par changement de coordonnées orthogonales. Plus précisément, si A est une matrice orthogonale, en posant $u_i = \sum_k A_{ik} e_k$ nous avons

$$\sum_i \frac{\partial^2 f}{\partial u_i^2} f = \Delta f. \tag{34.67}$$

Démonstration. Nous avons :

$$\frac{\partial f}{\partial u_i} = \sum_k A_{ik} \frac{\partial f}{\partial x_k}, \tag{34.68}$$

et donc

$$\sum_i \frac{\partial^2 f}{\partial u_i^2} = \sum_i \frac{\partial}{\partial u_i} \left(\sum_k A_{ik} \frac{\partial f}{\partial x_k} \right) = \sum_{ijk} (A^t)_{ji} A_{ik} \partial^2 f = \sum_{jk} (A^t A)_{jk} \partial^2_{jk} f. \tag{34.69}$$

En particulier si A est une matrice orthogonale, $(A^t A)_{jk} = \delta_{jk}$ et le résultat est prouvé. □

Les plus convaincus diront que $\Delta = \nabla \cdot \nabla$ et que le produit scalaire est invariant sous changement de coordonnées orthogonales.

Nous avons déjà déduit la discrétisation (34.65) du laplacien :

$$(\Delta_h u)(x_i, y_j) = \frac{1}{h^2} \big(-4u_{i,j} + u_{i+1,j} + u_{i-1,j} + u_{i,j-1} + u_{i,j+1} \big). \tag{34.70}$$

Nous allons maintenant en déduire une par l'idée de décomposer le laplacien dans la base $u = e_1 + e_2$, $v = e_1 - e_2$. Pour cela nous introduisons les opérations (le nombre h dont dépendent ces opérateurs est sous-entendu)

$$(D_u^+ f)(x, y) = \frac{f(x+h, y+h) - f(x, y)}{h} \tag{34.71a}$$

$$(D_v^+ f)(x, y) = \frac{f(x+h, y-h) - f(x, y)}{h} \tag{34.71b}$$

$$(D_u^- f)(x, y) = \frac{f(x, y) - f(x-h, y-h)}{h} \tag{34.71c}$$

$$(D_v^- f)(x, y) = \frac{f(x, y) - f(x-h, y+h)}{h}. \tag{34.71d}$$

Vu que

$$\partial_u^2 + \partial_v^2 = 2\Delta, \tag{34.72}$$

nous discrétisons le laplacien par

$$\Delta'_h = \frac{1}{2} \big(D_u^- D_u^+ + D_v^- D_v^+ \big). \tag{34.73}$$

Un peu de calcul donne :

$$(\Delta'_h f)(x, y) = \frac{1}{2h^2} \Big(-4f(x, y) + f(x+h, y+h) + f(x-h, y+h) \tag{34.74a}$$

$$+ f(x+h, y-h) + f(x-h, y-h) \Big). \tag{34.74b}$$

34.2.3 Résolution de la discrétisation en croix

Les équations (34.64) forment un système d'équations linéaires à résoudre. Certaines peuvent être simplifiées parce qu'elles « touchent » le bord. Nous verrons cela un peu plus tard.

Nous allons d'abord numéroter correctement les équations de façon à ne pas avoir deux mais un seul indice. Notre fonction de numérotation sera

$$\varphi(i, j) = (j-1)(N_x - 1) + i \tag{34.75}$$

avec $i = 1, \dots, N_x - 1$ et $j = 1, \dots, N_y - 1$. Cela correspond à numéroter les points de l'intérieur du quadrillage ligne par ligne en bas en haut et de gauche à droite. Avec cela les équations (34.64) vont être numérotées par un seul indice I allant de $\varphi(1, 1) = 1$ à $\varphi(N_x - 1, N_y - 1) = (N_x - 1)(N_y - 1)$.

Si $I = \varphi(i,j)$ alors nous avons vite

$$\varphi(i+1,j) = I+1 \tag{34.76a}$$
$$\varphi(i,j+1) = I+N_x-1 \tag{34.76b}$$
$$\varphi(i-1,j) = I-1 \tag{34.76c}$$
$$\varphi(i,j-1) = I-N_x+1. \tag{34.76d}$$

Nous posons $U_I = u_{\varphi^{-1}(I)}$, et l'équation (34.64) devient

$$\frac{1}{h^2}(-U_{I+1}+4U_I-U_{I-1}-U_{I+N_x-1}-U_{I-N_x+1}) = f_I. \tag{34.77}$$

Pour savoir la matrice représentant ce système, nous devons simplifier les équations qui doivent l'être. Par exemple avec $I=1$, le terme $U_{I-1}=U_0$ vaut $u_{0,1}=f_01$. Ce n'est donc pas réellement une inconnue de notre problème.

Nous voulons mettre les équations sous la forme du système

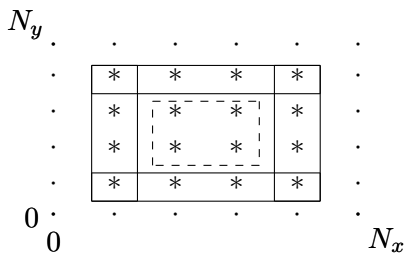
$$L_hU = F. \tag{34.78}$$

Sur la ligne numéro I de L_h, les éléments non nuls sont :

$$L_{I,I} = 4 \tag{34.79a}$$
$$L_{I,I+1} = -1 \tag{34.79b}$$
$$L_{I,I-1} = -1 \tag{34.79c}$$
$$L_{I,I+N_x-1} = -1 \tag{34.79d}$$
$$L_{I,I-N_x+1} = -1 \tag{34.79e}$$

pour peu qu'ils existent. Par exemple pour $I=1$, il n'y a pas d'éléments $L_{I,I-1}$. Les indices I et J de $L_{I,J}$ vont de 1 à $\varphi(N_x-1,N_y-1) = (N_y-1)(N_x-1)$.

Voici un dessin de notre situation :



À chaque élément du quadrillage correspond une équation.

— Aux points simples sur le bord, correspondent des équations triviales parce que la fonction u y est directement donnée par les conditions aux bords.

— Aux points étoilés entourés en traits continus correspondent des équations « incomplètes » parce que certains termes de l'équation (34.64) sont donnés par les conditions aux bords. Elle correspondent aussi aux lignes incomplète de la matrice L_h où certains éléments donnés en (34.79) n'existent pas.

  Le membre de droite de ces équations est par contre enrichi de ce qui à gauche est « donné ».

— Au points étoilés du centre entourés en traits discontinus correspondent des équations complètes.

Notons que f_{00} ne joue aucun rôle dans notre histoire parce que dans les équations (34.64), chaque point (i,j) du maillage n'est liée qu'aux quatre points situés « à côté ».

Proposition 34.11.
La matrice L_h est

(1) irréductible et à diagonale fortement dominante[5]*,*

(2) une M-matrice,

(3) inversible avec $L_h > 0$,

(4) symétrique,

(5) strictement définie positive.

Démonstration. On divise la preuve.

Irréductible Une matrice $n \times n$ dont les deux premières diagonales sont entièrement composées d'éléments non nuls est toujours irréductible. En effet, la première lie l'élément $(1, 2)$ à l'élément $(n-1, n)$ et donc permet de dire que tous les $i < j$ sont connectés.

La seconde diagonale lie l'élément $(n, n-1)$ à l'élément $(2, 1)$.

Diagonale fortement dominante En ce qui concerne la dominance de la diagonale, il faut sommer sur les lignes. Or chaque ligne contient (en valeur absolue) un 4 sur la diagonale et au plus quatre éléments qui valent 1. D'où

$$|L_{II}| \geqslant \sum_{J \neq I} |L_{IJ}|. \tag{34.80}$$

La première ligne n'est jamais complète : elle contient un 4 sur l'élément $(1, 1)$ au au maximum deux 1 plus à droite. Donc la matrice L_h est à diagonale fortement dominante.

M-matrice D'après ce que nous venons de voir (proposition 34.11), le théorème 33.163 fonctionne et L_h est une M-matrice[6].

Inverse strictement positif La proposition 33.164 nous assure qu'une M-matrice irréductible est d'inverse strictement positif. Donc $L_h^{-1} > 0$.

Symétrique La ligne numéro I est

$$\big(\dots, \underbrace{-1}_{I-N_x+1}, \dots, -1, 4, -1, \dots, \underbrace{-1}_{I+N_x-1}, \dots\big) \tag{34.81}$$

Prenons par exemple l'élément $(I, I - N_x + 1)$ qui vaut -1. Son symétrique est l'élément $(I - N_x + 1, I)$ qui se trouve sur la ligne $I - N_x + 1$. Sur cette dernière ligne nous avons un -1 sur la colonne $I - N_x + 1 + N_x - 1 = I$. Donc l'élément $(I - N_x + 1, I)$ vaut bien -1 et la matrice est symétrique.

Strictement définie positive Vu que la matrice L_h est symétrique, irréductible à diagonale fortement dominante (proposition 34.11), vu que ses éléments diagonaux sont strictement positifs (ils valent 4), la proposition 33.166 nous dit que L_h est strictement définie positive.

□

34.3 Consistance, convergence

34.3.1 Définitions, mise en place

Soit un ouvert $\Omega \subset \mathbb{R}^n$ et un opérateur différentiel L sur Ω. Nous considérons le problème qui consiste à trouver une fonction u sur Ω telle que

$$Lu = f \tag{34.82}$$

pour une fonction f donnée.

Problèmes et choses à faire

La définition suivante est une invention personnelle, n'est pas précise et mérite des commentaires de la part du lecteur.

5. Définition 33.157. Le cas 1×1 est discutablement à diagonale fortement dominante, il faut avouer.

6. Notons que c'est ici que nous sommes content d'avoir posé $-\Delta u = f$ dans le système (34.56), avec un signe négatif devant le laplacien. Sinon tous le signes auraient changé et la matrice $-L_h$ aurait été une M-matrice au lieu de L_h.

Définition 34.12 ([1]).
*Un **schéma numérique** de pas h pour $Lu = f$ est la donnée de*

(1) un nombre $h > 0$ supposé petit,

(2) une quantité N de points x_i dans Ω formant l'ensemble discret Ω_h,

(3) une matrice L_h de taille $N \times N$,

(4) une solution $u_h \colon \Omega_h \to \mathbb{R}$ de l'équation $(L_h u_h)(x_i) = f_i$ où nous avons posé $f_i = f(x_i)$.

34.13.
Évidemment pour qu'un schéma mérite le nom de schéma de pas h pour l'équation $Lu = f$, il faut que le nombre h soit lié au choix des points x_i, et que la matrice L_h soit liée à l'opérateur L. La définition n'impose pas formellement de tels liens, parce qu'il y a de nombreuses façons d'approximer une équation différentielle en un système linéaire, sans compter que même l'équation $(L_h u_h)_i = f_i$ peut se résoudre de beaucoup de façons, exacte ou approchées.

Cela pour dire que le lien entre la solution exacte u et la solution approchée n'a rien d'évident, et va dépendre des choix faits lors de la discrétisation et lors de la résolution du système linéaire. Nous allons supposer dans un premier temps que l'équation $L_h u_h = f$ est résolue exactement (nous avons un peu parlé de ces problème dans les sections 33.15 et suivantes).

Définition 34.14.
*L'erreur **de consistance** d'un schéma numérique est la fonction $\tau_h \colon \Omega_h \to \mathbb{R}$ définie par*

$$\tau_h(x_i) = (L_h u)_i - (Lu)(x_i). \tag{34.83}$$

Il y a un jeu de notation pas tout à fait évident dans la définition (34.83). En effet, L_h est une matrice, et ne s'applique donc à priori pas immédiatement à une fonction. Ce que signifie la notation $(L_h u)_i$ est que l'on applique la matrice L_h au vecteur $j \mapsto u(x_j)$ et que l'on prend la composante i du résultat.

Définition 34.15.
*Nous disons que le schéma est **consistant** avec l'opérateur différentiel L lorsque*

$$\lim_{h \to 0^+} \|\tau_h\| = 0 \tag{34.84}$$

où la norme $\|.\|$ est souvent la norme uniforme, c'est-à-dire $\|\tau_h\| = \max_i \tau_h(x_i)$.

Notons que le lien entre h et le choix des x_i fait partie de la définition des schéma. Sur un segment de longueur L, lorsque h n'est pas un diviseur de L, le schéma devrait expliquer ce que l'on fait pour que la limite (34.84) ait un sens.

Définition 34.16.
*Le schéma (Ω_h, L_h) est **consistant à l'ordre** p avec l'opérateur différentiel L pour la norme $\|.\|$ si il existe une constante C indépendante de h telle que*

$$\|\tau_h\| \leqslant C h^p. \tag{34.85}$$

Définition 34.17.
*L'**erreur de discrétisation** entre la solution u du problème $Lu = f$ et la solution approchée u_h sur Ω_h est la fonction*

$$\begin{aligned} e_h \colon \Omega_h &\to \mathbb{R} \\ x_i &\mapsto u(x_i) - u_i. \end{aligned} \tag{34.86}$$

où $u_i = u_h(x_i)$ est la solution approchée.

*Le schéma discret $(L_h u_h)(x_i) = f_i$ est **convergeant** si $\lim_{h \to 0} \|e_h\| = 0$. Si de plus is existe une constante C et $p > 0$ tels que*

$$\|e_h\| \leqslant C h^p, \tag{34.87}$$

*alors nous disons que le schéma est convergence à l'**ordre** p.*

Si l'erreur de consistance est petite, le *problème* est bien approximé par la système linéaire. Cela n'implique cependant pas que la solution trouvée soit bien approximée.

Exemple 34.18 (Deux opérateurs différentiels proches dont les solutions sont loin)**.**
Soit la partie $\Omega =]0,\infty[$, et les problèmes

$$\begin{cases} L_1 u = u' = 0 & (34.88a) \\ u(0) = 1 & (34.88b) \end{cases}$$

et

$$\begin{cases} L_2 v = v' - \epsilon v = 0 & (34.89a) \\ v(0) = 1. & (34.89b) \end{cases}$$

Les solutions exactes sont $u(x) = 1$ et $v(x) = e^{\epsilon x}$.

En ce qui concerne les opérateurs, quelle que soit la norme utilisée nous avons

$$\begin{aligned} \|L_1 - L_2\| &= \sup_{\|f\|=1} \|L_1(f) - L_2(f)\| && (34.90a) \\ &= \sup_{\|f\|=1} \|\epsilon f\| && (34.90b) \\ &= \epsilon. && (34.90c) \end{aligned}$$

Donc lorsque ϵ est petit, l'opérateur L_2 approxime bien l'opérateur L_1. Pour toutes les normes. Mais

$$\big|u(x) - v(x)\big| = |1 - e^{\epsilon x}|, \tag{34.91}$$

donc quel que soit ϵ nous avons $\|u - v\|_\infty = \infty$. Et d'ailleurs, quelle que soit la norme raisonnable que nous mettons sur l'espace des fonctions, avoir $\|u - v\| = \infty$ semble inévitable.

Donc deux opérateurs différentiels proches peuvent avoir des solutions lointaines. △

34.3.2 Exemple

Soit l'opérateur différentiel L donné par

$$Lu = -u'' + cu \tag{34.92}$$

où c est une fonction. Nous considérons sur $\Omega =]0,1[$ l'équation différentielle

$$Lu = 0. \tag{34.93}$$

En ce qui concerne la discrétisation, nous définissons le maillage $\Omega_h = \{x_i = ih\}$ avec $i = 0,\dots,N+1$. La solution approchée discrètement sera le vecteur v qui peut être vu comme fonction $v\colon \Omega_h \to \mathbb{R}$. Les nombres v_0 et v_{N+1} sont à priori donnés par les conditions aux bords. Pour les autres v_i nous avons les équations

$$(L_h v)_i = -\frac{v_{i+1} - 2v_i + v_{i-1}}{h^2} + c(x_i)v_i. \tag{34.94}$$

Cela est la définition de l'opérateur L_h, et le vecteur v solution de $L_h v = 0$ est la solution du problème au sens de la méthode des différences finies (pour peu qu'il existe, soit unique et tout ça).

Pour calculer l'erreur de consistence, nous considérons une fonction u et nous posons $u_i = u(x_i)$. Le vecteur (u_i) ainsi construit est approximé par v (on espère). Nous avons :

$$\tau_h(x_i) = -\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} - c(x_i)u_i - (Lu)(x_i). \tag{34.95}$$

Pour étudier cela nous développons $u_{i+1} = u(x_i + h)$ et $u_{i-1} = u(x_i - h)$ à l'ordre 4 : il existe $\alpha_i \in [x_i, x_i + h]$ et $\beta_i \in [x_i - h, x_i]$ tels que

$$u_{i+1} = u(x_i) + hu'(x_i) + \frac{h^2}{2}u''(x_i) + \frac{h^3}{3!}u^{(3)}(x_i) + \frac{h^4}{4!}u^{(4)}(\alpha_i) \tag{34.96}$$

et

$$u_{i-1} = u(x_i) - hu'(x_i) + \frac{h^2}{2}u''(x_i) - \frac{h^3}{3!}u^{(3)}(x_i) + \frac{h^4}{4!}u^{(4)}(\beta_i). \tag{34.97}$$

Après simplification de plusieurs termes,

$$\tau_h(x_i) = -\frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} - c_i u_i + u''(x_i) + c_i u_i = \frac{h^2}{4!}\big(u^{(4)}(\alpha_i) + u^{(4)}(\beta_i)\big). \tag{34.98}$$

Parler de la consistance du schéma demande d'étudier $\lim_{h\to 0^+} \|\tau_h\|$, et pour cela, il faut préciser la norme avec laquelle nous voulons travailler. L'ordre de consistance va dépendre de la norme utilisée.

Pour la norme $\|.\|_\infty$, les nombres $u^{(4)}(\alpha_i)$ et $u^{(4)}(\beta_i)$ se majorent par $\|u^{(4)}\|_\infty$ et nous avons

$$\|\tau_h\|_\infty \leqslant \frac{h^2}{12}\|u^{(4)}\|_\infty. \tag{34.99}$$

Nous avons consistance d'ordre 2.

Remarque 34.19.
La valeur de $\|\tau_h\|_\infty$ dépend de la fonction u sur laquelle nous la calculons. Cependant nous avons convergence $\|\tau_h\|_\infty \to 0$ pour toute fonction (de classe disons C^4).

La constante C pour laquelle nous avons $\|\tau_h\| \leqslant Ch^2$ et donc qui nous vaut de pouvoir dire que la consistance est d'ordre 2 ne dépend pas de h, ni des valeurs ponctuelles de u ou de ses dérivées, mais dépend des normes de u et de ses dérivées (en l'occurrence seulement de la norme de $u^{(4)}$.)

Étudions la consistance pour la norme L_1 :

$$\|\tau_h\|_1 = \sum_i |\tau_h(x_i)| \leqslant \frac{h^2}{12}\sum_i \|u^{(4)}\|_\infty \tag{34.100}$$

où nous avons majoré chacun des $u^{(4)}(\alpha_i)$ par $\|u^{(4)}\|_\infty$. Combien de termes dans la somme ? Nous avons $h = 1/(N-1)$ et donc $N = (1+h)/h$, ce qui donne

$$\|\tau_h\|_1 \leqslant N\frac{h^2}{12}\|u^{(4)}\|_\infty = (1+h)Ch. \tag{34.101}$$

La constante $1+h$ se majore par n'importe quelle constante strictement plus grande que 1. Nous pouvons donc la rentrer dans C et écrire

$$\|\tau_h\|_1 \leqslant Ch \tag{34.102}$$

et donc avoir la consistance à l'ordre 1.

34.3.3 Consistance, stabilité et convergence

Lemme 34.20.
Soit un opérateur différentiel L, soit u la solution de $Lu = f$ et un schéma numérique (L_h, Ω_h) pour cette équation. Nous notons u_h la solution de $L_h u_h = f$. Alors nous avons

$$L_h e_h = \tau_h \tag{34.103}$$

Et si de plus L_h est inversible,

$$\|e_h\| \leqslant \|L_h^{-1}\|\|\tau_h\|. \tag{34.104}$$

Démonstration. Par définition u_h est solution de $L_h u_h = f$ en tant que fonctions sur Ω_h. Nous avons donc

$$L_h e_h = L_h u_h - L_h u \tag{34.105}$$

où u doit être compris comme la restriction de u à Ω. En appliquant au point x_i,

$$(L_h e_h)(x_i) = \underbrace{(L_h u_h)(x_i)}_{=f_i} - (L_h u)(x_i), \tag{34.106}$$

mais $f_i = (Lu)(x_i)$ parce que u est solution de $Lu = f$. Donc

$$(L_h e_h)(x_i) = (Lu)(x_i) - (L_h u)(x_i) = \tau_h(x_i). \tag{34.107}$$

Si la matrice L_h est inversible nous avons $e_h = L_h^{-1}\tau_h$ et donc

$$\|e_h\| \leqslant \|L_h^{-1}\|\|\tau_h\| \tag{34.108}$$

par le lemme 10.10. □

Bien entendu, en tant qu'opérateur linéaire sur un espace de dimension finie, l'opérateur L_h^{-1} est borné pour chaque h. Mais si il n'y a pas une borne uniforme en h, alors le lemme 34.20 dit qu'il n'y a pas d'espoir de majorer $\|e_h\|$ de façon à passer à la limite $\lim_{h\to 0}\|L_h^{-1}\|$.

Définition 34.21.
Un schéma numérique est ***stable*** *si il existe une constante $C > 0$ indépendante de h telle que $\|L_h^{-1}\| \leqslant C$.*

Théorème 34.22.
En deux parties.

(1) Si un schéma discret est consistant et stable, alors il est convergeant.

(2) Si de plus il est consistant à l'ordre p, alors il est convergeant à l'ordre p.

Démonstration. Nous savons du lemme 34.20 (qui s'applique parce que l'inversibilité de L_h est dans la définition de la stabilité) que $\|e_h\| \leqslant \|L_h^{-1}\|\|\tau_h\|$ et que $\|L_h^{-1}\| \leqslant C$. En passant à la limite[7],

$$\lim_{h\to 0}\|e_h\| \leqslant C \lim_{h\to 0}\|\tau_h\| = 0. \tag{34.109}$$

La dernière limite est le fait que le schéma soit consistant. Le schéma est donc convergeant.

Si de plus il est consistant à l'ordre p, alors

$$\|e_h\| \leqslant C\|\tau_h\| \leqslant C' h^p \tag{34.110}$$

Il est donc également convergeant à l'ordre p. □

34.3.4 Exemple : schéma à cinq points, laplacien en croix

Nous avions développé le schéma dont l'opérateur sur Ω_h est (voir (34.64))

$$(L_h u_h)(x_i, y_j) = \frac{1}{h^2}\big(- u_{i+1,j} - u_{i,j+1} + 4u_{i,j} - u_{i-1,j} - u_{i,j-1}\big). \tag{34.111}$$

Proposition 34.23.
Le schéma est :

(1) consistant à l'ordre 2,

(2) stable pour la norme uniforme et

$$\|L_h^{-1}\|_\infty \leqslant \frac{1}{8}, \tag{34.112}$$

(3) convergeant à l'ordre 2 pour la norme $\|.\|_\infty$.

7. Toutes les limites $h \to 0$ sont en réalité des limites $h \to 0^+$, mais nous allégeons cette notation.

Démonstration. Cet opérateur avait été construit de telle sorte à avoir (voir (34.63))

$$(\Delta u)(x_i, y_j) = (L_h u)(x_i, y_i) + h^2 R(x, y, h) \tag{34.113}$$

où R peut être majoré indépendamment de h. En tant que fonctions sur Ω_h nous avons

$$\tau_h = \Delta u - L_h u = h^2 R(x, y, h), \tag{34.114}$$

et donc $\|\tau_h\|_\infty \leqslant Ch^2$, parce que le lemme 34.2(3) donne aussi

$$\|R\|_\infty \leqslant C \max\{\|\frac{\partial^4 u}{\partial x^4}\|_\infty, \|\frac{\partial^4 u}{\partial y^4}\|_\infty\}. \tag{34.115}$$

En ce qui concerne la stabilité nous allons utiliser le théorème 33.165. Nous considérons la fonction

$$g(x, y) = -\frac{1}{4}(x^2 + y^2), \tag{34.116}$$

qui vérifie $-\Delta g = 1$ sur le carré $[0, 1]^2$. Nous considérons le vecteur g_h d'indices $(i, j) \mapsto g_{ij} = g(x_i, y_j)$ sur lequel nous calculons L_h :

$$(L_h g_h)_{ij} = \frac{1}{h^2}(-g_{i+1,j} - g_{i,j+1} + 4g_{ij} - g_{i-1,j} - g_{i,j-1}); \tag{34.117}$$

en remplaçant les g par leurs valeurs en termes de x_i, x_{i-1}, x_{i+1}, y_j, y_{j-1} et y_{j+1}, et ne tenant compte du fait que $x_k = kh$ et $y_l = lh$ nous avons :

$$(L_h g_h)_{ij} = \frac{1}{4}\big((i+1)^2 - j^2 + i^2 + (j+1)^2 - 4i^2 - 4j^2 + (i-1)^2 + j^2 + i^2 + (j-1)^2\big) = 1. \tag{34.118}$$

Donc $L_h g_h = 1$.

Vu que L_h est une M-matrice (proposition 34.11(2)), le théorème 33.165 nous dit alors que L_h^{-1} vérifie

$$\|L_h^{-1}\|_\infty \leqslant \|g_h\|_\infty. \tag{34.119}$$

Mais

$$\|g_h\|_\infty \leqslant \|g\|_\infty = g\big(\frac{1}{2}, \frac{1}{2}\big) = \frac{1}{8}. \tag{34.120}$$

Notons que cela est bien une inégalité et non une égalité parce que rien n'assure que le point $(1/2, 1/2)$ soit sur le maillage ; donc rien n'assure que la valeur $g(1/2, 1/2)$ ne soit parmi les valeurs du vecteur discrétisé g_h.

Notre schéma numérique est stable et consistant à l'ordre 2 pour la norme $\|.\|_\infty$. Le théorème 34.22 dit alors que le schéma est convergeant à l'ordre 2 pour la même norme. □

34.4 Autres laplaciens

Nous avons vu le laplacien en croix (34.65)

$$(\Delta_h f)(x, y) = \frac{1}{h^2}\big(-4f(x, y) + f(x + h, y) + f(x - h, y) \tag{34.121a}$$

$$+ f(x, y + h) + f(x, y - h)\big) \tag{34.121b}$$

qui vérifie

$$\Delta_h f = \Delta f + Kh^2, \tag{34.122}$$

ainsi que le laplacien en carré (34.74)

$$(\Delta'_h f)(x, y) = \frac{1}{2h^2}\Big(-4f(x, y) + f(x + h, y + h) + f(x - h, y + h) \tag{34.123a}$$

$$+ f(x + h, y - h) + f(x - h, y - h)\Big) \tag{34.123b}$$

qui vérifie également $\Delta'_h f = \Delta f + Kh^2$.

À priori toutes combinaisons de la forme

$$a\Delta_h + b\Delta'_h \tag{34.124}$$

avec $a + b = 1$ est valable comme tentative de discrétiser le laplacien. Ce sont des schémas à 9 points. Évidemment la matrice L_h correspondante va être moins creuse, mais nous pouvons espérer ajuster a et b de telle sorte à obtenir une consistance d'un ordre supérieur à 2.

Nous allons développer les $\Delta_h f$ et $\Delta'_h f$ à l'ordre 4 (reste à l'ordre 6). Quelques remarques avant de commencer.

(1) Allez relire la proposition 11.289 et les notations qui vont avec pour comprendre les différentielles.

(2) Écrivez les formules du type (11.765) pour $d^2 f$ et $d^4 f$.

(3) Allez relire le développement de Taylor du théorème 11.379.

(4) À l'ordre zéro, il n'y a rien parce que le terme $-4f(x, y)$ compense les quatre termes d'ordre zéro des autres termes.

(5) Aux ordres impairs, il n'y a rien. En effet, prenons un nombre impair l et la formule

$$(d^l f)_x(h, \dots, h) = \sum_{i_1,\dots,i_l} h_{i_1} \dots h_{i_l} \frac{\partial^l f}{\partial x_{i_1} \dots \partial x_{i_l}}(x). \tag{34.125}$$

Nous avons

$$(d^l f)_x(h, \dots, h) + (d^l f)_x(-h, \dots, -h) = 0. \tag{34.126}$$

Or dans les expressions (34.123) et (34.121), les termes arrivent par pair opposées.

Commençons par calculer $h^2(\Delta_h f)(x, y)$.

Ordre 4 . Le premier terme est :

$$(d^2 f)_{(x,y)}\big((h, 0), (h, 0)\big) = h^2 (d^2 f)_{(x,y)}\big((1, 0), (1, 0)\big). \tag{34.127}$$

La formule (11.765) à peine adaptée permet de calculer ça explicitement.

Il y a encore les termes du même type avec $(0, 1)$, $(-1, 0)$ et $(0, -1)$.

Ordre 4 Cette fois, ce sont 4 termes du type

$$h^4 (d^4 f)_{(x,y)}\big((1, 0), (1, 0), (1, 0), (1, 0)\big) \tag{34.128}$$

à calculer.

Cela fait beaucoup de termes à calculer. Je vous laisse vous persuader que le programme suivant en Sage nous donne les coefficients.

```
#! /usr/bin/env python3
# -*- coding: utf8 -*-


x,y=var("x,y")

C=[]
C.append(x)
C.append(y)

def coef4(h):
    S=0
    for i in [0,1]:
```

```
        for j in [0,1]:
            for k in [0,1]:
                for l in [0,1]:
                    S=S+h[i]*h[j]*h[k]*h[l]*C[i]*C[j]*C[k]*C[l]
    return S

def coef2(h):
    S=0
    for i in [0,1]:
        for j in [0,1]:
            S=S+h[i]*h[j]*C[i]*C[j]
    return S

cross=[]
square=[]

cross.append(  [1,0] )
cross.append(  [-1,0] )
cross.append(  [0,1] )
cross.append(  [0,-1] )

square=[]
square.append(  [-1,-1] )
square.append(  [1,-1] )
square.append(  [-1,1] )
square.append(  [1,1] )


print("Cross scheme :")

K=sum(  coef2(v) for v in cross   )
L=sum(  coef4(v) for v in cross   )
print(K)
print(L)

print("square scheme :")

K=sum(  coef2(v) for v in square   )
L=sum(  coef4(v) for v in square   )
print(K)
print(L)
```

tex/sage/coefs.sage

Le résultat est que, en utilisant la formule

$$\partial_x^4 f + 2\partial_{xxyy}^4 f + \partial_y^4 f = \Delta\Delta f, \tag{34.129}$$

nous avons

$$(\Delta_h f)(x,y) = \frac{1}{2}(2\partial_x^2 f + 2\partial_y^2 f)(x,y) + \frac{1}{4!}2h^2(\partial_x^4 f + \partial_y^4 f)(x,y) + Kh^4 \tag{34.130a}$$
$$= (\Delta f)(x,y) + \frac{1}{12}h^2(\partial_x^4 f + \partial_y^4 f)(x,y) + Kh^4 \tag{34.130b}$$
$$= \Delta f + \frac{h^2}{12}\Delta\Delta f - \frac{h^2}{12}2\partial_{xxyy}^4 f + Kh^4 \tag{34.130c}$$
$$= \Delta f + \frac{h^2}{12}\Delta\Delta f - \frac{h^2}{6}\partial_{xxyy}^4 f + Kh^4 \tag{34.130d}$$

où K est une constante qui peut être majorée en termes des dérivées quatrièmes de f. En particulier la plus grande des normes supremum de ces dérivées.

Le même genre de calculs donnent

$$(\Delta_h' f)(x,y) = \frac{1}{2}\Big[\frac{1}{2}4\Delta f + \frac{h^2}{4!}(4\partial_x^4 f + 24\partial_x^2\partial_y^2 f + 4\partial_y^2 f)\Big] + Kh^4. \tag{34.131}$$

Ça donne :

$$(\Delta_h' f) = \Delta f + \frac{h^2}{12}\Delta\Delta f + \frac{h^2}{3}\partial_{xxyy}^4 f + Kh^4 \tag{34.132}$$

avec redéfinition du K ; nous ne le préciserons plus à chaque fois.

Nous avons donc le résultat proposé dans [517] :

$$a\Delta_h f + b\Delta_h' f = (a+b)\Delta f + (a+b)\frac{h^2}{12}\Delta^2 f + h^2\frac{1}{6}(a-2b)\partial_{xxyy}^4 f + Kh^4. \tag{34.133}$$

L'idée est d'appliquer ça à une fonction u qui vérifie l'équation différentielle $-\Delta u = f$ (attention au clash de notation pour f). Le mieux est de supprimer le terme en $\partial_{xxyy} f$ en demandant $a - 2b = 0$. Nous avons donc à résoudre le système

$$\begin{cases} a + b = 1 & (34.134a) \\ a - 2b = 0. & (34.134b) \end{cases}$$

Qui propose une décomposition PLU pour résoudre ce système linéaire ? Quelle que soit la manière, la solution est

$$a = \frac{2}{3}, \quad b = \frac{1}{3}. \tag{34.135}$$

Nous allons donc étudier la discrétisation à neuf points

$$L_h = \frac{2}{3}\Delta_h + \frac{1}{3}\Delta_h'. \tag{34.136}$$

En faisant quelques additions nous trouvons que l'opération

$$\begin{aligned}(L_h u)(x_i,y_j) = \frac{1}{6h^2}\Big(&-20u_{ij} + 4\big(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}\big) \\ &+ u_{i+1,j+1} + u_{i-1,j+1} + u_{i+1,j-1} + u_{i-1,j-1}\Big)\end{aligned} \tag{34.137}$$

vérifie

$$L_h f = \Delta f + \frac{h^2}{12}\Delta^2 f + Kh^4. \tag{34.138}$$

34.4.1 Travail avec le laplacien à 9 points

Nous allons écrire un schéma numérique pour l'équation différentielle $-\Delta u = f$ utilisant la discrétisation à 9 points du laplacien. Nous recopions ses propriétés fondamentales (34.136), (34.137), (34.138) :

$$L_h = \frac{2}{3}\Delta_h + \frac{1}{3}\Delta_h', \tag{34.139}$$

et

$$(L_h u)(x_i, y_j) = \frac{1}{6h^2}\Big(-20u_{ij} + 4\big(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}\big) \\ + u_{i+1,j+1} + u_{i-1,j+1} + u_{i+1,j-1} + u_{i-1,j-1}\Big), \tag{34.140}$$

et

$$L_h f = \Delta f + \frac{h^2}{12}\Delta^2 f + Kh^4. \tag{34.141}$$

Nous appliquons (34.141) à u et nous isolons Δu :

$$\Delta u = L_h u - \frac{h^2}{12}\Delta^2 u + Kh^4 = \frac{1}{6h^2}T_h u + \frac{h^2}{12}\Delta f + Kh^4 \tag{34.142}$$

où nous avons utilisé $\Delta^2 u = -\Delta f$ et avons noté

$$T_h u = -20u_{ij} + 4(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}) + u_{i+1,j+1} + u_{i-1,j+1} + u_{i+1,j-1} + u_{i-1,j-1}. \tag{34.143}$$

Nous imposons maintenant $\Delta u = -f$ en écrivant

$$\frac{1}{6h^2}T_h u = -f - \frac{h^2}{12}\Delta f + \alpha(h)h^4. \tag{34.144}$$

Une idée est de remplacer Δf par son approximation en croix (34.122) :

$$T_h u = -6h^2 f - \frac{h^4}{2}(\Delta_h f + Kh^2) + \alpha(h)h^6 \tag{34.145}$$

Avec quelques calculs nous trouvons le schéma numérique suivant :

$$20u_{ij} - 4(u_{i+1,j} + u_{i-1,j} + u_{i,j+1} - u_{i,j-1}) - u_{i+1,j+1} - u_{i-1,j+1} - u_{i+1,j-1} - u_{i-1,j-1} \tag{34.146a}$$

$$= \frac{h^2}{2}(8f_{ij} + f_{i+1,j} + f_{i-1,j} + f_{i,j+1} + f_{i,j-1}) + \alpha(h)h^6. \tag{34.146b}$$

En oubliant le terme en $\alpha(h)h^6$, nous obtenons un système d'équations linéaires.

Problèmes et choses à faire

Il me semble que ce schéma donne une convergence d'ordre 6. C'est correct ?

Chapitre 35

Variables aléatoires et théorie des probabilités

35.1 Espace de probabilité

Définition 35.1.
Une ***mesure de probabilité*** *sur un espace mesurable*[1] $(\Omega, \mathcal{A})$ *est une mesure positive telle que* $P(\Omega) = 1$. *Dans ce cas, le triple* $(\Omega, \mathcal{A}, P)$ *est un* ***espace de probabilité****.*

Un point $\omega \in \Omega$ est une **observation**, une partie mesurable $A \in \mathcal{A}$ est un **événement**. L'ensemble $A \cup B$ représente l'événement A ou B tandis que l'ensemble $A \cap B$ représente l'événement A et B.

Si les A_n sont des événements, nous avons défini en 7.146 limite supérieure et la limite inférieure de la suite A_n par

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geqslant 1} \bigcup_{k\geqslant n} A_k \tag{35.1}$$

et

$$\liminf_{n\to\infty} A_n = \bigcup_{n\geqslant 1} \bigcap_{k\geqslant n} A_k \tag{35.2}$$

Si $\omega \in \liminf A_n$, alors ω réalise tous les A_n sauf un nombre fini.

Nous avons

$$\limsup A_n = \{\omega \in \Omega \text{ tel que } \omega \in A_n \text{pour une infinité de } n\}. \tag{35.3}$$

Théorème 35.2 (Borel-Cantelli)**.**
Si

$$\sum_{n=1}^{\infty} P(A_n) < \infty \tag{35.4}$$

alors $P(\limsup A_n) = 0$.

Démonstration. La condition $\sum_{n\geqslant 1} P(A_n) < \infty$ signifie que la fonction

$$\varphi = \sum_{n\geqslant 1} \mathbb{1}_{A_n} \tag{35.5}$$

est P-intégrable. Par conséquent, elle est finie presque partout (au sens de P), c'est-à-dire

$$P(\varphi = \infty) = 0. \tag{35.6}$$

Les points ω sur lesquels $\varphi(\omega) = \infty$ sont ceux tels que

$$\sum_{n\geqslant 1} \mathbb{1}_{A_n}(\omega) = \infty, \tag{35.7}$$

1. Espace mesurable : 13.1, mesure positive : 13.20.

c'est-à-dire les ω qui appartiennent à une infinité d'ensembles A_n, ou encore les $\omega \in \limsup A_n$. Nous avons donc montré que

$$\{\omega \text{ tel que } \varphi(\omega) = \infty\} = \{\omega \in \Omega \text{ tel que } \omega \in A_n \text{pour une infinité de } n\} = \limsup A_n. \tag{35.8}$$

Or l'hypothèse signifie que la probabilité du membre de gauche est nulle. □

Corolaire 35.3.
Si $\sum_{n=1}^{\infty} P(\complement A_n) < \infty$, alors presque surement tous les B_n sont réalisés à l'exception d'un nombre fini.

35.2 Variables aléatoires

Définition 35.4.
*Une **variable aléatoire** est une application mesurable*

$$X\colon (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}or(\mathbb{R}^d)). \tag{35.9}$$

Nous convenons que $\mathbb{R}^1 = \bar{\mathbb{R}}$, c'est-à-dire que dans le cas où la variable aléatoire X est réelle, nous acceptons les valeurs $\pm\infty$.

Définition 35.5.
*Une variable aléatoire réelle X est **absolument continue** s'il existe une fonction positive et intégrable $f\colon \mathbb{R} \to \mathbb{R}$ telle que pour tout intervalle $I \subset \mathbb{R}$,*

$$P(X \in I) = \int_I f(t)dt. \tag{35.10}$$

*Nous disons alors que f est la **densité** de X.*

Cela ne devrait pas être sans rappeler la définition 16.20.

35.2.1 Indépendance

La définition suivante vient de l'instructive motivation de [518]. La définition d'indépendance de deux événements se généralise à n événements de la façon suivante.

Définition 35.6.
*Nous disons que les événements $A_1, \dots, A_n$ sont **indépendants** si pour tout choix $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ nous avons*

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}). \tag{35.11}$$

*Les sous tribus $\mathcal{A}_1, \dots, \mathcal{A}_n$ sont **indépendantes** si pour tout choix $A_i \in \mathcal{A}_i$, les événements A_i sont indépendants.*

Exemple 35.7.
Soit $\Omega = [0,1] \times [0,1]$ muni de la mesure de Lebesgue. Soient $A = [0,a] \times [0,1]$ et $B = [0,1] \times [0,b]$. Nous avons $P(A) = a$ et $P(B) = b$ ainsi que $P(A \cup B) = ab$. △

Lemme 35.8.
Les tribus $\mathcal{A}_1, \dots, \mathcal{A}_n$ sont indépendantes si et seulement si

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \dots P(A_n) \tag{35.12}$$

pour tout $A_i \in \mathcal{A}_i$.

Démonstration. L'implication dans le sens direct découle immédiatement des définitions.

Nous supposons avoir un choix $(A_i)_{i=1,\dots,n}$ avec $A_i \in \mathcal{A}_i$ et nous devons montrer que ces événements sont indépendants, c'est-à-dire que si $J \subset \{1,\dots,n\}$ alors les événements $(A_j)_{j\in J}$ sont indépendants. Sans perte de généralité, nous pouvons supposer que si $i \notin J$, $A_i = \Omega$. Alors nous avons

$$P\big(\bigcap_{j\in J} A_j\big) = P\big(\bigcap_{i=1}^{n} A_i\big) = \prod_{i=1}^{n} P(A_i) = \prod_{j\in J} P(A_j) \tag{35.13}$$

parce que $P(A_i) = P(\Omega) = 1$ lorsque i n'est pas dans J. □

Si A est un événement, la **tribu engendrée** par A est

$$\sigma(A) = \{\varnothing, A, \complement A, \Omega\}. \tag{35.14}$$

Soit $X\colon \Omega \to \mathbb{R}^d$ une variable aléatoire. Conformément à la définition 13.46, la **tribu engendrée** est

$$\mathcal{A}_X = \{X^{-1}(B) \text{ tel que } B \in \mathcal{B}\mathrm{or}(\mathbb{R}^d)\}. \tag{35.15}$$

Cela est la plus petite tribu sous tribu de $\mathcal{A}$ pour laquelle X est mesurable. Elle sera aussi (le plus) souvent notée $\sigma(X)$.

Définition 35.9.
*Nous disons que les variables aléatoires $X_k\colon \Omega \to \mathbb{R}^d$ sont **indépendantes** lorsque les tribus engendrées $\mathcal{A}_{X_1},\dots,\mathcal{A}_{X_n}$ le sont.*

Remarque 35.10.
Il n'a de sens de dire que X_1 et X_2 sont indépendants que si X_1 et X_2 sont des applications dont l'espace de départ est identique.

Si nous voulons modéliser le jet de deux pièce indépendantes, le mauvais choix est de faire $\Omega = \{0,1\}$, y mettre la mesure d'équiprobabilité, et de considérer les deux variables aléatoires

$$X_i(\omega) = \begin{cases} f & \text{si } \omega = 0 \\ p & \text{si } \omega = 1. \end{cases} \tag{35.16}$$

Ces deux variables sont évidemment pas indépendantes. Il faut poser $\Omega = \{0,1\}\times\{0,1\}$, y mettre la mesure d'équiprobabilité et poser

$$X_1(x,y) = \begin{cases} f & \text{si } x = 0 \\ p & \text{si } x = 1 \end{cases}, \tag{35.17}$$

$$X_2(x,y) = \begin{cases} f & \text{si } y = 0 \\ p & \text{si } y = 1 \end{cases}, \tag{35.18}$$

Ces variables aléatoires sont indépendantes. Par exemple

$$X_1^{-1}\{p\} = \{(1,0),(1,1)\} \tag{35.19a}$$
$$X_2^{-1}\{p\} = \{(0,1),(1,1)\} \tag{35.19b}$$

et on a bien

$$P\big(X_1^{-1}\{p\} \cap X_2^{-1}\{p\}\big) = P\{(1,1)\} = \frac{1}{4} \tag{35.20}$$

ainsi que

$$P\{X_p^{-1}(p)\} = \frac{1}{2} \tag{35.21a}$$

pour $i = 1$ et $i = 2$.

Proposition 35.11.
Soient $(X_k\colon \Omega \to \mathbb{R}^{d_k})$ *des variables aléatoires indépendantes.*

(1) Si $B_k \in \mathcal{B}or(\mathbb{R}^{d_k})$. *Alors*

$$P(X_k \in B_k \forall k \leqslant n) = P(X_1 \in B_1) \dots P(X_n \in B_n). \tag{35.22}$$

(2) Les événements $\{X_i \in B_i\}$ *sont indépendants.*

(3) Les tribus engendrées par des X_i *et d'autres sont indépendantes. Plus précisément, si* I *et* J *sont deux ensembles disjoints de* $\mathbb{N}$ *alors les tribus*

$$\sigma(\{X_i, i \in I\}) \tag{35.23}$$

et

$$\sigma(\{X_i, i \in J\}) \tag{35.24}$$

sont indépendantes.

Démonstration. Lorsque nous écrivons $X_i \in B_i$, nous parlons de l'événement

$$(X_i \in B_i) = \{\omega \in \Omega \text{ tel que } X_i(\omega) \in B_i\} = X_i^{-1}(B_i) \in \mathcal{A}_{X_i}. \tag{35.25}$$

Vu que par hypothèse les tribus $(\mathcal{A}_i)$ sont indépendantes, le lemme 35.8 nous montre que

$$P\big(\bigcap_{i=1}^{n} X_i \in B_i\big) = \prod_i P(X_i \in B_i). \tag{35.26}$$

Il reste à voir que l'ensemble $X_i^{-1}(B_i)$ fait partie de la tribu $\mathcal{A}$ de départ. Cela est la définition du fait que l'application X_i soit une variable aléatoire : elle doit être mesurable en tant qu'application

$$X_i\colon (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}\mathrm{or}(\mathbb{R}^d)). \tag{35.27}$$

Les affirmations (2) et (3) ne sont que des façons alternatives d'exprimer la même chose. □

Lemme 35.12.
Les événements $(A_i)_{i=0,\dots,n}$ *sont indépendants si et seulement si les événements que nous obtenons en remplaçant certains des* A_i *par* $\complement A_i$ *le sont.*

Démonstration. Sans perte de généralité, nous pouvons nous contenter de prouver que les événements $\complement A_0, A_1, \dots, A_n$ sont indépendants sous l'hypothèse que les événements $A_0, A_1, \dots, A_n$ sont indépendants. Soit I un sous-ensemble de $\{1, \dots, n\}$. Nous avons

$$\begin{aligned}
P\big(\complement A_0 \bigcap_{i\in I} A_i\big) &= P\big(\bigcap_{i\in I} A_i \backslash \bigcap_{i\in I} A_i \cap A_0\big) && (35.28a)\\
&= P\big(\bigcap_{i\in I} A_i\big) - P\big(\bigcap_{i\in I} A_i \cap A_0\big) && (35.28b)\\
&= P\big(\bigcap_{i\in I} A_i\big)\big(1 - P(\complement A_0)\big) && (35.28c)\\
&= P\big(\bigcap_{i\in I} A_i\big) P(\complement A_0). && (35.28d)
\end{aligned}$$

□

Proposition 35.13.
Les événements $(A_i)_{i=1,\dots,n}$ *sont indépendants si et seulement si les variables aléatoires associées* $\mathbb{1}_{A_1}, \dots, \mathbb{1}_{A_n}$ *le sont.*

Démonstration. La tribu engendrée par la variable aléatoire $\mathbb{1}_{A_k}$ est

$$\mathcal{A}_{\mathbb{1}_{A_k}} = \{\varnothing, A_k, \complement A_k, \Omega\}. \tag{35.29}$$

En effet si $1 \in B$, alors $A_i \subset \mathbb{1}_{A_i}^{-1}(B)$, et si $0 \in B$, alors $\complement A_i \subset \mathbb{1}_{A_i}^{-1}(B)$. Les éléments 0 et 1 sont tous deux soit dans B, soit hors de B. Cela donne les 4 possibilités énumérées dans (35.29).

Supposons que les événements (A_i) sont indépendants. Nous devons vérifier que les tribus le soient, c'est-à-dire que les événements A_i et $\complement A_j$ sont indépendants. Cela est une conséquence du lemme 35.12. □

Théorème 35.14 (Doob[270])**.**
Soit $X\colon \Omega \to \mathbb{R}^d$ une variable aléatoire. Une fonction $Y\colon \Omega \to \mathbb{R}^p$ est une variable aléatoire $\mathcal{A}_X$-mesurable si et seulement s'il existe une fonction borélienne $f\colon \mathbb{R}^d \to \mathbb{R}^p$ telle que $Y = f(X)$.

Proposition 35.15.
Soient des variables aléatoires $X_k\colon \Omega \to \mathbb{R}^{d_k}$ des variables aléatoires indépendantes et des fonctions boréliennes $f_k\colon \mathbb{R}^{d_k} \to \mathbb{R}^{p_k}$. Alors les variables aléatoires $f_k(X_k)$ sont indépendantes.

Démonstration. Le théorème 35.14 assure que les applications

$$f_k \circ X_k\colon \Omega \to \mathbb{R}^{d_k} \tag{35.30}$$

sont $\mathcal{A}_{X_k}$-mesurables. En particulier pour tout borélien $B \subset \mathbb{R}^{p_k}$, nous avons $X_k^{-1} \circ f_k^{-1}(B) \in \mathcal{A}_{X_k}$. Nous avons donc

$$\sigma(f_k \circ X_k) \subset \sigma(X_k), \tag{35.31}$$

et par conséquent les tribus $\sigma(f_k \circ X_k)$ sont indépendantes étant donné que les tribus $\sigma(X_k)$ le sont. □

Lemme 35.16 (Lemme de regroupement)**.**
Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité et $(\mathcal{A})_{i\in I}$ une famille de tribus indépendantes dans $\mathcal{A}$. Si $(M_j)_{j\in J}$ est une partition de I, alors les tribus

$$\mathcal{B}_j = \sigma\Big(\bigcup_{i\in M_j} \mathcal{A}_i\Big) \tag{35.32}$$

sont indépendantes.

Si les variables aléatoires $\{X_1, X_2, X_3, X_4, X_5\}$ sont indépendantes, et si f et g sont des fonctions mesurables, alors les variables aléatoires $f(X_2, x_3, X_5)$ et $g(X_1, X_4)$ sont indépendantes.

Une preuve a l'air d'être donnée dans [519].

35.2.2 Lois conjointes et indépendance

Définition 35.17.
*Deux événements A et B sont dits **indépendants** si*

$$P(A \cap B) = P(A)P(B). \tag{35.33}$$

Si nous considérons n variables aléatoires réelles $X_1, \dots, X_n\colon \Omega \to \mathbb{R}$, la loi du n-uplet $X = (X_1, \dots, X_n)$ est une variable aléatoire $X\colon \Omega \to \mathbb{R}^n$ appelée la **loi conjointe** des lois X_i. Dans ce cas, les variables aléatoires X_i elles-mêmes sont dites lois **marginales** de X.

Proposition 35.18.
Les variables aléatoires $\{X_i\}$ sont indépendantes si et seulement si

$$P_{(X_1,\dots,X_n)} = P_{X_1} \otimes \dots \otimes P_{X_n}. \tag{35.34}$$

Définition 35.19.
Soient $\{X_i\}_{1\leqslant i\leqslant n}$ des variables aléatoires réelles (pas spécialement indépendantes). La ***densité conjointe*** *de $X_1,\dots,X_n$ est la fonction $f\colon \mathbb{R}^n \to \mathbb{R}$ qui satisfait*

(1) $f(x_1,\dots,x_n) \geqslant 0$ pour tout $(x_1,\dots,x_n) \in \mathbb{R}^n$,

(2) $\int_{\mathbb{R}^n} f = 1$,

(3) pour tout $A_i \subset \mathbb{R}$ nous avons

$$P(\bigcap_{i=1}^n X_i \in A_i) = \int_{\prod_i A_i} f(x_1,\dots,x_n)dx_1\dots dx_n. \tag{35.35}$$

Proposition 35.20.
Si les variables aléatoires $X_1,\dots X_n$ sont indépendantes et ont des densités $f_{X_1},\dots,f_{X_n}$, alors la variable aléatoire conjointe $X = (X_1,\dots,X_n)$ a pour densité conjointe la fonction

$$f_X(x_1,\dots,x_n) = f_{X_1}(x_1)\dots f_{X_n}(x_n). \tag{35.36}$$

Démonstration. En partant de la définition de l'indépendance et de la fonction de densité conjointe, ainsi qu'en utilisant le théorème de Fubini,

$$\begin{aligned}\int_{A_1\times\dots\times A_n} f_X(x_1,\dots,x_n)dx_1\dots dx_n &= P(X_1\in A_1,\dots,X_n\in A_n)\\ &= P(X_1\in A_1)\dots P(X_n\in A_n)\\ &= \left(\int_{A_1} f_{X_1}(x_1)dx_1\right)\dots\left(\int_{A_n} f_{X_n}(x_n)dx_n\right)\\ &= \int_{A_1\times\dots\times A_n} f_{X_1}(x_1)\dots f_{X_n}(x_n)dx_1\dots dx_n.\end{aligned} \tag{35.37}$$

La fonction $(x_1,\dots,x_n)\mapsto f_{X_1}(x_1)\dots f_{X_n}(x_n)$ vérifie donc la condition (3) de la définition 35.19. La vérification des autres conditions est immédiate. □

La proposition suivante provient du fait que la mesure d'une loi conjointe est le produit des mesures lorsque les variables aléatoires sont indépendantes (proposition 35.18).

Proposition 35.21 ([270])**.**
Si les variables aléatoires réelles $X_1,\dots,X_n$ sont intégrables et indépendantes, alors leur produit est intégrable et l'espérance du produit est égal au produit des espérances :

$$E(X_1\cdots X_n) = E(X_1)\dots E(X_n). \tag{35.38}$$

35.2.3 Somme et produit de variables aléatoires indépendantes

Soient X et Y, deux variables aléatoires réelles indépendantes. Nous voudrions étudier la loi de la variable aléatoire $S = X + Y$. Nous commençons par calculer la fonction de répartition en utilisant le résultat de la proposition 35.20 :

$$F_{X+Y}(z) = P(X+Y\leqslant z) = \int_{x+y\leqslant z} f_{X,Y}(x,y)dx\,dy \tag{35.39a}$$

$$= \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} dy f_X(x)f_Y(y) \tag{35.39b}$$

$$= \int_{\mathbb{R}} \left(\int_{-\infty}^{z-x} f_Y(y)dy\right) f_X(x)dx \tag{35.39c}$$

$$= \int_{\mathbb{R}} F_Y(z-x)f_X(x)dx. \tag{35.39d}$$

Pour calculer la fonction de densité de S, nous dérivons la fonction de répartition :

$$f_{X+Y}(z) = \frac{dF_{X+Y}}{dz}(z) \tag{35.40a}$$

$$= \int_{\mathbb{R}} f_Y(z-x) f_X(x) dx, \tag{35.40b}$$

ce qui nous amène à dire que la densité de la somme est le produit de convolution [2] des densités :

$$f_{X+Y}(x) = \int_{\mathbb{R}} f_Y(x-t) f_X(t) dt, \tag{35.41}$$

ou encore $f_{X+Y} = f_X * f_Y$.

Notez que nous avons passé sous le silence la difficulté d'inverser la dérivée et l'intégrale. Un exemple sera donné au point 35.5.8.

Lemme 35.22.
Soient X et Y, deux variables aléatoires indépendantes. Alors

$$E(XY) = E(X)E(Y). \tag{35.42}$$

Démonstration. Par indépendance et par proposition 35.20, la fonction de densité conjointe de X et Y vaut $f_{X,Y} = f_X f_Y$. Par conséquent l'utilisation de Fubini sous la forme (13.814) entraine

$$E(XY) = \int_{\mathbb{R}\times\mathbb{R}} xy f_{X,Y}(x,y) dxdy = E(X)E(Y). \tag{35.43}$$

□

Nous dirons tout un tas de chose sur l'indépendance et la variance en 35.5.13, mais pour l'instant nous allons mentionner et démontrer déjà ceci :

Lemme 35.23.
Soient X et Y deux variables aléatoires indépendantes et identiquement distribuées. Alors

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y). \tag{35.44}$$

Démonstration. Par définition, $\mathrm{Var}(X+Y) = E\big([X+Y-E(X)-E(Y)]^2\big)$. En développant le carré et en utilisant le lemme 35.22,

$$\mathrm{Var}(X+Y) = E(X^2) - E(X)^2 + E(Y^2) - E(Y)^2 = \mathrm{Var}(X) + \mathrm{Var}(Y). \tag{35.45}$$

□

Exemple 35.24.
Deux variables aléatoires non indépendantes dont la covariance est nulle. Nous considérons la variable aléatoire

$$Z\colon \Omega \to \{(1,0),(-1,0),(0,1),(0,-1)\} \tag{35.46}$$

de loi uniforme. C'est-à-dire que $P\big(Z=z\big) = \frac{1}{4}$ pour tout z. Ensuite nous considérons les variables aléatoires $X = \mathtt{proj}_1 \circ Z$ et $Y = \mathtt{proj}_2 \circ Z$. Toute personne étant capable de compter jusqu'à 4 voit que

$$P(X=1) = P(X=-1) = \frac{1}{4} \tag{35.47a}$$

$$P(X=0) = \frac{1}{2}, \tag{35.47b}$$

2. Définition 26.52.

et les mêmes probabilités pour Y. De même $E(X) = E(Y) = 0$. Par conséquent

$$\mathrm{Cov}(X, Y) = E(XY) = 0 \tag{35.48}$$

parce que pour tout $\omega \in \Omega$ nous avons soit $X(\omega) = 0$ soit $Y(\omega) = 0$. Ces variables aléatoires X et Y ne sont donc pas corrélées.

Mais elles ne sont pas indépendantes pour autant, comme nous allons le voir pas plus tard qu'immédiatement. Nous avons

$$P(X = 0|Y = 0) = \frac{P(X = 0, Y = 0)}{P(Y = 0)} = 0 \tag{35.49}$$

parce que X et Y ne peuvent pas être simultanément nulles, tandis que

$$P(X = 0)P(Y = 0) = \frac{1}{4}. \tag{35.50}$$

△

35.2.4 Espérance

Définition 35.25.
*Soit $1 \leqslant p < \infty$. Le **moment** d'ordre p d'une variable aléatoire $X \in L^p(\Omega, \mathcal{A}, P)$ est l'espérance*

$$m_n(X) = E(X^n). \tag{35.51}$$

Définition 35.26.
*Si $X \in L^1(\Omega, \mathcal{A}, P)$ est à valeurs dans $\mathbb{R}^n$, nous définissons l'**espérance** de X par*

$$E(X) = \int_\Omega X dP \tag{35.52}$$

Lemme 35.27 ([1])**.**
Soit un espace de probabilité Ω ainsi qu'une variable aléatoire $X\colon \Omega \to \mathbb{R}$ prenant seulement les valeurs $\{y_k\}_{k\in\mathbb{N}}$. Alors

$$E(X) = \sum_{k=0}^{\infty} y_k P(X = y_k). \tag{35.53}$$

Démonstration. Vu que l'application $X\colon \Omega \to \mathbb{R}$ ne prend que les valeurs y_k, nous avons $\Omega = \bigcup_{k=0}^{\infty} X^{-1}(y_k)$, avec une union disjointe. Nous avons

$$\begin{aligned}
E(X) &= \int_{\bigcup_i X^{-1}(y_i)} X(\omega) dP(\omega) && \text{(35.54a)}\\
&= \sum_{i=0}^{\infty} \int_{X^{-1}(y_k)} \underbrace{X(\omega)}_{=y_k} dP(\omega) && \text{(35.54b)}\\
&= \sum_{k=0}^{\infty} y_k \int_{X^{-1}(y_k)} dP(\omega) && \text{(35.54c)}\\
&= \sum_{k=0}^{\infty} y_k P\big(X^{-1}(y_k)\big) && \text{(35.54d)}\\
&= \sum_{k=0}^{\infty} y_p P(X = y_k) && \text{(35.54e)}
\end{aligned}$$

Justifications :

— Pour (35.54b), nous avons utilisé la σ-additivité de l'intégrale de la proposition 13.187.

— Pour (35.54d), l'intégrale de la fonction 1 donne la mesure de la partie, c'est le lemme 13.157.

— Pour (35.54e), la notation $P(X = y)$ signifie $P\big(\{\omega \in \Omega \text{ tel que } X(\omega) = y\}\big)$; c'est la mesure de $X^{-1}(y)$.

□

Définition 35.28.
Si $E(X) = 0$ nous disons que la variable aléatoire est ***centrée****. La variable aléatoire $X - E(X)$ est la variable aléatoire centrée associée à X.*

Proposition 35.29 ([520])**.**
Si X et Y sont deux variables aléatoires (pas spécialement indépendantes), nous avons

$$E(X + Y) = E(X) + E(Y). \tag{35.55}$$

Nous donnons la preuve dans le cas de variables aléatoires indépendantes. Le cas plus général de variable aléatoires non indépendantes peut être trouvé dans [520].

Démonstration. Nous avons le calcul suivant :

$$\begin{aligned}
E(X + Y) &= \int_{\mathbb{R}} x f_{X+Y}(x) dx && (35.56a)\\
&= \int_{\mathbb{R}} x \int_{\mathbb{R}} f_Y(x - t) f_X(t) dt dx && (35.56b)\\
&= \int_{\mathbb{R}} f_X(t) \underbrace{\int_{\mathbb{R}} x f_Y(x - t) dx}_{=E(Y)+t} \, dt && (35.56c)\\
&= \int_{\mathbb{R}} f_X(t)\big(E(Y) + t\big) dt && (35.56d)\\
&= E(Y) + \int_{\mathbb{R}} t f_X(t) dt && (35.56e)\\
&= E(Y) + E(X) && (35.56f)
\end{aligned}$$

où nous avons utilisé la proposition 35.41 et le fait que l'intégrale sur $\mathbb{R}$ d'une densité vaut 1. □

Une application de l'inégalité de Hölder (proposition 26.33) est la suivante. Si X et Y sont des variables aléatoires intégrables alors

$$E(XY) \leqslant E(X^2)^{1/2} E(Y^2)^{1/2}. \tag{35.57}$$

En effet

$$E(XY) \leqslant \|XY\|_{L^1(\Omega)} \leqslant \|X\|_{L^2(\Omega)} \|Y\|_{L^2(\Omega)}. \tag{35.58}$$

35.2.5 Variance

Si $X \in L^2(\Omega, \mathcal{A}, P)$ alors nous définissons la **variance** de X par

$$\mathrm{Var}(X) = E\big([X - E(X)]^2\big). \tag{35.59}$$

Proposition 35.30.
La variance de la variable aléatoire X peut être exprimée par la formule

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 \tag{35.60}$$

où $X^2 = X \cdot X$ et $E(X)^2 = E(X) \cdot E(X)$ sont des produits scalaires dans $\mathbb{R}^d$.

Démonstration. De façon explicite, nous avons

$$E\big([X-E(X)]^2\big) = \int_\Omega \big(X(\omega)-E(X)\big)\cdot\big(X(\omega)-E(X)\big)dP(\omega) \tag{35.61}$$

où $E(X)\in\mathbb{R}^d$ est une constante. En développant le produit scalaire nous avons

$$\begin{aligned} E\big([X-E(X)]^2\big) &= E\big(X^2-2X\cdot E(X)+E(X)^2\big) && (35.62a)\\ &= E(X^2)-2E(X)^2+E(X)^2 && (35.62b)\\ &= E(X^2)-E(X)^2. && (35.62c)\end{aligned}$$

□

Nous définissons l'**écart-type** de X par

$$\sigma_X = \sqrt{\mathrm{Var}(X)}. \tag{35.63}$$

En d'autres termes,

$$\sigma_X = \|X-E(X)\|_{L^2}. \tag{35.64}$$

On définit encore la **moyenne quadratique** de X par

$$\|X\|_{L^2} = \big[E(X^2)\big]^{1/2}. \tag{35.65}$$

La variable aléatoire

$$\bar{V}_n = \frac{1}{n}\sum_i (X_i-\bar{X}_n)^2 \tag{35.66}$$

est la **variance empirique** de l'échantillon (X_i).

Lemme 35.31.
Si X est une variable aléatoire,

(1) $\mathrm{Var}(ax)=a^2\,\mathrm{Var}(X)$ *pour tout* $a\in\mathbb{R}$ *;*

(2) Si de plus Y est une variable aléatoire indépendante de X, alors $\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$.

Démonstration. Nous avons

$$\begin{aligned}\mathrm{Var}(X+Y) &= E(X^2+Y^2+2XY)-\big(E(X)+E(Y)\big)^2 && (35.67a)\\ &= E(X^2)+E(Y^2)+2E(XY)-E(X)^2-E(Y)^2+2E(X)E(Y). && (35.67b)\end{aligned}$$

Étant donné que X et Y sont indépendantes nous avons $E(XY)=E(X)E(Y)$ par le lemme 35.22.

□

Si les $X_1,\dots,X_n$ sont des variables aléatoires on considère la **moyenne empirique**

$$\bar{X}_n = \frac{X_1+\cdots+X_n}{n}. \tag{35.68}$$

35.2.6 Covariance

Soient X et Y, deux variables aléatoires réelles. Leur **covariance** est définie par

$$\mathrm{Cov}(X,Y) = E\Big[\big(X-E(X)\big)\big(Y-E(Y)\big)\Big] \tag{35.69}$$

L'idée est que la covariance devient grande si X et Y s'écartent de leurs moyennes dans le même sens. Il existe une formule alternative :

$$\mathrm{Cov}(X,Y) = E(XY)-E(X)E(Y) \tag{35.70}$$

En ce qui concerne les dimensions plus hautes, si $X\colon \Omega \to \mathbb{R}^d$ est un vecteur aléatoire de carré intégrable, nous définissons

$$\operatorname{Cov}(X) = E\Big[\big(X - E(X)\big) \otimes \big(X - E(X)\big)\Big] \tag{35.71}$$

où par $a \otimes b$ nous entendons la matrice $(a \otimes b)_{ij} = a_i b_j$. Cela peut aussi être noté $a^t b$ si l'on fait bien attention à qui est un vecteur colonne et qui est un vecteur ligne.

Proposition 35.32.
Si X et Y sont deux variables aléatoires non spécialement indépendantes, nous avons

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X,Y). \tag{35.72}$$

Démonstration. Il s'agit d'un calcul en partant de

$$\begin{aligned}\operatorname{Var}(X+Y) &= E\big((X+Y)^2\big) - E(X+Y)^2\\ &= E(X^2) + E(Y^2) + 2E(XY)\\ &\quad + \big(E(X)+E(Y)\big)^2 - 2E(X)^2 - 2E(X)E(Y)\\ &\quad - 2E(Y)E(X) - 2E(Y)^2.\end{aligned} \tag{35.73}$$

À partir d'ici il s'agit de recombiner tous les termes pour former la formule annoncée. □

Plus généralement nous avons la formule

$$\operatorname{Var}(\sum_i X_i) = \sum_i \operatorname{Var}(X_i) + 2\sum_{1\leqslant i<j\leqslant n} \operatorname{Cov}(X_i, X_j). \tag{35.74}$$

35.2.7 Probabilité conditionnelle : événements

Proposition-définition 35.33.
Soit $(\Omega, \mathcal{A}, P)$ un espace de probabilité et $B \in \mathcal{A}$ avec $P(B) > 0$. Alors avec la formule

$$P_B(A) = \frac{P(A\cap B)}{P(B)}, \tag{35.75}$$

*l'espace $(\Omega, \mathcal{A}, P_B)$ est un espace probabilisé. Nous notons $P(A|B)$ le nombre $P_B(A)$ et nous le nommons **probabilité conditionnelle** de A sachant B.*

Démonstration. On vérifie que $(\Omega, \mathcal{A}, P_B)$ est un espace de probabilité parce que $P_B(\Omega) = 1$ et

$$P_B(\bigcup_i A_i) = \sum_i P_B(A_i) \tag{35.76}$$

si les A_i sont deux à deux disjoints. □

Une conséquence immédiate de (35.75) est que si A et B sont des événements indépendants alors

$$P(A|B) = \frac{P(A\cap B)}{P(B)} = P(A). \tag{35.77}$$

La probabilité conditionnelle à B est quelque chose qui ne tient compte que de ce qui se passe dans B. Si K est un événement tel que $A \cap B = K \cap B$, alors

$$P(A|B) = P(K|B). \tag{35.78}$$

Théorème 35.34.
Soient $(B_n)_{n\geqslant 1}$ une partition finie de Ω telle que $P(B_i) > 0$. Soit $A \in \mathcal{A}$ tel que $P(A) > 0$.

(1) Si A, B et C sont des événements, alors

$$P(A \cap B|C) = P(A|B \cap C)P(B|C). \tag{35.79}$$

(2) Si $P(B) > 0$, alors $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$.

*(3) On a la **formule des probabilités totales** :*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i) = \sum_i P(A \cap B_i). \tag{35.80}$$

*(4) On a la **formule de Bayes** :*

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_i P(A|B_i)P(B_i)}. \tag{35.81}$$

Démonstration. (1) En développant le membre de droite,

$$\begin{aligned} P(A \cap B|C) &= \frac{P(A \cap B \cap C)}{P(B \cap C)} \frac{P(B \cap C)}{P(C)} \\ &= P(A \cap B|C). \end{aligned} \tag{35.82}$$

(2) C'est la définition de $P(A|B)$ et $P(B|A)$.

(3) Vu que les B_i forment une partition, nous avons

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i). \tag{35.83}$$

(4) En utilisant les deux premiers points, nous trouvons

$$\begin{aligned} P(A|B_k)P(B_k) &= P(A \cap B_k) \\ &= P(B_k|A)P(A) \\ &= P(B_k|A) \sum_i P(A|B_i)P(B_i). \end{aligned} \tag{35.84}$$

□

Lemme 35.35 ([1])**.**
Soient une variable aléatoire $X \colon \Omega \to \mathbb{R}$ ainsi qu'un réel $y \neq 0$ et qu'une partie $A \subset \Omega$ de mesure non nulle[3]*. Alors*

$$P(X\mathbb{1}_A = y) = P(X = y|A)P(A). \tag{35.85}$$

Démonstration. Juste pour être clair avec les notations,

— La notation $X = y$ désigne l'ensemble $\{\omega \in \Omega \text{ tel que } X(\omega) = y\} = X^{-1}(y)$

— De même $X\mathbb{1}_1 = y$ désigne l'ensemble $\{\omega \in \Omega \text{ tel que } X(\omega)\mathbb{1}_A(\omega) = y\}$. Comme ici $y \neq 0$, il s'agit des éléments de A tels que $X(\omega) = y$, ou encore $X^{-1}(y) \cap A$.

Nous avons :

$$P(X = y|A) = \frac{P(X = y \cap A)}{P(A)} \tag{35.86a}$$

$$= \frac{P(A\mathbb{1}_A = y)}{P(A)}. \tag{35.86b}$$

Justifications :

— Pour (35.86a), c'est la définition 35.33.

3. Ceci est exactement synonyme de « un événement de probabilité non nulle ».

— Pour (35.86b), vu que $y \neq 0$ nous avons

$$\{\omega \in \Omega \text{ tel que } X(\omega) = y\} \cap A = \{\omega \in \Omega \text{ tel que } (X\mathbb{1}_A)(\omega) = y\}. \tag{35.87}$$

□

Lemme 35.36 ([1]).
Soit un espace de probabilité $(\Omega, \mathcal{F}, P)$. Nous considérons un mesurable $B \in \mathcal{F}$ ainsi que des mesurables $(A_i)_{i\in\mathbb{N}}$ deux à deux disjoints. Alors

$$P\big(\bigcup_{i=0}^{\infty} A_i|B\big) = \sum_{i=0}^{\infty} P(A_i|B). \tag{35.88}$$

Démonstration. Nous commençons par la définition 35.33 de la probabilité conditionnelle :

$$P\big(\bigcup_i A_i|B\big) = \frac{P\big(\bigcup_i A_i \cap B\big)}{P(B)} = \frac{P\big(\bigcup_i (A_i \cap B)\big)}{P(B)}. \tag{35.89}$$

Ensuite, les A_i étant disjoints, les $A_i \cap B$ le sont aussi. Vu que P est une probabilité (et donc une mesure), elle vérifie la condition de la définition 13.20(2), qui donne ici

$$P\big(\bigcup_{i=0}^{\infty}(A_i \cap B)\big) = \sum_{i=0}^{\infty} P(A_i \cap B). \tag{35.90}$$

En remettant dans (35.89),

$$P\big(\bigcup_i A_i|B\big) = \sum_{i=0}^{\infty} \frac{P(A_i \cap B)}{P(B)} = \sum_{i=0}^{\infty} P(A_i|B). \tag{35.91}$$

□

35.2.8 Espérance conditionnelle

Théorème-définition 35.37 (Définition de l'espérance conditionnelle[424]).
Soit un espace de probabilité $(\Omega, \mathcal{A}, P)$ et une variable aléatoire intégrable $X : \Omega \to \mathbb{R}$. Pour chaque sous tribu $\mathcal{F}$ de $\mathcal{A}$, il existe une (presque partout) unique variable aléatoire $Y : \Omega \to \mathbb{R}$ telle que

(1) Y est $\mathcal{F}$-mesurable

(2) Y est P-intégrable

(3) pour tout $B \in \mathcal{F}$,

$$\int_B X dP = \int_B Y dP. \tag{35.92}$$

Cette variable aléatoire sera notée $E(X|\mathcal{F})$ pour des raisons qui apparaîtront plus tard.

Démonstration. Remarquons que prendre $Y = X$ ne fonctionne pas parce qu'en général si $\mathcal{O}$ est mesurable dans $\mathbb{R}$, alors $X^{-1}(\mathcal{O})$ est dans la tribu $\mathcal{A}$, mais n'est pas automatiquement dans la tribu $\mathcal{F}$. Il faudra donc un peu plus travailler.

Unicité Si Y_1 et Y_2 vérifient tous les deux les conditions, l'ensemble $\{Y_1 < Y_2\}$ est un élément de $\mathcal{F}$ et nous avons

$$\int_{\{Y_1<Y_2\}} X = \int_{Y_1<Y_2} Y_1 = \int_{Y_1<Y_2} Y_2. \tag{35.93}$$

En particulier nous avons $\int_{\{Y_1<Y_2\}}(Y_1 - Y_2) = 0$ et donc

$$(Y_1 - Y_2)\mathbb{1}_{Y_1-Y_2} = 0 \tag{35.94}$$

presque partout. Le corolaire 13.180 montre alors que $Y_1 - Y_2 \geqslant 0$ presque partout. De la même manière, l'ensemble $\{Y_2 < Y_1\}$ est dans $\mathcal{F}$ et nous trouvons que $Y_2 - Y_1 \geqslant 0$ presque partout. Par conséquent $Y_1 = Y_2$ presque partout.

Existence dans le cas de carré intégrable Nous supposons à présent que $X \in L^2(\Omega, \mathcal{A}, P)$ et nous considérons K, le sous-ensemble de $L^2(\Omega, \mathcal{A}, P)$ des fonctions $\mathcal{F}$-mesurables. Le théorème des projections 24.5 nous indique que

$$L^2(\Omega, \mathcal{A}, P) = K \oplus K^\perp \tag{35.95}$$

par la décomposition $X = \mathtt{proj}_K X + (X - \mathtt{proj}_K X)$. La variable aléatoire $Y = \mathtt{proj}_K X$ a les propriétés d'être $\mathcal{F}$-mesurable et $\langle Y - X, Z \rangle = 0$ pour tout $Z \in K$. Soit $A \in \mathcal{F}$, si nous considérons $Z = \mathbb{1}_A$, la dernière condition signifie que

$$\int_\Omega X \mathbb{1}_A = \int_\Omega Y \mathbb{1}_A, \tag{35.96}$$

ou encore

$$\int_A Y = \int_A X. \tag{35.97}$$

La variable aléatoire $Y = \mathtt{proj}_K(X)$ répond donc à la question lorsque $X \in L^2(\Omega, \mathcal{F}, P)$.

Existence en général Nous considérons maintenant que $X \in L^1(\Omega, \mathcal{A}, P)$. Quitte à décomposer X en deux fonctions positives X_+ et X_- telles que $X = X_+ + X_-$, nous pouvons supposer que X est positive. Par hypothèse $X \in L^1(\Omega, \mathcal{A}, P)$; pour chaque $n \in \mathbb{N}$ nous posons

$$X_n(\omega) = \min\{X(\omega), n\}. \tag{35.98}$$

Étant donné que la mesure P est une mesure de probabilité, les constantes sont intégrables et $X_n \in L^2(\Omega, \mathcal{A}, P)$. De plus la suite (X_n) est croissante et

$$\lim_{n\to\infty} X_n(\omega) = X(\omega). \tag{35.99}$$

Si nous notons encore K l'ensemble des variables aléatoires dans $L^2(\Omega, \mathcal{A}, P)$ qui sont $\mathcal{F}$-mesurables, pour chaque n nous avons donc la variable aléatoire

$$Y_n = \mathtt{proj}_K X_n = E(X_n|\mathcal{F}) \tag{35.100}$$

qui est $\mathcal{F}$-mesurable et telle que

$$\int_A X_n = \int_A Y_n \tag{35.101}$$

pour tout $A \in \mathcal{F}$. Nous voudrions prouver que la variable aléatoire $Y = \lim_n Y_n$ existe et est la solution au problème, c'est-à-dire est $E(X|\mathcal{F})$.

Commençons par prouver que $Y_n \geqslant 0$ presque partout. Pour cela nous remarquons que l'ensemble $\{Y_n < 0\}$ est mesurable et

$$0 \geqslant \int_{Y_n<0} Y_n = \int_{Y_n<0} X_n \geqslant 0. \tag{35.102}$$

La première inégalité est évidente et la dernière est due au fait que X_n est positive. Par conséquent

$$\int_{Y_n<0} Y_n = 0 \tag{35.103}$$

et le lemme 13.180 conclut que $P(Y_n < 0) = 0$.

Soit $Z\colon \Omega \to \mathbb{R}$ une variable aléatoire positive dans $L^2(\Omega, \mathcal{A}, P)$. Montrons que $\mathtt{proj}_K Z$ est encore positive. Pour cela nous considérons l'ensemble $A = \{\mathtt{proj}_K Z < 0\}$ et les inégalités

$$0 \leqslant \int_A Z = \int_A \mathtt{proj}_K Z \leqslant 0, \tag{35.104}$$

ce qui montre que $\int_A \mathtt{proj}_K Z = 0$ et par conséquent que $P\{\mathtt{proj}_K(Z) < 0\} = 0$. Cela nous montre que la projection depuis L^2 conserve la positivité.

Étant donné que $X_{n-1} - X_n \geqslant 0$ nous avons aussi

$$Y_{n-1} - Y_n \geqslant 0 \tag{35.105}$$

La suite de fonctions

$$n \mapsto Y_n = E(X_n|\mathcal{F}) \tag{35.106}$$

est croissante et vérifie le théorème de la convergence monotone :

$$\int_A X = \lim_{n\to\infty} \int_A X_n = \lim_{n\to\infty} \int_A E(X_n|\mathcal{F}) = \int_A \lim_{n\to\infty} E(X_n|\mathcal{F}) = \int_A Y. \tag{35.107}$$

Par conséquent $E(X|\mathcal{F})$ existe et

$$Y = \lim_{n\to\infty} E(X_n|\mathcal{F}) = E(X|\mathcal{F}). \tag{35.108}$$

□

35.38.
Vu la définition 35.37 nous pourrions croire que la variable aléatoire $E(X|\mathcal{F}) = X$ fait l'affaire. Il n'en est rien parce que la variable aléatoire X n'est pas spécialement $\mathcal{F}$-mesurable alors qu'il est requis que $E(X|\mathcal{F})$ le soit. Avec la tribu $\mathcal{F} = \{\emptyset, \Omega\}$, nous n'avons en général pas que $X^{-1}(B) \in \mathcal{F}$ pour tout borélien B.

Par contre si $\sigma(X)$ est la tribu engendrée par la variable aléatoire X, alors $E\big(X|\sigma(X)\big) = X$.

Définition 35.39.
*Soit Z une variable aléatoire. L'**espérance conditionnelle** « X sachant Z » est la variable aléatoire*

$$E(X|Z) = E(X|\sigma(Z)) \tag{35.109}$$

où $\sigma(Z)$ est la tribu engendrée par Z. Le membre de droite est une variable aléatoire définie en 35.37.

Définition 35.40.
Soient $A \in \mathcal{A}$ un événement et $\mathcal{F}$ une sous-tribu de $\mathcal{A}$. Nous définissons $P(A|\mathcal{F})$ par

$$P(A|\mathcal{F}) = E(\mathbb{1}_A|\mathcal{F}). \tag{35.110}$$

Notons que cela est une variable aléatoire et non un réel. Le membre de droite est l'espérance conditionnelle de la variable aléatoire $\mathbb{1}_A$ par rapport à $\mathcal{F}$ définie en 35.37.

Et l'espérance conditionnelle d'un événement par rapport à une variable aléatoire est :

$$E(A|X) = E\big(A|\sigma(A)\big). \tag{35.111}$$

Proposition 35.41.
Soit une espace probabilisé $(\Omega, \mathcal{A}, P)$ ainsi qu'une variable aléatoire X à valeurs dans $\mathbb{R}^d$, et un événement A. Alors

$$E\big(P(A|X)\big) = P(A). \tag{35.112}$$

Démonstration. Tout le point de la preuve est de remarquer que $E(\mathbb{1}_A) = E(\mathbb{1}_A|X)$.

La formule $E(\mathbb{1}_A) = E(\mathbb{1}_A|X)$ La notation $E(\mathbb{1}_A|X)$ est un raccourci pour écrire la variable aléatoire $E\big(\mathbb{1}_A|\sigma(X)\big)$. Cette dernière est l'application $\Omega \to \mathbb{R}^d$ telle que

$$\int_B E\big(\mathbb{1}_A|\sigma(X)\big) = \int_B \mathbb{1}_A \tag{35.113}$$

pour tout borélien B de $\mathbb{R}^d$ tout en étant $\sigma(X)$-mesurable. Comme expliqué en 35.38, il est tentant de dire $E\big(\mathbb{1}_A|\sigma(X)\big) = \mathbb{1}_A$, mais ce n'est pas le cas parce qu'il n'y a aucune raisons que $\mathbb{1}_A$ soit une application $\sigma(X)$-mesurable. Au niveau des espérances, par contre, l'égalité tient :

$$E\big(E(\mathbb{1}_A|X)\big) = \int_\Omega E(\mathbb{1}_A|X) = \int_\Omega \mathbb{1}_A = E(\mathbb{1}_A) \tag{35.114}$$

où nous avons utilisé le fait que Ω lui-même soit $\sigma(X)$-mesurable.

La preuve Nous avons alors

$$P(A) = E(\mathbb{1}_A) = E\big(E(\mathbb{1}_A|X)\big), \tag{35.115}$$

alors que $E(\mathbb{1}_A|X) = P(A|X)$. En mettant l'un dans l'autre :

$$P(A) = E\big(P(A|X)\big). \tag{35.116}$$

□

Proposition 35.42 (Transitivité de l'espérance conditionnelle)**.**
Si $\mathcal{B}_2 \subseteq \mathcal{B}_1 \subset \mathcal{A}$ alors

$$E\Big(E(X|\mathcal{B}_1)|\mathcal{B}_2\Big) = E(X|\mathcal{B}_2). \tag{35.117}$$

Démonstration. Si $B \in \mathcal{B}_2$, nous avons

$$\int_B E\big(E(X|\mathcal{B}_1)|\mathcal{B}_2\big)dP = \int_B E(X|\mathcal{B}_1)dP = \int_B dP. \tag{35.118}$$

La première égalité est la définition de l'espérance conditionnelle par rapport à $\mathcal{B}_2$. La seconde égalité est celle de l'espérance conditionnelle par rapport à $\mathcal{B}_1$ et le fait que $B \in \mathcal{B}_2 \subset \mathcal{B}_1$. Ce que nous avons prouvé est que

$$E\big(E(X|\mathcal{B}_1)|\mathcal{B}_2\big) \tag{35.119}$$

est une variable aléatoire $\mathcal{B}_2$-mesurable vérifiant la condition

$$\int_B E\big(E(X|\mathcal{B}_1)|\mathcal{B}_2\big) = \int_B E(X|\mathcal{B}_2) \tag{35.120}$$

pour tout $B \in \mathcal{B}_2$. C'est donc $E(X|\mathcal{B}_2)$ par la partie unicité du théorème 35.37. □

Proposition 35.43.
Soit $(\Omega, \mathcal{F}, P)$ un espace de probabilité, soit $\mathcal{A}$ une sous tribu de $\mathcal{F}$ et X, une variable aléatoire $\mathcal{F}$-mesurable et intégrable. Alors la variable aléatoire $E(X|\mathcal{A})$ du théorème 35.37 est l'unique (presque partout) variable aléatoire à être $\mathcal{A}$-mesurable telle que nous ayons

$$E\big(E(X|\mathcal{A})Y\big) = E(XY). \tag{35.121}$$

pour toute variable aléatoire Y $\mathcal{A}$-mesurable.

Démonstration. Supposons pour commencer que Y soit une fonction simple positive, alors $Y = \sum_{i=1}^n a_i \mathbb{1}_{E_i}$ et nous avons

$$\int_\Omega E(X|Y) = \sum_i a_i \int_{E_i} E(X|\mathcal{A}) \tag{35.122a}$$

$$= \sum_i a_i \int_{E_i} X \tag{35.122b}$$

$$= \int_\Omega XY. \tag{35.122c}$$

Maintenant si Y est mesurable et bornée, elle est limite croissante de fonctions étagées bornées (proposition 13.105) et le résultat tient par la convergence monotone, théorème 13.160.

Si Y n'est pas positive, nous séparons $Y = Y_+ - Y_-$.

Pour l'unicité, soit Z et Z' deux variables aléatoires telles que pour toute variable aléatoire Y,

$$\int_\Omega ZY = \int_\Omega XY = \int_\Omega Z'Y. \tag{35.123}$$

Si nous prenons $Y = \mathbb{1}_{\{Z \neq Z'\}}$, nous avons

$$0 = \int_\Omega (Z - Z')\mathbb{1}_{Z \neq Z'} = \int_{Z \neq Z'} Z - Z', \tag{35.124}$$

d'où le fait que $P(Z \neq Z') = 0$. □

Si X est une variable aléatoire dont la tribu engendrée est indépendante de la tribu $\mathcal{F}$, nous voudrions que la connaissance de $\mathcal{F}$ n'influence pas la connaissance de X, c'est-à-dire que

$$E(X|\mathcal{F}) = E(X). \tag{35.125}$$

Ce que nous avons est même mieux. Nous avons le lemme suivant.

Lemme 35.44 ([270]).
Les tribus $\mathcal{F}_1$ et $\mathcal{F}_2$ sont indépendantes si et seulement si

$$E(U|\mathcal{F}_1) = E(U) \tag{35.126}$$

pour toute variable aléatoire U étant $\mathcal{F}_1$-mesurable.

Ici, par $E(U)$ nous entendons la variable aléatoire constante prenant la valeur numérique $E(U)$ en tout point de Ω.

Démonstration. Si $\mathcal{F}_1$ et $\mathcal{F}_2$ sont indépendantes, alors pour tout $B \in \mathcal{F}_2$ nous avons

$$\begin{aligned}
\int_B U dP &= E(U\mathbb{1}_B) && (35.127a)\\
&= E(U)E(\mathbb{1}_B) && (35.127b)\\
&= E(U)\int_\Omega \mathbb{1}_B dP && (35.127c)\\
&= \int_B E(U)dP. && (35.127d)
\end{aligned}$$

Justifications.

— L'intégrale $\int_B UdP$ a un sens même si $B \in \mathcal{F}_2$ alors que U est $\mathcal{F}_1$-mesurable. Le supremum (13.424) définissant l'intégrale est tout de même bien défini, en particulier, l'ensemble sur lequel on prend le supremum est non vide.

— Pour (35.127b), la variable aléatoire U est $\mathcal{F}_1$-mesurable (donc la tribu engendrée par U est dans $\mathcal{F}_1$) alors que $\mathbb{1}_B$ est $\mathcal{F}_2$-mesurable. Les tribus engendrées étant indépendantes, les variables aléatoires le sont et nous pouvons décomposer l'espérance.

Ce que montre le calcul (35.127) est que $E(U)$ est une variable aléatoire $\mathcal{F}_2$-mesurable (parce que constante) dont l'intégrale sur chaque élément de $\mathcal{F}_2$ vaut l'intégrale de U. Par la partie unicité du théorème 35.37, nous déduisons que $E(U) = E(U|\mathcal{F}_2)$. □

Corolaire 35.45.
Si X est une variable aléatoire et si $\mathcal{F}$ est une tribu, alors

$$E\big(E(X|\mathcal{F})\big) = E(X). \tag{35.128}$$

Démonstration. Il suffit d'appliquer la définition (35.92) à $B = \Omega$:

$$E\big(E(X|\mathcal{F})\big) = \int_\Omega E(X|\mathcal{F})(\omega)dP(\omega) = \int_\Omega X(\omega)dP(\omega) = E(X). \tag{35.129}$$

□

Exemple 35.46.
Soient X_1, X_2 deux variables aléatoires à valeurs dans $\{0,1\}$ avec probabilité $1/2$ et indépendantes. Nous considérons $S = X_1 + X_2$. La situation est modélisée par l'espace

$$\Omega = \{(0,0),(0,1),(1,0),(1,1)\} \tag{35.130}$$

et les variables aléatoires

$$X_i(\omega_1, \omega_2) = \omega_i \tag{35.131a}$$
$$S(\omega_1, \omega_2) = \omega_1 + \omega_2. \tag{35.131b}$$

Pour vérifier que de cette manière nous avons bien que X_1 est indépendante de X_2, nous commençons par voir les tribus associées. Un ouvert de $\mathbb{R}$ soit contient 0 et 1, soit contient un seul des deux soit n'en contient aucun des deux. En appliquant X_1^{-1} à chacune de ces quatre situations nous voyons que la tribu $\sigma(X_1)$ est

$$\mathcal{F}_1 = \sigma(X_1) = \big\{\{(0,0),(0,1)\},\{(1,0),(1,1)\},\Omega,\varnothing\big\}. \tag{35.132}$$

De la même façon nous avons

$$\mathcal{F}_2 = \sigma(X_1) = \big\{\{(0,0),(1,0)\},\{(0,1),(1,1)\},\Omega,\varnothing\big\}. \tag{35.133}$$

Nous posons

$$A_0 = \{(0,0),(0,1)\} \tag{35.134a}$$
$$A_1 = \{(1,0),(1,1)\} \tag{35.134b}$$
$$B_0 = \{(0,0),(1,0)\} \tag{35.134c}$$
$$B_1 = \{(0,1),(1,1)\}. \tag{35.134d}$$

Étant donné que $A_i \cap B_j = (i,j)$, nous avons toujours que $P(A_i \cap B_j) = \frac{1}{4} = P(A_i)P(B_j)$. L'indépendance est donc assurée.

Calculons l'espérance conditionnelle $E(S|\mathcal{F}_1)$. Une fonction $\mathcal{F}_1$-mesurable doit être constante sur A_0 et A_1, donc l'espérance conditionnelle est une fonction constante sur A_0 et A_1 dont l'intégrale sur ces ensembles est égale à l'intégrale de S. Nous avons en particulier

$$\int_{A_0} E(S|\mathcal{F}_1) = \int_{A_0} S, \tag{35.135}$$

c'est-à-dire

$$E(S|\mathcal{F}_1)(0,0) + E(S|\mathcal{F}_1)(0,1) = S(0,0) + S(0,1) = 1. \tag{35.136}$$

Nous en concluons que $E(S|\mathcal{F}_1)(0,0) = E(S|\mathcal{F}_1)(0,1) = \frac{1}{2}$. Cela correspond à l'intuition que si on est au point $(0,1)$ ou au point $(0,0)$ en ne sachant que X_1, nous ne savons que le premier zéro, et donc l'espérance de la somme est $\frac{1}{2}$.

Un calcul très similaire montre que

$$E(S|\mathcal{F}_1)(1,0) = E(S|\mathcal{F}_1)(1,1) = \frac{3}{2}. \tag{35.137}$$

Cela correspond au fait qu'en ces points, nous ne savons que le fait que le premier tirage a donné 1, et donc que l'espérance est $\frac{3}{2}$.

Complétons ce tour d'horizon en mentionnant que la tribu engendrée par X_1 et X_2 est la tribu des parties de Ω, de telle façon que l'espérance conditionnelle de S sachant X_1 et X_2 est égale à S. △

Proposition 35.47 ([270]).
Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé et X, Y deux variables aléatoires sur Ω réelles. Soit $\mathcal{B}$ une sous-tribu de $\mathcal{A}$. Nous supposons que $X \in L^1(\Omega, \mathcal{A}, P)$, que $Z \in L^\infty(\Omega, \mathcal{B}, P)$ et que $XZ \in L^1(\Omega, P)$. Alors

$$E(ZX|\mathcal{B}) = ZE(X|\mathcal{B}) \tag{35.138}$$

presque surement.

Démonstration. Nous commençons par prouver que

$$\int_\Omega ZE(X|\mathcal{B}) = \int_\Omega ZX. \tag{35.139}$$

Si $Z = \mathbb{1}_B$ pour un ensemble $B \in \mathcal{B}$, alors cette égalité est vraie par définition de l'espérance conditionnelle[4]. Donc cette égalité est correcte tant que Z est une variable aléatoire $\mathcal{B}$-mesurable et étagée. Nous considérons alors, grâce au lemme 13.103, une suite Z_n de variables aléatoires étagées et $\mathcal{B}$-mesurables avec $|Z_n| < Z$. Pour chaque n nous avons donc

$$\int_\Omega Z_n X = \int_\Omega ZE(X|\mathcal{B}). \tag{35.140}$$

Notre idée est de passer à la limite. Vu que Z et Z_n sont bornées (et donc intégrables sur Ω), pour chaque n nous avons $|Z_n X| \leqslant M|X|$ où M majore Z et donc tous les Z_n de façon uniforme vis-à-vis de n. Tout cela pour dire que le théorème de la convergence dominée fonctionne et que

$$\lim_{n\to\infty} \int_\Omega Z_n X = \int_\Omega ZX. \tag{35.141}$$

D'autre part vu que $X \in L^1$ et que $\Omega \in \mathcal{B}$ nous avons l'égalité $\int_\Omega E(X|\mathcal{B}) = \int_\Omega X$, ce qui prouve que $|E(X|\mathcal{B})|$ est intégrable. Cela nous permet d'utiliser encore la convergence dominée avec l'inégalité $|Z_n E(X|\mathcal{B})| \leqslant |E(X|\mathcal{B})|$ pour écrire

$$\lim_{n\to\infty} \int_\Omega Z_n E(X|\mathcal{B}) = \int_\Omega ZE(X|\mathcal{B}). \tag{35.142}$$

En passant à la limite des deux côtés de (35.140) nous avons donc

$$\int_\Omega ZE(X|\mathcal{B}) = \int_\Omega ZX. \tag{35.143}$$

L'égalité (35.139) est prouvée.

Nous passons maintenant à la preuve de l'égalité demandée : $E(EX|\mathcal{B}) = ZE(X|\mathcal{B})$. Pour cela il faut montrer que pour tout $B \in \mathcal{B}$ nous avons

$$\int_B ZE(X|\mathcal{B}) = \int_B ZX. \tag{35.144}$$

Cela n'est rien d'autre que l'égalité (35.139) que nous venons de prouver avec $Z\mathbb{1}_B$ au lieu de Z. □

Proposition 35.48.
Soit une variable aléatoire réelle $X \in L^1(\Omega, \mathcal{A}, P)$. Pour toute variable aléatoire $Y\colon \Omega \to \mathbb{R}^d$, il existe une fonction borélienne $\mathcal{A}_Y$-mesurable $h\colon \mathbb{R}^d \to \mathbb{R}$ telle que

$$E(X|Y) = h \circ Y. \tag{35.145}$$

Démonstration. Nous utilisons le résultat de Doob (théorème 35.14). Par définition $E(X|Y)$ est une variable aléatoire réelle $\mathcal{A}_Y$-mesurable, et il existe une fonction borélienne $h\colon \mathbb{R}^d \to \mathbb{R}$ telle que $E(X|Y) = f \circ Y$. □

Cette fonction $h\colon \mathbb{R}^d \to \mathbb{R}$ nous permet de définir

$$E(X|Z = z) = h(z). \tag{35.146}$$

Cela est l'espérance conditionnelle d'une variable aléatoire par rapport à une valeur donnée d'une autre variable aléatoire.

4. Théorème 35.37.

Définition 35.49 (Espérence conditionnelle à un événement[521]).
Si X est une variable aléatoire et si A est un événement, nous définissons

$$E(X|A) = \frac{E(X\mathbb{1}_A)}{P(A)}. \tag{35.147}$$

Pour rappel, la définition de l'espérance d'une variable aléatoire est 35.26.

Lemme 35.50 ([1]).
Soit un espace de probabilité $(\Omega, \mathcal{F}, P)$. Nous considérons une variable aléatoire X à valeurs réelles, prenant ses valeurs dans la partie dénombrable

$$\{0\} \cup \{y_k\}_{k\in\mathbb{N}} \tag{35.148}$$

avec $y_k \neq 0$. Si A est mesurable dans Ω, alors

$$E(X|A) = \sum_{k=0}^{\infty} y_k P(X = y_k|A), \tag{35.149}$$

c'est à dire que nous prenons la somme sur les valeurs non nulles atteintes par X.

Démonstration. Nous partons de la définition 35.49 de l'espérance conditionnelle : $E(X|A) = E(X\mathbb{1}_A)/P(A)$. La variable aléatoire $X\mathbb{1}_A$ peut prendre les valeurs 0 et y_k. Nous pouvons utiliser le lemme 35.27 pour écrire

$$E(X\mathbb{1}_A) = 0 \times P(X\mathbb{1}_A = 0) + \sum_{k=0}^{\infty} y_k P(X\mathbb{1}_A = y_k) = \sum_{k=0}^{\infty} y_k P(X\mathbb{1}_A = y_k). \tag{35.150}$$

Afin de traiter le dernier terme, nous prouvons que $\{X\mathbb{1}_1 = y_k\} = X^{-1}(y_k) \cap A$. En effet si $\omega \in \{X\mathbb{1}_1 = y_k\}$, c'est que

$$X(\omega)\mathbb{1}_A(\omega) = y_k. \tag{35.151}$$

Mais dans notre cas, $y_k \neq 0$, donc $\mathbb{1}_A(\omega) = 1$, ce qui signifie $\omega \in A$. Donc $\omega \in A$ et $X(\omega) = y_k$, ce qui signifie $\omega \in X^{-1}(y_k) \cap A$. Nous avons donc

$$P(X\mathbb{1}_A = y_k) = P\big(X^{-1}(y_k) \cap A\big). \tag{35.152}$$

En utilisant (à l'envers) la définition de la probabilité conditionnelle 35.33,

$$P(X\mathbb{1}_A = y_k) = P\big(X^{-1}(y_k)|A\big)P(A). \tag{35.153}$$

En remettant ça dans (35.150) et dans la définition de $E(X|A)$,

$$E(X|A) = \frac{\sum_{k=0}^{\infty} y_k P(X\mathbb{1}_A = y_k)}{P(A)} = \frac{1}{P(A)} \sum_{k=0}^{\infty} P(X = y_k|A)P(A) = \sum_{k=0}^{\infty} y_k P(X = y_k|A) \tag{35.154}$$

□

35.2.9 Probabilité conditionnelle : tribu

Soit un espace de probabilisé $(\Omega, \mathcal{A}, P)$.

Lemme 35.51.
Soit $(B_i)_{i\in\mathbb{N}}$ une partition de Ω en éléments de $\mathcal{A}$ deux à deux disjoints tels que $P(B_i) \neq 0$. Soit $\mathcal{F}$ la tribu engendrée par les B_i. Une variable aléatoire réelle est $\mathcal{F}$-mesurable si et seulement si elle est constante sur chaque B_i.

Proposition 35.52.
Soit $(B_i)_{i\in\mathbb{N}}$ une partition de Ω en éléments de $\mathcal{A}$ deux à deux disjoints tels que $P(B_i)\neq 0$. Soit $\mathcal{F}$ la tribu engendrée par les B_i. Soit une variable aléatoire X. Alors nous avons :

$$E(X|\mathcal{F})=\sum_{i\in\mathbb{N}}\left(\frac{1}{P(B_i)}\int_{B_i}XdP\right)\mathbb{1}_{B_i}. \tag{35.155}$$

Démonstration. Si X est une variable aléatoire, alors la variable aléatoire $E(X|\mathcal{F})$ définie en 35.37 est une variable aléatoire $\mathcal{F}$-mesurable et elle est donc constante sur les ensembles B_i par le lemme 35.51 :

$$E(X|\mathcal{F})=\sum_{i\in\mathbb{N}}a_i\mathbb{1}_{B_i}. \tag{35.156}$$

Étant donné que, par construction, B_i est $\mathcal{F}$-mesurable, nous avons

$$\int_{B_i}XdP=\int_{B_i}E(X|\mathcal{F})=\sum_j a_j\int_{B_i}\mathbb{1}_{B_j}=\sum_j a_j\delta_{ij}P(B_j)=a_iP(B_i). \tag{35.157}$$

Par conséquent

$$a_i=\frac{1}{P(B_i)}\int_{B_i}XdP \tag{35.158}$$

et

$$E(X|\mathcal{F})=\sum_{i\in\mathbb{N}}\left(\frac{1}{P(B_i)}\int_{B_i}XdP\right)\mathbb{1}_{B_i}, \tag{35.159}$$

ce qu'il fallait. □

Notons que si $B\in\mathcal{A}$ alors la tribu engendrée par B est aussi celle engendrée par la partition $\{B,\complement B\}$ de Ω. Cette circonstance nous permet d'aller plus loin.

Proposition 35.53.
Soit un espace probabilisé $(\Omega,\mathcal{A},P)$ et un événement $B\in\mathcal{A}$ avec sa tribu engendrée $\mathcal{F}=\sigma(B)$. Alors

$$E(\mathbb{1}_A|\mathcal{F})=P(A|B)\mathbb{1}_B+P(A|\complement B)\mathbb{1}_{\complement B}. \tag{35.160}$$

Démonstration. Nous allons particulariser la formule (35.155). Si $B\in\mathcal{A}$ nous considérons la partition $\{B,\complement B\}$ de Ω et la tribu engendrée

$$\mathcal{F}=\{\emptyset,B,\complement B,\Omega\}. \tag{35.161}$$

La formule (35.155) devient

$$E(X|\mathcal{F})=\left(\frac{1}{P(B)}\int_B XdP\right)\mathbb{1}_B+\left(\frac{1}{P(\complement B)}\int_{\complement B}XdP\right)\mathbb{1}_{\complement B}. \tag{35.162}$$

Si nous considérons $A\in\mathcal{A}$, nous écrivons cette égalité avec $X=\mathbb{1}_A$ pour obtenir

$$E(\mathbb{1}_A|\mathcal{F})=\frac{P(A\cap B)}{P(B)}\mathbb{1}_B+\frac{P(A\cap\complement B)}{P(\complement B)}\mathbb{1}_{\complement B}=P(A|B)\mathbb{1}_B+P(A|\complement B)\mathbb{1}_{\complement B} \tag{35.163}$$

parce que nous avons reconnu la probabilité conditionnelle $P(A|B)$ de la définition 35.33. □

Remarque 35.54.
Les nombres $P\big(A|\sigma(B)\big)=P\big(\mathbb{1}_A|\sigma(B)\big)$ n'est pas la probabilité conditionnelle de A sachant B.

Il nous reste à définir la probabilité conditionnelle d'un événement relativement à une variable aléatoire.

Définition 35.55.
Si la variable aléatoire X est à valeurs discrètes, nous disons que $P(A|X)$ est la variable aléatoire de valeur

$$P(A|X)(\omega) = P(A|X = X(\omega)). \tag{35.164}$$

Dans le cas d'une variable aléatoire à valeurs continues, cette définition ne fonctionne pas parce que la condition $X = X(\omega)$ est souvent de probabilité nulle, tandis que c'est toujours une mauvaise idée de conditionner par rapport à un événement de probabilité nulle. C'est la base du paradoxe de Borel. La bonne définition du conditionnement de l'événement A par rapport à la variable aléatoire X est

Définition 35.56.
Si A est un événement et X une variable aléatoire à valeurs continues dans $\mathbb{R}$, nous définissons

$$P(A|X) = P(A|\sigma(X)) = E\big(\mathbb{1}_A|\sigma(X)\big). \tag{35.165}$$

La première égalité est une notation. La seconde est la définition.

Cette définition s'appuie également sur la définition 35.37.

Proposition 35.57.
Si X est une variable aléatoire et si A est un événement, alors

$$E\big(P(A|X)\big) = P(A). \tag{35.166}$$

Démonstration. Nous commençons par le cas discret, c'est-à-dire $X\colon \Omega \to \mathbb{N}$. Nous notons $p_k = P(X = k)$. En décomposant l'intégrale sur Ω par rapport à l'union disjointe

$$\Omega = \bigcup_{k\in\mathbb{N}} A_k = \bigcup_{k\in\mathbb{N}} \{\omega \in \Omega \text{ tel que } X(\omega) = k\}, \tag{35.167}$$

nous obtenons

$$\begin{aligned}
E\big(P(A|X)\big) &= \int_\Omega P(A|X)(\omega)dP(\omega) && \text{(35.168a)}\\
&= \sum_{k=0}^{\infty} \int_{A_k} P(A|X = X(\omega))dP(\omega) && \text{(35.168b)}\\
&= \sum_k \int_{A_k} \frac{P(A\cap X = k)}{P(X = k)} dP(\omega) \qquad \text{dans } A_k,\ X(\omega) = k && \text{(35.168c)}\\
&= \sum_k \frac{1}{p_k} P(A\cap X = k) \underbrace{\int_{A_k} 1dP(\omega)}_{=P(A_k)=p_k} && \text{(35.168d)}\\
&= \sum_k P(A\cap X = k) && \text{(35.168e)}\\
&= P(A). && \text{(35.168f)}
\end{aligned}$$

Nous devons maintenant prouver la propriété dans le cas où X prend des valeurs continues. Pour cela il suffit d'appliquer le corolaire 35.45 :

$$E\big(E(\mathbb{1}_A|\sigma(A))\big) = E(\mathbb{1}_A) = P(A). \tag{35.169}$$

□

35.2.10 Variables de Rademacher indépendantes

Une variable aléatoire de Rademacher est une variable aléatoire qui prend les valeurs 1 et -1 avec probabilité $\frac{1}{2}$. Nous pouvons en décrire une explicitement de la façon suivante. L'espace probabilité est à deux éléments : $\Omega = \{a, b\}$ avec la mesure $P(\{a\}) = P(\{b\}) = \frac{1}{2}$. La variable aléatoire est alors l'application $X\colon \Omega \to \mathbb{R}$ donnée par $X(a) = 1$ et $X(b) = -1$.

Soient X et Y deux variables aléatoires de Rademacher indépendantes. Cela donne $\Omega = \{a, b\}^2$ et

$$\begin{array}{llll} X(a,a) = 1 & X(a,b) = 1 & X(b,a) = -1 & X(b,b) = -1 \\ Y(a,a) = 1 & Y(a,b) = -1 & Y(b,a) = 1 & Y(b,b) = -1 \end{array} \tag{35.170}$$

Remarque 35.58.
Si une variable aléatoire d'un certain type est donnée par une application $X\colon \Omega \to \mathbb{R}$, pour construire des variables aléatoires indépendantes identiquement distribuées, il faut considérer les variables aléatoires sur (au moins) le produit $\Omega \times \Omega$ munie de la mesure produit.

Tribu du produit XY Quelle est la tribu de la variable aléatoire produit XY ? Le produit XY peut prendre les valeurs 1 et -1. Nous avons

$$(XY)^{-1}(1) = \{(a,a),(b,b)\}(XY)^{-1}(-1) \quad = \{(a,b),(b,a)\} \tag{35.171}$$

La tribu est donc

$$\sigma(XY) = \{\Omega, \varnothing, A, B\} \tag{35.172}$$

avec $A = \{(a,a),(b,b)\}$ et $B = \{(a,b),(b,a)\}$.

Calcul de $E(X|XY)$ La définition de l'espérance à calculer est le théorème 35.37. Pour chaque élément B de $\sigma(XY)$ nous avons besoin de $\int_B X = \int_B E(X|XY)$. Nous notons $V = E(X|XY)$ pour alléger la notation. Nous avons

$$4\int_A V = V(a,a) + V(b,b) \tag{35.173}$$

et

$$4\int_A X = X(a,a) + X(1,1) = 0. \tag{35.174}$$

Pourquoi le facteur 4 ? Parce que sur Ω nous avons la mesure produit de celle que dont nous avions parlé sur $\{a, b\}$. C'est la mesure d'équiprobabilité et donc chaque singleton a mesure 1/4. Pour plus de détails, il y a le théorème 13.210.

Nous en déduisons $V(a,a) + V(b,b) = 0$. Mais pour tout $t \in \mathbb{R}$ nous avons $V^{-1}(t) \in \sigma(XY)$ parce que la contrainte est que V soit XY-mesurable. En particulier

$$V^{-1}\big(V(a,a)\big) \tag{35.175}$$

est un mesurable qui contient (a,a). C'est donc soit Ω, soit $\{(a,a),(b,b)\}$. Dans les deux cas nous avons $V(a,a) = V(b,b)$ et nous en déduisons $V(a,a) = V(b,b) = 0$.

En faisant de même avec $\int_B V = V(a,b) + V(b,a)$ nous déduisons $V(a,b) = V(b,a) = 0$ et au final nous avons

$$E(X|XY) = 0. \tag{35.176}$$

Cette égalité signifie $E(X|XY)(\omega) = 0$ pour tout $\omega \in \Omega$.

Calcul de $E(X|X+Y)$ Il ne faudrait pas croire que, seulement parce que X a une espérance nulle, nous trouverons une espérance nulle quel que soit le conditionnement. Juste pour le plaisir, nous calculons $E(X|X+Y)$.

La variable aléatoire $X+Y$ peut prendre trois valeurs : -2, 0 et 2. La tribu engendrée par $X+Y$ doit en particulier contenir $A = \{(a,a)\}$, $B = \{(b,b)\}$ et $C = \{(a,b),(b,a)\}$.

Nous notons $V = E\big(X|\sigma(X+Y)\big)$. Vu que

$$\int_A V = \int_A V, \tag{35.177}$$

nous avons $V(a,a) = X(a,a) = 1$. Même chose pour B qui donne $V(b,b) = X(b,b) = -1$. En ce qui concerne l'intégrale sur C nous avons

$$V(a,b) + V(b,a) = X(a,b) + X(b,a) = 0. \tag{35.178}$$

Par ailleurs l'ensemble $V^{-1}\big(V(a,b)\big)$ est un ensemble mesurable qui doit au moins contenir (a,b). Vu la tribu que nous avons, cela doit également contenir (b,a), de telle sorte que $V(a,b) = V(b,a)$. La relation (35.178) nous permet alors de conclure que $V(a,b) = V(b,a) = 0$.

Quoi qu'il en soit, l'espérance conditionnelle $E(XY|X+Y)$ n'est pas nulle.

Calcul de $E(XY|\sigma(XY))$. Celle-là, elle est facile par 35.38 : c'est XY.

Nous aurions pu croire que si X et Y sont indépendantes, alors

$$E(XY|\mathcal{A}) = E(X|\mathcal{A})E(Y|\mathcal{A}). \tag{35.179}$$

L'exemple que nous venons de faire montre qu'il n'en est rien.

Exemple 35.59 ([522]).
Un autre exemple, peut-être plus simple, pour contredire l'équation (35.179). Soient X et Y des expériences indépendantes de pile ou face non truquées. Les résultats sont représentés par 0 et 1. Nous notons $\mathcal{A}$ la tribu engendrée par l'événement « les résultats des deux lancers sont différents » ; c'est-à-dire la tribu engendrée par l'événement $A = (1,0),(1,0)$. La variable aléatoire X et la tribu $\mathcal{A}$ sont indépendants (définition 35.6), donc, donc $E(X|\mathcal{A}) = E(X) = 1/2$. Pareil pour Y. En revanche, le produit XY est nul sur A donc $E(XY|\mathcal{A})$ aussi. Ça ne peut donc être égal à la constante $1/4 = (1/2)^2$. △

35.2.11 Un petit paradoxe

Attention : ce qui est écrit ici est ma réflexion personnelle sur le sujet. Merci de me dire si je me trompe.

Soit une famille dont vous savez seulement qu'il y a exactement deux enfants. Trois situations :

(1) Vous frappez, une fille ouvre la porte et dit « Bonjour, je suis l'aînée ». Quelle est la probabilité que l'autre enfant soit une fille ?
(2) Vous frappez, une fille ouvre la porte et dit « Bonjour ». Quelle est la probabilité que l'autre enfant soit une fille ?
(3) Vous demandez aux parents s'il y a au moins une fille, ils répondent « oui ». Quelle est la probabilité que les deux enfants soient des filles ?

Dans les trois cas l'intuition dit que la probabilité est 1/2. Il semble que de plus la (2) et la (3) soient les mêmes parce que l'on sait qu'il y a une fille et on se demande quelle est la probabilité qu'il y ait deux filles.

Nous allons voir ça de plus près.

35.2.11.1 « Bonjour, je suis l'aînée »

Résolution Si nous notons X_0 et X_1 les variables aléatoires donnant le sexe des deux enfants, ce sont des variables aléatoires indépendantes et identiquement distribuées, avec $P(X_i = f) = \frac{1}{2}$. La formule (35.75) de la probabilité conditionnelle ainsi que l'indépendance donnent :

$$P(X_1 = f|X_2 = f) = \frac{P(X_1 = f, X_2 = f)}{P(X_2 = f)}. \tag{35.180}$$

Le numérateur vaut $\frac{1}{4}$ et le dénominateur vaut $\frac{1}{2}$; le résultat vaut $\frac{1}{2}$. Fin de l'histoire.

Simulation Voici un petit programme qui simule la situation. Il retourne clairement 1/2.

```
#! /usr/bin/python3
# -*- coding: utf8 -*-

"""
Vous frappez à la porte d'une famille qui a deux enfants. Une ↩
   fille ouvre la porte et vous dit "Je suis l'aînée".
Quelle est la probabilité que l'autre soit une fille ?
"""


import random

def famille():
    """
    return a pair of 'f' and 'g'.
    """
    a=[random.choice( ['g','f'] )]
    a.append(random.choice( ['g','f'] ))
    return a

def toctoc():
    """
    - Create a family with two children.
    - Pick the second one, the elder.
    - if it is a 'g', return None.
    - if it is a 'f', return the other one.
    """
    F=famille()
    if F[1] != 'f':
        return None
    else :
        if F[0]=='f':
            return 1
        else :
            return 0

N_girl_opens=0
N_girl_other=0
for k in range(1,10000):
    res=toctoc()
    if res is not None:
        N_girl_opens = N_girl_opens+1
        N_girl_other = N_girl_other + res

proba=N_girl_other/N_girl_opens
print(proba)        # ~0.5, intuitively correct.
```

tex/sage/simul_famille_aine.py

35.2.11.2 « Bonjour »

Nous frappons à la porte, une fille ouvre en disant « bonjour », sans préciser si elle est la première ou la seconde. Quelle est la probabilité que l'autre soit une fille ? Naïvement on croirait que la probabilité est également $\frac{1}{2}$.

Un raisonnement moins naïf montre le contraire.

Et nous allons voir qu'un raisonnement encore moins naïf montre que la probabilité est bien $\frac{1}{2}$.

Premier raisonnement (incorrect) Voici le raisonnement qui est, à mon avis, faux. Vu que l'enfant qui ouvre la porte est une fille, la famille a une des compositions suivantes : fg, ff ou gf. Le cas où une fille ouvre la porte *et* que l'autre est également une fille est seulement le cas ff dont la probabilité est $\frac{1}{3}$.

Pour justifier cela nous considérons le couple de variables aléatoires (X_1, X_2) et le conditionnement $A = \{X_1 = f\} \cup \{X_2 = f\}$: évidemment $P(A) = \frac{3}{4}$. Nous calculons facilement la loi du couple (X_1, X_2) conditionné à A :

$$P(X_1 = f, X_2 = f|A) = \frac{P(\{X_1 = f, X_2 = f\} \cap A)}{P(A)} = \frac{1/4}{3/4} = \frac{1}{3}. \tag{35.181}$$

Donc sachant A, la probabilité que la famille soit constituée de deux filles est $\frac{1}{3}$.

Comment faire mieux ? Ce calcul semble être correct, mais il ne l'est pas. Ce raisonnement fait l'hypothèse implicite que l'espace probabilisé décrivant la situation contient deux variables aléatoires X_1 et X_2 représentant les deux enfants. Or nous avons bien trois événements aléatoires dans l'histoires : le sexe des deux enfants et le *choix* de l'enfant qui ouvre la porte.

Certes, nous pouvons penser que cette troisième variable aléatoire ne change rien. Oui oui, on peut le penser. Mais ici, on ne doit pas penser, on doit *démontrer*.

Nous allons donc rédiger un calcul complet, en introduisant toutes les variables aléatoires, et en décrivant correctement l'espace probabilisé Ω et la mesure de probabilité P.

Peut-être que ça ne changera rien. Ou peut-être pas. Mais au moins nous serons surs d'avoir résolu le problème correctement.

La vraie réponse Nous considérons les variables aléatoires $X_0, X_1 \colon \Omega_E \to \{f, g\}$ avec probabilité $\frac{1}{2}$. De plus nous considérons une nouvelle variable aléatoire qui donne le numéro de l'enfant qui ouvre la porte :

$$\sigma \colon \Omega_C \to \{1, 2\}. \tag{35.182}$$

Notre espace de probabilité est donc l'ensemble $\Omega = \{f, g\} \times \{f, g\} \times 0, 1$ sur lequel nous considérons la mesure d'équiprobabilité [5].

Nous introduisons les variables aléatoires [6]

$$\begin{aligned} X_1 \colon \Omega &\to \{f, g\} \\ (s_1, s_2, n) &\mapsto s_1 \end{aligned} \tag{35.183}$$

et

$$\begin{aligned} X_2 \colon \Omega &\to \{f, g\} \\ (s_1, s_2, n) &\mapsto s_2 \end{aligned} \tag{35.184}$$

et

$$\begin{aligned} \sigma \colon \Omega &\to \{1, 2\} \\ (s_1, s_2, n) &\mapsto n \end{aligned} \tag{35.185}$$

5. C'est une hypothèse forte faisant appel d'un part ce que l'on sait de la reproduction humaine, et d'autre part ce que l'on sait de la sociologie de deux enfants qui entendent une sonnette.

6. Sur Ω, sur $\{f, g\}$ et sur $\{0, 1\}$ nous mettons la tribu des parties. Vérifiez que X_1, X_2 et σ sont mesurables.

Nous devons calculer

$$P\big(X_{1-\sigma}=f|X_\sigma=f\big)=\frac{P\big(X_{1-\sigma}=f,X_\sigma=f\big)}{P(X_\sigma=f)}. \tag{35.186}$$

Pour être explicite jusqu'au bout, nous énumérons tous les éléments de Ω :

(1) $g,g,0$	(3) $g,f,0$	(5) $f,g,0$	(7) $f,f,0$
(2) $g,g,1$	(4) $g,f,1$	(6) $f,g,1$	(8) $f,f,1$.

Et tant qu'à être explicite, l'événement vulgairement noté $\{X_\sigma=f\}$ est la partie

$$\{X_\sigma=f\}=\{\omega\in\Omega \text{ tel que } X_{\sigma(\omega)}(\omega)=f\} \tag{35.187a}$$
$$=\{(s_1,s_2,n) \text{ tel que } X_n(s_1,s_2,n)=f\} \tag{35.187b}$$
$$=\{(s_1,s_2,n) \text{ tel que } s_n=f\}. \tag{35.187c}$$

Méditez la dernière égalité ; elle n'est pas totalement indispensable au raisonnement, mais elle est cool.

Nous avons

$$\{X_\sigma=f\}=\{(g,f,1),(f,g,0),(f,f,0),(f,f,1)\}. \tag{35.188}$$

et

$$\{X_{1-\sigma}=f\}\cap\{X_\sigma=f\}=\{(f,f,0),(f,f,1)\}. \tag{35.189}$$

Donc

$$P\big(X_{1-\sigma}=f,X_\sigma=f\big)=\frac{2}{8}=\frac{1}{4} \tag{35.190}$$

et

$$P(X_\sigma=f)=\frac{4}{8}=\frac{1}{2}. \tag{35.191}$$

Au final,

$$P\big(X_{1-\sigma}=f|X_\sigma=f\big)=\frac{1/4}{1/2}=\frac{2}{4}=\frac{1}{2}. \tag{35.192}$$

Simulation Vous avez encore un doute ? Faites tourner la simulation suivante :

```python
#! /usr/bin/python3
# -*- coding: utf8 -*-

"""
Vous frappez à la porte d'une famille qui a deux enfants. Une fille ouvre la porte.
Quelle est la probabilité que l'autre soit une fille ?
"""


import random

def famille():
    """
    return a pair of 'f' and 'g'
    """
    a=[random.choice( ['g','f'] )]
    a.append(random.choice( ['g','f'] ))
    return a

```

```
def toctoc():
    """
    - Create a family with two children.
    - Choose one (the one who opens the door)
    - if it is a 'g', return None.
    - if it is a 'f', return the other one.
    """
    F=famille()
    s=random.choice([0,1])
    if F[s] != 'f':
        return None
    else :
        t=(s+1)%2
        if F[t]=='f':
            return 1
        else :
            return 0

N_girl_opens=0
N_girl_other=0
for k in range(1,10000):
    res=toctoc()
    if res is not None:
        N_girl_opens = N_girl_opens+1
        N_girl_other = N_girl_other + res

proba=N_girl_other/N_girl_opens
print(proba)        # ~0.5, intuitively correct.
```

tex/sage/simul_famille_simple.py

Le faisant tourner, la réponse est sans appel : la fréquence observée est beaucoup plus proche de 0.5 que de 0.33 ou 0.66.

35.2.11.3 Le parent qui répond aux questions

Nous avons une famille de deux enfants dont nous savons qu'au moins un des deux est une fille. Quelle est la probabilité que la famille contienne deux filles ? Cela est à priori la même question que celle où une fille ouvre la porte sans dire si elle est l'aînée ou non.

Simulation Commençons par la simulation :

```
#! /usr/bin/python3
# -*- coding: utf8 -*-

"""
Vous savez qu'une famille a deux enfants.
Vous demandez à un parent si il y a une fille.
Réponse : Oui.
Question : quelle est la probabilité que ce soient deux filles ?
"""


import random
```

```python

def famille():
    """
    return a pair of 'f' and 'g'.
    """
    a=[random.choice( ['g','f'] )]
    a.append(random.choice( ['g','f'] ))
    return a

def toctoc():
    """
    - Create a family with two children.
    - If 'gg', there are no girls -> return None
    - If 'gf', there is a girl but the other is a boy -> 0
    - If 'fg', there is a girl but the other is a boy -> 0
    - If 'ff', there is a girl and the other is a girl -> 1
    """
    F=famille()
    if F==['g','g'] :
        return None
    if F==['g','f']:
        return 0
    if F==['f','g']:
        return 0
    if F==['f','f']:
        return 1

N_at_least_one_girl=0
N_two_girls=0
for k in range(1,10000):
    res=toctoc()
    if res is not None:
        N_at_least_one_girl=N_at_least_one_girl+1
        N_two_girls=N_two_girls+res

proba=N_two_girls/N_at_least_one_girl
print(proba)        # ~0.333. Beware !
```

tex/sage/simul_famille_une_fille.py

Et là, bing, la réponse est clairement plutôt 0.33 que 0.5.

Résolution Nous avons les variables aléatoires X_1 et X_2 qui valent 0 ou 1 suivant que l'enfant soit une fille ou un garçon ; ce sont des variables aléatoires indépendantes et identiquement distribuées. Nous définissons la variable aléatoire somme

$$S = X_1 + X_2 \tag{35.193}$$

qui compte le nombre de filles. La question est de calculer

$$P(S = 2|S \geqslant 1) = \frac{P(S = 2 \cap S \geqslant 1)}{P(S \geqslant 1)} = \frac{P(S = 2)}{P(S \geqslant 1)}. \tag{35.194}$$

L'événement $S = 2$ est réduit au singleton $\{ff\}$ et sa probabilité est $\frac{1}{4}$. Au contraire l'événement $S \geqslant 1$ est l'ensemble $\{fg, gf, ff\}$ et sa probabilité est $\frac{3}{4}$. Nous avons donc

$$P(S = 2|S \geqslant 1) = \frac{1/4}{3/4} = \frac{1}{3}. \tag{35.195}$$

Et là, la réponse est 1/3 et non 1/2 comme d'aucuns auraient pu le croire.

Précision Notons que l'événement $S \geqslant 1$ n'est pas le même que l'événement $X_\sigma = f$. En effet

$$S \geqslant 1 = \{(ff,1),(ff,2),(fg,1),(fg,2),(gf,1),(gf,2)\} \tag{35.196}$$

tandis que

$$\{X_\sigma = f\} = \{(ff,1),(ff,2),(fg,1),(gf,2)\}. \tag{35.197}$$

35.2.11.4 Conclusion

L'internet regorge de sites discutant du paradoxe des deux enfants[7].

Beaucoup insistent sur le fait que non seulement certaines informations apparemment anodines sont importantes, mais en plus *la façon* dont on obtient l'information est importante. Dans la situation « une fille ouvre », nous obtenons l'information « il y a au moins une fille » en en voyant une ; dans la situation « la parent dit qu'il y a au moins une fille », nous obtenons l'information « il y a au moins une fille » de façon plus « pure ».

Personnellement je ne souscris pas vraiment à cette façon de penser. Le fait est que la formule

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{35.198}$$

n'est pas seulement une formule dans laquelle il faut remplacer A par « la question » et B par « ce qu'on sait ». Il faut également remplacer P par « la bonne » mesure de probabilité.

Il est important de construire le bon espace de probabilité, avec la bonne mesure. Et pour cela, il faut bien s'assurer d'introduire une variable aléatoire pour chaque événement aléatoire se produisant dans l'histoire.

35.2.11.5 À propos des simulations

Si vous lisez ces lignes avec l'intention de passer l'agrégation en utilisant Sage à l'épreuve de modélisation, vous devez être capable de refaire les trois simulations. Les bouts de code donnés ici sont écrits pour python3 alors que Sage utilise Python2. Je ne vous dit pas si ça change quelque chose.

Allez oui, je vous dit. Si vous changez dans `simul_famille_une_fille.py` la première ligne pour utiliser python2 au lieu de python3, le résultat affiché sera `0` et non `0.333`. La raison est que dans Python2, l'opérateur `/` entre deux entiers est une **division entière**. Autrement dit : le résultat 0.33 est arrondi à zéro.

Solution : forcer python à interpréter le `/` comme une vraie division. Pour Sage, ça donne ceci comme début de programme :

```
#! /usr/bin/sage
# -*- coding: utf8 -*-

from __future__ import division
```

tex/frido/codeSnip_3.py

Importez toujours `division` de `__future__` .

Ah oui, et dernière remarque : pour autant que je le sache, le jour de l'oral, vous n'aurez que Sage en mode notebook. Je ne sais pas si l'import fonctionne aussi bien.

Sinon vous pouvez forcer la division dans les `float` de la façon suivante : `a/float(b)`.

7. Par exemple [523].

35.2.12 Inégalité de Jensen

Proposition 35.60 (Inégalité de Jensen)**.**
Soit g une fonction convexe[8] *sur $\mathbb{R}$ et une variable aléatoire $Y \in L^1(\Omega, \mathcal{A}, P)$ telle que $g \circ Y$ soit également L^1. Alors*

$$g\big(E(Y|\mathcal{F})\big) \leqslant E\big((g \circ Y)|\mathcal{F}\big). \tag{35.199}$$

Démonstration. La convexité de g et la proposition 16.89 nous donnent deux suites (a_n) et (b_n) dans $\mathbb{R}$ telles que pour tout $x \in \mathbb{R}$,

$$g(x) = \sup_{n\in\mathbb{N}}(a_n x + b_n). \tag{35.200}$$

Nous avons alors

$$a_n E(Y|\mathcal{F}) \overset{p.s.}{=} E\big(a_n Y + b_n|\mathcal{F}\big) \leqslant E(g \circ Y|\mathcal{F}). \tag{35.201}$$

L'inégalité est due au fait que $g \circ Y$ est le supremum sur les n de $a_n Y + b_n$. Pour chaque n, l'inégalité (35.201) est fausse sur un ensemble de mesure nulle $R_n \subset \Omega$. L'union

$$R = \bigcup_{n\in\mathbb{N}} R_n \tag{35.202}$$

est encore de mesure nulle. Sur $\Omega \backslash R$, nous avons

$$a_n E(Y|\mathcal{F}) + b_n \leqslant E(g \circ Y|\mathcal{F}). \tag{35.203}$$

Vu que cela est vrai presque partout et pour tout n nous passons a supremum et nous avons encore presque partout l'inégalité

$$\sup_{n\in\mathbb{N}} \big(a_n E(Y|\mathcal{F}) + b_n\big) \leqslant E(g \circ Y|\mathcal{F}). \tag{35.204}$$

□

Si nous ne nous intéressons pas à $E(Y|\mathcal{F})$ mais seulement à $E(Y)$, alors une démonstration plus simple est donnée sur Wikipédia[524].

35.2.13 Fonction de répartition

Définition 35.61.
Si X est une variable aléatoire réelle, nous définissons sa ***fonction de répartition*** *par*

$$\begin{aligned} F_X &: \mathbb{R} \to [0,1] \\ F_X(x) &= P(X \leqslant x). \end{aligned} \tag{35.205}$$

Remarque 35.62.
La fonction de répartition est discontinue en a si $P(X = a) > 0$. En particulier nous ne pouvons pas dire

$$P(X \geqslant a) = 1 - F_X(a). \tag{35.206}$$

35.2.14 Fonction caractéristique

Définition 35.63.
La ***fonction caractéristique*** *de la variable aléatoire $X : \Omega \to \mathbb{R}$ est la fonction réelle définie par*

$$\Phi_X(t) = E(e^{itX}). \tag{35.207}$$

8. Définition 16.76.

Une autre façon d'écrire la définition est

$$\Phi_X(t) = \int_{\mathbb{R}} e^{itX} dP_X(x), \tag{35.208}$$

ou encore, si X a une densité f_X,

$$\Phi_X(t) = \int_{\mathbb{R}} e^{itx} f_X(x) dx \tag{35.209}$$

Nous reconnaissons la transformée de Fourier :

$$\Phi_X(t) = \hat{f}_X(-t/2\pi). \tag{35.210}$$

La proposition suivante se déduit en utilisant le théorème de dérivation sous l'intégrale 16.21.

Proposition 35.64.
Soit X une variable aléatoire qui accepte un moment d'ordre $r \geqslant 1$. Alors la fonction caractéristique Φ_X est r fois continument dérivable et

$$\Phi_X^{(r)}(t) = E\big((iX)^r e^{itX}\big). \tag{35.211}$$

Démonstration. Nous étudions la fonction

$$\Phi(t) = \int_\Omega e^{itX(\omega)} dP(\omega). \tag{35.212}$$

Nous considérons la fonction

$$\begin{aligned} f\colon \mathbb{R} \times \Omega &\to \mathbb{R} \\ (t, \omega) &\mapsto e^{itX(\omega)}. \end{aligned} \tag{35.213}$$

et nous regardons si ce contexte vérifie les hypothèses du théorème 16.21.

(1) Étant donné que X est mesurable, f sera mesurable.
(2) La fonction $t \mapsto e^{itX(\omega)}$ est absolument continue pour chaque ω.
(3) Note : par rapport aux notations du théorème 16.21, nous avons ici $A = \mathbb{R}$. Prenons donc un intervalle (compact) $[a, b] \subset \mathbb{R}$ et calculons

$$\frac{\partial f}{\partial t}(t, \omega) = iX(\omega) e^{itX}, \tag{35.214}$$

et

$$\int_a^b \int_\Omega \left| iX(\omega) e^{itX(\omega)} \right| d\omega\, dt = \int_a^b \int_\Omega |X(\omega)| d\omega\, dt. \tag{35.215}$$

Par hypothèse X accepte un moment d'ordre 1, de sorte que l'intégrale par rapport à ω converge vers un nombre qui ne dépend pas de t. L'intégrale sur t ne pose alors aucun problèmes.

Par conséquent nous pouvons effectuer la première dérivation :

$$\Phi'(t) = \frac{d\Phi}{dt}(t) = \int_\Omega iX(\omega) e^{itX(\omega)} d\omega = E(iXe^{itX}) \tag{35.216}$$

et la fonction Φ' est absolument continue. Ce dernier point est important parce que c'est lui qui permet de faire la récurrence et passer à l'ordre deux.

Le résultat ressort alors en dérivant successivement l'expression (35.214). □

Exemple 35.65.
Sachant la fonction caractéristique de X, nous pouvons calculer les moments. Par exemple

$$E(X^2) = \Phi_X''(0). \tag{35.217}$$

△

Théorème 35.66.
Si $\Phi_X = \Phi_Y$, alors $P_X = P_Y$.

Notons que cela n'implique pas que $X = Y$. En effet X et Y peuvent même être définis sur des espaces probabilisés différents.

Dans le cas d'une variable aléatoire vectorielle, nous définissons $\Phi_X \colon \mathbb{R}^d \to \mathbb{R}$ par

$$\Phi_X(v) = E\big(e^{i\langle v,X\rangle}\big) \tag{35.218}$$

35.2.15 Fonction génératrice des moments, transformée de Laplace

Soit X une variable aléatoire. Sa **transformée de Laplace** ou **fonction génératrice des moments** est la fonction

$$M_X(t) = E(e^{tX}) \tag{35.219}$$

pour chaque t tel que cette espérance existe.

Théorème 35.67 ([525]).
Soit X une variable aléatoire réelle et

$$I_X = \{t \in \mathbb{R} \text{ tel que } E(e^{tX}) \text{ existe}\}. \tag{35.220}$$

La fonction

$$\begin{aligned} M_X \colon I_X &\to \mathbb{R} \\ t &\mapsto E(e^{tX}) \end{aligned} \tag{35.221}$$

*est la **transformée de Laplace** de X.*

(1) I_X est un intervalle contenant 0.

(2) Si I_X n'est pas réduit à $\{0\}$ alors M_X se développe en série entière

$$M_X(t) = \sum_{n=0}^{\infty} \frac{E(X^n)}{n!} t^n. \tag{35.222}$$

(3) Si X et Y sont des variables aléatoires indépendantes, alors $I_{X+Y} = I_X \cap I_Y$ et

$$M_{X+Y} = M_X M_Y \tag{35.223}$$

sur I_{X+Y}.

Démonstration. Le fait que 0 soit dans I_X est évident : $E(1) = 1$. Pour montrer que I_X est un intervalle nous prenons $z \in I_X$ et $0 < s < z$ ou $z < s < 0$, puis nous montrons que $s \in I_X$. Il faut remarquer que dans tous les cas,

$$e^{sX} \leqslant 1 + e^{zX}. \tag{35.224}$$

En effet soit sX et zX sont tous deux à gauche de zéro et alors ils sont tous deux plus petit que 1 ; soit ils sont tous deux à droite de 0 et alors $e^{zX} > e^{sX}$ par croissance de l'exponentielle. Nous avons donc dans tous les cas que

$$E(e^{sX}) = \int_{\mathbb{R}} f_X(x) e^{sX} dx \leqslant \int_{\mathbb{R}} f_X(x)(1 + e^{zx}) = 1 + E(e^{zX}). \tag{35.225}$$

Soit maintenant $a > 0$ tel que $[-a, a] \in I_X$. Étant donné que $e^{a|X|} < e^{aX} e^{-aX}$, l'espérance $E(e^{a|X|})$ existe toujours pour $|t|$. Nous avons

$$\left| M_X(t) - \sum_{n=0}^{N} \frac{E(X^n)}{n!} t^n \right| = \left| E\Big(e^{tX} - \sum_{n=0}^{N} \frac{X^n}{n!} t^n\Big) \right| \tag{35.226a}$$

$$= \left| E\Big(\sum_{n=N+1}^{\infty} \frac{X^n}{n!} t^n\Big) \right| \tag{35.226b}$$

$$\leqslant E\left(\sum_{n=N+1}^{\infty} \frac{|tX|^n}{n!}\right). \tag{35.226c}$$

Maintenant le but est de prendre la limite $N \to \infty$ en inversant la limite et l'espérance par le théorème de la convergence dominée (13.184). L'intégrale à traiter est

$$\lim_{N\to\infty} \int_\Omega \sum_{n=N+1}^{\infty} \frac{|tX(\omega)|^n}{n!} dP(\omega). \tag{35.227}$$

L'intégrante est uniformément borné (en N) par $e^{tX(\omega)}$, qui est intégrable par hypothèse (choix de t). Du coup

$$\lim_{N\to\infty} \int_\Omega \sum_{n=N+1}^{\infty} \frac{|tX(\omega)|^n}{n!} dP(\omega) = E\left(\lim_{N\to\infty} \sum_{n=N+1}^{\infty} \frac{|tX|^n}{n!} \right) = 0. \tag{35.228}$$

□

35.2.16 Loi d'une variable aléatoire

La **loi** de la variable aléatoire X, notée P_X est la mesure image de P par X, c'est-à-dire

$$P_X(B) = P(X \in B) \tag{35.229}$$

pour tout borélien $B \subset \mathbb{R}^d$. Note :

$$P(X \in B) = P\big(\{\omega \in \Omega \text{ tel que } X(\omega) \in B\}\big) = P\big(X^{-1}(B)\big). \tag{35.230}$$

En particulier P_X est une mesure de probabilité sur $\mathbb{R}^d$ parce que

$$P_X(\mathbb{R}^d) = P(\Omega) = 1. \tag{35.231}$$

Si Q est une mesure de probabilité sur $\mathbb{R}^d$, nous notons $X \sim Q$ si $P_X = Q$. Nous disons alors que « X suit la loi Q ».

La proposition suivante permet de calculer en pratique les intégrales qui définissent par exemple l'espérance mathématique d'une variable aléatoire.

Proposition 35.68 (Théorème de transfert[526])**.**
Si X est une variable aléatoire, alors

$$E(f \circ X) = \int_\Omega f\big(X(\omega)\big) dP(\omega) = \int_{\mathbb{R}^d} f(x) dP_X(x) \tag{35.232}$$

dès que $f\colon \mathbb{R}^d \to \bar{\mathbb{R}}$ est telle qu'une des deux intégrales existe. En particulier, ça marche si f est borélienne.

En utilisant cette proposition nous trouvons une formule pratique pour l'espérance d'une variable aléatoire réelle :

$$E(X) = \int_\Omega X(\omega) dP(\omega) = \int_{\mathbb{R}} x dP_X(x), \tag{35.233}$$

en vertu de la proposition 35.68 appliquée à la fonction $f(x) = x$.

Proposition 35.69.
Une variable aléatoire réelle X est intégrable si et seulement si $P(x = \pm\infty) = 0$ et

$$\int_{\mathbb{R}} |x| dP_X(x) < \infty. \tag{35.234}$$

Le lien entre la densité f_X de la variable aléatoire X et sa loi est

$$P_X(A) = \int_A f_X(x) dx \tag{35.235}$$

pour tout ensemble mesurable $A \subset \mathbb{R}$. Le lien entre la mesure de Lebesgue et celle de la loi de X est alors donné par

$$dP_X(x) = f_X(x) dx. \tag{35.236}$$

En particulier l'espérance de X peut être calculée à partir de sa densité via la formule

$$E(X) = \int_{\mathbb{R}} x dP_X(x) = \int_{\mathbb{R}} x f_X(x) dx. \tag{35.237}$$

35.2.17 Changement de variables

Théorème 35.70.
Soit $\mathcal{O}$, un ouvert de $\mathbb{R}^n$ et $\mathcal{O}'$ un ouvert de $\mathbb{R}^m$ ainsi qu'un difféomorphisme C^1 $\varphi\colon \mathcal{O} \to \mathcal{O}'$. Soit $X\colon \Omega \to \mathbb{R}^n$ une variable aléatoire prenant presque surement ses valeurs dans $\mathcal{O}$. Si nous supposons que X a la densité f_X, alors la variable aléatoire $Y = \varphi(X)$ accepte la densité $f_Y\colon \mathcal{O}' \to \mathbb{R}$ donnée par

$$f_Y(v) = f_X\big(\varphi^{-1}(v)\big)|J_{\varphi^{-1}}(v)|. \tag{35.238}$$

Démonstration. Nous devons vérifier la relation

$$P(Y \in B) = \int_B f_Y(v)dv \tag{35.239}$$

pour tout borélien $B \subset \mathcal{O}'$. Nous avons

$$\begin{aligned}
P(Y \in B) &= \int_{\mathbb{R}^m} \mathbb{1}_B(v)dP_Y(v) && (35.240a)\\
&= E(\mathbb{1}_B \circ Y) && (35.240b)\\
&= E\big((\mathbb{1}_B \circ \varphi) \circ X\big) && (35.240c)\\
&= \int_{\mathbb{R}^n} (\mathbb{1}_B \circ \varphi)(u)dP_X(u) && (35.240d)\\
&= \int_{\mathbb{R}^n} (\mathbb{1}_B \circ \varphi)(u)f_X(u)du. && (35.240e)
\end{aligned}$$

À ce niveau, nous utilisons la formule de changement de variables du théorème 13.255. Nous trouvons alors

$$P(Y \in B) = \int_{\mathbb{R}^m} \mathbb{1}_B\big(\varphi^{-1}(v)\big)f_X\big(\varphi^{-1}(v)\big)|J_{\varphi^{-1}}(v)|dv. \tag{35.241}$$

□

35.3 Convergence

Soient X_i des variables aléatoires réelles définies sur le même espace de probabilité $(\Omega, \mathcal{A}, P)$. Nous disons que X_i converge **presque surement** vers la variable aléatoire X et nous notons

$$X_n \xrightarrow{p.s.} X \tag{35.242}$$

si

$$P\big(\{\omega \in \Omega \text{ tel que } X_n(\omega) \to X(\omega)\}\big) = 1 \tag{35.243}$$

où la convergence $X_n(\omega) \to X(\omega)$ est la convergence usuelle dans $\mathbb{R}$.

Lemme 35.71.
Nous avons $X_n \xrightarrow{p.s.} X$ si et seulement s'il existe un événement $A \in \mathcal{A}$ tel que $P(A) = 1$ et tel que $X_n(\omega) \to X(\omega)$ pour tout $\omega \in A$.

Lemme 35.72.
Si X est une variable aléatoire à valeurs dans $\mathbb{R} \cup \{\infty\}$, alors

$$X \wedge n \xrightarrow{p.s.} X \tag{35.244}$$

Démonstration. Si $\omega \in \{X = \infty\}$ alors $(X \wedge n)(\omega) = n$ et d'accord. Si par contre $\omega \in \{X < \infty\}$ alors il existe N tel que si $n \geqslant N$ alors $n \geqslant T(\omega)$ et pour ces grandes valeurs de n nous avons $(T \wedge n)(\omega) = T(\omega)$. □

Définition 35.73 ([527]).
Nous disons que les variables aléatoires réelles X_n convergent ***en probabilité*** *vers la variable aléatoire X si pour tout $\eta > 0$, on a*

$$P\big(|X_n - X| \geqslant \eta\big) \to 0, \tag{35.245}$$

et on note

$$X_n \xrightarrow{P} X. \tag{35.246}$$

Définition 35.74 (Convergence en loi).
Nous disons que X_n converge vers X ***en loi*** *vers la variable aléatoire X et nous notons*

$$X_n \xrightarrow{\mathscr{L}} X \tag{35.247}$$

si pour toute fonction continue et bornée g nous avons

$$E\big(g(X_n)\big) \to E\big(g(X)\big) = \int g dP_X. \tag{35.248}$$

Proposition 35.75.
Deux autres caractérisations de la convergence en loi.

(1) Nous avons $X_n \xrightarrow{\mathscr{L}} X$ si et seulement si

$$\Phi_{X_n}(v) \to \Phi_X(v) \tag{35.249}$$

pour tout $v \in \mathbb{R}^d$. Ici Φ_X est la fonction caractéristique de X.

(2) Dans la définition de la convergence en loi nous pouvons indifféremment utiliser les fonctions continues et bornées, les fonctions continues à support compact ou les fonctions bornées uniformément continues.

Proposition 35.76.
Les types de convergence sont reliées par les implications suivantes :

$$\textit{presque sure} \;\Rightarrow\; \textit{en probabilité} \;\Rightarrow\; \textit{en loi.} \tag{35.250}$$

La convergence en loi n'implique pas la convergence en probabilité, et par conséquent pas non plus la convergence presque certaine.

Dans le cas particulier $d = 1$ nous avons quelques critères supplémentaires.

Proposition 35.77.
Supposons que les variables aléatoires X_n soient réelles, et notons F_n la fonction de répartition de X_n. Si $F_n(x) \to F(x)$ pour tout x dans l'ensemble des points de continuité de F, alors $X_n \xrightarrow{\mathscr{L}} X$.

Proposition 35.78.
Si les X_n sont des variables aléatoires réelles positives, et si X est une variable aléatoire positive, alors $X_n \xrightarrow{\mathscr{L}} X$ si les transformées de Laplace des fonctions de répartition convergent ponctuellement, c'est-à-dire si

$$E\big(e^{-\alpha X_n}\big) \to E\big(e^{-\alpha X}\big) \tag{35.251}$$

pour tout $\alpha \geqslant 0$.

Proposition 35.79.
Si les X_n et X sont des variables aléatoires réelles discrètes à valeurs dans $\{x_0, x_1, \ldots\}$ alors $X_n \xrightarrow{\mathscr{L}} X$ si et seulement si

$$P(X_n = x_k) \to P(X = x_k) \tag{35.252}$$

pour tout $k \in \mathbb{N}$.

Proposition 35.80 ([518]).
Soient X_n et X des variables aléatoires réelles. Nous avons

$$X_n \xrightarrow{\mathscr{L}} X \tag{35.253}$$

si et seulement si pour tout t où F_X est continue,

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t). \tag{35.254}$$

Proposition 35.81 ([518]).
Soit X_n une suite de variables aléatoires $\Omega \to \mathbb{R}^d$ et $a \in \mathbb{R}^d$. Si $X_n \xrightarrow{\mathscr{L}} a$, alors

$$X_n \xrightarrow{P} a. \tag{35.255}$$

Démonstration. Quitte à passer aux composantes, nous pouvons supposer que $d = 1$. Soit $\eta > 0$; nous savons que l'inégalité $|x| > a$ a pour solution $x > a$ ou $x < -a$. Dans notre cas,

$$\begin{aligned}
P\big(|X_n - a| > \eta\big) &= P(X_n - a > \eta) + P(X_n - a < -\eta) && (35.256a)\\
&= P(X_n > \eta + a) + P(X < a - \eta) && (35.256b)\\
&= 1 - P(X_n \leqslant \eta + a) + P(X_n \leqslant a - \eta) - P(X_n = a - \eta) && (35.256c)\\
&\leqslant 1 - F_{X_n}(\eta + a) + F_{X_n}(a - \eta) && (35.256d)
\end{aligned}$$

où la majoration est l'oubli du terme $P(X_n = a - \eta)$, lequel est positif ou nul et F_{X_n} est la fonction de répartition de X_n, définition 35.61. Nous allons utiliser la proposition 35.80. La fonction de répartition de la variable aléatoire constante $X = a$ est donnée par

$$F_a(t) = P(a \leqslant t) = \mathbb{1}_{[0,\infty[}(t - a). \tag{35.257}$$

Par conséquent, la convergence en loi $X_n \xrightarrow{\mathscr{L}} a$ nous montre que

$$F_{X_n}(t) \to \mathbb{1}_{[0,\infty[}(t - a) \tag{35.258}$$

pour tout $t \neq a$ parce que $t = 0$ est un point de discontinuité de $\mathbb{1}_{[0,\infty[}$. Nous avons par conséquent

$$P\big(|X_n - a| > \eta\big) = 1 - \mathbb{1}_{[0,\infty[}(\eta) + \mathbb{1}_{[0,\infty[}(-\eta) = 1 - 1 + 0 = 0 \tag{35.259}$$

parce que $\eta > 0$. □

Le lemme de Slutsky sera utilisé en combinaison avec la proposition 35.83 pour calculer des intervalles de confiance, voir par exemple ce qui se passe autour de l'équation (36.126).

Lemme 35.82 (Slutsky[528]).
Soient X_n et Y_n des suites de variables aléatoires réelles telles que

$$\begin{aligned}
X_n &\xrightarrow{\mathscr{L}} X\\
Y_n &\xrightarrow{P} a \in \mathbb{R}.
\end{aligned} \tag{35.260}$$

Alors $(X_n, Y_n) \xrightarrow{\mathscr{L}} (X, a)$.

Démonstration. Étant donné que $Y_n \xrightarrow{\mathscr{L}} a$, nous avons $Y_n \xrightarrow{P} a$ par la proposition 35.81. Soit une fonction $f\colon \mathbb{R}^2 \to \mathbb{R}^2$; nous devons prouver que

$$E\big(f(X_n, Y_n)\big) \to E\big(f(X, a)\big). \tag{35.261}$$

Soit $\epsilon > 0$. Nous avons

$$E\big(\|f(X_n, Y_n) - f(X, a)\|\big) \leqslant E\big(\|f(X_n, Y_n) - f(X_n, a)\|\big) + E\big(\|f(X_n, a) - f(X, a)\|\big). \tag{35.262}$$

La fonction $g(t) = f(t, a)$ étant continue et bornée, la convergence en loi $X_n \xrightarrow{\mathscr{L}} X$ donne

$$E\big(\|f(X_n, a) - f(X, a)\|\big) \to 0. \tag{35.263}$$

Étudions à présent le premier terme du membre de droite de (35.262). Pour tout $\eta > 0$ et toute variables aléatoires Z et Z' nous avons

$$E(Z) = E(Z\mathbb{1}_{|Z'|<\eta}) + E(Z\mathbb{1}_{|Z'|\geqslant\eta}). \tag{35.264}$$

Nous décomposons donc le premier terme de (35.262) en

$$\begin{aligned} E\big(\|f(X_n, Y_n) - f(X_n, a)\|\big) = {} & E\big(\|f(X_n, Y_n) - f(X_n, a)\|\mathbb{1}_{|Y_n-a|<\eta}\big) \\ & + E\big(\|f(X_n, Y_n) - f(X_n, a)\|\mathbb{1}_{|Y_n-a|\geqslant\eta}\big). \end{aligned} \tag{35.265}$$

Choisissons maintenant une valeur de η telle que

$$|(x, y) - (x', y')| < \eta \Rightarrow |f(x, y) - f(x', y')| \leqslant \epsilon. \tag{35.266}$$

Un tel η existe par l'uniforme continuité de f. Dans le premier terme, $|Y_n - a| < \eta$, par conséquent

$$\|(X_n, Y_n) - (X_n, a)\| = |Y_n - a| < \eta \tag{35.267}$$

et donc

$$\|f(X_n, Y_n) - f(X_n, a)\| \leqslant \epsilon. \tag{35.268}$$

Le premier terme devient donc

$$E\big(\|f(X_n, Y_n) - f(X_n, a)\|\mathbb{1}_{|Y_n-a|<\eta}\big) \leqslant \epsilon E(\mathbb{1}_{|Y_n-a|<\eta}) \leqslant \epsilon \tag{35.269}$$

parce que $E(\mathbb{1}_A) = P(A) \leqslant 1$. Pour le second terme de (35.265) nous effectuons la majoration

$$\|f(X_n, Y_n) - f(X_n, a)\| \leqslant 2\|f\|_\infty \tag{35.270}$$

tandis que la convergence $Y_n \xrightarrow{P} a$ entraine

$$P\big(|Y_n - a| \geqslant \eta\big). \tag{35.271}$$

□

Proposition 35.83 ([529]).
Soient X_i des variables aléatoires telles que

$$X_i \xrightarrow{\mathscr{L}} X \tag{35.272}$$

et h, une fonction mesurable sur l'espace d'arrivée de X_i. Soit C l'ensemble des points de continuité de h au sens

$$C = \{\omega \in \Omega \text{ tel que } h \text{ est continue en } X_i(\omega)\}. \tag{35.273}$$

Alors si $P(X \in C) = 1$, nous avons

$$h(X_i) \xrightarrow{\mathscr{L}} h(X). \tag{35.274}$$

Une conséquence de cette proposition couplée au lemme de Slutsky est le résultat suivant, qui est donné sous le nom de théorème de Slutsky sur wikipédia.

Corolaire 35.84.
En reprenant les notations du lemme de Slutsky, si

$$X_n \xrightarrow{\mathscr{L}} X \tag{35.275a}$$

$$Y_n \xrightarrow{P} a, \tag{35.275b}$$

alors

$$X_n + Y_n \xrightarrow{\mathscr{L}} X + a \tag{35.276a}$$
$$X_n Y_n \xrightarrow{\mathscr{L}} aX \tag{35.276b}$$
$$Y_n^{-1} X_n \xrightarrow{\mathscr{L}} a^{-1} X \tag{35.276c}$$
$$\tag{35.276d}$$

pourvu que a soit inversible.

Lemme 35.85 (Borel-Cantelli)**.**
Soit (A_n) une suite d'événements (avec $A_n \in \mathcal{A}$ pour tout n).

(1) Si $\sum_{n=0}^{\infty} P(A_n)$ converge, alors
$$P(A_n \text{ i.s.}) = 0. \tag{35.277}$$

(2) Si la somme $\sum_n P(A_n)$ diverge, et si de plus les A_i sont indépendants, alors
$$P(A_n \text{ i.s.}) = 1. \tag{35.278}$$

La notation $P(A_n \text{ i.s.})$ signifie « infiniment souvent », c'est-à-dire
$$P(A_n \text{ i.s.}) = P\Big(\bigcap_{N\in\mathbb{N}} \bigcup_{k\geqslant N} A_k\Big) = P(\limsup A_n) \tag{35.279}$$

Une façon de paraphraser le lemme de Borel-Cantelli est que nous avons l'alternative
$$P(\limsup A_n) = \begin{cases} 0 & \text{si } \sum_{n\geqslant 0} P(A_n) < \infty \\ 1 & \text{sinon.} \end{cases} \tag{35.280}$$

Proposition 35.86.
Soit X_n, une suite de variables aléatoires et X une variable aléatoires. Si
$$\sum_n P\big(\|X_n - X\| > \eta\big) < \infty \tag{35.281}$$

pour tout ϵ, alors $X_n \xrightarrow{p.s.} X$.

Démonstration. Fixons ϵ et considérons les événements $A_n = \|X_n - X\| > \epsilon$. L'hypothèse dit que
$$\sum_n P(A_{n,\epsilon}) < \infty \tag{35.282}$$

et le lemme de Borel-Cantelli implique que
$$P(\limsup \|X_n - X\| > \epsilon) = 0. \tag{35.283}$$

Un élément ω est dans $\limsup A_n$ s'il est contenu dans tous les A_n, par conséquent, pour chaque ϵ nous avons l'inclusion
$$\{\omega \in \Omega \text{ tel que } X_n(\omega) \to X(\omega)\} \subset \complement \limsup A_n. \tag{35.284}$$

Nous pouvons aller plus loin et écrire
$$\{\omega \in \Omega \text{ tel que } X_n(\omega) \to X(\omega)\} = \complement\{\omega \in \Omega \text{ tel que } \|X_n - X\| > \epsilon, \forall \epsilon > 0\}. \tag{35.285}$$

Or la probabilité de l'ensemble
$$\{\omega \in \Omega \text{ tel que } \|X_n - X\| > \epsilon\} \tag{35.286}$$

est 0 pour chaque ϵ, et par conséquent la probabilité du membre de droite de (35.285) est 1. □

Exemple 35.87.
Considérons une suite de 0 et de 1 dans laquelle le 1 arrive avec une probabilité p et le 0 avec une probabilité $1-p$. Une telle suite est modélisée par une suite de variables aléatoires de Bernoulli $(X_n)_{n\in\mathbb{N}}$ indépendantes de paramètre p.

Question : une telle suite contient elle une infinité de 1 ? Considérons les événements indépendants $A_n = \{X_n = 1\}$. Nous avons

$$\sum_n P(A_n) = \sum_n P(X_n = 1) = \sum_n p = \infty. \tag{35.287}$$

Par Borel-Cantelli et son expression (35.280), nous avons alors

$$P(\limsup A_n) = 1. \tag{35.288}$$

Donc une infinité d'événements A_n se produisent, et nous avons bien une infinité de 1 dans la suite.

Remarque : dans ce raisonnement nous pouvons considérer une probabilité non constante p_n tant que la série $\sum_n p_n$ diverge. △

Exemple 35.88.
À propos de maximum. La fonction $h\colon \mathbb{R}^d \to \mathbb{R}$ donnée par $h(x_1,\dots,x_d) = \max_i\{x_i\}$ est une fonction continue. Nous voudrions prouver que si on a une famille (finie en $i = 1,\dots,l$) de suites (en n) variables aléatoires $X_n^{(i)} \xrightarrow{p.s.} a$ convergeant toutes vers la même limite a, alors

$$M_n = \max_i\{X_n^{(i)}\} \xrightarrow{p.s.} a. \tag{35.289}$$

D'abord si nous avons l suite numériques $(x_n^{(i)})$, alors la suite

$$M_n = \max_i x_n^{(i)} \tag{35.290}$$

converge vers la même limite. En effet si $\epsilon > 0$ est donné, il suffit de prendre N_i l'entier tel que $|x_n^{(i)} - a| \leqslant \epsilon$ pour tout $n > N_i$. Et ensuite on prend $N > \max\{N_i\}$.

Si maintenant au lieu de suites numériques, nous avons des variables aléatoires, le résultat reste valable. Nous cherchons à prouver que

$$P\Big(\{\omega \in \Omega \text{ tel que } \max\{X_n^{(i)}(\omega)\} \to a\}\Big) = 1. \tag{35.291}$$

Par ce que nous venons de dire sur les suites numériques, un élément ω n'est pas dans cet ensemble seulement s'il y a un ∞ pour lequel $X_n^{(i)}$ ne converge pas vers a. Or cela, pour chaque i est un événement de probabilité zéro.

Les ω qui ne fonctionneront pas dans l'équation (35.291) sont ceux de la réunion d'un ensemble fini d'ensembles de probabilité nulle. C'est donc de probabilité nulle. △

35.4 Loi des grands nombres, théorème central limite

35.4.1 Loi des grands nombres

Lemme 35.89 (Inégalité de Markov[530])**.**
Soit une variable aléatoire $X \in L^p$ et $\epsilon > 0$. Nous avons

$$P(|X| \geqslant \epsilon) \leqslant \frac{1}{\epsilon^r} E(|X|^r). \tag{35.292}$$

Démonstration. Nous avons

$$E(|X|^r) \geqslant \int_{|X|\geqslant\epsilon} X(\omega)dP(\omega) \geqslant \epsilon^r \int_{|X|\geqslant\epsilon} dP = \epsilon^r P(|X| \geqslant \epsilon). \tag{35.293}$$

□

Corolaire 35.90 ([530]).
Soit ϕ, une fonction croissante et positive ou nulle sur l'intervalle I. Soit aussi une variable aléatoire $Y\colon \Omega \to \mathbb{R}$ telle que $P(Y \in I) = 1$. Alors pour tout $b \in I$ tel que $\phi(b) > 0$ nous avons

$$P(Y \geqslant b) \leqslant \frac{E\big[\phi(Y)\big]}{\phi(b)}. \tag{35.294}$$

Théorème 35.91 (Loi forte des grands nombres).
Soit (X_n) une suite de variables aléatoires réelles

(1) indépendantes et identiquement distribuées,

(2) intégrables (c'est-à-dire dans L^1),

alors

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s.} E(X_1). \tag{35.295}$$

Note : étant donné que les variables aléatoires sont identiquement distribuées, nous avons évidemment $E(X_1) = E(X_2) = \dots$

Problèmes et choses à faire

Est-ce que les variables aléatoires doivent vraiment être réelles ?

Corolaire 35.92.
Si les variables aléatoires réelles X_n sont

(1) indépendantes et identiquement distribuées,

(2) dans L^2

alors

$$\bar{X}_n \xrightarrow{P} E(X_1). \tag{35.296}$$

Démonstration. Nous voulons prouver que pour tout $\eta > 0$,

$$P\big(|\bar{X}_n - E(X_1)| > \eta\big) \to 0. \tag{35.297}$$

Remarquons d'abord que les variables aléatoires X_n étant identiquement distribuées, $E(\bar{X}_n) = E(X_1)$ parce que $E(X_i) = E(X_1)$ pour tout i. L'inégalité de Markov avec $r = 2$ nous donne

$$P\big(|\bar{X}_n - E(\bar{X}_n)| > \eta\big) \leqslant \frac{1}{\eta^2} E\big(|\bar{X}_n - E(\bar{X}_n)|^2\big) \tag{35.298a}$$

$$= \frac{1}{\eta^2}\operatorname{Var}(\bar{X}_n) \tag{35.298b}$$

$$= \frac{1}{n\eta^2}\operatorname{Var}(X_1) \tag{35.298c}$$

où nous avons utilisé la la proposition 35.31 : $\operatorname{Var}(\bar{X}_n) = \operatorname{Var}(X_1)/n$. Au final nous avons prouvé que

$$P\big(|\bar{X}_n - E(\bar{X}_n)| > \frac{1}{n\eta^2}\operatorname{Var}(X_1), \tag{35.299}$$

qui tend vers zéro lorsque $n \to \infty$. □

Proposition 35.93.
Soient X_n des variables aléatoires indépendantes et identiquement distribuées avec $X_n \geqslant 0$. Nous acceptons $E(X_1) = \infty$, c'est-à-dire que nous relaxons la condition $X_n \in L^1$ par rapport à la loi des grands nombres.

Alors

$$\bar{X}_n \xrightarrow{p.s.} E(X_1) \in [0, \infty]. \tag{35.300}$$

Démonstration. Si $E(X_1) < \infty$, nous sommes dans le cas de la loi des grands nombres. Pour chaque $N \in \mathbb{N}$ nous considérons la suite de variables aléatoires

$$X_n^{(N)} = \min(X_n, N). \tag{35.301}$$

Nous avons évidement $\bar{X}_n^{(N)} \leqslant \bar{X}_n$. Les variables aléatoires $X_n^{(N)}$ étant bornées par N, elles vérifient la loi des grands nombres pour chaque N séparément. Par conséquent nous avons pour chaque N la limite

$$\bar{X}_n^{(N)} \to E(X_1^{(N)}) \tag{35.302}$$

Nous supposons que $E(X_1) = \infty$, par conséquent $\lim_{N\to\infty} E(X_1^{(N)}) = \infty$. Soit $\eta > 0$ et choisissons N de telle manière à avoir

$$E(X_1^{(N)}) > \eta + 1. \tag{35.303}$$

La limite (35.302) nous permet de trouver n_0 tel que pour tout $n > n_0$ nous ayons $\bar{X}_n^{(N)} > \eta$. Au final,

$$\eta < \bar{X}_n^{(N)} \leqslant \bar{X}_n, \tag{35.304}$$

ce qui montre que $\bar{X}_n \to \infty$. □

Exemple 35.94.
La loi des grands nombres justifie la pratique courante d'approximer une grandeur physique par la moyenne empirique d'un grand nombre de mesures. △

Exemple 35.95.
Citons ici le dernier paragraphe de *Le mystère de Marie Roget* par Edgar Allan Poe, traduit par Charles Baudelaire [9].

> Rien, par exemple, n'est plus difficile que de convaincre le lecteur non spécialiste que, si un joueur de dés a amené les six deux fois coup sur coup, ce fait est une raison suffisante de parier gros que le troisième coup ne ramènera pas les six. Une opinion de ce genre est généralement rejetée tout d'abord par l'intelligence. On ne comprend pas comment les deux coups déjà joués, et qui sont maintenant complètement enfouis dans le Passé, peuvent avoir de l'influence sur le coup qui n'existe que dans le Futur. La chance pour amener les six semble être précisément ce qu'elle était à n'importe quel moment, c'est-à-dire soumise seulement à l'influence de tous les coups divers que peuvent amener les dés. Et c'est là une réflexion qui semble si parfaitement évidente, que tout effort pour la controverser est plus souvent accueilli par un sourire moqueur que par une condescendance attentive.

Dans le cours de la nouvelle, Edgar Poe cite et utilise la théorie des probabilités avec une justesse inaccoutumée dans la littérature. Mais dans ce paragraphe final, Poe montre de façon la plus formelle qu'il n'a *rien* compris à la loi des grands nombres. △

35.4.2 Théorème central limite

Lemme 35.96.
Soit $z_n \to z$ une suite convergente dans $\mathbb{C}$. Alors

$$\left(1 + \frac{z_n}{n}\right)^n \to e^z. \tag{35.305}$$

Théorème 35.97.
Si les variables aléatoires X_n sont

(1) indépendantes et identiquement distribuées de loi parente X,

(2) $X_1 \in L^2(\Omega, \mathcal{A})$,

9. Disponible sur https://fr.wikisource.org/wiki/Le_Mystère_de_Marie_Roget

alors nous notons $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, $m = E(X_1)$ *et* $\sigma^2 = \mathrm{Var}(X_1)$ *et nous avons*

$$\frac{\bar{X}_n - E(X)}{\sqrt{\mathrm{Var}(\bar{X}_n)}} = \frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \xrightarrow{\mathscr{L}} \mathscr{N}(0,1). \tag{35.306}$$

Démonstration. Nous allons écrire la démonstration dans le cas de variables aléatoires réelles. La proposition 35.75 dit que la suite X_n converge en loi vers X si et seulement si les fonctions caractéristiques convergent ponctuellement. Nous devons donc prouver, pour chaque[10] $t \in \mathbb{R}$, que

$$\Phi_{\frac{S_n - nm}{\sigma\sqrt{n}}}(t) \to \Phi_{\mathscr{N}(0,1)}(t). \tag{35.307}$$

Supposons dans un premier temps que $E(X_i) = 0$ et $\sigma(X_i) = 1$. Dans ce cas nous considérons la fonction

$$\Phi_{\frac{S_n}{\sqrt{n}}} = E\big(e^{i\frac{t}{\sqrt{n}}\sum_{k=1}^n X_k}\big) \tag{35.308a}$$

$$= \prod_{k=1}^n E\big(e^{i\frac{t}{\sqrt{n}}X_1}\big) \tag{35.308b}$$

$$= \prod_{k=1}^n \Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right) \tag{35.308c}$$

$$= \Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right)^n. \tag{35.308d}$$

Cette quantité est à priori complexe ; nous ne pouvons donc pas immédiatement passer au logarithme. Nous pouvons par contre utiliser un développement en puissances de t en nous servant de la proposition 35.64 et de l'hypothèse comme quoi $X_1 \in L^2$:

$$\Phi_{X_1}(t) = \Phi_{X_1}(0) + \Phi'_{X_1}(0)t + \Phi''_{X_1}(0)\frac{t^2}{2} + \alpha(t)t^2 \tag{35.309}$$

où α est une fonction qui a la propriété $\lim_{x\to 0}\alpha(x) = 0$.

En utilisant les hypothèses et la formule de dérivation de la fonction caractéristique,

$$\Phi_{X_1}(0) = 1 \tag{35.310a}$$

$$\Phi'_{X_1}(0) = E\big((iX)\big) = 0 \tag{35.310b}$$

$$\Phi''_{X_1}(0) = E(-X^2) = -\mathrm{Var}(X_1) = -1. \tag{35.310c}$$

Nous avons donc

$$\Phi_{X_1}\left(\frac{t}{\sqrt{n}}\right) = \underbrace{1 - \frac{1}{2}\frac{t^2}{n}}_{\in\mathbb{R}} + \underbrace{\frac{t^2}{n}\alpha\left(\frac{t}{\sqrt{n}}\right)}_{\in\mathbb{C}}, \tag{35.311}$$

de telle sorte que, en considérant une valeur fixée de t,

$$\Phi_{\frac{S_n}{\sqrt{n}}}(t) = \left(1 - \frac{\frac{t^2}{2} + \beta_n}{n}\right)^n \tag{35.312}$$

où $\beta_n = t^2\alpha(t/\sqrt{n})$. Nous avons bien entendu $\lim_{n\to\infty}\beta_n = 0$.

Nous pouvons appliquer le lemme 35.96 pour obtenir la limite

$$\lim_{n\to\infty}\Phi_{\frac{S_n}{\sqrt{n}}}(t) = e^{-t^2/2}. \tag{35.313}$$

La convergence (35.307) est par conséquent prouvée dans le cas où $E(X_i) = 0$ et $\mathrm{Var}(X_i) = 1$.

10. Chuck Norris peut *vraiment* le faire pour *chaque* $t \in \mathbb{R}$

Considérons maintenant des variables aléatoires avec $E(X_i) = m$ et $\mathrm{Var}(X_i) = \sigma^2$. Elles peuvent être écrites sous la forme

$$X_i = \sigma X_i + m \tag{35.314}$$

où X_i' est d'espérance nulle et de variance un. Nous avons alors

$$S_n = \sigma \sum_{i=1}^{n} X_i' + nm, \tag{35.315}$$

et

$$\frac{S_n - nm}{\sigma\sqrt{n}} = \frac{S_n'}{\sqrt{n}} \tag{35.316}$$

où $S_n' = \sum_i X_i'$. L'étude de la variable aléatoire

$$\frac{S_n - nm}{\sigma\sqrt{n}} \tag{35.317}$$

revient donc à celle de $S_n'/\sqrt{n}$ qui vient d'être effectuée. □

35.98.
À propos du théorème central limite 35.97. Si pour une certaine variable aléatoire X on a $E(X) = m$, alors nous n'avons pas forcément $P(X = m + a) = P(X = m - a)$. Est-ce que le théorème central limite permet cependant d'affirmer que dans un certaine mesure nous avons

$$P(\bar{X}_n = m + a) = P(\bar{X}_n = m - a) \tag{35.318}$$

lorsque n est grand ?

Tel quelle, l'équation (35.318) est en général fausse pour chaque n parce qu'il existe des distributions non symétriques. Mais bien entendu les deux membres tendent vers zéro pour $n \to \infty$. Mais cela n'est pas lié à la symétrie de la distribution gaussienne. C'est seulement le fait que la gaussienne n'a pas de masses ponctuelles.

Par contre, il y a effectivement une assurance de symétrie pour $\bar{X}_n$ lorsque $n \to \infty$. Le fait est que $X_n \xrightarrow{\mathscr{L}} X$ où X est une gaussienne. La fonction de répartition de X est continue partout et la proposition 35.80 nous dit que pour tout $x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(t). \tag{35.319}$$

Vu que le nombre $P\big(\bar{X}_n \in B(m + a, \delta)\big)$ peut être exprimé avec des sommes et différences F_{X_n}, nous avons

$$\lim_{n\to\infty} P\big(\bar{X}_n \in B(m + a, \delta)\big) = P\big(X \in B(m + a, \delta)\big). \tag{35.320}$$

Par symétrie de la gaussienne le membre de droite est égal à $P\big(X \in B(m - a, \delta)\big)$ et nous avons bien

$$\lim_{n\to\infty} P\big(\bar{X}_n \in B(m + a, \delta)\big) = \lim_{n\to\infty} P\big(\bar{X}_n \in B(m - a, \delta)\big). \tag{35.321}$$

Remarque 35.99.
Le théorème central limite s'applique quelle que soit la distribution des variables aléatoires X_i (dans les limites de hypothèses) ; en particulier il ne dit rien sur la moyenne des X_i. Il dit seulement que l'écart de la moyenne « mesurée » à la moyenne « théorique » est une variable aléatoire gaussienne si on a mesuré assez de fois.

Autrement dit, si la durée d'attente à la poste est de 5 minutes, et si j'y vais 2000 fois, alors la probabilité que ma moyenne d'attente soit de 4 minutes est la même que la probabilité qu'elle soit de 6 minutes [11].

11. Et en l'occurrence, cette probabilité est nulle parce qu'on est en train de parler de variable aléatoire continue, mais vous voyez l'idée.

Problèmes et choses à faire

La remarque 35.99 est une interprétation personnelle. J'aimerais avoir l'avis de quelqu'un de plus compétent.

Remarque 35.100.
Nous pouvons obtenir la limite (35.313) d'une façon alternative. Nous considérons la détermination du logarithme complexe sur $\mathbb{C}\backslash\mathbb{R}^-$; cela est une fonction analytique (théorème 25.55) vérifiant l'équation

$$e^{\ln(z)} = z \tag{35.322}$$

pour tout $z \in \mathbb{C}\backslash\mathbb{R}^-$ et le développement

$$\ln(1+z) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{z^n}{n}. \tag{35.323}$$

En particulier, $\ln(1+z) = z + z\alpha(z)$ où $\lim_{z\to 0}\alpha(z) = 0$. Nous reprenons à l'équation (35.312) en fixant t. Nous avons

$$\Phi_{\frac{S_n}{\sqrt{n}}}(t) = \exp\left[\ln\left(\Phi_{\frac{S_n}{\sqrt{n}}}(t)\right)\right] \tag{35.324a}$$

$$= \exp\left[n\ln\left(1+\frac{-\frac{t^2}{2}-\beta_n}{n}\right)\right] \tag{35.324b}$$

$$= \exp\left[-\frac{t^2}{2}-\beta_n+\left(-\frac{t^2}{2}-\beta_n\right)\alpha\left(\frac{-\frac{t^2}{2}-\beta_n}{n}\right)\right]. \tag{35.324c}$$

À la limite $n \to 0$ nous tombons sur $e^{-t^2/2}$.

Remarque 35.101.
Étant donné que la variable aléatoire

$$\frac{S_n - nm}{\sigma\sqrt{n}} \tag{35.325}$$

converge en loi vers $\mathscr{N}(0,1)$, nous avons la convergence des fonctions de répartition partout où la fonction de répartition de la normale est continue [12] (donc sur tout $\mathbb{R}$). En particulier,

$$\left|P\left(\frac{S_n - nm}{\sigma\sqrt{n}} \leqslant x\right) - \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy\right| \to 0. \tag{35.326}$$

Nous avons la borne de **Berry-Esséen** qui donne une estimation de la vitesse de convergence : si $X \in L^3$, alors il existe une constante C, indépendante de x, des X_i et de n telle que

$$\left|P\left(\frac{S_n - nm}{\sigma\sqrt{n}} \leqslant x\right) - \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy\right| \leqslant \frac{X\mu_3}{\sigma^3\sqrt{n}} \tag{35.327}$$

où $\mu_3 = E\big(|X_1 - m|^3\big)$ est le moment d'ordre 3 de X. La chose à retenir est que la convergence est à la vitesse de $1/\sqrt{n}$.

En dimension $d > 1$, nous avons encore un théorème central limite.

Théorème 35.102.
Si $d > 1$, et si nous avons des variables aléatoires X_n à valeurs dans $\mathbb{R}^d$ avec

(1) les X_n sont indépendantes et identiquement distribuées

(2) les X_n sont dans L^2.

Alors nous notons $X_1 = (X_1^{(1)}, \dots, X_1^{(d)})$. Nous avons

$$\frac{S_n - nm}{\sqrt{n}} \xrightarrow{\mathscr{L}} \mathscr{N}(0,\Sigma) \tag{35.328}$$

où Σ est ma matrice de covariance du vecteur aléatoire X_1 :

$$\Sigma = \big(\mathrm{Cov}(X_1^{(1)},\dots,X_1^{(d)})\big)_{i,j=1,\dots,d}. \tag{35.329}$$

12. Proposition 35.80.

35.4.3 Marche aléatoire

Nous considérons un mobile qui se déplace sur l'axe $\mathbb{Z}$. À chaque pas de temps, nous supposons qu'il va faire un pas à gauche avec une probabilité p et un pas à droite avec une probabilité $(1-p)$. Nous nous demandons quel est le mouvement du mobile sur le long terme.

La position S_n du mobile à l'instant n est donnée par

$$S_n = \sum_{i=1}^{n} X_i \tag{35.330}$$

où X_i est le pas effectué à l'instant i. Ce sont des variables de Bernoulli indépendantes avec

$$X_i \overset{\mathscr{L}}{=} p\delta_{-1} + (1-p)\delta_1 \tag{35.331}$$

c'est-à-dire

$$P(X_i = -1) = p \tag{35.332a}$$
$$P(X_i = 1) = 1 - p. \tag{35.332b}$$

Ces variables vérifient les hypothèses de la loi des grands nombres :

(1) elles sont indépendantes et identiquement distribuées,

(2) elles sont intégrables.

Pour le second point, le calcul est

$$\int_\Omega |X_i| dP = \int_{\mathbb{R}} |x| dP_X = \int_{\mathbb{R}} |x|(p\delta_{-1} + (1-p)\delta_1) = |1-p| + |p| = 1. \tag{35.333}$$

Nous avons par conséquent

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s.} E(X_1) = (1-2p) \tag{35.334}$$

et

$$\frac{S_n}{n} \to (1-2p). \tag{35.335}$$

Si $p \neq 1/2$ nous pouvons conclure que

(1) si $p > 1/2$, alors $S_n \xrightarrow{p.s.} -\infty$

(2) si $p < 1/2$, alors $S_n \xrightarrow{p.s.} \infty$.

De plus nous connaissons la vitesse de divergence : elle est linéaire. Le mobile suit essentiellement l'équation $p(n) = (1-2p)n$.

Remarque 35.103.
Cela ne traite pas le cas $p = 1/2$. Dans ce cas, nous pouvons simplement dire que $S_n = o(n)$.

35.5 Les lois usuelles

35.5.1 Loi de Bernoulli

Une expérience de Bernoulli consiste à tirer au hasard un 0 ou un 1 avec une probabilité p de tomber sur 1 et $1-p$ de tomber sur zéro. Il s'agit donc d'une expérience qui réussit ou qui rate.

Le cas typique est une urne avec des boules indiscernables blanches ou noires. La probabilité p est la proportion de blanches dans l'urne (avec remise entre les tirages). Dans ce cas, nous avons l'espace de probabilité $(\Omega, \mathcal{A}, P)$ où Ω représente l'ensemble des boules, $\mathcal{A}$ est l'ensemble des parties de Ω et P est l'équiprobabilité sur Ω. Une variable aléatoire est une application

$$\begin{aligned} X\colon \Omega &\to \{0,1\} \\ \omega &\mapsto \text{couleur de la boule } \omega. \end{aligned} \tag{35.336}$$

Nous notons $\mathscr{B}(1,p)$ la loi de Bernoulli. Elle a une expression très simple :

$$\mathscr{B}(0,1)\big(\{1\}\big) = p \tag{35.337a}$$
$$\mathscr{B}(0,1)\big(\{0\}\big) = 1-p \tag{35.337b}$$

Une variable aléatoire réelle est de **Bernoulli** de paramètre p $(0 < p < 1)$ si

$$X\colon \Omega \to \mathbb{R} \tag{35.338}$$

avec $P(x = 1) = p$ et $P(X = 0) = 1 - p$. En tant que mesure sur $\mathbb{R}$, nous avons

$$P_X = p\delta_1 + (1-p)\delta_0. \tag{35.339}$$

Une fonction h qui réalise le supremum de la formule (13.424) est par exemple une fonction en escalier qui vaut en x le plus petit entier plus grand ou égal à x. L'espérance d'une loi de Bernoulli est alors

$$E(x) = p. \tag{35.340}$$

Étant donné que la variable aléatoire X prend seulement les valeurs 0 et 1, nous avons pour tout ensemble mesurable B

$$P_{X^2}(B) = P(X^2 \in B) = P(X \in B), \tag{35.341}$$

et par conséquent $P_{X^2} = P_X$ et $E(X^2) = E(X)$. Nous trouvons donc la variance

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p). \tag{35.342}$$

35.5.2 Loi binomiale

Une expérience binomiale consiste à répéter n expériences de Bernoulli de paramètre p et de compter le nombre de réussites. Une telle expérience peut être réalisée selon la procédure suivante.

Soit une urne contenant N boules dont une proportion p de 1 et $1-p$ de 0. Une expérience binomiale de paramètres n et p consistera à prendre n boules *avec remise* et à compter le nombre de 1 obtenus.

En termes d'espaces probabilisé, nous avons Ω qui est l'ensemble des tuples de taille n à valeurs dans $\{0,1\}$, la tribu $\mathcal{A}$ est l'ensemble des parties de Ω, et la probabilité P est l'équiprobabilité :

$$P(\omega) = \frac{1}{N^n} \tag{35.343}$$

si il y a N boules dans l'urne. Nous construisons alors la variable aléatoire

$$X(\omega) = \sum_{i=1}^{n} \omega_i \tag{35.344}$$

où ω est une suite de taille n de 0 et de 1.

Calculons $P(X = k)$. Il s'agit de considérer tous les sous-ensembles de taille n de Ω contenant exactement k fois 1. Il y a $\binom{n}{k}$ manière de décider lesquelles des n boules seront blanches. Ensuite, chaque boule blanche peut être choisie parmi les m boules disponibles, et chaque boule noire peut être choisie parmi les $(N-m)$ disponibles. Nous avons donc

$$P(X = k) = \binom{n}{k}\frac{m^k(N-m)^{n-k}}{N^k}. \tag{35.345}$$

En effet la mesure de probabilité sur Ω est la mesure de comptage renormalisée par le cardinal de Ω qui vaut N^m. Étant donné que $p = m/N$, nous transformons facilement (35.345) en

$$P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}. \tag{35.346}$$

Une variable aléatoire de loi binomiale étant une somme de variables aléatoires de Bernoulli indépendantes, l'espérance [13] et la variance [14] s'obtiennent en sommant les espérances et variances termes à terme :

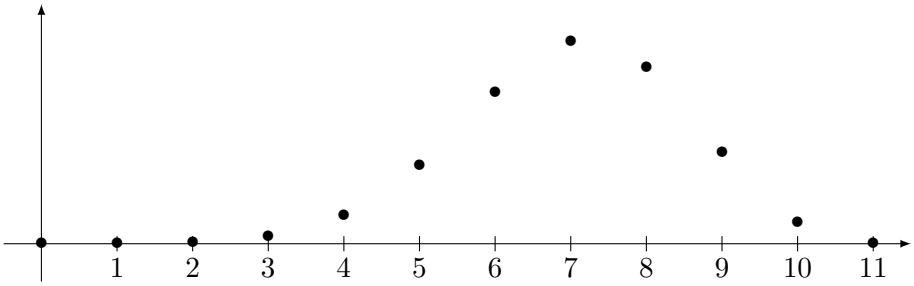
$$E(X) = np \tag{35.347}$$

et

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = np(1-p) \tag{35.348}$$

en vertu de (35.342).

La loi binomiale lorsque $p = 0.7$ et $n = 10$.



35.5.3 Loi multinomiale

La loi multinomiale $\mathscr{M}(n; k; p_1, \dots, p_k)$ consiste à effectuer n épreuves d'une démarche aléatoire qui peut avoir k issues différentes avec probabilités $p_1, \dots, p_k$. Les variables aléatoires multinomiales sont N_i avec les contraintes

$$\sum_{i=1}^{k} N_i = n \tag{35.349a}$$

$$\sum_{i=1}^{k} p_i = 1. \tag{35.349b}$$

La fonction de probabilité multinomiale est

$$P(N_1 = n_1, \dots, N_k = n_k) = \frac{n!}{n_1! \dots n_k!} p_1^{n_1} \dots p_k^k. \tag{35.350}$$

Chacune des N_i est une binomiale de probabilité p_i.

35.5.4 Loi géométrique

Soit (X_n) une suite indépendante et identiquement distribuée de lois de Bernoulli de paramètre p. Alors la variable aléatoire

$$Z = \inf\{n \geqslant 1 \text{ tel que } X_n = 1\} \tag{35.351}$$

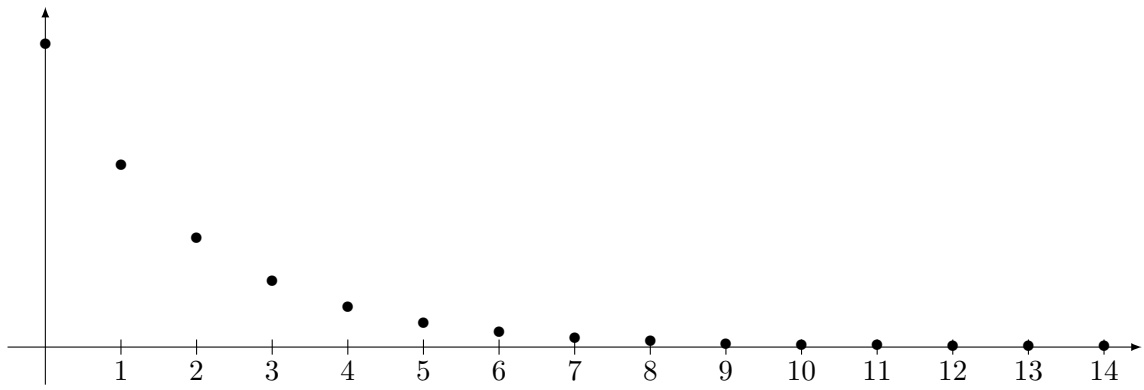
est une loi géométrique de paramètre p.

La loi géométrique compte donc le nombre d'expériences de Bernoulli à effectuer avant que le premier succès soit au rendez-vous. Nous avons

$$P(Z = k) = P(X_k = 1)P(X_1, \dots, X_{k-1} = 0) = p(1-p)^{k-1} \tag{35.352}$$

La loi géométrique de paramètre $p = 0.2$.

13. Valable même sans indépendance, proposition 35.29
14. Lemme 35.23.



Note : si p est trop grand, on ne voit vite plus rien parce que la probabilité d'attendre longtemps est vite très faible.

35.5.5 Loi de Poisson

Une variable aléatoire Z suit une **loi de Poisson** de paramètre λ, notée $\mathscr{P}(\lambda)$ si

$$P(Z = k) = e^{-\lambda}\frac{\lambda^k}{k!} \tag{35.353}$$

pour tout $k \in \mathbb{N}$.

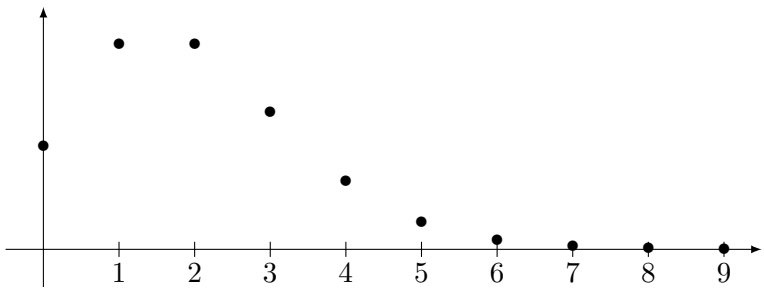
La loi de Poisson est une loi de probabilité discrète qui décrit le comportement du nombre d'évènements se produisant dans un laps de temps fixé, si ces évènements se produisent avec une fréquence moyenne connue et indépendamment du temps écoulé depuis l'évènement précédent.

Si un événement se produit en moyenne p fois par seconde, la probabilité d'observer l'événement k fois durant n secondes est donnée par $P(Z = k)$ où Z est une loi de Poisson de paramètre $\lambda = pn$.

Théorème 35.104 (wikipedia)**.**
L'espérance et la variance d'une variable aléatoire de Poisson sont λ :

$$E(X) = \lambda \tag{35.354a}$$
$$\mathrm{Var}(X) = \lambda. \tag{35.354b}$$

La loi de Poisson de paramètre $\lambda = 2$.



35.5.6 Loi exponentielle

La loi de exponentielle représente le temps qu'il faut attendre pour qu'une particule se désintègre si elle a en permanence une probabilité [15] λdt de se désintégrer entre t et $t + dt$. L'espérance est donc $1/\lambda$.

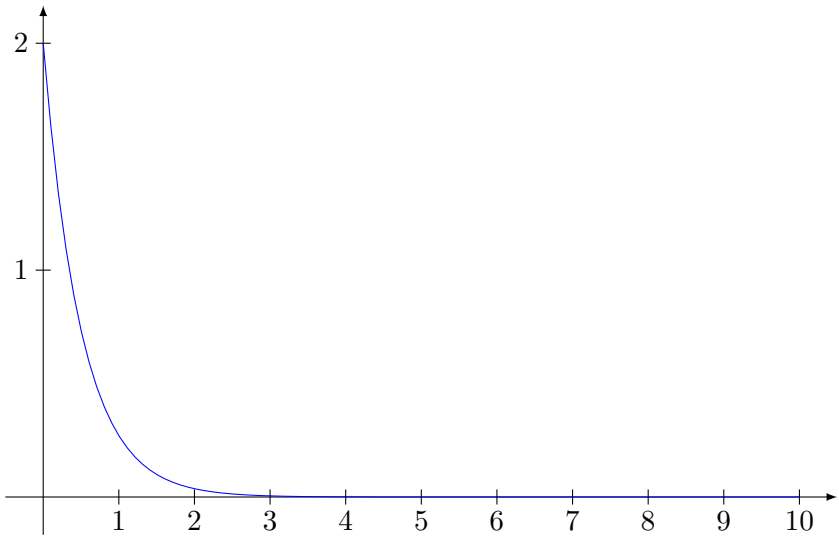
Il se passe donc en moyenne λ événements par seconde. La proposition 35.111 nous montrera que le nombre d'événements se produisant en une seconde suit une loi de Poisson de paramètre λ.

15. Étant donné que λ n'est pas limité à 1, en réalité ce n'est pas une probabilité. Je suis preneur d'une bonne interprétation physique de ce paramètre.

Plus formellement, la loi exponentielle de paramètre λ, notée $\mathscr{E}(\lambda)$ est la loi de densité

$$f_X : x \mapsto \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geqslant 0 \\ 0 & \text{sinon.} \end{cases} \tag{35.355}$$

Densité de la loi exponentielle de paramètre $\lambda = 2$.



Proposition 35.105.
Si $X \sim \mathscr{E}(\lambda)$, alors la fonction de répartition de X est donnée par

$$F(x) = P(X \leqslant x) = \begin{cases} 1 - e^{-\lambda x} & \textit{si } x \geqslant 0 \\ 0 & \textit{sinon.} \end{cases} \tag{35.356}$$

La fonction caractéristique est donnée par

$$E(e^{itX}) = \frac{\lambda}{\lambda - it}. \tag{35.357}$$

L'espérance et la variance sont données par

$$\begin{aligned} E(X) &= \frac{1}{\lambda} \\ \mathrm{Var}(X) &= \frac{1}{\lambda^2}. \end{aligned} \tag{35.358}$$

Démonstration. Pour la fonction caractéristique,

$$E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} \lambda e^{-\lambda x} \mathbb{1}_{[0,\infty[}(x) dx \tag{35.359a}$$

$$= \lambda \int_0^{\infty} e^{x(-\lambda + it)} dx \tag{35.359b}$$

$$= \lim_{A \to \infty} \left[\frac{e^{x(-\lambda+it)}}{it - \lambda} \right]_{x=0}^{x=A} \tag{35.359c}$$

$$= \lim_{A \to \infty} \frac{e^{A(it-\lambda)}}{it - \lambda} - \frac{1}{it - \lambda}. \tag{35.359d}$$

Le premier terme est nul parce que si on prend la norme,

$$\left| \frac{e^{A(-\lambda+it)}}{-\lambda + it} \right| = \frac{e^{-\lambda A}}{|it - \lambda|} \to 0. \tag{35.360}$$

En ce qui concerne l'espérance nous faisons le calcul suivant :

$$E(X) = \int_{\mathbb{R}} x f_X(x) dx = \lambda \int_{\mathbb{R}^+} x e^{-\lambda x} dx = \frac{1}{\lambda}. \tag{35.361}$$

Pour la variance, nous utilisons la formule (35.60). Nous avons

$$E(X^2) = \int_{\mathbb{R}} x^2 f_X(x) = \int_0^\infty x^2 \lambda e^{-\lambda x} = \frac{2}{\lambda^2}. \tag{35.362}$$

Donc $\mathrm{Var}(X) = E(X^2) - E(X)^2 = \frac{1}{\lambda^2}$. □

La loi exponentielle est une loi **sans mémoire** en ce sens que

$$P(X > x + y | X > y) = P(X > x). \tag{35.363}$$

En effet nous utilisons la règle de la probabilité conditionnelle

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{35.364}$$

Ici,

$$P(X > x + y | X > y) = \frac{P(X > x + y)}{P(X > y)} = e^{-\lambda x}. \tag{35.365}$$

La proposition suivante montre que la loi exponentielle est à peu près la seule à être sans mémoire. D'où son importance dans l'étude des machines dont les pièces ne subissent pas d'usure.

Proposition 35.106.
Soit X, une variable aléatoire admettant une densité continue f par rapport à la mesure de Lebesgue. Si elle est sans mémoire, alors elle est exponentielle.

Démonstration. Nous posons $\varphi(x) = P(X \geqslant x)$. Cela est la fonction de répartition de X (à part que cette dernière est $1 - \varphi(x)$ mais c'est pas grave), et est donnée par

$$\varphi(x) = 1 - \int_0^x f(t) dt. \tag{35.366}$$

Cette dernière intégrale vérifie les hypothèses du théorème 35.106, de telle sorte que φ soit une fonction dérivable et $\varphi'(x) = f(x)$.

D'autre part en utilisant la définition de la probabilité conditionnelle la propriété de ne pas avoir de mémoire donne

$$\varphi(x + y) = \varphi(x)\varphi(y) \tag{35.367}$$

et de plus $\varphi(0) = 0$. Calculons la dérivée de φ :

$$\varphi'(x) = \lim_{\epsilon \to 0} \frac{\varphi(x + \epsilon) - \varphi(x)}{\epsilon} = \varphi(x) \lim_{\epsilon \to 0} \frac{\varphi(\epsilon) - 1}{\epsilon} = \varphi(x)\varphi'(0). \tag{35.368}$$

Donc φ vérifie l'équation différentielle de l'exponentielle.

□

Exemple 35.107.
Une machine a une durée de vie représentée par une variable aléatoire suivant une loi de Poisson de paramètre λ. Soit T_y la variable aléatoire qui représente la temps de vie restant sachant que la machine a déjà vécu un temps y. Nous voulons trouver la fonction de répartition de T_y. Nous avons

$$P(T_y > x) = P(X > x + y | X > y) = P(X > x) = e^{-\lambda x}. \tag{35.369}$$

Dans ce cas, la loi de T_y ne dépend pas de y. Cela signifie que la machine ne vieilli pas et surtout que le modèle n'est pas réaliste. △

Proposition 35.108.
Si $X \sim \mathscr{E}(\lambda)$ et $Y \sim \mathscr{E}(\mu)$ sont indépendantes, alors

(1) $P(X < Y) = \frac{\lambda}{\lambda+\mu}$
(2) $P(X > Y) = \frac{\mu}{\lambda+\mu}$
(3) $P(X = Y) = 0$
(4) $\min(X, Y) \sim \mathscr{E}(\lambda + \mu)$.

De plus les variables aléatoires exponentielles ont une propriété d'absence de mémoire :

$$P(X > t + s|X > s) = P(X > t) = e^{-\lambda t}. \tag{35.370}$$

Démonstration. Étant donné que X et Y sont indépendantes, la densité conjointe est le produit des densités (35.20). Nous avons donc

$$P(X > Y) = \int_D \lambda e^{-\lambda x} \mu e^{-\mu y} dxdy \tag{35.371}$$

où D est le domaine $D = \{(x, y) \in \mathbb{R}^2 \text{ tel que } x, y > 0, x > y\}$. Nous avons donc

$$P(X > Y) = \lambda\mu \int_0^\infty dx \int_0^x dy e^{-\lambda x} e^{-\mu y} = \frac{\mu}{\lambda + \mu}. \tag{35.372}$$

```
sage: var('a,b')
(a, b)
sage: f(x,y)=exp(-a*x)*exp(-b*y)
sage: assume(a>0)
sage: assume(b>0)
sage: a*b*f.integrate(y,0,x).integrate(x,0,oo)
(x, y) |--> a*b/(a^2 + a*b)
```

Pour trouver la loi de $\min(X, Y)$, nous écrivons

$$\begin{aligned}
P\big(\min(X, Y) > t\big) &= P(X > t, Y > t) && (35.373a)\\
&= P(X > t)P(Y > t) \qquad \text{par indépendance} && (35.373b)\\
&= \big(1 - F_X(t)\big)\big(1 - F_Y(t)\big) && (35.373c)\\
&= e^{-(\lambda+\mu)t} \mathbb{1}_{[0,\infty[}(t) && (35.373d)\\
&= 1 - F_Z(t) && (35.373e)
\end{aligned}$$

où $Z \sim \mathscr{E}(\lambda + \mu)$. □

35.5.7 Approximation de la binomiale par une Poisson

Proposition 35.109.
Soit (X_n) une suite de variables aléatoires avec $X_n \sim \mathscr{B}(n, p_n)$ telle que np_n converge vers une constante $\lambda > 0$. Alors $X_n \xrightarrow{\mathscr{L}} \mathscr{P}(\lambda)$.

Démonstration. Commençons par écrire la loi binomiale sous une forme plus adaptée au passage à la limite :

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n_k} = \frac{n(n-1)\dots(n-k+1)}{k!} p^k (1 - p)^{n-k}. \tag{35.374}$$

Le produit au numérateur contient k termes dans lesquels nous mettons n en évidence. Nous trouvons

$$P(X = k) = \frac{(np)^k \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\dots\left(1 - \frac{k-1}{n}\right)}{k!} p^k (n - p)^{n-k}. \tag{35.375}$$

Lorsque nous passons à la limite, tous les facteurs du type $1-l/n$ tendent vers 1 ainsi que $(1-p_n)^{-k}$. Les facteurs dont la limite n'est pas 1 sont donc

$$P(X_n = k) \simeq \frac{(np_n)^k}{k!}(1-p_n)^k. \tag{35.376}$$

Nous avons

$$\lim_{n\to\infty}(1-p_n)^n = \lim_{n\to\infty}\left(1-\frac{np_n}{n}\right)^n = e^{-\lambda}. \tag{35.377}$$

La thèse est alors obtenue en remettant les morceaux ensemble. □

Exemple 35.110.
Considérons un serveur informatique qui reçoit des requêtes. Toutes les 10^{-3} s il reçoit une requête avec une probabilité $p = 0.05$. La variable aléatoire qui consiste à donner le nombre de requêtes effectivement effectuées en une seconde suit une loi binomiale $\mathscr{P}(1000, p)$.

Déterminons la probabilité que le serveur reçoive 20 requêtes en une seconde. Nous approximons $\mathscr{B}(1000, 0.05)$ par $\mathscr{P}(50)$, et la réponse est

$$e^{-50}\frac{50^{20}}{20!} \simeq 7 \cdot 10^{-7}. \tag{35.378}$$

△

35.5.8 Loi de Poisson et loi exponentielle

Soient $X_1, \dots, X_n$ des variables aléatoires réelles indépendantes de loi exponentielle de paramètre λ. En utilisant le produit de convolution, nous pouvons trouver la fonction de densité de la somme (voir point 35.2.3). Commençons avec deux variables aléatoires X et Y. Les densités sont

$$f_X(x) = \mathbb{1}_{[x\geqslant 0]}\lambda e^{-\lambda x} \tag{35.379a}$$
$$f_Y(y) = \mathbb{1}_{[y\geqslant 0]}\lambda e^{-\lambda y}, \tag{35.379b}$$

et la densité conjointe est alors

$$f_{X+Y}(x) = \int_{\mathbb{R}} \mathbb{1}_{[x-t\geqslant 0]}\lambda e^{-\lambda(x-t)}\mathbb{1}_{[t\geqslant 0]}\lambda e^{-\lambda t}dt \tag{35.380a}$$
$$= \lambda^2 e^{-\lambda x}\int_0^x 1\,dt \tag{35.380b}$$
$$= x\lambda^2 e^{-\lambda x}. \tag{35.380c}$$

Par récurrence si $S = X_1 + \cdots + X_n$ nous trouvons

$$f_S(x) = x^{n-1}\lambda^n e^{-\lambda x}. \tag{35.381}$$

Proposition 35.111 ([531]).
Soit $(T_k)_{k\in\mathbb{N}}$ une suite de variables aléatoires indépendantes de loi $\mathscr{E}(\lambda)$. Nous considérons la variable aléatoire $S_n = \sum_{i=1}^n X_i$ et pour chaque $t \in \mathbb{R}^+$ nous considérons

$$N_t = \max\{n \geqslant 1 \text{ tel que } \sum_{i=1}^n T_i \leqslant t\}. \tag{35.382}$$

Alors $N_t \sim \mathscr{P}(\lambda t)$.

Démonstration. Ce que nous devons calculer est

$$P(N_t = k) = P(S_n \leqslant t \leqslant S_{n+1}). \tag{35.383}$$

Nous introduisons la variable aléatoire $V_{n=1} = (X_1, \dots, X_{n+1})$ ainsi que l'ensemble

$$A_{n+1} = \{x \in \mathbb{R}^{n+1} \text{ tel que } x_1 + \cdots + x_n \leqslant t \leqslant x_1 + \cdots + x_n + x_{n+1}\}. \tag{35.384}$$

Le problème est donc de calculer

$$P(S_n \leqslant t \leqslant S_{n+1}) = P(V_{n+1} \in A_{n+1}^+) = \int_{A_{n+1}^+} f_{n+1}(x)dx \tag{35.385}$$

où A_{n+1}^+ est la partie de A_{n+1} dans laquelle $x_i \geqslant 0$ pour tout i et f_{n+1} est la fonction de densité conjointe des variables aléatoires X_i. Nous effectuons le changement de variables

$$s_k = \sum_{i \leqslant k} x_i \tag{35.386a}$$

$$x_k = s_k - s_{k-1} \tag{35.386b}$$

dont le déterminant vaut 1. D'autre part par indépendance des variables aléatoires X_i, la fonction de partition jointe f_{n+1} s'exprime sous la forme

$$f_{n+1}(x_1, \ldots, x_{n+1}) = f_{X_1}(x_1) \ldots f_{X_{n+1}}(x_{n+1}) \tag{35.387a}$$

$$= \lambda^{n+1} e^{-\lambda(x_1 + \cdots + x_{n+1})} \tag{35.387b}$$

$$= \lambda^{n+1} e^{-\lambda s_{n+1}}. \tag{35.387c}$$

En ce qui concerne les bornes de l'intégrale dans les variables s_i, nous voulons que tous les x_i soient positifs, par conséquent $s_1 \geqslant 0$ et ensuite l'équation $x_k = s_k - s_{k-1}$ demande $s_k \geqslant s_{k-1}$. Les bornes sont donc données par l'ensemble

$$0 \leqslant s_1 \leqslant s_2 \leqslant \ldots \leqslant s_n \leqslant t \leqslant s_{n+1}, \tag{35.388}$$

c'est-à-dire $B_n \times]t, \infty[$ où

$$B_n = \{(s_1, \ldots, s_n) \text{ tel que } 0 \leqslant s_1 \leqslant s_2 \leqslant \ldots \leqslant s_n \leqslant t\}. \tag{35.389}$$

Le théorème de Fubini nous permet de décomposer l'intégrale :

$$P(S_n \leqslant t \leqslant S_{n+1}) = \int_{B_n \times]t,\infty[} \lambda^{n+1} e^{-\lambda s_{n+1}} ds_1 \ldots ds_{n+1} \tag{35.390a}$$

$$= \lambda^{n+1} \left(\int_{B_n} ds_1 \ldots ds_n \right) \underbrace{\left(\int_t^\infty e^{-\lambda s_{n+1}} ds_{n+1} \right)}_{=\lambda^{-1} e^{-\lambda t}} \tag{35.390b}$$

$$= \lambda^n e^{-\lambda t} \operatorname{Vol}(B_n) \tag{35.390c}$$

où $\operatorname{Vol}(B_n)$ est le volume de B_n qui reste à calculer. L'ensemble $C^n = [0, t]^n$ se décompose en cellules disjointes (à ensemble de mesure nulle près) de la forme

$$C_\sigma = \{0 \leqslant s_{\sigma(1)} \leqslant s_{\sigma(2)} \leqslant \ldots \leqslant s_{\sigma(n)} \leqslant t\} \tag{35.391}$$

pour chaque permutation $\sigma \in S_n$. Il y a exactement $n!$ telles cellules dans C^n. Par conséquent

$$t^n = \operatorname{Vol}(C^n) = n! \operatorname{Vol}(C_\sigma) = n! \operatorname{Vol}(B_n) \tag{35.392}$$

et $\operatorname{Vol}(B_n) = \frac{t^n}{n!}$. Finalement nous avons

$$P(n_t = n) = P(S_n \leqslant t \leqslant S_{n+1}) = \frac{(\lambda t)^n}{n!} e^{-(\lambda t)}. \tag{35.393}$$

□

35.5.9 Loi normale

La loi normale de paramètres m et $\sigma > 0$, notée $\mathscr{N}(m, \sigma^2)$ est la loi donnée par la densité

$$\gamma_{m,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right]. \tag{35.394}$$

Proposition 35.112.
Si la variable aléatoire réelle X suit une loi normale $\mathscr{N}(m, \sigma^2)$, alors nous avons $E(X) = m$ et $\mathrm{Var}(X) = \sigma^2$.

Démonstration. L'espérance d'une variable aléatoire se calcule à partir de la formule (35.237) :

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} x \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] dx \tag{35.395a}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \sigma \int_{\mathbb{R}} (\sigma u + m) e^{-u^2/2} \tag{35.395b}$$

où nous avons effectué le changement de variable $u = (x-m)/\sigma$. Nous utilisons ensuite l'intégrale remarquable

$$\int_{\mathbb{R}} e^{-u^2/2} du = \sqrt{2\pi}. \tag{35.396}$$

En ce qui concerne la variance, nous avons le même genre de calculs. □

La **loi normale réduite** est la densité

$$\gamma(x) = \gamma_{0,1}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \tag{35.397}$$

La variable aléatoire X suit la loi $\mathscr{N}(m, \sigma^2)$ si et seulement si la variable aléatoire $Z = \frac{X-m}{\sigma}$ suit la loi normale réduite $\mathscr{N}(0, 1)$.

Proposition 35.113.
La fonction caractéristique de la distribution normale $\mathscr{N}(m, \sigma^2)$ est

$$\Phi_{\mathscr{N}(m,\sigma^2)}(t) = \exp\left(itm - \frac{\sigma^2 t^2}{2}\right). \tag{35.398}$$

Démonstration. En suivant la formule (35.209), l'intégrale à calculer est

$$\Phi_{\mathscr{N}(m,\sigma^2)}(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2} dx. \tag{35.399}$$

Nous reconnaissons une transformée de Fourier. Afin de la calculer sans encombres, nous passons par les fonctions intermédiaires suivantes :

$$\begin{aligned} g(x) &= e^{-\frac{1}{2}x^2} \\ h(x) &= g\left(\frac{x}{\sigma}\right) \\ k(x) &= h(x-m). \end{aligned} \tag{35.400}$$

La fonction caractéristique que nous cherchons est $\frac{1}{\sigma\sqrt{2\pi}}\hat{k}(t)$. Les formules liées à la transformée de Fourier nous donnent

$$\hat{k}(t) = \hat{h}(t) e^{itm} \tag{35.401a}$$

$$\hat{h}(t) = \sigma \hat{g}(\sigma t) \tag{35.401b}$$

$$\hat{g}(t) = \int_{\mathbb{R}} e^{-itx} e^{-\frac{1}{2}x^2} dx = \sqrt{2\pi} e^{-t^2/2}. \tag{35.401c}$$

Attention : l'intégrale à calculer est une transformée de Fourier *inverse*, d'où la formule (35.401a) qui a un signe de différence avec la formule usuelle. En recombinant toutes ces expressions nous trouvons

$$\Phi_{\mathcal{N}(m,\sigma^2)} = e^{-\sigma^2 t^2/2} e^{itm}, \tag{35.402}$$

ce qu'il nous fallait. □

Exemple 35.114.
Une espérance qui sert de temps en temps est celle de $X = e^{\beta Z}$ lorsque $Z \sim \mathcal{N}(0,1)$. Elle se calcule en remarquant que $x^2 - 2\beta x = (x-\beta)^2 - \beta^2$, donc

$$E(e^{\beta Z}) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{\beta x} e^{-x^2/2} dx \tag{35.403a}$$
$$= \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{1}{2}(x-\beta)^2} e^{\beta^2/2} \tag{35.403b}$$
$$= \frac{e^{\beta^2/2}}{\sqrt{2\pi}}\int_{e^{-y^2/2}} dy \tag{35.403c}$$
$$= e^{\beta^2/2}. \tag{35.403d}$$

△

35.5.10 Vecteurs gaussiens

Source : [270, 532].

Définition 35.115.
*Un vecteur aléatoire $X\colon \Omega \to \mathbb{R}^d$ est un **vecteur gaussien** si toutes les combinaisons linéaires de ses composantes sont des variables aléatoires normales. En d'autres termes, X est un vecteur gaussien si pour tout vecteur u, la variable aléatoire $u \cdot X$ est gaussienne.*

Le vecteur moyenne d'un vecteur gaussien est $E(X) = \big(E(X_1), \dots, E(X_n)\big)$ et sa matrice de **variance-covariance** est la matrice

$$K_X = \mathrm{Var}(X) = E\Big[\big(X - E(X)\big) \otimes \big(X - E(X)\big)\Big] \tag{35.404}$$

où l'opération $\otimes$ est celle introduite autour de l'équation (10.284). Cela n'est rien d'autre que la matrice de covariance de la variable aléatoire $X\colon \Omega \to \mathbb{R}^d$.

Lemme 35.116.
Si les variables aléatoires réelles $X_1, \dots, X_n$ sont des variables aléatoires réelles gaussiennes indépendantes, alors le vecteur $X = (X_1, \dots, X_n)$ est un vecteur gaussien.

Démonstration. Nous devons montrer que si X et Y sont des variables aléatoires gaussiennes indépendantes, alors $X+Y$ est encore gaussienne. L'indépendance nous assure les égalités suivantes pour la fonction caractéristique :

$$\Phi_{X+Y}(t) = E\big(e^{it(X-Y)}\big) = E\big(e^{itX}\big)E\big(e^{itY}\big). \tag{35.405}$$

Dans le cas où X et Y sont gaussiens nous trouvons

$$\Phi_{X+Y}(t) = \exp\left(im_1 t - \frac{\sigma_1^2 t^2}{2}\right)\exp\left(im_2 t - \frac{\sigma_2^2 t^2}{2}\right) = \exp\left(i(m_1+m_2)t - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}\right). \tag{35.406}$$

Étant donné que la loi d'une variable aléatoire est entièrement déterminée par sa fonction caractéristique (théorème 35.66), nous déduisons que $X+Y$ est une normale de moyenne m_1+m_2 et de variance $\sigma = \sigma_1^2 + \sigma_2^2$. □

Proposition 35.117.
La fonction caractéristique d'un vecteur gaussien est donnée par

$$\Phi_X(u) = \exp\left(iu \cdot E(X) - \frac{1}{2}u \cdot K_X u\right) \tag{35.407}$$

où K_X est la matrice de covariance de X.

Démonstration. Nous considérons la variable aléatoire réelle gaussienne $u \cdot X$. Son espérance $m = E(u \cdot X) = u \cdot E(X)$. Nous commençons par établir la formule suivante :

$$u^t K_X u = u \cdot K_X u = E\Big[\big([X - E(X)] \cdot u\big)^2\Big]. \tag{35.408}$$

Utilisant la linéarité de l'espérance,

$$\sum_{kl} E(A_{kl})u_k u_l = \sum_{kl} E(A_{kl}u_k u_l), \tag{35.409}$$

nous trouvons

$$\begin{aligned}
K_X(u,u) &= E\Big([X - E(X)] \otimes [X - E(X)]\Big)(u,u) && \text{(35.410a)}\\
&= E\Big(\big([X - E(X)] \otimes [X - E(X)]\big)(u,u)\Big) && \text{(35.410b)}\\
&= E\Big(\big([X - E(X)] \cdot u\big)\big([X - E(X)] \cdot u\big)\Big) && \text{(35.410c)}\\
&= E\left(\big([X - E(X)] \cdot u\big)^2\right). && \text{(35.410d)}
\end{aligned}$$

Par la linéarité du produit scalaire et de l'espérance,

$$[X - E(X)] \cdot u = u \cdot X - E(u \cdot X), \tag{35.411}$$

ce qui nous ramène à la variable aléatoire $u \cdot X$. Nous avons alors

$$K_X(u,u) = E\left[\big([X - E(X)] \cdot u\big)^2\right] = \mathrm{Var}(u \cdot X). \tag{35.412a}$$

Nous avons donc obtenu une forme pour la variance de la variable aléatoire $u \cdot X$. Étant donné que $u \cdot X$ est gaussienne de moyenne $m = u \cdot E(X)$ et de variance $\sigma^2 = K_X(u,u)$, nous avons

$$\Phi_{u \cdot X}(t) = \exp\big(itm - \frac{1}{2}\sigma^2 t^2\big). \tag{35.413}$$

Par ailleurs nous avons $\Phi_X(u) = \Phi_{u \cdot X}(1)$ parce que

$$\Phi_X(u) = E\big(e^{iu \cdot X}\big) = \Phi_{u \cdot X}(1). \tag{35.414}$$

En utilisant la forme (35.413) pour $\Phi_{u \cdot X}$ nous trouvons

$$\Phi_X(u) = \Phi_{u \cdot X}(1) = \exp\big(im - \frac{1}{2}\sigma^2\big) = \exp\Big(iE(u \cdot X) - \frac{1}{2}u \cdot KXu\Big). \tag{35.415}$$

□

Théorème 35.118.
Soit $X = (X_1, \dots, X_d)$, un vecteur gaussien. Les composantes sont indépendantes si et seulement si elles sont non corrélées.

Démonstration. Nous savons que les variables aléatoires indépendantes sont non corrélées. Nous devons donc surtout prouver le contraire. Le fait que les variables aléatoires X_i soient non corrélées signifie que la matrice de covariance est

$$K_X = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_d^2 \end{pmatrix} \tag{35.416}$$

où $\sigma_k^2 = \mathrm{Var}(X_k)$. Notons $m_k = E(X_k)$. Si $u \in \mathbb{R}^d$, nous avons en vertu de la proposition 35.117 que

$$\Phi_X(u) = \exp\Big(i(u_1 m_1 + \cdots + u_d m_d) - \frac{1}{2}(\sigma_1^2 u_1^2 + \cdots + \sigma_d^2 u_d^2)\Big) \tag{35.417a}$$

$$= \exp\big(iu_1 m_1 - \frac{1}{2}\sigma_1^2 u_1^2\big) \ldots \exp\big(iu_d m_d - \frac{1}{2}\sigma_d^2 u_d^2\big) \tag{35.417b}$$

$$= \Phi_{X_1}(u_1) \ldots \Phi_{X_d}(u_d). \tag{35.417c}$$

Les variables aléatoires X_i sont donc indépendantes parce que la fonction caractéristique se factorise. □

Exemple 35.119.
Nous donnons à présent un exemple de deux variables aléatoires gaussiennes qui ne forment pas un vecteur gaussien. Pour ce faire nous devons chercher des variables aléatoires non indépendantes. Soit $Y \sim \mathscr{N}(0,1)$ et ϵ une variable aléatoire (indépendante de Y) donnée par $P(\epsilon = -1) = \frac{1}{2}$, $P(\epsilon = 1) = \frac{1}{2}$. Nous considérons le vecteur $(Y, \epsilon Y)$.

D'abord montrons que ϵY est une variable aléatoire gaussienne. Soit A un borélien de $\mathbb{R}$. Nous avons

$$P(\epsilon Y \in A) = P(\epsilon = -1, Y \in A) + P(\epsilon = -1, -Y \in A). \tag{35.418}$$

Par indépendance et par symétrie de Y [16] nous trouvons

$$P(\epsilon Y \in A) = P(\epsilon = 1)P(Y \in A) + P(\epsilon = -1)P(-Y \in A) \tag{35.419a}$$

$$= \frac{1}{2}P(Y \in A) + \frac{1}{2}P(-Y \in A) \tag{35.419b}$$

$$= P(Y \in A). \tag{35.419c}$$

Nous avons donc $Y \sim \epsilon Y$, et donc ϵY est gaussienne.

En ce qui concerne la covariance, nous savons que $E(Y) = E(\epsilon) = 0$, donc

$$\mathrm{Cov}(Y, \epsilon Y) = E(Y \cdot \epsilon Y) = E(\epsilon Y^2) = E(\epsilon)E(Y^2) = 0. \tag{35.420}$$

Note : $E(Y^2) = 1$.

Les variables aléatoires Y et ϵY ne sont pas indépendantes. En effet si elles l'étaient, Y serait aussi indépendante de $(\epsilon Y)^2 = Y^2$, alors que Y et Y^2 ne sont pas indépendantes. Donc $X = (Y, \epsilon Y)$ n'est pas un vecteur gaussien. △

Théorème 35.120 ([270]).
Soit $m \in \mathbb{R}^d$ et $K \in \mathbb{M}(d, \mathbb{R})$ une matrice symétrique et positive. Alors il existe un vecteur gaussien de moyenne m et de matrice de covariance K.

Démonstration. Nous effectuons la preuve avec $m = 0$. Nous choisissons l'espace probabilisé $(\Omega, \mathcal{A}) = (\mathbb{R}^r, \mathcal{B}\mathrm{or}(\mathbb{R}^r))$ où r est le rang de K muni de la probabilité de densité

$$\gamma(u) = \frac{1}{(2\pi)^{r/2}} \exp\big(-\frac{1}{2}\|u\|^2\big). \tag{35.421}$$

16. C'est-à-dire que $P(Y \in A) = P(Y \in -A) = P(-Y \in A)$.

Nous considérons la variable aléatoire

$$\begin{aligned} Y_0 \colon \Omega &\to \mathbb{R}^r \\ u &\mapsto u. \end{aligned} \tag{35.422}$$

C'est une variable aléatoire gaussienne de loi $P_{Y_0} = \gamma\lambda_d$ où λ_d est la mesure de Lebesgue sur $\mathbb{R}^d$. Sa densité (35.421) s'écrit comme le produit de r gaussiennes indépendantes ; sa matrice de covariance est donc $\mathbb{1}_{r\times r}$.

Étant donné que K est symétrique et positive, il existe une matrice $d \times r$ telle que $K = AA^t$. Pour voir cela, remarquons qu'il existe une matrice $d\times d$ qui fait le travail. En effet K se diagonalise par une orthogonale (théorème 9.150) :

$$K = ADA^t = A\sqrt{D}\sqrt{D}A^t \tag{35.423}$$

où D est une matrice diagonale contenant $d-r$ zéros et $\sqrt{D}$ est la matrice que l'on imagine. Donc la matrice $L = A\sqrt{D}$ est une matrice telle que $LL^t = K$. Maintenant, étant donné que les $d-r$ dernières lignes de D sont vides, les $d-r$ dernières lignes de L n'ont pas d'importance et peuvent être choisies nulles, voire même ne pas exister. La matrice $A \in \mathbb{M}_{d\times r,\mathbb{R}}$ qui réalise $AA^t = K$ est la « troncature » de L à ses r premières lignes.

Nous considérons la variable aléatoire $Y = AY_0 \colon \Omega \to \mathbb{R}^d$. Étant donné que AY_0 est une transformation linéaire d'un vecteur gaussien, c'est un vecteur gaussien. Nous avons encore $m(Y) = 0$ et

$$\operatorname{Var}(Y) = E(Y \otimes Y) = E\big((AY_0) \otimes (AY_0)\big) = AE(Y_0 \otimes Y_0)A^t = AA^t = K \tag{35.424}$$

parce que $E(Y_0 \otimes Y_0) = \mathbb{1}$. Nous avons utilisé les formules du produit tensoriel introduit en (10.284), et en particulier la formule (10.287). □

Lorsqu'une matrice symétrique et positive K est donné, nous avons créé un vecteur gaussien de covariance K en créant $X = AY$ où Y est le vecteur gaussien « le plus simple » et où A est donné par $AA^t = K$. La proposition suivante montre l'inverse : un vecteur gaussien X peut se réduire au vecteur gaussien « le plus simple » en utilisant la transformation $Y = A^{-1}X$ où A est encore donnée par $AA^t = K$. Ce résultat nous permettra de voir les vecteurs gaussiens généraux comme des « changements de coordonnées » par rapport au vecteur gaussien simple $Y = (Y_1, \dots, Y_d)$ avec $Y_i \sim \mathcal{N}(0,1)$.

Proposition 35.121.
Soit $Y = (Y_1, \dots, Y_n)$ le vecteur gaussien formé des variables aléatoires indépendantes $Y_i \sim \mathcal{N}(0,1)$. Soit K une matrice symétrique et positive et la matrice A telle que $AA^t = K$. Alors le vecteur $X = AY$ est gaussien de covariance K.

Démonstration. Nous montrons que la covariance de $X = AY$ est donnée par K. Nous avons

$$\begin{aligned} K_X &= E\big([X - E(X)] \otimes [X - E(X)]\big) && (35.425a)\\ &= E\big(A(Y - E(Y)) \otimes A(Y - E(Y))\big) && (35.425b)\\ &= E\Big(A\big([Y - E(Y)] \otimes [Y - E(Y)]\big)A^t\Big) && (35.425c)\\ &= E(AA^t) && (35.425d)\\ &= K_X. && (35.425e) \end{aligned}$$

Nous avons utilisé le lemme 10.126 ainsi que le fait que

$$[Y - E(Y)] \otimes [Y - E(Y)] = K_Y = \mathbb{1}. \tag{35.426}$$

Notons que $E(Y) = 0$, mais cela ne joue pas ici. □

Théorème 35.122.
Un vecteur gaussien $X \colon \Omega \to \mathbb{R}^d$ possède une densité par rapport à la mesure de Lebesgue si et seulement si sa matrice de covariance est inversible. Dans ce cas nous avons la densité

$$f_X(x) = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{|\det(K_X)|}} \exp\left(-\frac{1}{2}[X - E(X)] \cdot K_X^{-1}[X - E(X)]\right). \tag{35.427}$$

Démonstration. Nous supposons que $E(X) = 0$. En utilisant la proposition 35.121, nous posons $X = AY$ où Y est un vecteur gaussien $Y \sim \mathscr{N}(0, \mathbb{1})$ et $AA^t = K_X$. Si K n'est pas inversible, alors A n'est pas inversible non plus. Notons $r < d$ le rang de A. Étant donné que $X = AY$, la variable aléatoire X prend presque surement ses valeurs dans l'image de A, c'est-à-dire dans un sous-espace de dimension $r < d$ de $\mathbb{R}^d$. Ce sous-espace est de mesure nulle pour la mesure de Lebesgue, mais de mesure 1 pour la mesure P_X. La mesure P_X ne peut donc pas avoir de densité par rapport à celle de Lebesgue.

Supposons maintenant que K_X soit inversible. La matrice A l'est aussi. Nous anticipons l'utilisation du théorème de changement de variable 13.255. Ici le changement de variable sera la transformation linéaire A dont le jacobien vaut

$$|\det(A^{-1})| = \frac{1}{\det A} = \frac{1}{\sqrt{|\det K_X|}}. \tag{35.428}$$

Soit γ la densité de Y. Nous posons

$$f_X(x) = \frac{1}{\sqrt{|\det K_X|}}\gamma(A^{-1}X) \tag{35.429}$$

où γ est le produit des densités de d gaussiennes usuelles $\mathscr{N}(0,1)$. Nous allons d'abord montrer que cette formule est bien la fonction (35.427) et ensuite que X admet f_X comme densité. Nous avons

$$\gamma(A^{-1}) = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\|A^{-1}x\|^2\right), \tag{35.430}$$

et

$$\begin{aligned}
\|A^{-1}x\|^2 &= \langle A^{-1}x, A^{-1}x\rangle && (35.431a)\\
&= \langle (A^{-1})^t A^{-1}x, x\rangle && (35.431b)\\
&= \langle (AA^t)^{-1}x, x\rangle && (35.431c)\\
&= x \cdot K_X^{-1}x, && (35.431d)
\end{aligned}$$

ce qui nous donne bien la formule (35.427). Nous vérifions maintenant que f_X est bien une densité pour X. Soit B un borélien de $\mathbb{R}^d$. Nous avons d'abord

$$P(X \in B) = P(AY \in B) = P(Y \in A^{-1}B). \tag{35.432}$$

Ici nous avons utilisé le fait que A était bijectif. Nous avons ensuite

$$P(X \in B) = \int_{A^{-1}B} \gamma(t)dy = \int_B \gamma\big(A^{-1}x\big)|J_{A^{-1}}(x)|dx = \int_B f_X(x)dx. \tag{35.433}$$

C'est ici que nous avons utilisé le théorème de changement de variable 13.255. □

35.5.11 Variable aléatoire de Rademacher

Une variable aléatoire **de Rademacher** est une variable aléatoire de loi

$$\epsilon \sim \delta_0 - \delta_1, \tag{35.434}$$

sur l'ensemble $\Omega = \{0, 1\}$. C'est la variable aléatoire qui prend valeur 1 ou -1 avec probabilité $\frac{1}{2}$.

Parmi les propriétés évidentes de cette variable aléatoire nous avons $E(\epsilon) = 0$ et $E(\epsilon^2) = 1$.

Proposition 35.123 (Inégalité de Khintchine[54])**.**
Soient $r_1, \dots, r_n$ des variables aléatoires indépendantes et identiquement distribuées de Rademacher et une combinaison linéaire

$$X = \sum_{i=1}^n a_i r_i \tag{35.435}$$

avec $a_i \in \mathbb{R}$. Alors

$$\|X\|_2 \leqslant \sqrt{e}E\big(|X|\big). \tag{35.436}$$

Démonstration. Pour rappel la définition est que

$$\|X\|_2 = \sqrt{E\big(|X|^2\big)}. \tag{35.437}$$

Vu que c'est de la quantité $E\big(|X|\big)$ que nous voulons parler, nous notons l'inégalité

$$E\big(|X|\big) = \int_\Omega \big|X(\omega)\big|dP(\omega) \geqslant \Big|\int_\Omega X(\omega)dP(\omega)\Big| = \big|E(X)\big|. \tag{35.438}$$

Nous supposons que $\sum_{j=1}^n a_j^2 = 1$; sinon nous multiplions les a_j parce qu'il faut pour l'avoir et ce facteur sortira des deux côtés de (35.436).

Nous allons passer par la variable aléatoire intermédiaire

$$Y = \prod_{j=1}^n (1 + ia_j r_j) \tag{35.439}$$

et pour presque tout $\omega \in \Omega$ nous avons

$$\begin{aligned}
|Y(\omega)| &= \prod_{j=1}^n |1 + ia_j r_j(\omega)| && (35.440a)\\
&= \prod_{j=1}^n \sqrt{1 + a_j^2 \underbrace{r_j(\omega)^2}_{=1}} && (35.440b)\\
&= \prod_{j=1}^n \sqrt{1 + a_j^2} && (35.440c)\\
&\leqslant \prod_{j=1}^n \sqrt{e^{a_j^2}} && (35.440d)\\
&= \sqrt{\prod_{j=1}^n e^{a_j^2}} && (35.440e)\\
&= \sqrt{e^{\sum a_j^2}} && (35.440f)\\
&= \sqrt{e}. && (35.440g)
\end{aligned}$$

Donc $\|Y\|_\infty \leqslant \sqrt{e}$ où $\|.\|_\infty$ est la norme supremum sur Ω. Cela nous permet de donner une première inégalité à propos de $E(|X|)$. D'abord

$$\big|E(XY)\big| = \Big|\int_\Omega X(\omega)Y(\omega)dP(\omega)\Big| \leqslant \int_\Omega \big|X(\omega)\big|\|Y\|_\infty dP(\omega) = \|Y\|_\infty E\big(|X|\big), \tag{35.441}$$

ensuite en remplaçant $\|Y\|_\infty$ par la majoration que nous venons de donner de $|Y(\omega)|$,

$$\sqrt{e}E\big(|X|\big) \geqslant \big|E(XY)\big|. \tag{35.442}$$

Il nous reste à prouver que $\big|E(XY)\big| \geqslant \|X\|_2$.

Pour ce faire nous commençons par noter que pour chaque j nous avons $E(r_j) = 0$ et $E(ia_j r_j^2) = ia_j$; en utilisant l'indépendance des r_j et le lemme 35.22 nous avons alors

$$\begin{aligned}
E(r_jY) &= E\Big(r_j(1 + ia_j r_j)\prod_{k\neq j}(1 + ia_k r_k)\Big) && (35.443a)\\
&= E\big(r_j(1 + ia_j r_j)\big)\prod_{k\neq j} E(1 + ia_k r_k) && (35.443b)\\
&= ia_j \prod_{k\neq j}\big(1 + ia_k E(r_k)\big) && (35.443c)\\
&= ia_j. && (35.443d)
\end{aligned}$$

Par conséquent, en utilisant la proposition 35.29 dans le cas non indépendant,

$$E(XY) = \sum_{j=1}^{n} a_j E(r_j Y) = \sum_j a_j i a_j = i \sum_j a_j^2 = i. \tag{35.444}$$

Nous pouvons compléter l'équation (35.442) en

$$\sqrt{e} E\big(|X|\big) \geqslant \big|E(XY)\big| = 1, \tag{35.445}$$

et nous nous empressons de montrer que $\|X\|_2 = 1$. En effet $\|X\|_2 = \sqrt{E(|X|^2)}$ alors que

$$\begin{aligned}
\|X\|_2^2 &= E\big((\sum_i a_i r_i)^2\big) && (35.446a)\\
&= E\Big(\sum_k a_k^2 r_k^2 + 2\sum_{i\neq j} a_i a_j r_i r_j\Big) && (35.446b)\\
&= \sum_k a_k^2 + 2\sum_{i\neq j} a_i a_j E(r_i r_j) && (35.446c)\\
&= 1 && (35.446d)
\end{aligned}$$

□

35.5.12 Loi de Student

Définition 35.124.
La loi χ^2 à d degrés de liberté est la loi de la variable aléatoire $Y_1^2 + \cdots + y_n^2$ si les (Y_i) sont des variables aléatoires normales indépendantes centrées et réduites.

*La loi de **Student** à d degrés de liberté est la loi de la variable aléatoire*

$$\frac{X}{\sqrt{K/d}} \tag{35.447}$$

où $X \sim \mathscr{N}(0,1)$ et $K \sim \chi^2(d)$ sont des variables aléatoires indépendantes. Cette loi est notée $\mathscr{T}(d)$

Nous avons une illustration de la densité de la loi $\chi^2(10)$ à la figure 35.1.

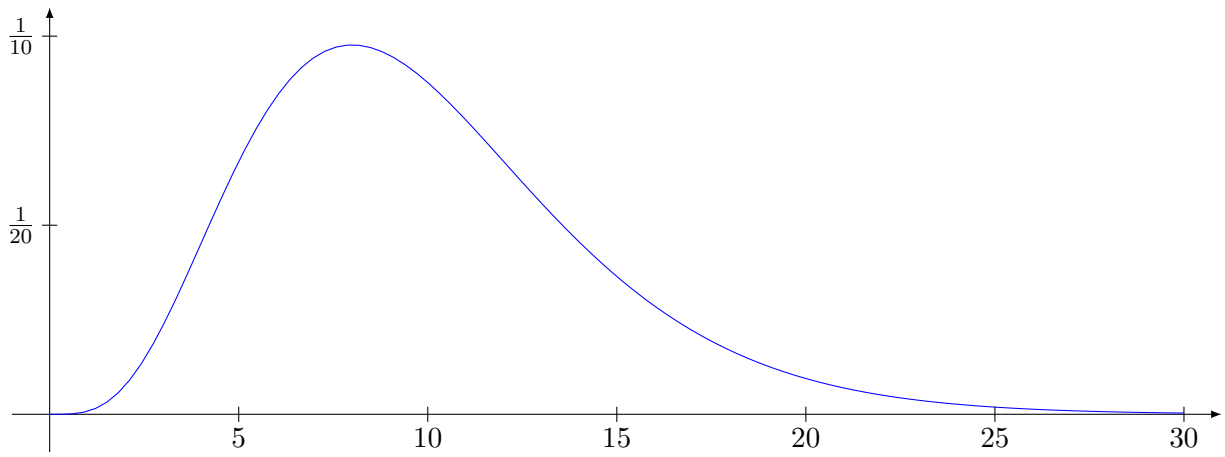


FIGURE 35.1 – La densité de $\chi^2(10)$.

L'importance de cette loi sera dans le théorème de Cochrane 36.15.

35.5.13 Indépendance, covariance et variance de somme

Si X et Y sont des variables aléatoires réelles, nous avons défini la covariance par (35.69) :

$$\mathrm{Cov}(X,Y) = E\Big[\big(X - E(X)\big)\big(Y - E(Y)\big)\Big] \tag{35.448}$$

Une clef est le lemme 35.22 qui dit que lorsque X et Y sont indépendantes, alors $E(XY) = E(X)E(Y)$. Nous avons les liens suivants.

(1) Si X et Y sont indépendantes, alors $\mathrm{Cov}(X,Y) = 0$.

(2) La réciproque n'est pas vraie par l'exemple 35.24.

(3) Si Z est un vecteur gaussien, les composantes Z_i sont indépendantes si et seulement si la matrice de covariance est diagonale, c'est-à-dire que les Z_i sont deux à deux non corrélées ; c'est le théorème 35.118.

(4) En ce qui concerne la variance d'une somme,

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y). \tag{35.449}$$

Donc lorsque X et Y ne sont pas corrélées, la variance est sympa avec la somme. En particulier lorsqu'elles sont indépendantes, mais pas seulement.

35.6 Estimation des grands écarts

Si S_n est une somme de variables aléatoires de Bernoulli X_i indépendantes de probabilité p, l'espérance de S_n/n est p, et nous voudrions savoir quelle est la probabilité d'avoir un taux de succès un peu plus important :

$$P\big(\frac{S_n}{n} \geqslant p + \epsilon\big). \tag{35.450}$$

La loi des grands nombres nous permet de dire que ça ne va pas être très grand. En effet si nous posons

$$Y_i = X_i - p - \epsilon \tag{35.451}$$

et

$$Z_n = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{S_n}{n} - p - \epsilon, \tag{35.452}$$

la loi des grands nombres (théorème 35.91) nous indique que

$$Z_n \xrightarrow{p.s.} E(Y_1) = -\epsilon. \tag{35.453}$$

La proposition 35.76 sur le lien entre les types de convergence nous donne immédiatement la convergence en loi. C'est-à-dire que pour tout $\eta > 0$,

$$P\Big(|Z_n + \epsilon| \geqslant \eta\Big) \to 0. \tag{35.454}$$

En prenant $\eta = \frac{\epsilon}{2}$,

$$P(Z_n = 0) \leqslant P\big(|Z_n + \epsilon| \geqslant \frac{\epsilon}{2}\big) \to 0. \tag{35.455}$$

Tout cela montre que

$$\lim_{n\to\infty} P\big(\frac{S_n}{n} \geqslant p + \epsilon\big) = 0. \tag{35.456}$$

Autrement dit, la probabilité de tomber à une distance fixée de la moyenne tend vers zéro lorsque le nombre d'essais augmente. Rien d'étonnant.

Le théorème suivant nous indique la vitesse de convergence. Elle est exponentielle et le coefficient est donné en fonction de p et de ϵ.

Théorème 35.125 ([54]).
Soient des variables aléatoires $X_i \sim \mathscr{B}(p)$ et

$$S_n = \sum_{i=1}^{n} X_i \sim \mathscr{B}(n,p). \tag{35.457}$$

Pour $\epsilon \in]0, 1-p[$, nous définissons

$$h_+(\epsilon) = (p+\epsilon)\ln\left(\frac{p+\epsilon}{p}\right) + (1-p-\epsilon)\ln\left(\frac{1-p-\epsilon}{1-p}\right). \tag{35.458}$$

Alors

(1) $h_+(\epsilon) > 0$.

(2) Pour tout $n \geqslant 1$, nous avons

$$P\left(\frac{S_n}{n} \geqslant p+\epsilon\right) \leqslant e^{-nh_+(\epsilon)}. \tag{35.459}$$

(3) L'estimation (35.459) est optimale au sens que

$$\lim_{n\to\infty} \frac{1}{n} \ln\left(P\big(\frac{S_n}{n} \geqslant p+\epsilon\big)\right) = -h_+(\epsilon). \tag{35.460}$$

Démonstration. Pour tout $t \geqslant 0$ nous avons :

$$\begin{aligned}
P\left(\frac{S_n}{n} \geqslant p+\epsilon\right) &= P\big(S_n \geqslant np + n\epsilon\big) && (35.461a)\\
&\leqslant E\Big(e^{t(S_n - np - n\epsilon)}\Big) && (35.461b)\\
&= e^{-nt(p+\epsilon)} E\left(e^{tS_n}\right) && (35.461c)\\
&= e^{-nt(p+\epsilon)} \sum_{k=0}^{\infty} e^{tk} P(S_n = k) && (35.461d)\\
&= e^{-nt(p+\epsilon)} \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} && (35.461e)\\
&= e^{-nt(p+\epsilon)} \big((1-p) + pe^t\big)^n && (35.461f)\\
&= \exp\Big(-n\big(t(p+\epsilon) - \ln(1-p+pe^t)\big)\Big). && (35.461g)
\end{aligned}$$

Justifications :

— L'inégalité (35.461b) est l'inégalité de Markov (corolaire 35.90) avec $\phi(x) = e^{tx}$.

— La ligne (35.461d) est une utilisation du théorème de transfert 35.68.

Nous posons maintenant

$$h = \sup_{t>0} \big(t(p+\epsilon) - \ln(1-p+pe^t)\big). \tag{35.462}$$

Pour cette valeur de h nous avons

$$P\left(\frac{S_n}{n} \geqslant p+\epsilon\right) \leqslant e^{-nh}. \tag{35.463}$$

Nous considérons la fonction

$$\begin{aligned}
g\colon \mathbb{R}^+ &\to \mathbb{R}\\
t &\mapsto t(p+\epsilon) - \ln(1-p+pe^t).
\end{aligned} \tag{35.464}$$

Elle vérifie $g(0) = 0$ et

$$g'(t) = (p+\epsilon) - \frac{pe^t}{1-p+pe^t} \tag{35.465a}$$

$$g'(0) = (p+\epsilon) - \frac{p}{1-p+p} = \epsilon > 0. \tag{35.465b}$$

Par conséquent sur un voisinage de $t = 0$ la fonction g est strictement croissante et nous concluons que g prend (au moins) quelques valeurs strictement positives. Du coup nous avons

$$h = \|g\|_\infty > 0. \tag{35.466}$$

Nous cherchons maintenant pour quelle valeur de t est réalisé le maximum de g. D'abord résoudre $g'(t) = 0$ donne

$$t_0 = \ln\left(\frac{(p+\epsilon)(1-p)}{p(1-p-\epsilon)}\right) \tag{35.467}$$

Vu que $g'(t) \to p + \epsilon - 1 < 0$, nous avons $\lim_{t\to\infty} g(t) = -\infty$ et donc le t_0 trouvé est bien un maximum et non un minimum.

Il est maintenant loisible de calculer une valeur pour h : il suffit de calculer $g(t_0)$. La calcul n'est pas très compliqué et donne

$$h = g(t_0) = (p+\epsilon)\ln\left(\frac{p+\epsilon}{p}\right) + (p+\epsilon-1)\ln\left(\frac{1-p}{1-p-\epsilon}\right), \tag{35.468}$$

ce qui est bien $h = h_+(\epsilon)$. Cela démontre les points (1) et (2).

Nous montrons à présent l'aspect optimal de l'estimation. Nous savons déjà que

$$\frac{1}{n}\ln\left(P\big(\frac{S_n}{n} \geqslant p+\epsilon\big)\right) \leqslant h_+(\epsilon). \tag{35.469}$$

Nous posons $k_n = \lceil n(p+\epsilon)\rceil$. Vu que $p + \epsilon < 1$ et que $S_n \leqslant n$, nous avons

$$P\big(\frac{S_n}{n} \geqslant p+\epsilon\big) = P\big(S_n \geqslant n(p+\epsilon)\big) \geqslant P(S_n = k_n) = \binom{n}{k_n} p^{k_n}(1-p)^{n-k_n}. \tag{35.470}$$

C'est maintenant que nous utilisons la formule de Stirling (lemme 19.202) pour chacune des factorielles intervenant dans le coefficient binomial. Nous trouvons :

$$P\left(\frac{S_n}{n} \geqslant p+\epsilon\right) \geqslant P(S_n = k_n) = \clubsuit = \frac{\left(\frac{n}{e}\right)^n \sqrt{2\pi n}\alpha(n) p^{k_n}(1-p)^{n-k_n}}{\left(\frac{k_n}{e}\right)^{k_n}\sqrt{2\pi k_n}\alpha(k_n)\left(\frac{n-k_n}{e}\right)^{n-k_n}\sqrt{2\pi(n-k_n)}}. \tag{35.471}$$

Nous savons que $k_n = n(p+\epsilon) + \sigma(n)$ avec σ borné par 1. Par conséquent $n - k_n \to \infty$ et nous pouvons regrouper les coefficients en α en

$$\beta(n) = \frac{\alpha(n)}{\alpha(k_n)\alpha(n-k_n)} \to 1. \tag{35.472}$$

Nous remarquons aussi que les e se simplifient. Nous récrivons $\clubsuit$ sous la forme

$$\clubsuit = \underbrace{\frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{k_n(n-k_n)}}\beta(n)}_{=A(n)} \frac{n^n p^{k_n}(1-p)^{n-k_n}}{k_n^{k_n}(n-k_n)^{n-k_n}} \tag{35.473a}$$

$$= A(n)n^n\left(\frac{p}{k_n}\right)^{k_n}\left(\frac{1-p}{n-k_n}\right)^{n-k_n} \tag{35.473b}$$

$$= A(n)\left(\frac{np}{k_n}\right)^{k_n}\left(\frac{n(1-p)}{n-k_n}\right)^{n-k_n}. \tag{35.473c}$$

Nous passons au logarithme et nous étudions $\frac{1}{n}\ln\big(P(S_n=k_n)\big)$. Nous avons les termes suivants à étudier :

$$\begin{aligned}\frac{1}{n}\ln\big(P(S_n=k_n)\big) &= -\frac{1}{2n}\ln(2\pi)+\frac{1}{2n}\ln\left(\frac{n}{k_n(n-k_n)}\right)\\ &\quad + k_n\ln\left(\frac{np}{k_n}\right)+(n-k_n)\ln\left(\frac{n(1-p)}{n-k_n}\right)+\frac{1}{n}\ln\big(\alpha(n)\big).\end{aligned}\tag{35.474}$$

Nous étudions terme à terme la limite de cela lorsque $n\to\infty$.

(1) Le terme $\frac{1}{2n}\ln(2\pi)$ ne pose pas de problèmes. Il tend vers zéro.

(2) Si nous remplaçons k_n par $n(p+\epsilon)+\sigma(n)$ nous voyons que ce qui est dans le logarithme est majoré par $\frac{1}{P(n)}$ pour un certain polynôme P. Ce terme est dans le cas $\frac{\ln(P(n))}{n}$ qui tend vers zéro lorsque $n\to\infty$.

(3) Pour ce terme nous remplaçons k_n par $n(p+\epsilon)+k_n-n(p+\epsilon)$. Nous devons alors étudier la limite de

$$(p+\epsilon)\ln\left(\frac{np}{k_n}\right)+\frac{k_n-n(p+\epsilon)}{n}\ln\left(\frac{np}{k_n}\right).\tag{35.475}$$

Ce qui est dans les logarithmes est encadré de la façon suivante :

$$\frac{n(p+\epsilon)}{np}\leqslant\frac{k_n}{np}\leqslant\frac{n(p+\epsilon)+1}{np}.\tag{35.476}$$

Donc la limite de k_n/np est $(p+\epsilon)/p$. Les logarithmes restent bornés. Pour le second terme de (35.475), le numérateur du coefficient est borné par 1. Donc le second terme tend vers zéro et le tout tend vers

$$(p+\epsilon)\ln\left(\frac{p}{p+\epsilon}\right).\tag{35.477}$$

(4) Nous devons enfin étudier le dernier terme. La combinaison $\frac{n-k_n}{n}$ s'étudie de la façon suivante :

$$\frac{n-k_n}{n}=\frac{n-n(p+\epsilon)+n(p+\epsilon)-k_n}{n}=\frac{n(1-p-\epsilon)+n(p+\epsilon)-k_n}{n}\to 1-p-\epsilon\tag{35.478}$$

parce que $n(p+\epsilon)-k_n$ est borné par 1. Sachant cela, notre terme a pour limite

$$\frac{n-k_n}{n}\ln\left(\frac{n(1-p)}{n-k_n}\right)\to(1-p-\epsilon)\ln\left(\frac{1-p}{1-p-\epsilon}\right).\tag{35.479}$$

En remettant tous les morceaux bouts à bout,

$$\frac{1}{n}P(S_n=k_n)\to(1+\epsilon)\ln\left(\frac{p}{p+/e}\right)+(1-p-\epsilon)\ln\left(\frac{1-p}{1-p-\epsilon}\right)=-h_+(\epsilon).\tag{35.480}$$

Étant donné que nous avions déjà prouvé que $P\big(\frac{S_n}{n}\geqslant p+\epsilon\big)\geqslant P(S_n=k_n)$, nous avons

$$\liminf_{n\to\infty}\frac{1}{n}\ln\Big(P\big(\frac{S_n}{n}\geqslant p+\epsilon\big)\Big)\geqslant\lim_{n\to\infty}\frac{1}{n}P(S_n=k_n)=-h_+(\epsilon).\tag{35.481}$$

En combinant avec (35.469) nous trouvons que

$$\lim_{n\to\infty}\frac{1}{n}\ln\Big(P\big(\frac{S_n}{n}\geqslant p+\epsilon\big)\Big)=h_+(\epsilon).\tag{35.482}$$

Pour cette dernière déduction nous utilisons le fait que si (a_n) est une suite telle que $a_n\leqslant l$ et $\liminf_{n\to\infty}a_n=l$, alors (a_n) admet une limite qui vaut l. □

35.7 Simulations de réalisations de variables aléatoires

Le générateur de base que possède un système informatique est un générateur de nombres pseudo-aléatoires de nombres entiers entre 0 et $m-1$ généré par une suite du type

$$x_{n+1} = (ax_n + b) \mod m. \tag{35.483}$$

35.7.1 Générateur uniforme

35.7.1.1 Première méthode

Une première façon de générer une variable aléatoire de loi uniforme sur $]0,1[$ est de diviser par $m-1$ la suite de x_n. En effet nous avons la proposition suivante.

Proposition 35.126.
Si (Y_n) est une suite de variables aléatoires indépendantes et identiquement distribuées de loi uniforme sur $\{0,\dots,n-1\}$. Alors

$$\frac{Y_n}{n} \xrightarrow{\mathscr{L}} \mathscr{U}[0,1]. \tag{35.484}$$

Démonstration. Nous prouvons la convergence en loi en passant par la fonction de répartition et la proposition 35.77. La fonction de répartition de la densité $\mathscr{U}[0,1]$ est

$$F_X(x) = x\mathbb{1}|_{[0,1]}. \tag{35.485}$$

La fonction de répartition de la variable aléatoire (discrète) $X_n = \frac{Y_n}{n}$ est

$$P(\frac{Y_n}{n} \leqslant x) = P(Y_n \leqslant nx) = \frac{\lfloor nx \rfloor}{n} \tag{35.486}$$

où $\lfloor a \rfloor$ est le plus grand entier inférieur à a. Nous avons évidemment

$$\lim_{n\to\infty} \frac{\lfloor nx \rfloor}{n} = x, \tag{35.487}$$

ce qui montre la convergence des fonctions de répartitions et donc la convergence en loi qui nous intéresse. □

35.7.1.2 Seconde méthode

Soit (Y_n) une suite de variables aléatoires indépendantes et identiquement distribuées selon la loi $Y_n \sim \mathscr{U}\{0,\dots,m-1\}$. Alors la série de variables aléatoires

$$Z = \sum_{k=0}^{\infty} \frac{Y_k}{m^{k+1}} \tag{35.488}$$

est une série qui converge presque surement parce que Y_k est borné par m. Avec probabilité zéro nous avons $Z = \sum_k 1/m^k$ qui converge. Nous avons

$$Z \sim \mathscr{U}[0,1]. \tag{35.489}$$

L'argument pour montrer cette loi est qu'en base m, la variable aléatoire Z a un développement décimal $Z = 0.Y_1Y_2Y_3Y_4\cdots$.

35.7.2 Simulation par inversion

Nous cherchons maintenant à simuler une loi X de fonction de répartition F.

Définition 35.127.
Soit $f\colon \mathbb{R} \to [0,1]$ une fonction croissante, continue à droite et telle que

$$\lim_{x\to-\infty} f(x) = 0 \tag{35.490a}$$

$$\lim_{x\to\infty} f(x) = 1. \tag{35.490b}$$

*L'**inverse généralisé** de f, notée f^{-1} est la fonction définie par*

$$f^{-1}(t) = \inf\{x \text{ tel que } f(x) \geqslant t\}. \tag{35.491}$$

Remarque 35.128.
L'inverse généralisé d'une fonction bijective est la vraie fonction réciproque usuelle.

Proposition 35.129.
Soit f une fonction admettant un inverse généralisé f^{-1}. Alors nous avons $f^{-1}(t) \leqslant a$ si et seulement si $t \leqslant f(a)$.

La continuité à droite joue pour démontrer cette proposition.

Proposition 35.130.
Si F est la fonction de répartition de la variable aléatoire X et si V est une variable aléatoire de loi uniforme $\mathscr{U}[0,1]$, alors $F^{-1}(U)$ a la même loi que X.

Démonstration. Nous montrons que les fonctions de répartition de X et de $F^{-1}(U)$ sont identiques. En utilisant la proposition 35.129, nous avons

$$P\big(F^{-1}(U) \leqslant y\big) = P\big(U \leqslant F(y)\big) \tag{35.492a}$$

$$= F(y) \tag{35.492b}$$

$$= P(X \leqslant y). \tag{35.492c}$$

Donc $F^{-1}(U)$ est la fonction de répartition de X. □

La difficulté de la méthode par inversion est qu'il faut être capable de calculer l'inverse de la fonction de répartition de la loi à simuler.

35.7.2.1 Loi exponentielle

La loi exponentielle est une loi qui peut être simulée par inversion. La fonction de répartition vaut

$$F(x) = 1 - e^{\lambda x}, \tag{35.493}$$

et l'inverse vaut

$$F^{-1}(x) = -\frac{1}{\lambda}\ln(1-y). \tag{35.494}$$

Par conséquent, une bonne formule pour simuler une loi exponentielle est

$$-\frac{1}{\lambda}\ln(1-U). \tag{35.495}$$

Notez que U étant uniforme, nous pouvons tout autant prendre $-\ln(U)/\lambda$.

35.7.3 Algorithme de Box-Muller

Il s'agit de simuler une loi gaussienne. La proposition est la suivante.

Proposition 35.131.
Si U et V sont des variables aléatoires indépendantes de même loi uniforme sur $[0,1]$, alors le couple

$$(X,Y) = \big(\sqrt{-2\ln(U)}\cos(2\pi V), \sqrt{-2\ln(U)}\sin(2\pi V)\big) \tag{35.496}$$

vérifie

(1) X est indépendante de Y

(2) X et Y sont de loi $\mathscr{N}(0,1)$.

Démonstration. Nous allons montrer la proposition en utilisant les fonctions tests. Soit donc $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$ une fonction bornée et mesurable. Soient Z et W, deux variables aléatoires indépendantes de loi $\mathscr{N}(0,1)$. Nous allons montrer que

$$F\big(\varphi(X,Y)\big) = E\Big[\varphi\big(\sqrt{-2\ln(U)}\cos(2\pi V), \sqrt{-2\ln(U)}\sin(2\pi V)\big)\Big]. \tag{35.497}$$

Par indépendance de U et V, la densité du couple est le produit des densités, donc en passant aux coordonnées polaires,

$$\begin{aligned}
\diamondsuit &= E\big[\varphi(Z,W)\big] && (35.498a)\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{\infty}^{\infty}\varphi(x,y)e^{-x^2/2}e^{-y^2/2}dxdy && (35.498b)\\
&= \frac{1}{2\pi}\int_0^{2\pi}d\theta\int_0^{\infty}\varphi(r\cos\theta, r\sin\theta)e^{-r^2/2}rdr. && (35.498c)
\end{aligned}$$

Nous posons $u = e^{-r^2/2}$ et $v = \frac{\theta}{2\pi}$. En particulier $r = \sqrt{-2\ln(u)}$ et

$$\begin{aligned}
\diamondsuit &= \int_0^1\int_0^1\varphi\big(\sqrt{-2\ln(u)}\cos(2\pi v), \sqrt{-2\ln(u)}\sin(2\pi v)\big)dudv && (35.499a)\\
&= E\Big(\varphi\big(\sqrt{-2\ln(U)}\cos(2\pi V), \sqrt{-2\ln(U)}\sin(2\pi V)\big)\Big) && (35.499b)
\end{aligned}$$

parce que mesure $dudv$ est la densité de la loi uniforme. □

En pratique, la formule

$$(x,y) \mapsto \big(\sqrt{-2\ln x}\cos(2\pi y), \sqrt{-2\ln x}\sin(2\pi y)\big) \tag{35.500}$$

est une façon d'obtenir deux gaussiennes à partir de deux variables uniformes.

35.7.4 Méthode du rejet

La méthode du rejet permet de simuler des lois à densité. Soit f la densité de la loi à simuler. Nous faisons les hypothèses suivantes.

(1) Il existe une densité g d'une variable aléatoire facile à simuler.

(2) Il existe un $k \geqslant 0$ tel que $f(x) \leqslant kg(x)$.

Remarque 35.132.
Le k de la seconde hypothèse est nécessairement plus grand que 1. En effet,

$$1 = \int f \leqslant k\int g = k \tag{35.501}$$

parce que f et g sont des densités et ont donc une intégrale égale à 1.

Proposition 35.133.
Soient (X_n) et (U_n) des suites de variables aléatoires indépendantes au sens où non seulement les X_i et U_k sont indépendants entre eux, mais de plus X_i est indépendant de U_j pour tout i et j. Nous supposons que les X_i sont indépendantes et identiquement distribuées, de densité g et que les U_i sont indépendantes et identiquement distribuées de loi uniforme.

Nous introduisons la variable aléatoire à valeurs dans $\mathbb{N}$

$$p(\omega) = \inf\{n \geqslant 0 \text{ tel que } \alpha\big(X_n(\omega)\big) \geqslant U_n(\omega)\} \tag{35.502}$$

où α est la fonction définie par

$$\alpha(x) = \begin{cases} \frac{f(x)}{kg(x)} & si\ g(x) \neq 0 \\ 0 & si\ g(x) = 0. \end{cases} \tag{35.503}$$

Alors la variable aléatoire Y définie par

$$Y(\omega) = X_{p(\omega)}(\omega) \tag{35.504}$$

admet f pour densité.

Démonstration. D'abord étant donné que $f(x) \leqslant kg(x)$ nous avons $\alpha(x) \in [0,1]$. Nous pouvons à priori avoir $p(\omega) = \infty$, ce qui rendrait caduque la définition de $Y(\omega)$. Montrons donc pour commencer que $P(p = \infty) = 0$. En utilisant l'indépendance nous avons

$$P\big(\alpha(X_n) < U_n, \forall n\big) = \lim_{N\to\infty} \prod_{i=1}^{N} P\big(\alpha(X_i) < U_i\big) \tag{35.505a}$$

$$= \lim_{N\to\infty} P\big(\alpha(X_1) < U_1\big)^N. \tag{35.505b}$$

Pour conclure nous devons prouver que $P\big(\alpha(X_1) < U_1\big) < 1$. Pour cela nous calculons

$$P\big(\alpha(X_1) < U_1\big) = \int_{\mathbb{R}} dx \int_0^1 du \mathbb{1}_{\alpha(x)<u} g(x) \tag{35.506a}$$

$$= \int_{\mathbb{R}} g(x)\big(1 - \alpha(x)\big)dx \tag{35.506b}$$

$$= \int_{\mathbb{R}} \left(g(x) - \frac{f(x)}{k}\right) \tag{35.506c}$$

$$= 1 - \frac{1}{k} \tag{35.506d}$$

$$< 1. \tag{35.506e}$$

L'équation (35.505) nous permet donc de conclure que $P\big(\alpha(X_n) < U_n, \forall n\big) = 0$. Par conséquent la variable aléatoire $Y(\omega) = X_{p(\omega)}(\omega)$ a un sens.

Nous devons maintenant prouver que Y a bien f pour densité. Pour cela nous considérons un ensemble mesurable $A \in \mathcal{B}\mathrm{or}(\mathbb{R})$ et nous montrons que $P(Y \in A) = \int_A f(x)dx$. Nous avons

$$P(Y \in A) = P(X_p \in A) = \sum_{i=1}^{\infty} P(X_j \in A, p = j). \tag{35.507}$$

Par ailleurs nous avons

$$\begin{aligned} P(X_j \in A, p = j) &= P\big(X_j \in A, \alpha(X_j) \geqslant U_j, \alpha(X_m) < U_m\ \forall m \leqslant j-1\big) \\ &= P\big(X_j \in A, \alpha(X_j) \geqslant U_j\big) P\big(\alpha(X_1) < U_1\big)^{j-1} \\ &= P\big(X_j \in A, \alpha(X_j) \geqslant U_j\big) \left(1 - \frac{1}{k}\right)^{j-1}. \end{aligned} \tag{35.508}$$

Étant donné que $g(x)dx$ est la densité de X_j et que du est la densité de U, nous avons

$$P\big(X_j \in A, \alpha(X_j) \geqslant U_j\big) = \int_{\mathbb{R}} \int_0^1 \mathbb{1}_{x\in A} \mathbb{1}_{\alpha(x)\geqslant u} g(x) du dx \tag{35.509a}$$

$$= \int_{\mathbb{R}} g(x) \mathbb{1}_{x\in A} \underbrace{\int_0^1 \mathbb{1}_{\alpha(x)\geqslant u} du}_{\alpha(x)}\ dx. \tag{35.509b}$$

$$= \int_{\mathbb{R}} \mathbb{1}_{x\in A} \frac{f(x)}{k} dx \tag{35.509c}$$

$$= \frac{1}{k} P(X \in A). \tag{35.509d}$$

En remplaçant dans l'équation (35.508) nous trouvons

$$P(X_j \in A, p = j) = \frac{1}{k} P(X \in A) \left(1 - \frac{1}{k}\right)^{j-1}. \tag{35.510}$$

Et enfin, l'équation (35.507) donne

$$P(Y \in A) = \frac{1}{k} P(X \in A) \sum_{i=1}^{\infty} \left(1 - \frac{1}{k}\right)^{j-1} = P(X \in A). \tag{35.511}$$

□

35.7.5 Simuler une loi géométrique à l'ordinateur

Si (X_n) est une suite de variables aléatoires indépendantes et identiquement distribuées avec $X_n \sim \mathscr{B}(p)$, alors

$$Z = \min\{k \geqslant 1 \text{ tel que } X_k = 1\} \sim \mathscr{G}(p). \tag{35.512}$$

Nous avons alors $P(Z = k) = (1 - p)^k p$.

Si nous avons un générateur de lois de Bernoulli de paramètre p, alors nous on simulons jusqu'à obtenir 1 et nous comptons combien de simulations ont été nécessaires.

35.7.6 Simuler une loi exponentielle à l'ordinateur

Nous pouvons utiliser la méthode de l'inversion. Étant donné que la fonction de répartition de la loi exponentielle est $F(x) = 1 - e^{-\lambda x}$, nous avons $F^{-1}(y) = \frac{1}{\lambda} \ln(1 - y)$. Par conséquent à partir d'un générateur uniforme U, nous pouvons calculer

$$F^{-1}(U) = \frac{1}{\lambda} \ln(U) \tag{35.513}$$

qui suivra une loi exponentielle d'espérance $1/\lambda$.

35.7.7 Simuler une loi de Poisson à l'ordinateur

Nous savons du point 35.5.8 que si les T_i sont des variables aléatoires indépendantes et identiquement distribuées de loi $\mathscr{E}(\lambda)$, alors nous avons

$$\max\{n \geqslant 1 \text{ tel que } \sum_i T_i \leqslant 1\} \sim \mathscr{P}(\lambda). \tag{35.514}$$

La façon usuelle pour créer une loi exponentielle est d'avoir un générateur de loi uniforme U_i et d'écrire que

$$-\frac{1}{\lambda} \ln(U_i) \sim \mathscr{E}(\lambda). \tag{35.515}$$

Nous devons donc faire la somme de telles variables aléatoires et voir à partir de quel moment la somme dépasse 1. Le calcul est le suivant :

$$-\sum_{i=1}^{n} \frac{1}{\lambda} \ln(U_i) \leqslant 1 \tag{35.516}$$

implique

$$\prod_{i=1}^{n} U_i \leqslant e^{-\lambda}. \tag{35.517}$$

En pratique, la variable aléatoire qui se comporte comme une loi de Poisson de paramètre λ est

$$N = \max\{n \geqslant 1 \text{ tel que } \prod_{i=1}^{n} U_i \geqslant e^{-\lambda}\}. \tag{35.518}$$

Nous générons donc des nombres aléatoires entre 1 et 1 et nous effectuons le produit jusqu'à ce qu'il passe en dessous de $e^{-\lambda}$. À ce moment, nous retournons le nombre de nombres qu'il a fallu générer.

35.8 Sage

Nous allons montrer maintenant quelques trucs importants dans l'utilisation de Sage pour réaliser des petits graphiques.

Remarque 35.134.
Dans ce qui suit, nous allons parler de « Sage », mais en réalité nous allons surtout parler du module `scipy` qui fait partie des modules hyper-usuels de Python. Les remerciements vont donc au moins autant du côté de l'équipe de `scipy` que vers celle de Sage.

35.8.1 Loi exponentielle

Il faut savoir que la définition d'une loi continue retourne automatiquement la loi centrée réduite. Pour avoir une loi exponentielle de moyenne donnée, il faut donc préciser de façon plus maligne que ce que l'on croit.

```
from scipy import stats

X=stats.expon(scale=5)
print(X.mean())      # retourne 5

P=plot( X.pdf,x,0,10 )
show(P)              # Affiche le graphique
```

tex/frido/code_sage1.py

35.8.2 Inverser des lois

Pour trouver des intervalles de confiance, il faut souvent calculer des inverses de loi. Bien entendu Sage le fait. Ce que sage connaît, c'est l'inverse de la fonction de survie. Autrement dit si `X` est une variable aléatoire, `X.sf` est la fonction $x \mapsto 1 - P(X < x)$ et `X.isf` en est l'inverse. Pour résoudre $P(X < \xi) = \alpha$, il faut résoudre $F(\xi) = \alpha$, c'est-à-dire

$$1 - F(\xi) = 1 - \alpha, \tag{35.519}$$

ce qui se fait de la façon suivante : le programme suivant donne pour une loi normale centrée réduite la valeur de ξ pour laquelle $P(N < \xi) = 0.05$:

```
from scipy import stats

N=stats.norm
print N.mean()       # 0
print N.var()        # 1

xi = N.isf(0.95)
print xi             # -1.64485

N.cdf(xi)            # Vérification : 0.05

# Graphiques de la fonction de densité et la cumulative.
P=plot(N.cdf,x,-10,10)
Q=plot(N.pdf,x,-10,10,color="red")
show(P+Q)
```

tex/frido/code_sage2.py

35.9 Monte-Carlo

Nous voudrions calculer une valeur approchée de l'intégrale

$$I = \int_a^b f(x)dx. \tag{35.520}$$

Les méthodes classiques consistent à discrétiser l'intervalle $[a, b]$ et en calculant une somme de la forme $\sum_i w_i f(x_i)$.

L'idée de Monté Carlo est de remplacer le découpage déterministe x_i par des variables aléatoires X_i en trois étapes.

(1) Pour cela nous commençons par écrire l'intégrale comme une espérance : $I = E(X)$ où X est une variable aléatoire définie sur un espace probabilisé $(\Omega, \mathcal{F}, P)$ à déterminer. Une contrainte est évidemment d'avoir $X \in L^1(\Omega, \mathcal{F}, P)$.

(2) Nous générons une suite indépendantes et identiquement distribuée de variables aléatoires (X_n) de même loi que X et la loi (forte) des grands nombres implique que

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{p.s.} E(X) = I. \tag{35.521}$$

(3) Le dernier point sera de donner un intervalle de confiance.

Exemple 35.135.
Nous voudrions déterminer de façon approchée l'intégrale $I = \int_0^1 f(x)dx$. Si $U \sim \mathscr{U}[0, 1]$, alors

$$I = E\big(f(U)\big) \tag{35.522}$$

et il suffit de faire

$$I \simeq \frac{1}{n}\sum_{k=1}^n f(U_i). \tag{35.523}$$

où mes U_i sont indépendantes et identiquement distribuées de loi $\mathscr{U}[0, 1]$. △

Exemple 35.136.
Supposons que la fonction à intégrer se présente sous la forme $f(x) = h(x)g(x)$ avec $g \geqslant 0$ et telle que l'intégrale $\int_{\mathbb{R}} g$ existe. Notons

$$c = \int_{\mathbb{R}} g \tag{35.524}$$

et

$$I = \int_{\mathbb{R}} h(x)c\frac{g(x)}{c}dx. \tag{35.525}$$

Nous avons alors $I = E\big(ch(Y)\big)$ où Y admet la densité $g(x)c$. △

Passons au cas de plusieurs variables et considérons l'intégrale

$$I = \int_{[0,1]^1} f(x_1, \ldots, x_d)dx_1 \ldots dx_d. \tag{35.526}$$

Nous écrivons

$$I = E\big(f(U_1, \ldots, U_d)\big) \tag{35.527}$$

où les U_i sont de loi uniformes sur $[0, 1]$. En pratique, nous générons une suite de variables aléatoires de (Z_k) de lois uniformes que nous regroupons par paquets :

$$V_k = (Z_{dk}, Z_{dk+1}, \ldots, Z_{d(k+1)-1}). \tag{35.528}$$

Ces variables aléatoires V_k sont indépendantes et identiquement distribuées de loi $\mathscr{U}[0, 1]^d$. Ensuite la loi des grands nombres nous indique que

$$I \sim \frac{1}{n}\sum_{k=1}^n f(V_k). \tag{35.529}$$

35.9.1 Intervalle de confiance

35.9.1.1 Principe

Nous supposons que nous travaillons sur une approximation de Monte-Carlo telle que la variable aléatoire choisie soit dans L^2. La loi des grands nombres nous dit que $\bar{X}_n \sim I$ tandis que le théorème central limite nous enseigne que

$$\frac{\bar{X}_n - E(X)}{\sigma/\sqrt{n}} \xrightarrow{\mathscr{L}} \mathscr{N}(0,1). \tag{35.530}$$

Par conséquent

$$P\left(\frac{\bar{X}_n - E(X)}{\sigma/\sqrt{n}} \in [-u,u]\right) \simeq P(-u \leqslant Z \leqslant u) \tag{35.531}$$

où $Z \sim \mathscr{N}(0,1)$. En remplaçant $E(X)$ par I et en effectuant les manipulations usuelles, nous trouvons que $P(I \in J_\alpha) = 1-\alpha$ si

$$J_\alpha = \left[\bar{X}_n - u_\alpha \frac{\sigma}{\sqrt{n}}, \bar{X}_n + u_\alpha \frac{\sigma}{\sqrt{n}}\right] \tag{35.532}$$

où σ^2 est la variance de X. Si σ n'est pas connue, alors nous le remplaçons par un estimateur

$$S'_n = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X}_n)^2 \tag{35.533}$$

et nous considérons l'intervalle

$$J'_\alpha = \left[\bar{X}_n - u_\alpha \frac{S'_n}{\sqrt{n}} \bar{X}_n + u_\alpha \frac{S'_n}{\sqrt{n}}\right]. \tag{35.534}$$

Il y a deux façons de faire diminuer la longueur de l'intervalle de confiance : augmenter n ou diminuer σ. Pour le second point, le choix de X dans $I = E(X)$ est essentiel.

Exemple 35.137.
Soit à calculer

$$I = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{\beta z} e^{-z^2/2} dz \tag{35.535}$$

avec $\beta > 0$. Nous introduisons la variable aléatoire $X = e^{\beta Z}$ avec $Z \sim \mathscr{N}(0,1)$. Nous avons alors

$$I = E(X). \tag{35.536}$$

Par ailleurs l'intégrale demandée vaut $e^{\beta^2/2}$. En appliquant les formules vues plus haut nous trouvons

$$J_\alpha = \left[\bar{X}_n - u_\alpha \frac{\sigma}{\sqrt{n}}, \bar{X}_n + u_\alpha \frac{\sigma}{\sqrt{n}}\right] \tag{35.537}$$

où

$$\sigma^2 = \mathrm{Var}(X) = E(e^{2\beta Z}) - E(e^{\beta z})^2 = e^{2\beta^2} - e^{\beta^2}. \tag{35.538}$$

Nous avons utilisé la formule (35.403). Si nous choisissons $\beta = 2$, nous trouvons $\sigma^2 \simeq 2926$. Donc si nous voulons une longueur de J_α plus petite que 10^{-2} tout en demandant $\alpha = 0.05$ (ce qui implique $u_\alpha = 1.96$), nous devons avoir

$$1.96\frac{2973}{\sqrt{n}} < 10^{-2}, \tag{35.539}$$

c'est-à-dire environ $n = 10^{11}$, ce qui soit dit en passant est très largement au delà des capacités de la commande rand de scilab. △

Nous allons maintenant voir quelques méthodes pour réduire la variance.

35.9.1.2 Échantillonnage préférentiel

Nous devons calculer $I = \int_{\mathbb{R}} f(x)dx$. Pour cela nous introduisons artificiellement une densité g et nous écrivons

$$I = \int_{\mathbb{R}} \frac{f(x)}{g(x)} g(x)dx = E\left(\frac{f(Y)}{g(Y)}\right) \tag{35.540}$$

où Y est de densité g. Il faut essayer de trouver g de telle sorte à ce que

$$\operatorname{Var}\left(\frac{f(Y)}{g(Y)}\right) \tag{35.541}$$

soit la plus petite possible.

35.9.1.3 Méthode de la variable de contrôle

Soit $I = E(X)$. Nous introduisons une variable aléatoire Z et nous écrivons

$$I = E(X - Z) + E(Z). \tag{35.542}$$

Il faut alors choisir Z de telle sorte que $E(Z)$ soit calculable et que $X - Z$ ait une variance plus faible. En particulier, Z ne peut pas être indépendante de X.

35.9.1.4 Variables antithétiques

Soit $I = \int_0^1 f(x)dx$. La première idée (exemple 35.135) est d'écrire

$$I = E\big(f(U)\big) \tag{35.543}$$

où $U \sim \mathscr{U}[0,1]$, mais nous n'avons pas de garanties sur la variance de $f(U)$. Nous pouvons écrire

$$I = E\big(f(U)\big) = E\Big[\frac{1}{2}\big(f(U) - f(1-U)\big)\Big]. \tag{35.544}$$

Ici $1-U$ est encore une variable aléatoire uniforme sur $[0,1]$, mais il se fait que la variable aléatoire

$$Z = \frac{1}{2}\big(f(U) + f(1-U)\big) \tag{35.545}$$

a une variance inférieure à $\operatorname{Var}\big(f(U)\big)$. En effet, $f(U)$ et $f(1-U)$ ne sont pas indépendantes, par conséquent le résultat du lemme 35.23 n'est pas valide, par contre la proposition 35.32 reste vraie et nous avons

$$\operatorname{Var}(Z) = \frac{1}{4}\operatorname{Var}\big(f(U)\big) + \frac{1}{4}\operatorname{Var}\big(f(1-U)\big) + \frac{1}{2}\operatorname{Cov}\big(f(U), f(1-U)\big). \tag{35.546}$$

Nous avons $\operatorname{Var}\big(f(1-U)\big) = \operatorname{Var}\big(f(U)\big)$. En ce qui concerne le terme avec la covariance, nous lui appliquons l'équation (35.58) :

$$\begin{aligned}
\operatorname{Cov}\big(f(U), f(1-U)\big) &= E\Big((f(U) - I)(f(1-U) - I)\Big) && \text{(35.547a)}\\
&\leqslant E\big((f(U) - I)^2\big)^{1/2} E\big((f(1-U) - I)^2\big)^{1/2} && \text{(35.547b)}\\
&= \operatorname{Var}\big(f(U)\big) && \text{(35.547c)}
\end{aligned}$$

où nous avons utilisé le fait que $E\big(f(U)\big) = E\big(f(1-U)\big) = I$. Au final nous avons bien obtenu

$$\operatorname{Var}(Z) \leqslant \operatorname{Var}\big(f(U)\big). \tag{35.548}$$

35.10 Résultats qui se démontrent avec des variables aléatoires

35.10.1 Nombres normaux

Tout nombre $x \in [0,1[$ admet un unique[17] développement en base $b \geqslant 2$:

$$x = \sum_{n=1}^{\infty} \frac{\epsilon_n(x)}{b^n} \tag{35.549}$$

avec $\epsilon_n(x) \in \mathcal{A} = \{0, \dots, b-1\}$.

Soit $k \geqslant 1$ et $r \in \mathcal{A}^k$; nous posons

$$N_x(r,n) = \text{Card}\big\{i \in \{1,\dots,n-k+1\} \text{ tel que } \epsilon_1(x) = r_1, \dots, \epsilon_{i+k-1} = r_k\big\}. \tag{35.550}$$

C'est le nombre d'occurrences du motif r (de longueur k) dans les n premières décimales de x.

Définition 35.138.
*Un nombre $x \in [0,1[$ est **normal** en base b si pour tout $r \in \{0,\dots,b-1\}^k$ nous avons*

$$\frac{N_x(b,n)}{n} \to \frac{1}{b^k}. \tag{35.551}$$

Un nombre est normal s'il est normal en toute base.

Proposition 35.139 ([533, 54])**.**
Au sens de la mesure de Lebesgue, presque tous les nombres de $[0,1[$ sont normaux.

Démonstration. Pour $x \in [0,1[$, nous notons $\epsilon_n(x)$ son développement en base b. Cela nous donne des variables aléatoires $\epsilon_i \colon [0,1[\to \mathcal{A}$ dont la loi de probabilité est donnée par

$$P(\epsilon_1 = d) = P\big([\frac{d}{b}, \frac{d+1}{b}[\big) = \frac{1}{b} \tag{35.552}$$

parce que l'intervalle $[\frac{d}{b}, \frac{d+1}{d}[$ est l'ensemble des nombres de $[0,1[$ dont la première décimale est d. Pour la loi des ϵ_i, il faut un peu plus découper, mais ça donne le même résultat : $P(\epsilon_i = d) = 1/b$. Ces variables aléatoires sont indépendantes et identiquement distribuées. Nous considérons aussi la variable aléatoire

$$\begin{aligned} N(r,n) \colon [0,1[&\to \mathbb{N} \\ x &\mapsto N_x(r,n) \end{aligned} \tag{35.553}$$

Pour un $r \in \mathcal{A}$ fixé, nous définissons encore la variable aléatoire

$$\begin{aligned} X_j \colon \mathcal{A} &\to \{0,1\} \\ x &\mapsto \begin{cases} 1 & \text{si } \epsilon_j(x) = b \\ 0 & \text{sinon.} \end{cases} \end{aligned} \tag{35.554}$$

Les variables aléatoires X_j sont des variables aléatoires de Bernoulli indépendantes et identiquement distribuées de paramètre $E(X_j) = P(X_j = 1) = P(\epsilon_1 = b) = \frac{1}{b}$. Nous pouvons utiliser dessus la loi forte des grands nombres (théorème 35.91). Pour dire que

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p.s.} E(X_1) = \frac{1}{b}. \tag{35.555}$$

Mais en réalité nous avons aussi $\sum_{j=1}^{n} X_j = N(r,n)$ parce que en appliquant à $x \in [0,1[$:

$$\sum_{j=1}^{n} \begin{cases} 1 & \text{si } \epsilon_j(x) = r \\ 0 & \text{sinon} \end{cases} = \text{Card}\big\{i \in \{1,\dots,n\} \text{ tel que } \epsilon_i(x) = r\big\} = N_x(r,n), \tag{35.556}$$

17. Nous excluons 1 parce que son développement en puissances négatives de b est zéro.

de sorte que l'équation (35.555) nous dit exactement que pour tout $r \in \mathcal{A}$,

$$\lim_{n\to\infty} \frac{N_x(r,n)}{n} = \frac{1}{b} \tag{35.557}$$

pour presque tout $x \in [0,1[$.

Il reste à prouver la même chose pour tout $r \in \mathcal{A}^k$. Voyons avec $k = 2$ et $r = (u,v) \in \mathcal{A}^2$. Nous posons

$$Y_j = \mathbb{1}_{\{\epsilon_j = u, \epsilon_{j+1} = v\}}, \tag{35.558}$$

et $N(r,n) = \sum_{j=1}^{n-1} Y_j$. Les Y_i sont encore des binomiales de paramètre $\frac{1}{b^2}$, mais elles ne sont pas indépendantes. En effet pour avoir $Y_1(x) = Y_2(x) = 1$, il faut que les trois premières décimales de x soit en même temps de la forme $uv.$ et $.uv$, donc

$$P(Y_1, Y_2 = 1) = b^3 \delta_{u,v} \tag{35.559}$$

alors que $P(Y_1 = 1)P(Y_2 = 2) = b^4$. Nous pouvons contourner ce problème en remarquant que les ϵ_i, eux, sont indépendants. Donc le lemme de regroupement 35.16 nous dit que la famille $\{Y_{2n}\}$ est une famille de variables aléatoires indépendantes (et idem pour la famille Y_{2n-1}). En effet, les variables aléatoires Y_{2n} correspondent à la partition 23, 45, 67, etc.

Nous appliquons la loi des grands nombres sur les deux familles indépendamment :

$$\frac{1}{n}\sum_{j=1}^{n} Y_{2j-1} \xrightarrow{p.s.} \frac{1}{b^2} \tag{35.560}$$

et

$$\frac{1}{n}\sum_{j=1}^{n} Y_{2j} \xrightarrow{p.s.} \frac{1}{b^2} \tag{35.561}$$

Pour rappel, le but pour l'instant est d'établir la limite $\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} Y_j = \frac{1}{b^2}$. Nous allons l'établir séparément pour les termes pairs et impairs de la suite. Pour les pairs :

$$\frac{1}{2n}\sum_{j=1}^{2n} Y_j = \frac{1}{2}\left(\frac{1}{n}\sum_{j=1}^{n} Y_{2j-1}\right) + \frac{1}{2}\left(\frac{1}{n}\sum_{j=1}^{n} Y_{2j}\right) \xrightarrow{p.s.} \frac{1}{b^2} \tag{35.562}$$

Pour les impairs [18],

$$\frac{1}{2n-1}\sum_{j=1}^{2n-1} Y_j = \frac{n}{2n-1}\left(\frac{1}{n}\sum_{j=1}^{n} Y_{2j-1}\right) + \frac{n}{2n-1}\left(\frac{1}{n}\sum_{j=1}^{n} Y_{2j}\right) \to \frac{1}{b^2} \tag{35.563}$$

parce que les deux parenthèses convergent vers $\frac{1}{b^2}$ alors que les coefficients devant convergent vers $\frac{1}{2}$.

Au final nous avons bien

$$\frac{N(r,n)}{n} = \frac{1}{n}\sum_{j=1}^{n-1} Y_j = \frac{n-1}{n}\left(\frac{1}{n-1}\sum_{j=1}^{n-1} Y_j\right) \to \frac{1}{b^2} \tag{35.564}$$

tant que $r \in \mathcal{A}^2$.

Pour prouver la même chose avec $r \in \mathcal{A}^k$, il suffit de faire le même raisonnement en divisant en plus de paquets : $\{Y_{kj+m}\}_{m=1,\dots,k-1}$ sont indépendants et nous utilisons k fois la loi des grands nombres.

Donc pour toute base b nous savons que les nombres normaux en base b forment un ensemble de mesure nulle dans $[0,1[$. Il reste à voir que leur union reste de mesure nulle. Cela est vrai parce que nous avons une union dénombrable et qu'une union dénombrable d'ensembles de mesure nulle est de mesure nulle par le lemme 13.27. □

18. Ici dans [54], la seconde somme va jusqu'à $n-1$ et je ne comprends pas pourquoi.

Remarque 35.140.
Un nombre x est normal en base b si et seulement si la suite $u_k = xb^k$ est équirépartie modulo 1 sur $[0,1]$ (c'est-à-dire quel la suite des parties fractionnelles des u_k est équirépartie). Pour le nombre $0.2357873\ldots$, nous parlons de la suite $0.2357873\ldots$; $0.357873\ldots$; $0.57897\ldots$ etc. C'est la suite des queues de suites de la suite de ses décimales [19].

35.10.2 Théorème de Bernstein

Théorème 35.141 (Théorème de Bernstein[54])**.**
Soit $f \in C^0\big([0,1],\mathbb{C}\big)$ et son module de continuité

$$\begin{aligned}\omega\colon [0,1] &\to \mathbb{R}\\ h &\mapsto \sup\{|f(u)-f(v)| \text{ tel que } |u-v|<h\}.\end{aligned} \tag{35.565}$$

*Pour $n \geqslant 0$ nous définissons le n^e **polynôme de Bernstein** de f par*

$$B_n(f)(x) = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right). \tag{35.566}$$

Alors il existe C tel que pour tout $n \geqslant 1$:

(1)

$$\|f - B_n(f)\|_\infty \leqslant C\omega\left(\frac{1}{\sqrt{n}}\right). \tag{35.567}$$

(2)

$$B_n(f) \xrightarrow{unif} f \tag{35.568}$$

sur $[0,1]$.

(3) L'inégalité (35.567) est optimale : il existe une fonction $g \in C^0\big([0,1],\mathbb{C}\big)$ et $\delta > 0$ tels que pour tout $N \geqslant 1$, $\|g - B_n(g)\|_\infty \geqslant \frac{\delta}{\sqrt{n}}$. Cette fonction peut être choisie Lipschitzienne. Une telle fonction est donnée par exemple par $g(x) = |x - \frac{1}{2}|$.

(4) Les polynômes forment une partie dense dans $\Big(C^0\big([0,1]\big), \|.\|_\infty\Big)$.

Démonstration. Soit $x \in [0,1]$ et une suite de variables aléatoires de Bernoulli indépendantes [20] et identiquement distribuées $(X_i)_{i\geqslant 1}$ de paramètre x. Nous notons $S_n = \sum_{k=1}^n X_k$.

(1) Pour cette histoire de convergence, il faut majorer la quantité $\big|f(x) - B_n(f)(x)\big|$. Pour cela il y a trois astuces. La première est de se souvenir que $E\big(f(x)\big) = f(x)$, et la seconde est que le théorème de transfert 35.68 appliqué à $x \mapsto f(x/n)$ donne [21]

$$E\left(f\big(\frac{S_n}{n}\big)\right) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) P(S_n = k) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}, \tag{35.569}$$

c'est-à-dire que

$$B_n(f)(x) = E\left(f\big(\frac{S_n}{n}\big)\right). \tag{35.570}$$

Et enfin la troisième astuce est d'utiliser le lemme 10.165 pour avoir

$$\omega\left(|x - \frac{S_n}{n}|\right) = \omega\left(\frac{1}{\sqrt{n}}|\sqrt{n} - \frac{S_n}{\sqrt{n}}|\right) \leqslant \left(\sqrt{n}|x - \frac{S_n}{n}| + 1\right)\omega\big(\frac{1}{\sqrt{n}}\big). \tag{35.571}$$

19. C'est pas trop bien dit, mais on se comprend, non ?
20. Définition 35.9.
21. Nous avons aussi utilisé la formule de l'espérance pour les variables aléatoires discrètes.

À partir de là nous pouvons un peu calculer :

$$\begin{aligned}
\big|f(x)-B_n(f)(x)\big| &= \Big|E\Big(f(x)-f(\frac{S_n}{n})\Big)\Big| && (35.572a)\\
&\leqslant E\Big(|f(x)-f(\frac{S_n}{n})|\Big) && (35.572b)\\
&\leqslant E\Big(\omega\Big(|x-\frac{S_n}{n}|\Big)\Big) && (35.572c)\\
&\leqslant \omega\Big(\frac{1}{\sqrt{n}}\Big)E\Big(|\sqrt{n}x-\frac{S_n}{n}|+1\Big). && (35.572d)
\end{aligned}$$

Le dernier facteur peut être récrit sous la forme

$$E\Big(\sqrt{n}|x-\frac{S_n}{n}|+1\Big)=\sqrt{n}E\Big(|x-\frac{S_n}{n}|\Big)+1, \tag{35.573}$$

et c'est là que nous pouvons utiliser l'inégalité de Hölder 26.33 :

$$E\big(|X|\big)=\|X\|_1\leqslant\|X\|_2 \tag{35.574}$$

où $\|X\|_2$ désigne

$$\|X\|_2=\sqrt{E\big(|X|^2\big)}. \tag{35.575}$$

Nous pouvons donc écrire

$$\big|f(x)-B_n(f)(x)\big|\leqslant\omega\Big(\frac{1}{\sqrt{n}}\Big)\Big(\sqrt{n}\|x-\frac{S_n}{n}\|_2+1\Big). \tag{35.576}$$

Nous étudions maintenant de plus près la quantité $\|x-\frac{S_n}{n}\|_2$. D'abord

$$E\Big(|x-\frac{S_n}{n}|^2\Big)=x^2-2\frac{x}{n}E(S_n)+\frac{1}{n^2}E(S_n^2). \tag{35.577}$$

Ensuite nous savons l'espérance de S_n (qui vaut $E(S_n)=nx$) par (35.347) et le lemme 35.22 nous permet de calculer $E(S_n^2)$ par indépendance des X_i qui composent S_n. Nous avons alors

$$\begin{aligned}
E\Big(|x-\frac{S_n}{n}|^2\Big) &= x^2-2x^2+\frac{1}{n^2}\sum_{1\leqslant i\neq j\leqslant n}E(X_i)E(X_j)+\frac{1}{n^2}\sum_{i=1}^{n}E(X_i^2) && (35.578a)\\
&= -x^2+\frac{n^2-n}{n^2}x^2+\frac{nx}{n^2} && (35.578b)\\
&= \frac{x(1-x)}{n}. && (35.578c)
\end{aligned}$$

Quelques justifications :

— $E(X_i)=E(X_i^2)=x$ parce que X_i est une variable aléatoire de Bernoulli de paramètre x.

— La première somme contient tous les couples (i,j) sauf les diagonaux ; il y en a donc n^2-n.

En recombinant le tout,

$$\begin{aligned}
\big|f(x)-B_n(f)(x)\big| &\leqslant \omega\Big(\frac{1}{\sqrt{n}}\Big)\Big(\sqrt{n}\sqrt{\frac{x(1-x)}{n}}+1\Big) && (35.579a)\\
&= \omega\Big(\frac{1}{\sqrt{n}}\Big)\big(\sqrt{x(1-x)}+1\big) && (35.579b)\\
&\leqslant \frac{3}{2}\omega\Big(\frac{1}{\sqrt{n}}\Big). && (35.579c)
\end{aligned}$$

La dernière majoration est une rapide étude de la fonction $x(1-x)$.

Étant donné que les majorations (35.579) sont valables pour tout x, en passant au supremum nous avons

$$\|f - B_n(f)\|_\infty \leqslant \frac{3}{2}\omega\left(\frac{1}{\sqrt{n}}\right) \to 0. \tag{35.580}$$

Ceci prouve les deux premiers points du théorème.

(2) Fait.

(3) Nous considérons la fonction

$$g(x) = \|x - \frac{1}{2}j\| \tag{35.581}$$

et nous vérifions qu'elle vérifie toutes les conditions. D'abord si $u, v \in [0,1]$ alors

$$\big|g(u) - g(v)\big| \leqslant |u - v| \tag{35.582}$$

et donc $\omega(h) \leqslant h$, ce qui signifie que g est 1-Lipschitz. Le principe de cette partie est de montrer que $\|g - B_n(g)\|_\infty$ est plus grand que d'autres trucs (et non plus petit que d'autres trucs comme d'habitude). Nous commençons par

$$\|g - B_n(g)\|_\infty \geqslant g(\frac{1}{2}) - B_n(g)(\frac{1}{2}). \tag{35.583}$$

Très vite nous nous rendons compte que $g(1/2) = 0$. Ensuite nous nous souvenons que

$$B_n(g)(\frac{1}{2}) = E\left(g(\frac{S_n}{n})\right) = E\left(|\frac{S_n}{n} - \frac{1}{2}|\right) = \frac{1}{2n}E\big(|2S_n - n|\big). \tag{35.584}$$

si nous posons $\epsilon_i = 2X_i - 1$, alors les ϵ_i sont des variables aléatoires de Rademacher indépendantes et identiquement distribuées qui satisfont à $2S_n - n = \sum_{i=1}^n \epsilon_i$. Nous utilisons la proposition 35.123 :

$$\|f - B_n(f)_\infty\| \geqslant \frac{1}{2n}E\big(|\sum_i \epsilon_i|\big) \geqslant \frac{1}{2n\sqrt{e}}\|\sum_{i=1}^n \epsilon_j\|_2. \tag{35.585}$$

Calculons ce qui est dans la norme :

$$\|\sum_{j=1}^n \epsilon_j\|_2^2 = E\left(\big(\sum_{j=1}^n \epsilon_j\big)^2\right) = \sum_{1\leqslant i\neq j\leqslant n} E(\epsilon_i)E(\epsilon_j) + \sum_{i=1}^n E(\epsilon_i^2) = 0 + n = n. \tag{35.586}$$

Nous finissons alors notre travail de majoration :

$$\|f - B_n(f)_\infty\| \geqslant \frac{1}{2n\sqrt{e}}\|\sum_{i=1}^n \epsilon_j\|_2 \geqslant \frac{1}{2\sqrt{n}\sqrt{e}} \geqslant \frac{1}{2\sqrt{e}}\omega\left(\frac{1}{\sqrt{n}}\right). \tag{35.587}$$

(4) Nous avons trouvé une suite de polynômes qui converge uniformément vers un élément arbitraire de $L^2\big([0,1]\big)$. Cela prouve la densité.

□

Corolaire 35.142.
Dans $\mathbb{R}$, si $I = [a,b]$ alors les polynômes forment une partie dense dans $\big(C^0(I), \|.\|_\infty\big)$.

Démonstration. Nous supposons que $b > a$. Le cas $a = b$ est assez facile parce que l'espace des fonctions sur $\{a\}$ est de dimension 1.

Nous considérons une bijection affine $\varphi\colon [0,1] \to [a,b]$ telle que $\varphi(0) = a$ et $\varphi(1) = b$. Soit $f \in C^0(I)$.

Si $g = f \circ \varphi$, alors le théorème de Bernstein 35.141 nous donne une suite de polynômes g_k sur $[0,1]$ tels que

$$g_k \xrightarrow{unif} g. \tag{35.588}$$

Nous considérons $f_k = g_k \circ \varphi^{-1}$ qui est encore un polynôme parce que φ^{-1} est affine. Étant donné que φ^{-1} est une bijection, si h est une fonction sur $[0,1]$, nous avons

$$\sup_{x\in[a,b]} \|(h\circ\varphi^{-1})(x)\| = \sup_{y\in[0,1]} \|h(y)\|. \tag{35.589}$$

Cela nous permt le calcul suivant :

$$\begin{aligned}
\|f_k - f\|_\infty &= \|g_k\circ\varphi^{-1} - g\circ\varphi^{-1}\| && (35.590a)\\
&= \|(g_k - g)\circ\varphi^{-1}\| && (35.590b)\\
&= \sup_{x\in[a,b]} \|(g_k - g)\big(\varphi^{-1}(x)\big)\| && (35.590c)\\
&= \sup_{y\in[0,1]} \|(g_k - g)(y)\| && (35.590d)\\
&= \|g_k - g\|_\infty. && (35.590e)
\end{aligned}$$

Nous avons donc

$$\lim_{k\to\infty} \|f_k - f\|_\infty = 0, \tag{35.591}$$

ce qui prouve la densité. □

Chapitre 36

Statistiques

36.1 Notations et hypothèses

Nous notons X le caractère à étudier, et Ω l'ensemble des individus. Le caractère à étudier est vu comme une fonction sur Ω :

$$X : \Omega \to \mathbb{R}, \mathbb{N}, \{0, 1\}, \ldots \tag{36.1}$$

Les **statistiques descriptives** sont les techniques pour présenter et résumer les données : diagrammes, graphiques, indicateurs numériques : moyenne, écart-type, médiane, ...

Nous faisons les hypothèses suivantes :

(1) Chaque observation x_i est la réalisation de la variable aléatoire X qui sera de loi inconnue μ.

(2) Le n-uple $(x_1, \ldots, x_n)$ est la réalisation de $(X_1, \ldots, X_n)$ qui est l'échantillon de taille n.

(3) Les variables aléatoires X_i sont indépendantes et identiquement distribuées, de loi commune μ. La loi μ est la **loi parente** de l'échantillon.

Exemple 36.1.
Un échantillon de taille 1 consisterait à tirer au sort une personne dans une population et mesurer sa taille. △

Exemple 36.2.
Une échantillon de taille n consisterait à tirer au sort n personnes dans une population et de mesurer leurs tailles. △

L'**inférence statistique** est l'art de dégager des informations sur la population à partir d'informations partielles : intervalles de confiance, estimateurs, test d'hypothèses, ...

En théorie des probabilités, nous connaissons la loi de la variable aléatoire X et nous en déduisons des informations sur le réalisations de X : valeur la plus probable, moyenne, intervalle dans lequel $X(\omega)$ a le plus de chance d'appartenir. En statistique, au contraire, la loi est inconnues et nous cherchons des informations sur la loi à partir d'un échantillon de données numériques observées.

36.2 Modèle statistique

Une **modèle statistique** est un triplet

$$\mathcal{S} = \Big[(\Omega, \mathcal{F}, P), (X_\theta)_{\theta\in\Theta}, (\mu_\theta)_{\theta\in\Theta}\Big] \tag{36.2}$$

où $(\Omega, \mathcal{F}, P)$ est un espace probabilisé, (X_θ) est une famille de variables aléatoires définies sur Ω et telles que pour tout $\theta \in \Theta$, la variable aléatoire X_θ suive la loi μ_θ. Les μ_θ sont des mesures sur les boréliens de $\mathbb{R}$ et pour tout $B \in \mathcal{B}\mathrm{or}(\mathbb{R})$ nous avons

$$P(X_\theta \in B) = \mu_\theta(B). \tag{36.3}$$

Remarque 36.3.
D'une certaine manière, l'introduction de μ_θ dans la définition est redondante parce que ces mesures sont déjà contenues dans la données des variables aléatoires X_θ.

Exemple 36.4 (Modèle statistique gaussien)**.**
Si nous savons que les variables aléatoires X_i suivent une loi gaussienne, alors nous considérons $\Theta = \mathbb{R} \times \mathbb{R}^+$ et $\theta = (m, \sigma^2 j)$. Dans ce cas, $\mu_\theta = \mathscr{N}(m, \sigma^2)$ et le but de la statistique est de déterminer la valeur de θ qui correspond à une population en partant de l'observation d'un échantillon. △

Définition 36.5.
*Si $\Theta \subset \mathbb{R}^k$, nous disons que le modèle statistique est un modèle **paramétrique**.*

Le modèle gaussien est un modèle paramétrique : dès que m et σ^2 sont déterminés, la loi du phénomène X est connue.

Définition 36.6.
*Pour chaque $\theta \in \theta$, nous disons qu'un **échantillon** de taille n associé à un modèle statistique $\big[(\Omega, \mathcal{F}, P), (X_\theta), (\mu_\theta)\big]$ est un vecteur $\big(X_{\theta,1}, \dots, X_{\theta,n}\big)$ de taille n de variables aléatoires indépendantes et identiquement distribuées de la même loi que la variable aléatoire X_θ. La loi μ_θ est la **loi parente** de l'échantillon.*

Définition 36.7.
*Un **modèle d'échantillonnage** sur le modèle statistique $\mathcal{S}$ est une famille $(X_{\theta,1}, \dots, X_{\theta,n})_{\theta\in\Theta}$ d'échantillons de taille $n \geqslant 1$.*

Nous noterons souvent $(X_1, \dots, X_n)$ à la place de $(X_{\theta,1}, \dots, X_{\theta,n})$ un échantillon, mais il faut se souvenir que les X_i suivent toujours la même loi donnée par θ. La loi du vecteur $(X_1, \dots, X_n)$ est $\mu_\theta \otimes \dots \otimes \mu_\theta$ et est définie sur l'espace $(\Omega^n, \mathcal{F} \otimes \dots \otimes \mathcal{F}, P^{\otimes n})$.

Remarque 36.8.
Le travail du statisticien est de proposer un modèle statistique $\mathcal{S}$ à priori. Si nous étudions la taille d'une population, nous allons choisir un modèle gaussien. Plus le modèle est précis, plus l'espace Θ est petit mais plus il y a de risques que le vérité soit hors de l'ensemble considéré.

Exemple 36.9.
Soit X une variable aléatoire de carré intégrable que l'on sait simuler. Afin d'évaluer la moyenne μ de X, nous pouvons considérer la moyenne empirique des simulations : $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ où les variables aléatoires X_i sont indépendantes, identiquement distribuées et de même loi que X.

La loi des grands nombres nous enseigne que $\bar{X}_n \to \mu$. De plus,

$$\lim_{n\to\infty} P\left(\mu \in \left[\bar{X}_n - \frac{a\sigma}{\sqrt{n}}, \bar{X}_n + \frac{a\sigma}{\sqrt{n}}\right]\right) = \int_{-a}^{a} e^{-x^2/2}\frac{dx}{\sqrt{2\pi}}. \tag{36.4}$$

En effet, la condition sur μ est équivalente à

$$-a \leqslant \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leqslant a, \tag{36.5}$$

tandis que le théorème central limite nous enseigne que la variable aléatoire $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$ se comporte comme $\mathscr{N}(0,1)$ lorsque n est grand. Dans ce cas, nous avons que

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \in [-a, a]\right) = \frac{1}{\sqrt{2\pi}}\int_{-a}^{a} e^{-x^2/2}dx. \tag{36.6}$$

Notons que dans ce calcul nous avons utilisé le fait que $\mu = E(X_1)$.

Montrons que la suite

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2 \tag{36.7}$$

converge presque surement vers σ^2. Le théorème central limite implique que

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p.s.} E(X_1^2) \tag{36.8}$$

et que

$$\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 \xrightarrow{p.s.} E(X_1)^2. \tag{36.9}$$

La différence converge donc presque surement vers σ^2 en vertu de la proposition 35.30.

Nous avons également $E(\sigma_n^2) = \sigma^2$. En effet, sachant que $E(X_i) = E(X_1) = \mu$ et que $E(X_i^2) = E(X_1^2) = \mu^2 + \sigma^2$,

$$E(\sigma_n^2) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) - \frac{1}{n^2}\left(\sum_{i=1}^{n} E(X_i^2)\right) + \sum_{i\neq j} E(X_iX_j)) \tag{36.10a}$$

$$= \sigma^2 + \mu^2 - \frac{1}{n^2}\big(n(\sigma^2+\mu^2) + (n^2-n)\mu^2\big) \tag{36.10b}$$

$$= \sigma^2 - \frac{1}{n}\sigma^2, \tag{36.10c}$$

dont la limite $n \to \infty$ donne bien σ^2.

Nous voudrions à présent montrer que

$$\frac{\bar{X}_n - \mu}{\sigma_n/\sqrt{n}} \xrightarrow{\mathscr{L}} \mathscr{N}(0,1). \tag{36.11}$$

Vu que le théorème central limite donne une convergence en loi, nous pouvons utiliser le lemme de Slutsky pour montrer que

$$\left(\frac{\bar{X}_n - \mu}{1/\sqrt{n}}, \sigma_n^2\right) \xrightarrow{\mathscr{L}} (\sigma Z, \sigma^2) \tag{36.12}$$

où $Z \sim \mathscr{N}(0,1)$. En vertu de la proposition 35.83 appliqué à la fonction $f\colon \mathbb{R}^2 \to \mathbb{R}$,

$$f(x,y) = \begin{cases} \frac{x}{\sqrt{y}} & \text{si } y \neq 0 \\ 0 & \text{sinon} \end{cases} \tag{36.13}$$

nous avons la convergence en loi

$$f\left(\frac{\bar{X}_n - \mu}{1/\sqrt{n}}, \sigma_n^2\right) \xrightarrow{\mathscr{L}} f(\sigma Z, \sigma^2), \tag{36.14}$$

c'est-à-dire

$$\frac{\bar{X}_n - \mu}{\sigma_n/\sqrt{n}} \xrightarrow{\mathscr{L}} Z. \tag{36.15}$$

Afin d'être complet, précisons que

$$P\big((\sigma Z, \sigma) \in \mathbb{R} \times \{0\}\big) = 0. \tag{36.16}$$

△

Proposition 36.10.
Soit X_i des variables aléatoires indépendantes et identiquement distribuées de loi parente $\mathscr{B}(n,p)$. Alors si $\bar{X}_n$ désigne la moyenne empirique,

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{\mathscr{L}} Z \sim \mathscr{N}(0,1). \tag{36.17}$$

Démonstration. Cela est une application de la loi des grands nombres, tu théorème central limite, du lemme de Slutsky et de la proposition 35.83.

D'abord, la loi des grands nombres nous indique que $\bar{X}_n \to p$ parce que p est l'espérance de Bernoulli. Ensuite nous avons

$$\frac{\bar{X}_n - E(X)}{\sqrt{\mathrm{Var}(X)}/\sqrt{n}} = \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \tag{36.18}$$

parce que la variance d'une loi de Bernoulli est $p(1-p)$. Le théorème central limite nous indique par conséquent que

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow{\mathscr{L}} Z \sim \mathscr{N}(0,1). \tag{36.19}$$

Le lemme de Slutsky implique alors la convergence du couple :

$$\left(\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}, \bar{X}_n\right) \xrightarrow{\mathscr{L}} (Z, p). \tag{36.20}$$

Nous appliquons maintenant la proposition 35.83 avec la fonction

$$f(x,y) = \frac{\sqrt{p(1-p)}x}{\sqrt{y(1-y)}} \tag{36.21}$$

qui est une fonction dont l'ensemble des points de discontinuité est $C = \{0\}$. Étant donné que $P(\bar{X}_n = 0) = 0$, la proposition s'applique et nous avons

$$f\left(\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}, \bar{X}_n\right) \xrightarrow{\mathscr{L}} f(Z, p), \tag{36.22}$$

c'est-à-dire

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{\mathscr{L}} Z \sim \mathscr{N}(0,1). \tag{36.23}$$

□

36.3 Modèles d'échantillonnages

Soit X, une variable aléatoire sur $(\Omega, \mathcal{F}, P)$. Un **échantillon** de taille n pour X est une suite de n variables aléatoires $(X_1, \dots, X_n)$ définies sur $(\Omega, \mathcal{F}, P)$ indépendantes et de même loi que X. Nous disons que la loi de X est la **loi parente** de la suite X_i.

Définition 36.11.
Soit

$$\mathcal{S} = \Big[(\Omega, \mathcal{F}, P), (X_\theta), (\mu_\theta)\Big]_{\theta\in\Theta}, \tag{36.24}$$

un modèle statistique. Un ***modèle d'échantillonnage*** *de taille n associée au modèle statistique $\mathcal{S}$ est la donnée d'une famille de n-échantillons $(X_{\theta,1}), \dots, X_{\theta,n}$ telle que pour tout $\theta \in \Theta$, l'échantillon $(X_{\theta,i})$ soit de variable parente X_θ.*

La **moyenne empirique** du n-échantillon (X_i) est la variable aléatoire

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{36.25}$$

La proposition suivante signifie que la moyenne empirique est une « bonne » façon d'approcher la variable aléatoire.

Proposition 36.12.
Soit X, une variable aléatoire dans $L^2(\Omega)$ (c'est-à-dire $E(X^2) < \infty$) d'espérance m et de variance σ^2. Alors

(1) $E(\bar{X}_n) = m$ *et* $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$.

(2) Nous avons les convergences

$$\bar{X}_n \xrightarrow{p.s.} m \tag{36.26a}$$

$$\frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \xrightarrow{\mathscr{L}} Z \sim \mathscr{N}(0,1). \tag{36.26b}$$

(3) Si X est de loi $\mathscr{N}(m, \sigma^2)$, alors $\bar{X}_n \sim \mathscr{N}(m, \frac{\sigma^2}{n})$, c'est-à-dire

$$\frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \sim \mathscr{N}(0,1). \tag{36.27}$$

Remarque 36.13.
L'intérêt de cette proposition en statistique descriptive expérimentale est le suivant. La taille moyenne des français est un nombre m qui existe, mais qui est largement hors de portée de l'expérience (mesurer $65 \cdot 10^6$ personnes risque de prendre un sacré temps). Si on mesure seulement n personnes dont les tailles sont $(x_i)_{i=1,\dots,n}$ (ici x_i est un *nombre expérimental*, pas une variable aléatoire), alors on peut calculer la moyenne $\bar{x}_n$ de ces n personnes-là. La proposition indique que si n est assez grand, alors $\bar{x}_n$ donne une bonne idée de m.

Ne pas confondre X_n qui est une variable aléatoire, c'est-à-dire une application mesurable, qui nous sert à démontrer des théorèmes en mathématique, avec x_n qui est un nombre mesuré sur le terrain, qui a une existence *physique* bien définie, mais aucun statut mathématique.

Si on croit que toute cette histoire de variables aléatoires, de tribu et de mesures décrit effectivement la réalité, alors on peut croire que le comportement de la suite $\bar{X}_n$ décrit bien le comportement de la suite $\bar{x}_n$ (cette dernière n'étant même pas une suite parce qu'on n'a jamais qu'un nombre fini de mesures expérimentales).

La **variance empirique** d'un échantillon est la variable aléatoire

$$V_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2. \tag{36.28}$$

La **variance empirique corrigée** est la variable aléatoire

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2. \tag{36.29}$$

Lemme 36.14.
La variance corrigée et la variance empirique ont comme espérances :

$$E(V_n^2) = \frac{n-1}{n}\sigma^2 \tag{36.30a}$$

$$E(S_n^2) = \sigma^2. \tag{36.30b}$$

Démonstration. Nous commençons par calculer l'espérance de la variance non corrigée. La première

étape est de la récrire sous la forme

$$\begin{aligned} V_n^2 &= \frac{1}{n}\sum_k (X_k^2 - 2X_k\bar{X}_n + \bar{X}_n^2) \\ &= \frac{1}{n}\sum_k X_k^2 - \frac{2}{n}\bar{X}_n \underbrace{\sum_k X_k}_{=n\bar{X}_n} + \frac{1}{n}\underbrace{\sum_k \bar{X}_n^2}_{=n\bar{X}_n^2} \\ &= \frac{1}{n}\sum_k X_k^2 - 2\bar{X}_n^2 + \bar{X}_n^2 \\ &= \frac{1}{n}\sum_k X_k^2 - \bar{X}_n^2. \end{aligned} \tag{36.31}$$

Nous calculons séparément l'espérance de ces deux termes. Si X est la loi parente des X_i, en utilisant l'indépendance des X_i nous trouvons

$$\begin{aligned} E\left(\frac{1}{n}\sum_k X_k^2\right) &= \frac{1}{n}\sum_{k=1}^n E(X_k^2) \\ &= E(X^2) \\ &= E(X)^2 - \mathrm{Var}(X). \end{aligned} \tag{36.32}$$

Nous devons à présent calculer l'espérance de $\bar{X}_n^2$:

$$E(\bar{X}_n^2) = E(\bar{X}_n)^2 + \mathrm{Var}(\bar{X}_n). \tag{36.33}$$

En utilisant le lemme 35.23,

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_k X_k\right) \tag{36.34a}$$

$$= \frac{1}{n^2}\sum_k \mathrm{Var}(X_k) \tag{36.34b}$$

$$= \frac{1}{n}\mathrm{Var}(X). \tag{36.34c}$$

Par conséquent

$$E(V_n^2) = \mathrm{Var}(X)\left(1 - \frac{1}{n}\right) = \frac{n-1}{n}\mathrm{Var}(X). \tag{36.35}$$

En ce qui concerne la variance corrigée,

$$S_n^2 = \frac{1}{n-1}\sum_k (X_k - \bar{X}_n)^2 = \frac{n}{n-1}V_n^2, \tag{36.36}$$

par conséquent $E(S_n^2) = \frac{n}{n-1}E(V_n^2) = \mathrm{Var}(X)$. □

Théorème 36.15 (Théorème de Cochran[270])**.**
Soient (X_i) des variables aléatoires gaussiennes indépendantes de loi $\mathscr{N}(m, \sigma^2)$ avec $\sigma > 0$. Alors

(1) $\bar{X}_n \sim \mathscr{N}(m, \frac{\sigma^2}{n})$,

(2) $\left(\frac{n-1}{\sigma^2}\right) S_n^2 = \left(\frac{n}{\sigma^2}\right) \bar{V}_n^2 \sim \chi^2(n-1)$,

(3) les variables aléatoires $\bar{X}_n$ et $\bar{V}_n$ sont indépendantes et

$$\frac{\bar{X}_n - m}{S_n/\sqrt{n}} = \frac{\bar{X}_n - m}{\sqrt{\bar{V}_n/(n-1)}} \sim \mathscr{T}(n-1). \tag{36.37}$$

Nous pouvons aussi écrire le dernier résultat en termes de la variance corrigée S_n, l'estimateur sans biais de la variance parce que

$$\sqrt{\frac{\bar{V}_n}{n-1}} = \frac{1}{\sqrt{n}} S_n \tag{36.38}$$

en vertu de la définition (36.29).

Proposition 36.16.
Soit X une variable aléatoire de variance $\mathrm{Var}(X) = \sigma$. Si $E(X^4) < \infty$, alors

(1) $S_n^2 \xrightarrow{p.s.} \sigma^2$.

(2)

$$\frac{S_n^2 - \sigma^2}{\sqrt{\frac{\mu^4 - \sigma^4}{n}}} \xrightarrow{\mathscr{L}} Z \sim \mathscr{N}(0,1) \tag{36.39}$$

où $\mu^4 = E(X^4)$ est le moment d'ordre 4 de X.

Théorème 36.17.
Si $(X_1, \ldots, X_n)$ est un n-échantillon de loi parente $\mathscr{N}(m, \sigma^2)$, alors

(1) Les variables aléatoires

$$\frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \quad \text{et} \quad (n-1)\frac{S_n^2}{\sigma^2} \tag{36.40}$$

sont indépendantes.

(2) La loi de $(n-1)\frac{S_n^2}{\sigma^2}$ est $\chi^2(n-1)$.

Si un échantillon vérifie ces deux propriétés, alors les X_i sont de loi $\mathscr{N}(m, \sigma^2)$.

L'inégalité de Markov donne une borne supérieure à la probabilité qu'une variable aléatoire positive soit plus grande ou égale à une constante.

Théorème 36.18 (Inégalité de Markov)**.**
Soit X une variable aléatoire à valeurs dans $\mathbb{R}^d$ et $\varphi\colon \mathbb{R}^d \to [0, \infty[$. Alors

$$P\big(\varphi(X) \geqslant a\big) \leqslant \frac{E\big(\varphi(X)\big)}{a} \tag{36.41}$$

pour tout $a > 0$.

Démonstration. Calculons le second membre :

$$\begin{aligned}
\frac{E\big(\varphi(X)\big)}{a} &= \int_\Omega \frac{\varphi(X)}{a} dP \\
&= \int_{\varphi(X) \geqslant a} \underbrace{\frac{\varphi(X)}{a}}_{\leqslant 1} dP + \int_{\varphi(X) < a} \frac{\varphi(X)}{a} dP \\
&\geqslant \int_{\varphi(X) \leqslant a} dP \\
&= P\big(\varphi(X) \leqslant a\big).
\end{aligned} \tag{36.42}$$

D'où l'inégalité voulue. □

36.4 Estimation ponctuelle

Nous considérons un modèle statistique

$$\mathcal{S} = \big[(\Omega, \mathcal{F}, P), (X_\theta), (\mu_\theta)\big]_{\theta \in \Theta} \tag{36.43}$$

et pour tout θ nous notons $X = (X_{\theta,1}, \dots, X_{\theta,n})$ un échantillon de loi parente μ_θ. Tant que nous travaillerons avec un θ fixé, nous écrirons $X = (X_1, \dots, X_n)$ sans expliciter la paramètre θ. Nous noterons

$$E_\theta\big(\varphi(X_1, \dots, X_n)\big) = \int_{\mathbb{R}^n} \varphi(x_1, \dots, x_n) d\mu_\theta^{\otimes n}(x_1, \dots, x_n). \tag{36.44}$$

Dans cette notation nous plaçons le θ sur l'espérance, tandis qu'en réalité le θ devrait être sur chaque X_1. Tant qu'aucune confusion n'est possible nous ferons toujours cet abus d'écriture.

Le but de la théorie de l'estimation est de déduire la valeur de θ (et donc la loi μ_θ) à partir d'un échantillon de loi parente θ.

Nous posons les hypothèses suivantes.

(1) Le modèle statistique $\mathcal{S}$ est paramétré, c'est-à-dire que $\Theta \subset \mathbb{R}^d$ avec le plus souvent $d = 1, 2$. Typiquement les paramètres seront la moyenne et la variance.

(2) Le modèle statistique est **identifiable**, c'est-à-dire que pour tout couple $(\theta_1, \theta_2) \in \Theta^2$, si $\theta_1 \neq \theta_2$, alors $\mu_{\theta_1} \neq \mu_{\theta_2}$.

(3) Le modèle $\mathcal{S}$ est **dominé** par la mesure de Lebesgue si les lois μ_θ sont continues et par la mesure de comptage si les lois μ_θ sont discrètes.

Exemple 36.19.
Quelques familles identifiables :

— La famille des lois exponentielles $\big(\mathscr{E}(\lambda)\big)_{\lambda>0}$ est identifiable.

— Les lois gaussiennes $\big(\mathscr{N}(m, \sigma^2)\big)_{m\in\mathbb{R},\sigma^2>0}$ sont également identifiables.

En réalité il est assez compliqué de trouver un exemple de modèle non identifiable à moins de la faire exprès. Par exemple en paramétrant les lois exponentielles de la façon suivante : $\big(\mathscr{E}(\sin(\lambda))\big)_{\lambda\in\mathbb{R}}$. Cette famille n'est pas identifiable. △

Le corolaire 13.200 ainsi que l'hypothèse de modèle dominé implique que les lois ont des densités. Si la loi μ_θ est discrète, nous notons

$$p(x, \theta) = \mu_\theta(\{x\}) \tag{36.45}$$

la densité de μ_θ par rapport à la mesure de comptage. Si La loi μ_θ est continue, nous notons

$$p(x, \theta) = f_\theta(x) \tag{36.46}$$

la densité par rapport à la mesure de Lebesgue.

Si μ_θ est une loi discrète et si $(X_1, \dots, X_n)$ est un échantillon de taille n, alors pour tout $(x_1, \dots, x_n) \in \mathbb{R}^n$ nous avons

$$p_n(x_1, \dots, x_n; \theta) = \mu_\theta^{\otimes n}\big(\{x_1, \dots, x_n\}\big) = \mu_\theta(\{x_1\}) \dots \mu_\theta(\{x_n\}) = p(x_1, \theta) \dots p(x_n, \theta). \tag{36.47}$$

La première et la dernière égalité sont des notations ; la seconde est une conséquence de l'indépendance des X_i contenues dans l'échantillon. Pour une loi continues, nous adoptons la même notation. Le vecteur aléatoire $(X_1, \dots, X_n)$ admet la densité

$$(x_1, \dots, x_n) \mapsto p_n(x_1, \dots, x_n; \theta) = f_\theta(x_1) \dots f_\theta(x_n) = p(x_1, \theta) \dots p(x_n, \theta). \tag{36.48}$$

Exemple 36.20.
Soit $(X_1, \dots, X_n)$ un n-échantillon de loi $\mathscr{B}(1, \theta)$ avec $\theta \in]0, 1[$. C'est une loi discrète portée par l'ensemble $\{0, 1\}$. Nous avons

$$p(x, \theta) = \begin{cases} 0 & \text{si } 0 \neq x \neq 1 \\ 1 - \theta & \text{si } x = 0 \\ \theta & \text{si } x = 1. \end{cases} \tag{36.49}$$

De façon plus condensée nous pouvons écrire

$$p(x,\theta) = \theta^x(1-\theta)^{1-x}\mathbb{1}_{\{0,1\}}(x). \tag{36.50}$$

Pour tout $(x_1,\dots,x_n) \in \mathbb{R}^p$, la densité du n-échantillon est donnée par

$$p_n(x_1,\dots,x_n;\theta) = \theta^{x_1+\dots+x_n}(1-p)^{1-\sum_i(1-x_i)}\mathbb{1}_{\{0,1\}^n}(x_1,\dots,x_n). \tag{36.51}$$

△

36.5 Statistiques et estimateurs

Définition 36.21.
*Une **statistique** sur un modèle d'échantillonnage est une variable aléatoire fonction de l'échantillon $(X_1,\dots,X_n)$ ne dépendant pas de θ*[1]*. C'est-à-dire une application borélienne $T\colon \mathbb{R}^n \to \mathbb{R}$ ne dépendant pas de θ. La statistique associée à cette application est $S = T(X_1,\dots,X_n)$.*

Les fonctions $T(X_1,\dots,X_n)$ données par $\sum_i X_i$, $e^{\sum X_i}$ sont des statistiques. La constante $\frac{1}{2}$ est également une statistique (mais elle est moins intéressante).

Un **estimateur** est une statistique qui prend ses valeurs dans Θ. Nous la noterons

$$\hat{\theta}_n = \theta(X_1,\dots,X_n). \tag{36.52}$$

La fonction $\hat{\theta}_n$ est borélienne à valeurs dans Θ.

Exemple 36.22.
Soit un n-échantillon de loi $\mathscr{B}(1,\theta)_{\theta\in[0,1]}$. Les fonctions $\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$ et $\hat{\varphi}_n = \frac{1}{2}$ sont des estimateurs. Cependant nous devinons que la première va être plus intéressante que la seconde. △

Pour la suite, nous travaillerons avec des estimateurs de carré intégrable, c'est-à-dire que

$$E_\theta\big(|\hat{\theta}_n(X_1,\dots,X_n)|^2\big) < \infty \tag{36.53}$$

pour tout $\theta \in \Theta$.

36.5.1 Qualité des estimateurs

Définition 36.23.
*Une estimateur est **convergent** ou **consistant** si pour tout $\theta \in \Theta$, la suite de variables aléatoires $\hat{\theta}_n(X_1,\dots,X_n)$ converge en probabilité vers θ.*

En d'autres termes, l'estimateur $\hat{\theta}_n$ est convergeant si pour tout $\theta \in \Theta$ et pour tout $\eta > 0$,

$$\lim_{n\to\infty} P\big(|\hat{\theta}_n(X_1,\dots,X_n) - \theta| > \eta\big) = 0. \tag{36.54}$$

La probabilité dans le membre de gauche est donnée par

$$\mu_\theta^{\otimes n}\Big(\big\{(x_1,\dots,x_n) \in \mathbb{R}^n \text{ tel que } |\hat{\theta}_n(x_1,\dots,x_n) - \theta| > \eta\big\}\Big). \tag{36.55}$$

Soit $\hat{\theta}_n$ un estimateur pour θ. Nous cherchons à minimiser l'erreur commise en remplaçant θ par $\hat{\theta}_n$. Nous introduisons donc le **risque quadratique** de l'estimateur $\hat{\theta}_n$ par

$$R(\hat{\theta}_n,\theta) = E_\theta\big((\hat{\theta}_n - \theta)^2\big). \tag{36.56}$$

Nous disons qu'un estimateur $\hat{\theta}_{n,1}$ est préférable à $\hat{\theta}_{n,2}$ si pour tout $\theta \in \Theta$ nous avons

$$R(\hat{\theta}_{n,1},\theta) < R(\hat{\theta}_{n,2},\theta). \tag{36.57}$$

1. Parce que d'habitude c'est ce qu'on cherche à estimer.

Lemme 36.24.
Une formule alternative pour le risque quadratique :

$$R(\hat{\theta}_n, \theta) = \mathrm{Var}(\hat{\theta}_n) + \big(E_\theta(\hat{\theta}_n) - \theta\big)^2 \tag{36.58}$$

Démonstration. Nous avons

$$E_\theta\Big((\hat{\theta}_n - \theta)^2\Big) = \mathrm{Var}(\hat{\theta}_n - \theta) - E_\theta(\hat{\theta}_n - \theta)^2. \tag{36.59}$$

D'une part $\mathrm{Var}(\hat{\theta}_n - \theta) = \mathrm{Var}(\hat{\theta}_n)$ et d'autre part $E_\theta(\hat{\theta}_n - \theta)^2 = \big[E(\hat{\theta}_n) - \theta\big]^2$. Par conséquent

$$R(\hat{\theta}_n, \theta) = \mathrm{Var}(\hat{\theta}_n) - \Big(E(\hat{\theta}_n) - \theta\Big)^2. \tag{36.60}$$

□

Le **biais** de l'estimateur $\hat{\theta}_n$ est la quantité

$$E_\theta(\hat{\theta}_n) - \theta. \tag{36.61}$$

À ce niveau, nous rappelons que nous écrivons E_θ l'espérance calculée en supposant la valeur θ pour le paramètre des différentes variables aléatoires entrant dans le calcul. Voir la discussion autour de la définition (36.44).

Exemple 36.25.
Dans le cadre de la proposition 35.111, nous voulons savoir si $\frac{N_t}{t}$ est un estimateur sans biais de λ. Pour ce faire nous calculons

$$E_\lambda\left(\frac{N_t}{t}\right) = \frac{1}{t}E_\lambda(N_t) = \lambda \tag{36.62}$$

parce que $E(N_t) = \lambda t$. Ici nous avons calculé $E(N_t)$ en prenant λ pour valeur du paramètre du processus de Poisson, alors que en principe c'est justement le paramètre que nous voulons estimer. △

Exemple 36.26.
La moyenne empirique $\bar{X}_n$ est un estimateur sans billet de la moyenne. L'estimateur

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_i)^2 \tag{36.63}$$

est un estimateur sans billet de la variance. △

Un estimateur sans biais n'est pas toujours de meilleur qualité qu'un estimateur avec biais. En effet ce que nous voulons est de se donner un (petit) intervalle I autour de la bonne valeur de θ et de maximiser $P(\hat{\theta}_n \in I)$. Sur la figure 36.1, c'est l'estimateur biaisé rouge tombe plus souvent sur le bon intervalle que l'estimateur non biaisé bleu.

Nous allons maintenant étudier quelques manières de construire des estimateurs convergeant [2]. Il vont évidemment s'appuyer sur la loi des grands nombres.

36.5.2 Méthode des moments

Sans surprises, un bon estimateur pour la moyenne est

$$\hat{\theta}_n(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i. \tag{36.64}$$

2. Là je ne suis vraiment pas sûr de l'orthographe. Comme partout où j'utilise le verbe « converger » d'ailleurs.

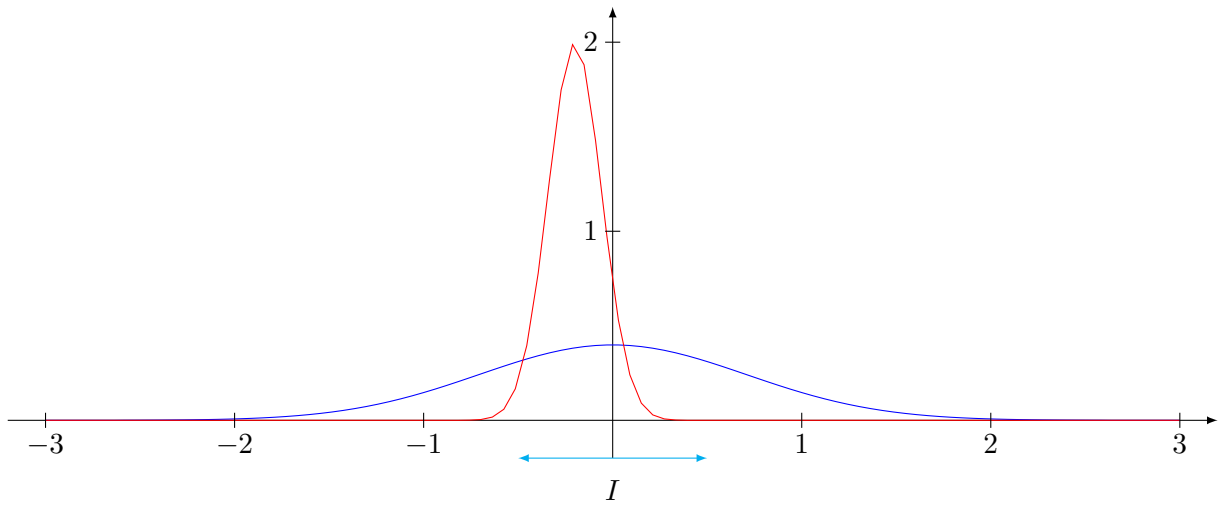


FIGURE 36.1 – Un estimateur sans biais et un avec biais.

Plus généralement, nous supposons qu'il existe une fonction borélienne[3] $M\colon \mathbb{R} \to \mathbb{R}$ telle que

$$E_\theta\big(|M(X)|\big) < \infty \tag{36.65}$$

où X est la loi parente de l'échantillon. Supposons également que la fonction

$$h(\theta) = E_\theta\big(M(X)\big) \tag{36.66}$$

soit inversible et continue sur Θ. Dans ce cas, pour estimer le paramètre θ, nous considérons l'estimateur

$$\hat{\theta}_n = h^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} M(X_i)\right). \tag{36.67}$$

Cela est un estimateur convergeant. En effet, la loi des grands nombres dit que

$$\frac{1}{n}\sum_i M(X_i) \xrightarrow{p.s.} E_\theta\big(M(X)\big). \tag{36.68}$$

En composant avec la fonction h, nous avons

$$\hat{\theta}_n \xrightarrow{p.s.} h^{-1}\Big(E_\theta\big(M(X)\big)\Big) = \theta. \tag{36.69}$$

Dans cette construction, $M(X)$ est le moment de X que l'on souhaite déterminer.

Exemple 36.27.
Soit $(X_1, \dots, X_n)$ un échantillon de loi $\mathscr{E}(\lambda)$. Construire $\hat{\lambda}_n$. Pour une loi exponentielle,

$$E(X) = \frac{1}{\lambda}. \tag{36.70}$$

Nous devons donc déterminer le moment d'ordre 1 de X (c'est-à-dire sa moyenne). Nous considérons donc la fonction $M(x) = x$; par conséquent

$$h(\lambda) = E(X) = \frac{1}{\lambda} \tag{36.71}$$

et $h^{-1}(\theta) = 1/\theta$. L'estimateur que nous considérons pour λ est finalement

$$\hat{\theta}_n = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i}. \tag{36.72}$$

△

3. Définition 13.42.

36.5.3 Méthode de substitution

Supposons que nous connaissions un estimateur convergeant $\hat{\theta}_n \to \theta$. Si $g\colon \mathbb{R}^d \to \mathbb{R}$ est une fonction continue, alors

$$g(\hat{\theta}_n) \to g(\theta). \tag{36.73}$$

36.5.4 Méthode du maximum de vraisemblance

Exemple 36.28.
Nous désirons contrôler la qualité d'une chaine de production ; pour cela nous prélevons un échantillon de 10 pièces, et nous en trouvons 3 défectueuses. Que dire de la proportion de pièces défectueuses ?

Évidemment, le plus probable est que la proportion de pièces défectueuses soit de 1/3. Analysons en détail comment nous arrivons à ce résultat. Nous considérons que le fait de tirer 10 pièces revient à une expérience binomiale de paramètres 10 et de probabilité p inconnue. Dans ce cas, la probabilité d'observer exactement 3 pièces défectueuses est de

$$L(p) = P(X = 3) = \binom{10}{3} p^3 (1-p)^7. \tag{36.74}$$

Le maximum de $L(p)$ est $p = 3/10$.

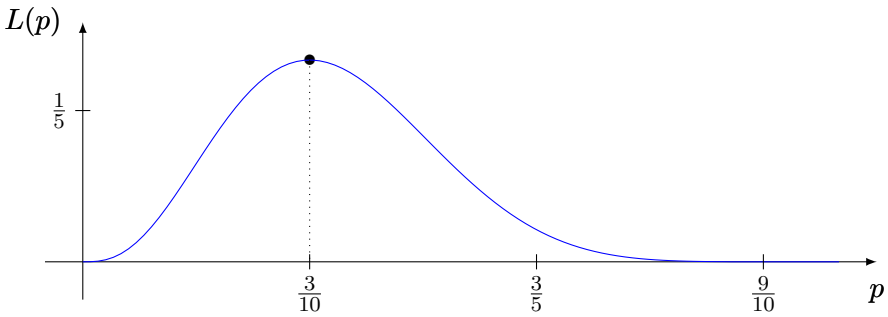


FIGURE 36.2 – La fonction de vraisemblance de l'exemple 36.28.

△

Soit $(x_1, \ldots, x_n)$, une réalisation de l'échantillon $(X_1, \ldots, X_n)$. L'application

$$\theta \mapsto p_n(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} p(x_i, \theta) \tag{36.75}$$

est la **vraisemblance** de l'échantillon. Nous définissons $\hat{\theta}_n$ par

$$p_n(x_1, \ldots, x_n; \hat{\theta}_n) = \sup_{\theta \in \Theta} p_n(x_1, \ldots, x_n; \theta). \tag{36.76}$$

Remarque 36.29.
Nous passons sous le silence le fait que la fonction sup soit une fonction mesurable, et que par conséquent $\hat{\theta}_n$ soit bien une variable aléatoire.

La variable aléatoire $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ est l'**estimateur de maximum de vraisemblance** de θ.

36.5.5 Exemples sous forme d'exercices

Exemple 36.30.
Soit $(X_1, \ldots, X_n)$ un échantillon de loi $\mathscr{B}(1, \theta)$ avec $\theta \in [0, 1]$. Trouver l'estimateur de maximum de vraisemblance de θ.

En utilisant la densité de la loi multinomiale,

$$p_n(x_1,\dots,x_n;\theta)=\prod_i p(x_i,\theta)=\prod_i \theta^{x_i}(1-\theta)^{1-x_i}\mathbb{1}_{\{0,1\}^n}(x_1,\dots,x_n). \tag{36.77}$$

En passant au logarithme, si $x_i\in\{0,1\}$,

$$L_n(\theta)=\sum_i x_i\ln(\theta)+(1-x_i)\ln(1-\theta) \tag{36.78}$$

En passant à la dérivée, nous trouvons que l'extrémum est donné par

$$\theta=\frac{1}{n}\sum_{i=1}^n x_i. \tag{36.79}$$

△

Exemple 36.31.
Soit $(X_1,\dots,X_n)$ un échantillon de loi $\mathscr{E}(\lambda)$ avec $\theta>0$. Trouver l'estimateur de maximum de vraisemblance de θ.

Nous avons

$$p_n(x_1,\dots,x_n)=\prod_{i=1}^n \theta_i e^{-\theta x_i}. \tag{36.80}$$

En passant au logarithme la fonction à minimiser est

$$L(\theta)=n\ln(\theta)-\theta\sum_i x_i. \tag{36.81}$$

La minimisation donne

$$\theta=\frac{n}{\sum_i x_i}, \tag{36.82}$$

c'est-à-dire

$$\hat{\theta}_n(X_1,\dots,X_n)=\frac{n}{\sum_i X_i}=\frac{1}{\bar{X}_n}. \tag{36.83}$$

Ce résultat n'est pas étonnant vu que le paramètre λ de la loi exponentielle est l'inverse de la moyenne.

△

Exemple 36.32.
Soit $(X_1,\dots,X_n)$, un échantillon de loi parente uniforme sur $[0,\theta]$ avec $\theta>0$.

(1) Montrer que la fonction de vraisemblance est donnée par

$$L(x_1,\dots,x_n;\theta)=\frac{1}{\theta^n}\mathbb{1}_{[0,\infty[}\big(\min(x_1,\dots,x_n)\big)\mathbb{1}_{[\max(x_1,\dots,x_n),\infty[}(\theta). \tag{36.84}$$

(2) Déterminer l'estimateur du maximum de vraisemblance de θ.

Lorsque nous parlons de paramètres qui peuvent prendre un spectre continu de valeurs, il est inutile de calculer la *probabilité* parce qu'elle est nulle. Le système de maximum de vraisemblance fonctionne avec les densités. Dans notre cas, la fonction de vraisemblance est le produit des densités :

$$L(x_1,\dots,x_n;\theta)=\prod_{i=1}^n p(x_i;\theta) \tag{36.85a}$$

$$=\prod_i \frac{1}{\theta}\mathbb{1}_{[0,\theta]}(x_i) \tag{36.85b}$$

$$=\frac{1}{\theta^n}\mathbb{1}_{[0,\theta]^n}(x_1,\dots,x_n). \tag{36.85c}$$

De cette expression nous voyons que $\min\{x_1,\dots,x_n\}$ doit être positif en même temps que le maximum doit être plus petit que θ. Cette seconde condition peut s'écrire $\mathbb{1}_{[\max\{x_1,\dots,x_n\},\infty[}(\theta)$. Au final nous avons

$$L(x_1,\dots,x_n;\theta)=\frac{1}{\theta^n}\mathbb{1}_{[0,\infty[}\big(\min(x_1,\dots,x_n)\big)\mathbb{1}_{[\max\{x_1,\dots,x_n\},\infty[}(\theta). \tag{36.86}$$

Il n'est évidemment pas possible de dériver explicitement cette expression. Par contre pour que cette fonction soit non nulle, il faut obligatoirement $\theta\geqslant\max\{x_1,\dots,x_n\}$. Par conséquent elle prend son maximum pour $\theta=\max\{x_1,\dots,x_n\}$.

La conclusion est que l'estimateur de maximum de vraisemblance de θ est $\hat{\theta}_n=\max_i\{X_i\}$. △

Exemple 36.33.
Déterminer l'estimateur de maximum de vraisemblance de la moyenne pour un modèle statistique dans laquelle la famille de probabilités est

$$(\mu_\theta)=\{\mathscr{N}(\theta,\sigma^2)\text{ tel que }\theta\in\in\mathbb{R}\}. \tag{36.87}$$

Nous supposons que σ est connu.

La densité de la variable aléatoire conjointe $(X_1,\dots,X_n)$ au point $(x_1,\dots,x_n)$ est le produit des densités, donc

$$\prod_{i=1}^n\gamma_{m,\sigma}(x_1,\dots,x_n)=\frac{1}{\sigma^n(2\pi)^{n/2}}\exp-\frac{1}{2}\left(\sum_i\left(\frac{x_i-m}{\sigma}\right)^2\right). \tag{36.88}$$

Étant donné que le but est de minimiser, nous pouvons oublier le facteur et passer au logarithme :

$$L_0(\bar{x})=-\frac{1}{2}\sum_{i=0}^n\left(\frac{x_i-m}{\sigma}\right)^2. \tag{36.89}$$

Nous pouvons également supprimer le $\frac{1}{2}$ et le $1/\sigma^2$. La fonction à minimiser devient

$$L(x_1,\dots,x_n)=-\sum_i(x_i-m)^2, \tag{36.90}$$

dont la dérivée vaut $2nm-2\sum_i x_i$. Par conséquent nous avons un minimum avec

$$m=\frac{1}{n}\sum_{i=1}^n x_i. \tag{36.91}$$

△

36.5.6 Estimation d'une fonction de répartition

Théorème 36.34 (Glivenko-Cantelli[534])**.**
Soient X_i des variables aléatoires indépendantes et identiquement distribuées suivant une loi dont la fonction de distribution est F (inconnue). Nous définissons l'estimateur

$$F_n(x)=\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{X_i\leqslant x}. \tag{36.92}$$

Alors

$$P\big(\lim_{n\to\infty}\|F_n-F\|_\infty=0\big)=1. \tag{36.93}$$

Autrement dit, pour presque tout $\omega\in\Omega$,

$$\lim_{n\to\infty}\|F_n(\omega,.)-F\|_\infty=0. \tag{36.94}$$

C'est-à-dire qu'il y a presque certainement convergence en probabilité.

Notons que de façon générale lorsqu'on parle d'estimateurs, partout où il y a un « n » dans une variable aléatoire, il y a une dépendance sous-entendue en ω.

Proposition 36.35.
Pour presque tout x, l'estimateur $F_n(x)$ est sans biais par rapport à $F(x)$:

$$E\big(F_n(x)\big) = F(x). \tag{36.95}$$

Démonstration. C'est juste un calcul :

$$E\big(F_n(x)\big) = \frac{1}{n}\sum_{i=1}^{n} E\big(\mathbb{1}_{\{X_i \leqslant x\}}\big) = \frac{1}{n}\sum_{i=1}^{n} F(x) = F(x). \tag{36.96}$$

□

36.5.7 Exemples sous forme d'exercices

Exemple 36.36.
Dans cet exercice nous construisons un estimateur biaisé qui présente un risque quadratique inférieur à un estimateur non biaisé.

Soit un modèle statistique dont la famille de lois est $\{\mathcal{N}(m,\sigma^2)\}_{\theta\in]0,\infty[}$ où m est un paramètre réel connu. En ce qui concerne la variance nous considérons

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2. \tag{36.97}$$

(1) Montrer que $\hat{\sigma}_n^2$ est un estimateur sans biais de θ.
(2) Montrer que le risque quadratique de $\hat{\sigma}_n^2$ est

$$R(\hat{\sigma}_n^2, \theta) = \frac{2\theta^2}{n}. \tag{36.98}$$

(3) Considérer la famille d'estimateurs $c\hat{\sigma}_n^2$ avec $c > 0$. Déterminer la valeur de c qui minimise le risque quadratique de l'estimateur.

(1) Nous calculons

$$\begin{aligned} E(\hat{\sigma}_n^2) &= \frac{1}{n}\sum_i E\big[(X_i - m)^2\big] \\ &= \frac{1}{n}\sum_i \mathrm{Var}(X_i - m) - \frac{1}{n}\sum_i E(X_i - mj^2) \\ &= \frac{1}{n}\sum_i \mathrm{Var}(X_i) \\ &= \sigma^2. \end{aligned} \tag{36.99}$$

(2) Par la formule (36.60) nous savons que le risque d'un estimateur sans biais est donné par sa variance :

$$R(\hat{\theta}_n, \theta) = \mathrm{Var}(|\hat{\sigma}_n^2 - \theta|) = \mathrm{Var}(\hat{\sigma}_n^2). \tag{36.100}$$

Étant donné que les variables aléatoires X_i sont indépendantes et identiquement distribuées, le lemme 35.23 nous enseigne que la variance de la somme est la somme des variances. Nous avons donc à calculer

$$\begin{aligned} \mathrm{Var}(|\hat{\sigma}_n^2 - \theta|) &= \mathrm{Var}(\hat{\sigma}_n^2) \\ &= \frac{1}{n^2}\sum_i \mathrm{Var}\left[(X_i - m)^2\right] \\ &= \frac{1}{n^2}\sum_i E\big[(X_i - m)^4\big] - E\big[(X_i - m)^2\big]^2. \end{aligned} \tag{36.101}$$

Ces espérances ne sont pas très compliquées à calculer en utilisant la fonction caractéristique donnée par la proposition 35.113 :

$$\Phi_{X-m}(t) = E\big(e^{it(X-m)}\big) = e^{-itm}E(e^{itX}) = e^{-itm}\Phi_X(t) = \exp\left(-\frac{\sigma^2t^2}{2}\right). \tag{36.102}$$

Nous avons $\Phi^{(4)}(0) = 3\sigma^4$ et $\Phi''(0) = -\sigma^2$. Par conséquent

$$\mathrm{Var}(\hat{\sigma}_n^2) = \frac{1}{n^2}\sum_i 2\theta^2 = \frac{2\theta^2}{n}. \tag{36.103}$$

Notez ici que $\theta = \sigma^2$.

(3) En tenant compte du fait que $\mathrm{Var}(\hat{\sigma}_n^2) = \frac{2\theta^2}{n}$ et $E(\hat{\sigma}_n^2) = \theta$, nous avons

$$E\big((c\hat{\sigma}_n^2 - \theta)^2\big) = \mathrm{Var}(c\hat{\sigma}_n^2 - \theta) + E(c\hat{\sigma}_n^2 - \theta)^2 \tag{36.104a}$$

$$= 2\frac{c^2\theta^2}{n} + (c-1)^2\theta^2. \tag{36.104b}$$

La dérivée (par rapport à c) de cela s'annule pour $c_0 = \frac{n}{n+2}$. Notons que nous n'avons pas tout à fait démontré que cela est bien un minimum. Calculons cependant le risque quadratique de notre estimateur pour cette valeur de c. Pour cela nous reportons $c = c_0$ dans l'expression (36.104b) :

$$E\big(\frac{n}{n+2}\hat{\sigma}_n^2\big) = \frac{2\theta^2}{n+2}. \tag{36.105}$$

Cela est effectivement plus petit que $R(\hat{\sigma}_n^2, \theta)$.

Nous avons ainsi construit un estimateur biaisé qui a un risque quadratique plus petit que l'estimateur non biaisé. △

Exemple 36.37.
Nous considérons la famille de probabilités $\mu_\theta = \mathscr{N}(\theta)$ où $\theta = (m, \sigma^2) \in \mathbb{R} \times]0, \infty[$. Déterminer l'estimateur de maximum de vraisemblance du paramètre $\theta = (m, \sigma^2)$.

Pour chaque observation x_i nous avons une densité gaussienne. Le produit donne

$$p\big(x_1, \dots, x_n; (m, \sigma^2)\big) = \frac{1}{\sigma^2(2\pi)^{n/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_i (x_i - m)^2\right]. \tag{36.106}$$

En passant au logarithme et en supprimant des facteurs inutiles à la minimisation,

$$L(m, \sigma) = -n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_i (x_i - m)^2. \tag{36.107}$$

L'annulation de la dérivée par rapport à m donne immédiatement $m = \frac{1}{n}\sum_i x_i$. L'annulation de la dérivée par rapport à σ donne

$$-n\sigma^2 + \sum_i (x_i - m)^2 = 0 \tag{36.108}$$

et donc

$$\sigma^2 = \frac{1}{n}\sum_i (x_i - \bar{x}_n). \tag{36.109}$$

L'estimateur de maximum de vraisemblance du couple $\theta = (m, \sigma^2)$ est donc

$$\hat{\theta}_n = \left(\frac{1}{n}\sum_i X_i, \frac{1}{n}\sum_i (X_i - \bar{X}_n)^2\right). \tag{36.110}$$

△

36.5.8 Espérance et variance d'un estimateur

Soit $T_n = T_(X_1, \dots, X_n)$ un estimateur du paramètre θ dans un modèle d'échantillonnage. Les moyennes et variances de l'estimateur sont les variables aléatoires

$$m_{\theta,n} = E_\theta(T_n) = \int_{\mathbb{R}^n} T_n(x_1, \dots, x_n) d\mu_\theta^{\otimes n}(x1, \dots, x_n), \tag{36.111a}$$

$$\mathrm{Var}_\theta(T_n) = \int_{\mathbb{R}^n} [T_n(x_1, \dots, x_n) - m_{\theta,n}]^2 d\mu_\theta^{\otimes n}(x1, \dots, x_n), \tag{36.111b}$$

Lemme 36.38.
Si l'estimateur T_n satisfait

$$\lim_{n\to\infty} E_\theta(T_n) = \theta \tag{36.112a}$$

$$\lim_{n\to\infty} \mathrm{Var}_\theta(T_n) = 0, \tag{36.112b}$$

alors il est convergeant.

Démonstration. Nous utilisons l'inégalité de Markov (théorème 36.18) et nous introduisons l'espérance de l'estimateur :

$$P\big(|T_n(X_1, \dots, X_n) - \theta| > \epsilon\big) \leqslant \frac{1}{\epsilon} E\big(T_n(X_1, \dots, X_n) - \theta\big) \tag{36.113a}$$

$$\leqslant \frac{1}{\epsilon} E\big(|T_n - m_{n,\theta}|\big) + \frac{1}{\epsilon} E\big(|m_{n,\theta} - \theta|\big) \tag{36.113b}$$

$$\tag{36.113c}$$

Le second terme est l'espérance d'une constante. Nous majorons le premier terme en utilisant le fait que $\|.\|_1 \leqslant \|.\|_2$ (voir la remarque 26.34 après l'inégalité de Hölder) :

$$P\big(|T_n(X_1, \dots, X_n) - \theta| > \epsilon\big) \leqslant \frac{1}{\epsilon} E\big(|T_n - m_{n,\theta}|^2\big)^{1/2} + \frac{1}{\epsilon}|E_\theta(T_n) - \theta| \tag{36.114a}$$

$$= \frac{1}{\epsilon}\,\mathrm{Var}(T_n)^{1/2} + \frac{1}{\epsilon}|E_\theta(T_n) - \theta|. \tag{36.114b}$$

Les deux termes tendent séparément vers zéro par hypothèse. Nous avons par conséquent la convergence en probabilité $T_n \to \theta$. □

36.6 Estimation par intervalle de confiance

Nous voudrions estimer la proportion d'individus dans une population ayant un certain caractère déterminé par une variable booléenne : chaque individu a ou non le caractère étudié. L'échantillon sera donc une suite de 0 et de 1.

Pour tout $i \in \{1, \dots, n\}$ nous notons

$$x_i = \begin{cases} 1 & \text{si le } i\text{ème individu a la caractère} \\ 0 & \text{sinon.} \end{cases} \tag{36.115}$$

et $\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i$. Notre modèle statistique sera

$$\mathcal{S} = \Big[(\Omega, \mathcal{F}, P), (X_\theta), B(1, \theta)\Big] \tag{36.116}$$

où Ω est l'ensemble des individus étudiés, P est la manière de choisir les individus lors du sondage (essentiellement c'est une loi uniforme) et X_θ est la variable aléatoire

$$X(\omega) = \begin{cases} 1 & \text{si } \omega \text{ a le caractère} \\ 0 & \text{sinon} \end{cases} \tag{36.117}$$

Cela est une variable aléatoire de distribution $\mathscr{B}(1, p)$ où p est inconnu. Ici, $\Theta = [0, 1]$ est l'ensemble des p possibles.

Remarque 36.39.
Nous supposons que Ω est la population entière et que la variable aléatoire est l'opinion de la personne ω. En cela, nous considérons que le tirage de l'échantillon est sans remise. Le fait que nous modélisions par une variable aléatoire de Bernoulli signifie que nous considérons l'approximation dans laquelle la population globale est grande.

Nous supposons que nous ayons un échantillon $(X_1, \dots, X_n)$ dont nous avons observé une réalisation $(x_1, \dots, x_n)$ de fréquence $\bar{x}_n$. Nous voudrions déterminer un intervalle dans lequel $\bar{X}_n$ a de fortes chances de se trouver. Plus précisément nous considérons un petit α et nous cherchons ϵ tel que

$$P\big(p \in [\bar{X}_n - \epsilon, \bar{X}_n + \epsilon]\big) = 1 - \alpha. \tag{36.118}$$

Typiquement, $\alpha = 5\%$. Le nombre α est le **niveau de confiance** que nous nous fixons à priori.

Si nous trouvons un intervalle I tel que $P(p \in I) = 1 - \alpha$, nous disons que l'intervalle est **exact**, si nous avons $P(p \in I) \geqslant 1 - \alpha$, nous disons que l'intervalle est **par excès**.

Il y a deux points de départs pour trouver l'intervalle. Le plus simple est d'utiliser le théorème central limite et considérer

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow{\mathscr{L}} \mathscr{N}(0,1). \tag{36.119}$$

La seconde est d'utiliser la loi exacte : $n\bar{X}_n = \sum_i X_i \sim \mathscr{B}(n,p)$.

Bien entendu la seconde donne lieu à des calculs plus compliquée.

Remarque 36.40.
Dans certaines vraies vies (par exemple en médecine), la taille des échantillons est très réduite. Dans ce cas le théorème central limite n'a aucun sens et les calculs exact s'imposent.

De plus dans de nombreux cas de la vraie vie, nous avons un ordinateur à disposition pour calculer avec la loi exacte. L'utilisation du théorème central limite dans le but de produire un intervalle de confiance semble donc de plus en plus être une survivance du passé.

Dans la suite, nous allons supposer que n est suffisamment grand pour justifier l'approximation normale du théorème central limite 35.97. Si Z est une variable aléatoire normale centrée réduite, notre premier essai est de faire

$$\begin{aligned}
1 - \alpha &= P\big(p \in [\bar{X}_n - \epsilon, \bar{X}_n + \epsilon]\big) && (36.120a)\\
&= P\left(\frac{-\epsilon\sqrt{n}}{\sqrt{p(1-p)}} \leqslant \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \leqslant \frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) && (36.120b)\\
&\simeq P\left(-\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} \leqslant Z \leqslant \frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) && (36.120c)\\
&= 2P\left(Z \leqslant \frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) - 1. && (36.120d)
\end{aligned}$$

La dernière ligne utilise la symétrie de la distribution $\mathscr{N}(0,1)$. Le nombre ϵ que nous cherchons vérifie donc

$$P\left(Z \leqslant \frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) = 1 - \frac{\alpha}{2}. \tag{36.121}$$

De nos jours, les ordinateurs donnent la loi de répartition inverse des normales. Cela nous fournit un nombre t_α tel que

$$\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} = t_\alpha \tag{36.122}$$

où t_α est le nombre tel que $P(Z \leqslant t_\alpha) = 1 - \alpha/2$.

Conclusions de ce premier essai :

(1) Le problème est que nous ne pouvons pas déduire ϵ de l'équation (36.122) parce que p est inconnu.

(2) Cela ruine notre premier essai et nous demande de trouver mieux.

(3) L'astuce est évidemment de remplacer p par $\bar{X}_n$, mais il faut le justifier.

Méthode Slutsky Le point de départ du premier essai infructueux était la convergence

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \xrightarrow{\mathscr{L}} N \sim \mathscr{N}(0,1) \tag{36.123}$$

donnée par le théorème central limite 35.97. Ce que nous voudrions en réalité est la convergence

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{\mathscr{L}} N \tag{36.124}$$

La loi des grands nombres nous donne

$$\bar{X}_n(1-\bar{X}_n) \xrightarrow{p.s.} p(1-p). \tag{36.125}$$

Par conséquent le lemme de Slutsky implique la convergence en loi du couple :

$$\left(\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}, \sqrt{\bar{X}_n(1-\bar{X}_n)}\right) \xrightarrow{\mathscr{L}} \left(N, \sqrt{p(1-p)}\right). \tag{36.126}$$

À ce point des opérations nous pouvons utiliser la proposition 35.83 au couple avec la fonction

$$f(x,y) = \frac{x}{y}\sqrt{p(1-p)} \tag{36.127}$$

dont la probabilité d'être continue est 1 ($y = 0$ est de mesure nulle dans $\mathbb{R}^2$). La conclusion du théorème est que

$$\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{\mathscr{L}} \mathscr{N}(0,1). \tag{36.128}$$

C'est à partie de là que nous pouvons construire notre intervalle de confiance :

$$1-\alpha = P(\bar{X}_n - \epsilon \leqslant p \leqslant \bar{X}_n + \epsilon) \tag{36.129a}$$

$$= P\left(\frac{\sqrt{n}\epsilon}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \geqslant \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \geqslant \frac{-\sqrt{n}\epsilon}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}\right)) \tag{36.129b}$$

$$\simeq P\left(\frac{\sqrt{n}\epsilon}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \geqslant N \geqslant \frac{-\sqrt{n}\epsilon}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}\right)). \tag{36.129c}$$

Nous cherchons maintenant dans les tables le ξ qui fait

$$P(-\xi \leqslant N \leqslant \xi) = 1-\alpha \tag{36.130}$$

puis nous cherchons ϵ de telle sorte à avoir

$$\frac{\sqrt{n}\epsilon}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} = \xi. \tag{36.131}$$

Dans cette équation tout est connu à part le ϵ qui se découvre.

Méthode piétonne Nous remarquons que l'événement

$$-\frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} \leqslant \sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \leqslant \frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}} \tag{36.132}$$

est le même que l'événement

$$\left|\sqrt{n}\frac{\bar{X}_n - p}{\sqrt{p(1-p)}}\right| \leqslant t_\alpha. \tag{36.133}$$

Vu que t_α est positif, cela est encore le même événement que

$$n\frac{(\bar{X}_n - p)^2}{p(1-p)} \leqslant t_\alpha^2 \tag{36.134}$$

ou encore

$$p^2(n + t_\alpha^2) - p(2n\bar{X}_n + t_\alpha^2) + n\bar{X}_n^2 \leqslant 0. \tag{36.135}$$

Les racines du polynôme du membre de gauche sont

$$p_\pm = \frac{2n\bar{X}_n + t_\alpha^2 \pm \sqrt{(2n\bar{X}_n + t_\alpha)^2 - 4(n + t_\alpha^2)n\bar{X}_n^2}}{2(n + t_\alpha^2)}. \tag{36.136}$$

Le but étant d'effectuer une limite $n \to \infty$, nous factorisons d'abord n. Après simplification

$$p_\pm = \frac{\bar{X}_n + \frac{t_\alpha}{2n} \pm t_\alpha\sqrt{\frac{t_\alpha^2}{4n^2} + \frac{\bar{X}_n(1-\bar{X}_n)}{n}}}{1 + \frac{t_\alpha}{n}}. \tag{36.137}$$

Étant donné que nous considérons que n est grand, nous allons négliger les termes en $\frac{1}{n}$ en faisant attention à ce que le terme en $\frac{1}{n}$ sous la racine est en réalité $1/\sqrt{n}$ et ne doit pas être négligé. Nous trouvons, à cette approximation, que

$$p \in \Big[\bar{X}_n - t_\alpha\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + t_\alpha\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}\Big] \tag{36.138}$$

avec une probabilité $1 - \alpha$.

36.6.1 Région de confiance

Soit un n-échantillon $(X_1, \dots, X_n)$ de loi parente μ_θ. Nous supposons $\Theta \subset \mathbb{R}$ avec Θ ouvert. Soit $\alpha \in [0, 1]$ un intervalle de confiance et une application mesurable

$$\begin{aligned} \Lambda\colon \mathbb{R}^n &\to \mathcal{B}\mathrm{or}(\Theta) \\ (x_1, \dots, x_n) &\mapsto \Lambda(x_1, \dots, x_n). \end{aligned} \tag{36.139}$$

On appelle **région de confiance exact** au niveau de confiance $1 - \alpha$ une région aléatoire $\Lambda(x_1, \dots, x_n)$ telle que

$$P\big(\theta \in \Lambda(x_1, \dots, x_n)\big) = 1 - \alpha. \tag{36.140}$$

Si $d = 1$, la région $\Lambda(x_1, \dots, x_n)$ est un intervalle.

36.6.2 Fonction pivotale

Soit $\hat{\theta}_n$, un estimateur de θ. Une fonction v sur $\Theta \times \Theta$ est **pivotale** pour θ si la loi de la variable aléatoire $v(\hat{\theta}_n, \theta)$ ne dépend pas de θ. Elle est **asymptotiquement pivotale** si

$$v(\hat{\theta}_n, \theta) \xrightarrow{\mathscr{L}} \xi \tag{36.141}$$

où ξ est une variable aléatoire indépendante de θ.

En pratique, nous essayons de faire apparaitre une variable aléatoire de loi connue qui ne dépend pas du paramètre que l'on recherche. Si la variance est connue et si l'échantillon est grand, le théorème central limite nous permet d'introduire une loi normale centrée réduite.

Exemple 36.41.
Soit $X_1, \dots, X_n$ des variables aléatoires de loi parente $\mathscr{N}(m, \sigma^2)$. Une fonction asymptotiquement pivotale pour m est

$$v(z_1, z_2) = \frac{z_1 - z_2}{\sigma/\sqrt{n}} \tag{36.142}$$

parce que la variable aléatoire

$$v(\bar{X}_n, m) = \frac{\bar{X}_n - m}{\alpha/\sqrt{n}} \tag{36.143}$$

tend vers $\mathscr{N}(0, 1)$ qui ne dépend pas de m. △

Exemple 36.42.
Si (X_n) est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne m et d'écart type σ que nous supposons inconnus. Le fonction suivante est asymptotiquement pivotale pour m :

$$v(\bar{X}_n, m) = \frac{\bar{X}_n - m}{\sigma/\sqrt{n}}. \tag{36.144}$$

△

Soit $(X_1, \dots, X_n)$ un échantillon de loi $\mathscr{N}(m, \sigma_0^2)$ avec σ_0^2 connu. Nous cherchons un intervalle de confiance $1 - \alpha$ pour m. Pour cela nous allons utiliser une fonction asymptotiquement pivotale, à savoir

$$\frac{\bar{X}_n - m}{\sigma_0^2/\sqrt{n}} \sim \mathscr{N}(0, 1). \tag{36.145}$$

Nous devrions chercher des valeurs z_+ de z_- telles que

$$P\left(z_- \leqslant \frac{\bar{X}_n - m}{\sigma_0/\sqrt{n}} \leqslant z_+\right) = 1 - \alpha. \tag{36.146}$$

Pour des raisons de symétries (de la courbe gaussienne), nous allons chercher un intervalle symétrique : $z_- = -z_+$. Le nombre à chercher est donc le z_α tel que

$$P\big(|Z| \leqslant z_\alpha\big) = 1 - \alpha. \tag{36.147}$$

Si nous demandons $\alpha = 5\%$, la réponse est $z_\alpha = 1.96$, c'est-à-dire que

$$P\left(-1.96 \leqslant \frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \leqslant 1.96\right) = 0.95. \tag{36.148}$$

Nous avons donc

$$P\left(m \in \big[\bar{X}_n - \frac{1.96\sigma}{\sqrt{n}}, \bar{X}_n + \frac{1.96\sigma}{\sqrt{n}}\big]\right) = 0.95. \tag{36.149}$$

Supposons maintenant que nous avons observé 100 valeurs numériques avec $\bar{x}_n = 12$ et $\sigma = 1$. La réalisation de l'intervalle de confiance pour m au niveau de confiance 0.95 est :

$$\big[12 - 0.196, 12 + 0.196\big]. \tag{36.150}$$

Cet intervalle est à interpréter de la façon suivante : si nous recommençons un grand nombre de fois le sondage, la moyenne tombera 95% des fois dans l'intervalle ainsi calculé. Mais il faut bien comprendre que la probabilité

$$P\left(m \in \big[12 - 0.196, 12 + 0.196\big]\right) \tag{36.151}$$

vaut zéro ou un.

Exemple 36.43.
Soit un échantillon $(X_1, \dots, X_n)$ de loi parente $\mathscr{N}(m, \sigma^2)$ où m et σ^2 sont inconnus. Déterminer un intervalle de confiance exact symétrique au niveau de confiance $1 - \alpha$.

Déterminer un intervalle de confiance pour σ^2.

Nous savons que la moyenne empirique est un estimateur de la moyenne :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{36.152}$$

Nous cherchons un intervalle du type $I = [\bar{X}_n - \epsilon, \bar{X}_n + \epsilon]$ pour lequel $P(m \in I) = 1 - \alpha$. Nous savons que la variable aléatoire

$$\frac{\bar{X}_n - m}{\alpha/\sqrt{n}} \tag{36.153}$$

suit une loi $\mathscr{N}(0, 1)$, mais la variance est inconnue. La subtilité à savoir est que la variable aléatoire

$$Z = \frac{\bar{X}_n - m}{S_n/\sqrt{\sigma}} \tag{36.154}$$

où $S_n^2 = \sum_i (X_i - \bar{X}_n)^2/(n-1)$ suit une loi de Student à n degrés de liberté $\mathscr{T}(n-1)$ en vertu du théorème de Cochran 36.15. Comme il est usuel de le faire, nous inversons l'intervalle :

$$1 - \alpha = P\left(-\epsilon \leqslant \bar{X}_n - m \leqslant \epsilon\right) \tag{36.155a}$$

$$= P\left(-\frac{\epsilon\sqrt{n}}{S_n} \leqslant Z \leqslant \frac{\epsilon\sqrt{n}}{S_n}\right). \tag{36.155b}$$

Les valeurs se trouvent dans des tables ; par exemple pour $n = 10$ et $\alpha = 5\%$ nous trouvons

$$\frac{\epsilon\sqrt{n}}{S_n} = 2.262. \tag{36.156}$$

Plus généralement nous notons $t_{n-1,1-\frac{\alpha}{2}}$ le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi $\mathscr{T}(n-1)$, c'est-à-dire le nombre tel que

$$P(Z \leqslant t_{n-1,1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} \tag{36.157}$$

si $Z \sim \mathscr{T}(n-1)$. L'intervalle de confiance est alors donné par

$$I = \left[\bar{X}_n - \frac{t_{n-1,1-\frac{\alpha}{2}} S_n}{\sqrt{n}}, \bar{X}_n + \frac{t_{n-1,1-\frac{\alpha}{2}} S_n}{\sqrt{n}}\right]. \tag{36.158}$$

Cela est un intervalle exact pour m au niveau de confiance $1 - \alpha$.

Nous pouvons aussi trouver un intervalle asymptotique en utilisant le théorème central limite :

$$\frac{\bar{X}_n - m}{S_n/\sqrt{n}} \xrightarrow{\mathscr{L}} T \tag{36.159}$$

avec $T \sim \mathscr{N}(0, 1)$.

Rappel : dire que I_n est un **intervalle de confiance asymptotique** signifie que

$$\lim_{n \to \infty} P(m \in I_n) = 1 - \alpha. \tag{36.160}$$

En ce qui concerne la variance σ^2, l'intervalle de confiance se construit en utilisant la partie (2) du théorème de Cochran 36.15. Nous introduisons la variable aléatoire pivot

$$Z = (n-1)\frac{S_n^2}{\sigma^2} \tag{36.161}$$

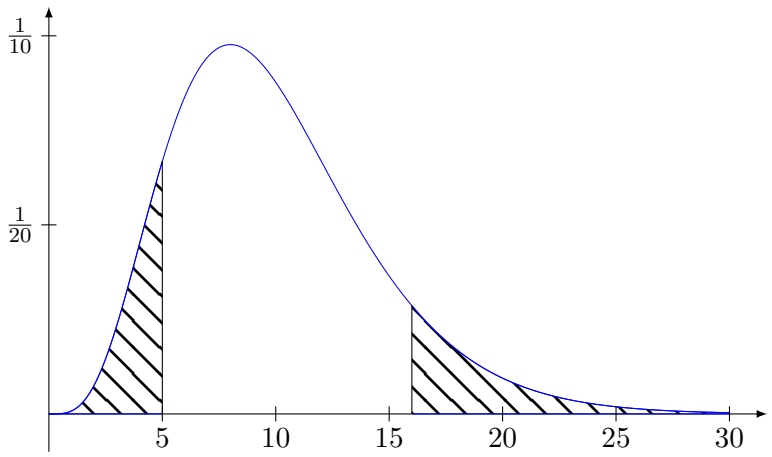


FIGURE 36.3 – L'intervalle de confiance pour une χ^2.

qui suit une loi $\chi^2(n-1)$. Cette loi n'étant pas symétrique (voir figure 35.1), nous n'allons pas chercher un intervalle de confiance symétrique. Nous cherchons c_1 et c_2 tels que

$$\begin{cases} P\big(\sigma^2 \in [c_1, c_2]\big) = 1-\alpha & (36.162a) \\ P\big(\sigma^2 \in [0, c_1]\big) = \dfrac{\alpha}{2} & (36.162b) \\ P\big(\sigma^2 \in [c_2, \infty[\big) = \dfrac{\alpha}{2} & (36.162c) \end{cases}$$

La situation est représentée à la figure 36.3.

Le construction des nombres c_1 et c_2 passe par la relation

$$P(c_1 \leqslant \sigma^2 \leqslant c_2) = P\left(\frac{(n-1)S_n^2}{c_2} \leqslant Z \leqslant \frac{(n-1)S_n^2}{c_1}\right). \tag{36.163}$$

Nous notons $t_{n-1,\frac{\alpha}{2}}$ et $t_{n-1,1-\frac{\alpha}{2}}$ les quantiles donnés sur la figure 36.3, c'est-à-dire

$$P(Z \leqslant t_{n-1,\frac{\alpha}{2}}) = \frac{\alpha}{2} P(Z \geqslant t_{n-1,1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}. \tag{36.164a}$$

Ce que nous obtenons est

$$\frac{(n-1)S_n^2}{c_2} = t_{n-1,\frac{\alpha}{2}} \tag{36.165a}$$

$$\frac{(n-1)S_n^2}{c_1} = t_{n-1,1-\frac{\alpha}{2}}, \tag{36.165b}$$

et par conséquent l'intervalle de confiance pour σ^2 est

$$I = \left[\frac{(n-1)S_n^2}{t_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S_n^2}{t_{n-1,1-\frac{\alpha}{2}}}\right] \tag{36.166}$$

avec $P(\sigma^2 \in I) = 95\%$.

Remarque 36.44.
Il n'est pas clair à priori que la longueur de l'intervalle I décroisse avec n parce qu'il y a n dans les t au numérateur.

△

36.6.3 Sondage de proportion

Une utilisation classique des statistiques est d'interpréter une proportion donnée par un sondage. Nous considérons une élection avec deux candidats A et B. Nous avons interrogés $n = 2500$ personnes et nous avons obtenus 51% pour le candidat A et 49% pour le candidat B. Que peut on dire ?

La modélisation de cette situation est que nous avons des variables aléatoires $X_i \sim \mathscr{B}(p_A)$ et que nous en avons observés n avec une moyenne

$$\bar{x}_n = 0.51. \tag{36.167}$$

La loi de $\bar{X}_n$ est une binomiale. Sa densité n'est pas symétrique, mais si n est grand, elle le devient. Nous cherchons un intervalle

$$I = [\bar{X}_n - \epsilon, \bar{X}_n + \epsilon] \tag{36.168}$$

tel que $P(p_A \in I) = 1 - \alpha$. Pour cela nous considérons le fait que $n = 2500$ est grand et nous utilisons la limite de la proposition 36.10 :

$$Z_n = \sqrt{n}\frac{\bar{X}_n - p_A}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \xrightarrow{\mathscr{L}} Z \sim \mathscr{N}(0,1). \tag{36.169}$$

La variable aléatoire Z_n est asymptotiquement pivotale et normale centrée réduite. Nous cherchons donc un intervalle symétrique pour $\bar{X}_n - p_A$:

$$1 - \alpha = P(-\epsilon \leqslant \bar{X}_n - p_A \leqslant \epsilon), \tag{36.170}$$

c'est-à-dire, si n est grand,

$$1 - \alpha = P\left(-\epsilon\frac{\sqrt{n}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} \leqslant Z_n \leqslant \epsilon\frac{\sqrt{n}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}}\right) \tag{36.171}$$

où Z_n est une normale centrée réduite. Nous trouvons ainsi, via les tables que

$$\frac{\epsilon\sqrt{n}}{\sqrt{\bar{X}_n(1-\bar{X}_n)}} = 1.96 \tag{36.172}$$

si nous voulons un intervalle à 5%. Par conséquent nous avons $\epsilon = 1.96\frac{\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{nj}}$ et l'intervalle de confiance est

$$I_C = \left[\bar{X}_n - \frac{1.96\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}, \bar{X}_n + \frac{2.96\sqrt{\bar{X}_n(1-\bar{X}_n)}}{\sqrt{n}}\right]. \tag{36.173}$$

La propriété de cet intervalle est que

$$\lim_{n\to\infty} P(p_A \in I_c) = 1 - \alpha. \tag{36.174}$$

Remarque 36.45.
À quel moment avons nous fait une hypothèse sur la taille de la population globale ? En modélisant les sondés par des variables de Bernoulli et leur somme par une binomiale, nous supposons que le sondage est *avec remise*, sinon, elles ne seraient pas indépendantes. En supposant les sondés indépendants, nous avons donc fait comme si la population totale était infinie.

36.7 Estimer une densité lorsqu'on ne sait rien

Nous supposons avoir une série d'observations issues d'un processus complexe dont nous n'avons aucune idée de la loi parente, et nous voudrions nous faire une idée de la densité de cette loi inconnue.

Nous observons une suite de réalisations que nous modélisons comme étant des variables aléatoires $(X_1, \dots, X_n)$ de loi parente (inconnue) μ. Notre but est de trouver un estimateur $\hat{\mu}$ de μ. Par simplicité nous allons nous restreindre aux lois admettant une densité par rapport à la mesure de Lebesgue. C'est-à-dire que nous allons estimer μ par une suite de lois $\hat{\mu}_n$ qui sont toutes des lois acceptant une densité par rapport à la mesure de Lebesgue.

36.7.1 Distance entre des mesures

Si ν_1 et ν_2 sont deux mesures de densité sur $\mathbb{R}$, la **distance** entre ν_1 et ν_2 est définie par

$$d(\nu_1, \nu_2) = \sup_{A \in \mathcal{B}\mathrm{or}(\mathbb{R})} \big|\nu_1(A) - \nu_2(A)\big| \tag{36.175}$$

où $\mathcal{B}\mathrm{or}(\mathbb{R})$ désigne l'ensemble des boréliens de $\mathbb{R}$.

Théorème 36.46 (de Scheffé[535]).
Si f_1 et f_2 sont les densités de ν_1 et ν_2 par rapport à la mesure de Lebesgue, alors

$$d(\nu_1, \nu_2) = \int_{\mathbb{R}} (f_1 - f_2)_+ = \frac{1}{2} \int_{\mathbb{R}} |f_1 - f_2| = \frac{1}{2} \|f_1 - f_2\|_1 \tag{36.176}$$

où f_+ est la partie positive de f (pour la décomposition $f(x) = f_+(x) - f_-(x)$).

Démonstration. La dernière égalité est simplement une notation usuelle ; nous devons seulement prouver les deux premières. Pour la première nous commençons par prouver que le borélien réalisant le supremum est

$$B = \{x \in \mathbb{R} \text{ tel que } f_1(x) \geqslant f_2(x)\}. \tag{36.177}$$

En effet si A est un borélien nous avons

$$\nu_1(A) - \nu_2(A) = \int_A f_1 - f_2 \leqslant \int_{A \cap B} f_1 - f_2 \leqslant \int_B f_1 - f_2 = \int_B (f_2 - f_2)_+ = \int_{\mathbb{R}} (f_1 - f_2)_+ \tag{36.178}$$

Justifications :

— $f_1 - f_2$ négative sur $A \cap \complement B$.

— Vu que $f_1 - f_2 \geqslant 0$ sur B, l'intégrale augmente si on augmente le domaine.

— Sur B nous avons $f_1 - f_2 = (f_1 - f_2)_+$.

Donc pour tout borélien A nous avons

$$d(\nu_1, \nu_2) \leqslant \int_{\mathbb{R}} (f_1 - f_2)_+. \tag{36.179}$$

Mais pour $A = B$ nous avons égalité :

$$\nu_1(B) - \nu_2(B) = \int_B f_1 - f_2 = \int_B (f_1 - f_2)_+ = \int_{\mathbb{R}} (f_1 - f_2). \tag{36.180}$$

Pour la seconde égalité nous savons que f_1 et f_2 s'intègrent toutes deux à 1 (parce que ce sont des densités de probabilité), donc

$$\int_{\mathbb{R}} f_1 - f_2 = 0. \tag{36.181}$$

En particulier nous avons

$$\int_{\mathbb{R}} (f_1 - f_2)_+ = \int_{\mathbb{R}} (f_1 - f_2)_-, \tag{36.182}$$

ce qui donne bien

$$\int_{\mathbb{R}} (f_1 - f_2)_+ = \frac{1}{2} \int_{\mathbb{R}} |f_1 - f_2|. \tag{36.183}$$

□

36.7.2 Estimateur par fenêtres glissantes

Nous considérons les estimations suivantes de la fonction de répartition :

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \leqslant x\}}, \tag{36.184}$$

et un nombre h_n qui sera la taille de la fenêtre glissante. Nous avons en tête de faire $\lim_{n\to\infty} h_n = 0$. Nous considérons ceci comme estimateur de la densité inconnue f des variables aléatoires X_i :

$$\hat{f}_n(x) = \frac{F_n(x+h_n) - F_n(x-h_n)}{2h_n}. \tag{36.185}$$

L'idée sous-jacente est de prendre la dérivée de la fonction de répartition comme densité.

Lemme 36.47 ([535]).
Pour tout $h_n > 0$, l'estimateur $\hat{f}_n$ est une densité de probabilité.

Démonstration. D'abord $\hat{f}_n$ est bien à valeurs positives ou nulle. Ensuite devons parler de son intégrale. Pour le numérateur nous avons

$$F_n(x+h_n) - F_n(x-h_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \in B(x,h_n)}. \tag{36.186}$$

En réalité cette égalité est valable seulement presque partout parce qu'elle n'est pas valable en $x = x \pm h_n$, mais cela ne va pas nous ennuyer dans la mesure où nous avons dans l'idée d'intégrer cela sur $\mathbb{R}$. Avant de nous lancer dans l'intégrale nous remarquons que $X_i \in B(x,h_n)$ est la même chose que $x \in B(X_i,h_n)$, c'est-à-dire que

$$\{X_i \in B(x,h_n)\} = \{x \in B(X_i,h_n)\}. \tag{36.187}$$

Donc

$$\int_{\mathbb{R}} \hat{f}_n = \frac{1}{2h_n}\frac{1}{n}\sum_{i=1}^{n} \underbrace{\int_{\mathbb{R}} \mathbb{1}_{B(X_i,h_n)}}_{2h_n} = \frac{1}{2nh_n}\sum_{i=1}^{n} 2h_n = 1. \tag{36.188}$$

□

Lemme 36.48 ([535]).
L'estimateur $\hat{f}_n$ est déjà pas mal parce que

$$\lim_{h_n\to 0} E\big(\hat{f}_n(x)\big) \to f(x) \tag{36.189}$$

pour presque tout $x \in \mathbb{R}$.

Démonstration. Nous commençons par nous rappeler le fait que $F_n(x)$ est un estimateur sans biais de $F(x)$ (proposition 36.35). Donc

$$E\big(\hat{f}_n(x)\big) = \frac{F(x+h_n) - F(x-h_n)}{2h_n}. \tag{36.190}$$

Nous devons prendre la limite de cela lorsque $h_n \to 0$, c'est-à-dire considérer la dérivée de F. Attention : rien ne dit que F soit dérivable, si ce n'est la proposition 16.24 qui indique qu'elle est dérivable presque partout avec f comme dérivée.

La limite $h_n \to 0$ dans (36.190) nous donne donc bien presque partout

$$\lim_{h_n\to 0} E\big(\hat{f}_n(x)\big) = f(x). \tag{36.191}$$

□

Proposition 36.49 ([535]).
Si la suite (h_n) est telle que $h_n \to 0$ et $nh_n \to \infty$, alors pour presque tout $x \in \mathbb{R}$ nous avons les convergences

$$\hat{f}_n(x) \overset{L^2(P)}{\to} f(x) \tag{36.192}$$

et

$$\hat{f}_n(x) \overset{P}{\to} f(x). \tag{36.193}$$

Démonstration. Il faut d'abord comprendre ce que signifie la convergence $L^2(P)$ pour presque tout x. Pour cela il faut comprendre que $\hat{f}_n(x)$ est en soi une variable aléatoire et est en réalité une fonction $\omega \mapsto \hat{f}_n(x,\omega)$. Ce que nous allons montrer est que pour presque tout x (maintenant fixé), cette variable aléatoire converge vers une constante (par rapport à ω) et que cette constante est $f(x)$.

La convergence $X_n \overset{L^2(P)}{\to} X$ signifie $E\big(|X_n - X|^2\big) \to 0$, c'est-à-dire

$$\int_\Omega \big|X_n(\omega) - X(\omega)\big|^2 dP(\omega) \to 0. \tag{36.194}$$

En faisant une décomposition biais-variance nous devons donc étudier

$$E\Big[\big(\hat{f}_n(x) - f(x)\big)^2\Big] = E\big[\hat{f}_n(x) - f(x)\big]^2 + \operatorname{Var}\big(\hat{f}_n(x) - f(x)\big) \tag{36.195}$$

Ici $f(x)$ doit être vue comme la variable aléatoire constante sur Ω. Par le lemme 36.48 et la proposition 35.29 le terme de biais converge vers zéro lorsque $n \to \infty$.

Pour traiter le terme de biais, nous savons déjà que

$$2nh_n\hat{f}_n(x) = \sum_{i=1}^n \mathbb{1}_{\{X_i \in B(x,h_n)\}}, \tag{36.196}$$

où le membre de droite (et donc aussi celui de gauche) est est une variable aléatoire binomiale comptant le nombre de succès de l'expérience $X_i \in B(x,h_n)$ en n essais. Nous notons $p_{x,n} = P\big(X_i \in B(x,h_n)\big)$. Si μ est la loi parente des X_i, alors

$$p_{n,x} = P\big(X_i \in B(x,h_n)\big) = \mu\big(B(x,h_n)\big) = F(x+h_n) - F(x-h_n) \tag{36.197}$$

où F est la fonction de répartition (parente) des X_i.

Alors la variance de ladite binomiale est donnée par (35.348), c'est-à-dire $np_{x,n}(1-p_{x,n})$. Nous avons alors

$$\operatorname{Var}\big(2nh_n\hat{f}_n(x)\big) = bp_{x,n}(1-p_{x,n}) \tag{36.198}$$

et

$$\operatorname{Var}\big(\hat{f}_n(x)\big) = \frac{1}{4n^2h_n^2}np_{x,n}(1-p_{x,n}). \tag{36.199}$$

Nous pouvons faire la majoration $t(1-t) \leqslant t$ qui est valable pour tout t et écrire

$$\operatorname{Var}(\hat{f}_n(x)) \leqslant \frac{1}{4nh_n}\frac{p_{x,n}}{h_n}. \tag{36.200}$$

Le premier facteur tend vers zéro parce que nous avons supposé que $nh_n \to \infty$. Pour le second facteur, il faut remarquer que l'expression (36.197) nous donne presque partout

$$\lim_{h_n \to 0} \frac{p_{n,x}}{h_n} = 2f(x), \tag{36.201}$$

qui est constant et certainement borné.

Nous avons maintenant prouvé que pour presque tout x nous avons $\hat{f}_n(x) \overset{L^2(P)}{\to} f(x)$. Montrons que cela implique la convergence en loi, c'est-à-dire que pour tout $\eta > 0$, nous avons la limite

$$P\big(|\hat{f}_n(x) - f(x)| > \eta\big) \to 0. \tag{36.202}$$

Si cela n'était pas vrai, nous aurions un nombre $\eta_0 > 0$ et $\epsilon > 0$ tel que pour tout n à partir d'une certaine taille,

$$P\Big(|\hat{f}_n(x) - f(x)|^2 > \eta_0\Big) > \epsilon, \tag{36.203}$$

et en particulier en notant A l'événement $|\hat{f}_n(x) - f(x)|^2 > \eta_0^2$, $P(A) > \epsilon$. Alors

$$\int_\Omega \big|\hat{f}_n(x,\omega) - f(x,\omega)\big|^2 dP(\omega) \geqslant \int_A \big|\hat{f}_n(x,\omega) - f(x,\omega)\big|^2 dP(\omega) \geqslant \int_A \eta_0^2 = \eta_0^2 P(A). \tag{36.204}$$

Cela signifie que

$$\|\hat{f}_n(x) - f(x)\|_{L^2(P)} \geqslant \eta_0 P(A), \tag{36.205}$$

ce qui contredit la première convergence démontrée. □

Note : l'hypothèse $nh_n \to \infty$ revient à dire que nous voulons que chaque boîte contienne de plus en plus d'observations. Si nous avions $nh_n \to 0$, alors avec n qui augmente, la majorité des boîtes deviendraient vides, ce qui reviendrait à une perte d'information.

36.8 Test d'hypothèses, prise de décision

36.8.1 Exemple : qualité des pièces d'usine

Une usine fabrique des composantes électronique garantis un an. Le constructeur ne veut pas accepter que plus de 5% des pièces tombent en panne avant un an.

Nous supposons que la durée de vie T d'une pièce soit une variable aléatoire suivant une loi exponentielle de paramètre λ (qui est l'inverse de la moyenne : $\theta = 1/\lambda$). Le fabriquant veut donc s'assurer que

$$0.95 \leqslant P(T \geqslant 1), \tag{36.206}$$

ou encore

$$P(T \geqslant 1) = \int_0^\infty \frac{1}{\theta} e^{-x/\theta} dx = e^{-1/\theta} \geqslant 0.95, \tag{36.207}$$

donc le fabriquant doit s'assurer que

$$\theta \geqslant \frac{1}{\ln\left(\frac{1}{0.95}\right)}. \tag{36.208}$$

Nous posons donc $\theta_0 = 19.5$ et nous prenons comme modèle de décision que si $\theta < \theta_0$, alors la chaine de production doit être revue, et si $\theta > \theta_0$, alors l'usine peut continuer son travail.

Ce dont nous disposons n'est pas de θ, mais d'une estimation de θ à partir d'un échantillon. Cela étant il faudra aussi pouvoir estimer la probabilité de faire un mauvais choix.

36.8.2 Exemple : la résistance d'un fil

Un artisan a besoin d'un fil qui a une résistance à une traction de 100 g en moyenne. Si la résistance est trop faible, le fil casse ; si elle est trop grande, c'est trop rigide et ça ne convient pas.

Remarque 36.50.
Dans l'exemple précédent, avoir $\theta > \theta_0$ ne dérange pas. Si la durée de vie moyenne est de 2 ans, le directeur de l'usine ne sera pas malheureux. Ici par contre l'artisan cherche une valeur précise et a donc une borne vers le haut et vers le bas.

L'artisan reçoit un lot de fils et souhaite savoir s'il est conforme. Pour cela, il prend 4 fils au hasard et mesure une moyenne de 112 g. Est-ce que cela est cohérent avec une moyenne de 100 g ?

Nous faisons l'hypothèse que ma résistance des fils suit une loi normale $\mathcal{N}(m, \sigma^2)$ avec m inconnu. Pour la simplicité nous supposons que σ est connu et vaut 10.

Nous devons prendre une décision entre deux hypothèses. L'hypothèse H_0 sera de dire que le lot a une résistance de 100 g et l'hypothèse alternative sera que le lot a une résistance différente.

Les 4 observations sont quatre variables aléatoires $(X_i)_{i=1,2,3,4}$, et le nombre 112 est une réalisation de la variable aléatoire

$$\bar{X}_4 = \frac{1}{4}(X_1 + X_2 + X_3 + X_4). \tag{36.209}$$

Nous supposons que H_0 est vraie, et nous calculons quelle est l'intervalle autour de $m = 100$ qui a 95% de chances de contenir la moyenne observée. Si 112 est dedans, nous acceptons H_0 et si 112 est hors de cet intervalle, nous refusons H_0.

Compte tenue de l'hypothèse H_0, nous avons

$$\frac{\bar{X}_4 - 100}{\frac{10}{\sqrt{4}}} = \frac{\bar{X}_4 - 100}{5} \sim \mathscr{N}(0,1). \tag{36.210}$$

Nous commençons à connaitre par cœur l'intervalle de confiance à 95% d'une loi normale centrée réduite ; le quantile est à 1.96, c'est-à-dire

$$P\left(-1.96 \leqslant \frac{\bar{X}_4 - 100}{5} \leqslant 1.96\right) = 0.95, \tag{36.211}$$

ou encore

$$P\big(\bar{X}_4 \in [90.2, 109.8]\big). \tag{36.212}$$

Il y a donc moins de 5% de chances que la moyenne de ces quatre fils tombent en dehors de l'intervalle $[90.2, 109.8]$. L'artisan doit donc rejeter l'hypothèse et considérer que le lot est mauvais.

La région

$$]-\infty, 90.2] \cup [109.8, \infty[\tag{36.213}$$

est la **région de rejet**, ou **région critique**.

Ici, le nombre 5% représente le risque de refuser H_0 alors qu'elle était vraie. Notons que nous ne pouvons pas calculer le risque d'accepter H_0 alors qu'elle est fausse. En effet, si H_0 est fausse, nous ne savons pas quelles sont les valeurs de $\bar{X}_4$ acceptables parce qu'il y a une infinité de possibilités pour m qui soient alternatifs à $m = 100$.

Évidemment si la vraie moyenne est $100+10^{-7}$, l'hypothèse H_0 sera acceptée, mais nous n'avons aucun moyen de savoir si elle est vraie ou non.

36.8.3 Vocabulaire et théorie

Nous avons un modèle d'échantillonnage paramétrique $(X_{\theta,1}, \dots, X_{\theta,n})$ de taille n et de paramètre inconnu θ, de loi parente μ_θ appartenant à une famille paramétrique de lois $(\mu_\theta)_{\theta\in\Theta}$.

Soient H_0 et H_1 deux ensembles disjoints tels que $\Theta = H_0 \cup H_1$. L'ensemble H_0 sera nommé **hypothèse nulle** et l'ensemble H_1 sera l'**hypothèse alternative**.

Pour l'exemple des fils, nous avions $H_0 = \{100\}$ et $H_1 = \mathbb{R}\backslash\{100\}$. Si une hypothèse est réduite à un singleton, nous parlons d'hypothèse **simple** et sinon c'est une hypothèse **composite** ou **multiple**. Faire un tests consiste à déterminer une région critique.

Définition 36.51.
*Un **test** est une application mesurable δ qui à $(x_1, \dots, x_n) \in \mathbb{R}^n$ associe*

$$\delta(x_1, \dots, x_n) \in \{0, 1\}. \tag{36.214}$$

Si $\delta(x_1, \dots, x_n) = 0$ on accepte l'hypothèse H_0 pour l'échantillon $x_1, \dots, x_n$, et si $\delta(x_1, \dots, x_n) = 1$, alors on rejette H_0 et on choisit H_1. L'ensemble

$$W = \{(x_1, \dots, x_n) \in \mathbb{R}^n \text{ tel que } \delta(x_1, \dots, x_n) = 1\} \tag{36.215}$$

*est la **région de rejet** ou la **région critique**.*

L'ensemble $W = \delta^{-1}(1)$ est un borélien de $\mathbb{R}^n$ parce que δ est mesurable. L'événement auquel nous sommes intéressés est l'événement

$$R = \{(X_{\theta,1}, \dots, X_{\theta,n}) \in W\}. \tag{36.216}$$

Exemple 36.52.
Pour l'exemple de 36.8.2 nous avions

$$\delta(x_1, \dots, x_4) = \mathbb{1}_{\complement[90.2,109.8]}(\bar{x}_4). \tag{36.217}$$

△

36.8.4 Risque de première et seconde espèce

Le modèle de décision que nous avons introduit comprend deux façons de se tromper. Soit nous rejetons H_0 alors qu'elle est vraie (c'est le **risque de première espèce**), soit nous acceptions H_0 alors qu'elle est fausse (risque de **seconde espèce**). Nous pouvons formaliser ces concepts de la façon suivante.

Nous considérons un test de région critique W. Le risque de première espèce, noté α est la fonction

$$\begin{aligned} \alpha\colon H_0 &\to [0,1] \\ \theta &\mapsto P\big((X_{\theta,1}, \dots, X_{\theta,n}) \in W\big). \end{aligned} \tag{36.218}$$

Il s'agit de la probabilité de rejeter H_0 alors qu'elle es vraie. Le risque de seconde espèce est la fonction

$$\begin{aligned} \beta\colon H_1 &\to [0,1] \\ \theta &\mapsto P\big((X_{\theta,1}, \dots, X_{\theta,n}) \notin W\big). \end{aligned} \tag{36.219}$$

C'est la probabilité d'accepter H_0 alors qu'elle est fausse.

Définition 36.53.
*Soit δ un test de région critique W. La **puissance** du test est la fonction*

$$\begin{aligned} \eta\colon H_1 &\to [0,1] \\ \theta &\mapsto P\big((X_{\theta,1}, \dots, X_{\theta,n}) \in W\big). \end{aligned} \tag{36.220}$$

*La **courbe d'efficacité** du test est la fonction*

$$\begin{aligned} h\colon \Theta &\to [0,1] \\ \theta &\mapsto P\big((X_{\theta,1}, \dots, X_{\theta,n}) \notin W\big). \end{aligned} \tag{36.221}$$

La puissance d'un test est la probabilité de rejeter H_0 lorsque H_1 est vraie. Plus la puissance est grande, mieux c'est. La courbe d'efficacité du test est la probabilité d'accepter H_0 pour une certaine valeur de θ.

Soit un test δ. Une statistique $T_\theta = T_n(X_{\theta,1}, \dots, X_{\theta,n})$ est une **variable de décision** pour δ si $\complement W$ peut s'écrire d'une des façons suivantes

$$\complement W = \begin{cases} \{(x_1, \dots, x_n) \in \mathbb{R}^n \text{ tel que } T_n(x_1, \dots, x_n) < c\} & \text{test unilatéral à droite} \\ \{(x_1, \dots, x_n) \in \mathbb{R}^n \text{ tel que } T_n(x_1, \dots, x_n) > c\} & \text{test unilatéral à gauche} \\ \{(x_1, \dots, x_n) \in \mathbb{R}^n \text{ tel que } c_1 \leqslant T_n(x_1, \dots, x_n) < c_2\} & \text{test bilatéral.} \end{cases} \tag{36.222}$$

Le plus souvent la variable de décision sera la moyenne : $T_n(x_1, \dots, x_n) = \frac{1}{n}\sum_{i=1}^n x_i$. Les valeurs c, c_1, c_2 sont des **valeurs critiques**.

En ce qui concerne les notations, ici T_n représente la valeur mesurée sur un n-échantillon (d'où l'indice n) alors que T_θ est la valeur *théorique* de T lorsque θ est la vraie valeur du paramètre qu'on veut estimer.

Pour un test unilatéral à gauche, nous fixons la valeur critique c de telle manière a avoir

$$P(T_\theta > c) \leqslant \alpha. \tag{36.223}$$

Pour un test unilatéral à gauche, nous fixons c de telle manière a avoir

$$P(T_\theta < c) \leqslant \alpha \tag{36.224}$$

et pour un test bilatéral nous fixons c_1 et c_2 de telle façon à avoir

$$P(T_\theta > c_2) = P(T_\theta < c_1) \leqslant \frac{\alpha}{2}. \tag{36.225}$$

36.8.5 Modèle paramétrique de loi gaussienne

Soit un modèle statistique paramétrique de lois gaussiennes $\mathscr{N}(m,1)$ de moyenne m inconnue avec $m \in \mathbb{R}^+$. Nous avons $\Theta = [0,\infty[$.

Nous observons un échantillon de taille $n = 36$. Avec un risque de première espèce de 5% nous voulons estimer l'hypothèse $H_0 = \{0\}$ contre l'hypothèse $H_1 =]0,\infty[$. De notre échantillon nous construisons la variable aléatoire

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{36.226}$$

dans laquelle les X_i sont les éléments de l'échantillon, elles sont indépendantes et identiquement distribuées de loi $\mathscr{N}(m,1)$ avec m inconnu.

Si $\bar{X}_n$ est proche de zéro nous acceptons H_0, sinon nous la rejetons. La région de rejet s'écrit donc sous la forme

$$W = \{(x_1,\dots,x_n) \in \mathbb{R}^n \text{ tel que } \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i > u\} \tag{36.227}$$

dans lequel il faut fixer le u pour satisfaire au risque de première espèce de 5%. La contrainte est d'avoir

$$P(\bar{X}_n > u) = \alpha \tag{36.228}$$

si H_0 est vérifiée. Cela revient à dire que dans 5% des cas où H_0 est correcte, nous la rejetterons. Si H_0 est vraie alors $\bar{X}_n$ est une moyenne de gaussiennes de moyennes m et nous avons

$$\frac{\bar{X}_n - m}{\sigma/\sqrt{n}} \sim \mathscr{N}(0,1) \tag{36.229}$$

avec $m = 0$ et $\sigma = 1$. L'équation (36.228) devient donc

$$\alpha = P\left(\frac{\bar{X}_n}{1/\sqrt{n}} > \frac{u}{1/\sqrt{n}}\right) = P(T > \sqrt{n}u) \tag{36.230}$$

où $T \sim \mathscr{N}(0,1)$. Avec $n = 36$ et $\alpha = 5\%$ nous trouvons

$$u = \frac{1.645}{6} \simeq 0.274 \tag{36.231}$$

La règle de décision est donc la suivante : si $\bar{x}_n > 0.274$ alors nous rejetons H_0, et sinon nous l'acceptons.

Calculons la puissance de ce test (définition 36.53). C'est la fonction donnée par

$$\begin{aligned} \eta\colon H_1 &\to \mathbb{R} \\ m &\mapsto P\big((X_{1,m},\dots,X_{n,m}) \in W\big) = P\left(\frac{1}{n}\sum X_i > u\right). \end{aligned} \tag{36.232}$$

Dans ce calcul, les X_i sont d'une loi normale $\mathscr{N}(m,1)$, et non $\mathscr{N}(0,1)$. En retranchant m et en divisant par $1/\sqrt{n}$ nous trouvons

$$\eta(m) = P\left(\frac{\frac{1}{n}\sum X_i - m}{1/\sqrt{t}} > \frac{u-m}{1/\sqrt{n}}\right) \tag{36.233a}$$
$$= P(T > \sqrt{n}(u-m)) \tag{36.233b}$$
$$= P(T > 16.45 - 6m) \tag{36.233c}$$
$$= 1 - \Phi(1.645 - 6m) \tag{36.233d}$$

où Φ est la fonction de répartition de $\mathscr{N}(0,1)$. La fonction η a les propriétés suivantes :

$$\lim_{m\to-\infty} \eta(m) = 0 \tag{36.234a}$$
$$\lim_{m\to\infty} \eta(m) = 1 \tag{36.234b}$$
$$\eta(0) = \frac{5}{100}. \tag{36.234c}$$

Remarque 36.54.
Si nous regardons $m = 0.001$, le risque de seconde espèce est quasiment de 90%. En effet le risque de seconde espèce est d'accepter H_0 alors qu'il est faux. Lorsque $m = 0.001$, l'hypothèse H_0 est fausse, mais la probabilité qu'on l'accepte est grande. D'ailleurs les conséquences de l'accepter à tort ne sont peut-être pas si grandes que cela.

36.9 Tests paramétriques

La proposition suivante montre le lien entre région de confiance et les tests.

Proposition 36.55.
Soit $\Lambda(X_1,\dots,X_n)$ une région de confiance par excès de niveau de confiance $1-\alpha$. Alors il existe un tests pur de niveau α pour tester $H_0 = \{\theta_0\}$ de région de rejet

$$W_n = \{x = (x_1,\dots,x_n) \in \mathbb{R}^n \text{ tel que } \theta_0 \notin \Lambda(x_1,\dots,x_n)\}. \tag{36.235}$$

Démonstration. L'hypothèse sur Λ signifie qu'avec les observations $(X_1,\dots,X_n)$, il y a une forte probabilité (plus grande que $1-\alpha$) que θ soit dans $\Lambda(X_1,\dots,X_n)$. Avec ou sans H_0 nous avons donc

$$P(\theta \in \Lambda) \geqslant 1-\alpha. \tag{36.236}$$

Supposons maintenant l'hypothèse H_0, alors

$$P\big((X_1,\dots,X_n) \in W_n\big) = P\big(\theta_0 \notin \Lambda(X_1,\dots,X_n)\big) \leqslant \alpha. \tag{36.237}$$

□

Remarque 36.56.
Soit W_n la région de confiance d'un test de niveau α pour tester $H_0 = \{\theta_0\}$. Alors

$$\Lambda = \{x \in \mathbb{R}^n \text{ tel que } x \notin W_n\} \tag{36.238}$$

est une région de confiance $1-\alpha$ pour θ.

Exemple 36.57.
Soit $(X_1,\dots,X_n)$ un échantillon de loi parente $\mathscr{N}(\theta,1)$ avec $\theta \in \{\theta_0,\theta_1\}$. Nous supposons $\theta_0 < \theta_1$. Nous voulons tester $H_0 = \{\theta_0\}$ contre $H_1 = \{\theta_1\}$. Nous proposons le test suivant. La variable de décision sera $\bar{X}_n$ et la région de rejet sera

$$W = \{(x_1,\dots,x_n) \in \mathbb{R}^n \text{ tel que } \frac{1}{n}\sum_i x_i > \frac{\theta_1+\theta_2}{2}\}. \tag{36.239}$$

(1) Donner le risque de première espèce de ce test.
(2) Soit $0 < \alpha < 1$. Pour quelle valeur de n le tests a-t-il un risque de première espèce égal à α ?
(3) Donner la puissance du test.

Les réponses peuvent être exprimées en termes de la fonction de répartition F de la loi normale centrée réduite.

Le risque de première espèce est donné par

$$\alpha = P\left(\frac{1}{n}\sum_i X_i > \frac{\theta_0 + \theta_1}{2}\right) \tag{36.240}$$

où les X_i sont des variables aléatoires indépendantes et identiquement distribuées de loi parente $\mathscr{N}(\theta_0, 1)$. Cela est la probabilité d'être dans la région de rejet alors que l'hypothèse H_0 est vraie. La formule (36.240) se transforme en

$$\alpha = P\left(\frac{\frac{1}{n}\sum X_i - \theta_0}{1/\sqrt{n}} > \frac{\frac{\theta_0+\theta_1}{2} - \theta_0}{1/\sqrt{n}}\right) \tag{36.241a}$$

$$= P(T > \sqrt{n}\frac{\theta_1 - \theta_0}{2}). \tag{36.241b}$$

En termes de la fonction de répartition nous avons alors

$$\alpha = 1 - F\big(\sqrt{n}\frac{\theta_1 - \theta_0}{2}\big) \tag{36.242}$$

Il s'agit maintenant de trouver le nombre n qui réalise cette égalité. Pour cela nous utilisons l'inverse F^{-1} de la fonction de répartition de la normale :

$$n = \left(\frac{2}{\theta_1 - \theta_0}F^{-1}(1-\alpha)\right)^2. \tag{36.243}$$

Le risque de seconde espèce est la possibilité d'accepter H_0 lorsque H_1 est vraie, c'est-à-dire

$$\beta = P\left(\frac{1}{n}\sum_i X_i < \frac{\theta_0 + \theta_1}{2}\right) \tag{36.244}$$

sous l'hypothèse H_1. Dans le calcul de (36.244) nous prenons donc $X_i \sim \mathscr{N}(\theta_1, 1)$. Le résultat est que

$$\beta = F\big(\sqrt{n}\frac{\theta_0 - \theta_1}{2}\big). \tag{36.245}$$

Remarque 36.58.
L'expression (36.243) diminue lorsque θ_0 et θ_1 s'éloignent, ce qui est normal : plus les nombres à discerner sont éloignés, moins l'échantillon à prendre pour réaliser le travail doit être grand.

Notons aussi que $\theta_0 - \theta_1 < 0$, par conséquent augmenter n diminue la valeur de

$$\beta = F\big(\sqrt{n}\frac{\theta_0 - \theta_1}{2}\big) \tag{36.246}$$

△

<++>

36.10 Tests d'adéquation

Soit $X_1, \dots, X_n$ un échantillon de loi parente X finie prenant ses valeurs dans $\{a_1, \dots, a_k\}$. La loi de X est donnée par les nombres

$$p_i = P(X = a_i) \tag{36.247}$$

pour $i = 1, \dots, k$. Nous introduisons l'**effectif empirique**, la variable aléatoire N_i qui compte le nombre de fois que a_i est observé dans l'échantillon. La **fréquence empirique** est la variable aléatoire

$$F_i = \frac{N_i}{n}. \tag{36.248}$$

Nous savons que la loi de N_i est $\mathscr{B}(n, p_i)$, et la loi des grands nombres dit que

$$F_i \xrightarrow{p.s.} p_i \tag{36.249}$$

pour chaque i. Le théorème central limite nous indique de plus que

$$\frac{N_i - np_i}{\sqrt{np_i(1-p_i)}} \xrightarrow{\mathscr{L}} T \sim \mathscr{N}(0,1). \tag{36.250}$$

Nous considérons un cas où les p_i sont inconnus. Ils peuvent être approchés par $N_i \simeq np_i$. Le théorème de Pearson nous indique comment.

Théorème 36.59 (Théorème de Pearson)**.**
Nous avons

$$\sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i} \xrightarrow{\mathscr{L}} K \sim \chi^2(k-1) \tag{36.251}$$

où la distribution $\chi^2(l)$ est la somme des carrés de l gaussiennes centrées réduites indépendantes.

Démonstration. Nous commençons par le cas $k = 2$. Dans ce cas nous avons $N_2 = n - N_1$ et $p_1 + p_2 = 1$. La sommes que nous regardons est

$$\frac{(N_1 - np_1)^2}{np_1} + \frac{(N_2 - np_2)^2}{np_2} = \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_1 - np_1)^2}{n(1-p_1)} \tag{36.252a}$$

$$= \frac{(N_1 - np_1)^2}{np_1(1-p_1)}. \tag{36.252b}$$

Étant donné que N_1 est une variable aléatoire binomiale nous avons

$$\frac{N_1 - np_1}{\sqrt{np_1(1-p_1)}} \xrightarrow{\mathscr{L}} T \sim \mathscr{N}(0,1). \tag{36.253}$$

Par conséquent la limite de (36.252b) est

$$\left(\frac{N_1 - np_1}{\sqrt{np_1(1-p_1)}}\right)^2 \xrightarrow{\mathscr{L}} T^2 \simeq \chi^2(1). \tag{36.254}$$

Cela conclut le cas $k = 2$.

Passons à présent au cas général. Le k-uple $(N_1, \dots, N_k)$ est une variable aléatoire multinomiale de loi

$$\mathscr{M}(n; k; p_1, \dots, p_k). \tag{36.255}$$

Nous introduisons les variables aléatoires U_i données par $U_i \colon \Omega \to \mathbb{R}^k$ avec

$$P\big(U_i = (0, \dots, 1, \dots, 0)\big) = p_i; \tag{36.256}$$

c'est le vecteur aléatoire qui prend ses valeurs dans les vecteurs de la base canonique de $\mathbb{R}^k$ et qui prend la valeur e_i avec probabilité p_i. Par construction nous avons

$$(N_1, \dots, N_k) = \sum_{i=1}^{n} U_i. \tag{36.257}$$

Nous allons étudier la fonction caractéristique de $(N_1, \dots, N_k)$ définie par l'équation (35.218) :

$$\begin{aligned}\Phi_{(N_1,\dots,N_k)} \colon \mathbb{R}^k &\to \mathbb{C}\\ e_j &\mapsto E(e^{ie_j \cdot N}) = E(e^{iN_j}).\end{aligned} \tag{36.258}$$

Plus généralement,

$$\Phi_{(N_1,\dots,N_k)}(t_1, \dots, t_k) = E\big(e^{i\langle t,N\rangle_{\mathbb{R}^k}}\big). \tag{36.259}$$

Nous avons

$$e^{i\langle t,N\rangle} = e^{i\sum_j \langle t,U_j\rangle} = \prod_j e^{i\langle t,U_j\rangle} \tag{36.260}$$

et vu que les U_i sont indépendantes et identiquement distribuées nous pouvons écrire U_1 à la place de U_j de façon à avoir

$$\Phi_{(N_1,\dots,N_k)}(t_1, \dots, t_k) = \prod_j E\big(e^{i\langle t,U_1\rangle}\big) \tag{36.261a}$$

$$= \prod_j \sum_l p_l e^{i\langle t,e_l\rangle} \tag{36.261b}$$

$$= \prod_j \sum_l p_l e^{it_l} \tag{36.261c}$$

$$= \left(\sum_{l=1}^k p_l e^{it_l}\right)^n. \tag{36.261d}$$

Nous allons montrer que

$$\lim_{n\to\infty} \Phi_{(N_1,\dots,N_n)}(t_1, \dots, t_k) = e^{-\frac{1}{2}\sum_{j=1}^n t_j^2} - \left(\sum_{j=1}^k t_j\sqrt{p_j}\right)^2. \tag{36.262}$$

Pour ce faire, nous allons effectuer un développement limité. D'abord nous introduisons les variables aléatoires

$$\alpha_j = \frac{N_j - np_j}{\sqrt{np_j}} \tag{36.263}$$

et nous calculons

$$\Phi_{(\alpha_1,\dots,\alpha_k)}(t_1, \dots, t_k) = E\left[\exp\left(i\langle t, (\frac{N_1 - np_1}{\sqrt{np_1}}, \dots, \frac{N_k - np_k}{\sqrt{np_k}})\rangle\right)\right] \tag{36.264}$$

Étant donné que n et p_j sont des variables déterministes, nous pouvons les sortir de l'espérance. Nous avons alors

$$\Phi_{(\alpha_1,\dots,\alpha_k)}(t_1, \dots, t_k) = \exp\left(-i\sum_{j=1}^k t_j\sqrt{np_j}\right)\Phi_{(N_1,\dots,N_k)}\left(\frac{t_1}{\sqrt{np_1}}, \dots, \frac{t_k}{\sqrt{np_k}}\right) \tag{36.265}$$

parce que

$$E\left(e^{t_j\frac{N_j - np_j}{\sqrt{np_j}}}\right) = e^{\frac{-t_j np_j}{\sqrt{np_j}}} E\left(e^{t_j N_j/\sqrt{np_j}}\right) \tag{36.266}$$

En remplaçant (36.261) dans (36.265) nous trouvons

$$\Phi_{(\alpha_1,\dots,\alpha_k)}(t_1, \dots, t_k) = \exp\left(-i\sum_{j=1}^k t_j\sqrt{np_j}\right)\underbrace{\left(\sum_{j=1}^k p_j e^{i\frac{t_j}{\sqrt{np_j}}}\right)^n}_{A} \tag{36.267}$$

Nous analysons maintenant le terme A. Nous écrivons l'égalité $A = A + 1 - 1$ en tenant compte de $\sum_j p_j = 1$ sous la forme

$$A = \left(1 + \sum_{j=1}^{k} p_j\big(\exp(it_j/\sqrt{np_j}) - 1\big)\right)^n, \tag{36.268}$$

Nous avons alors

$$\ln(A) = n \ln\left[1 + \sum_{j=1}^{\infty} p_j\big(e^{it_j/\sqrt{np_j}} - 1\big)\right] \tag{36.269}$$

Nous développons l'exponentielle en

$$e^{it_j/\sqrt{np_j}} - 1 = \frac{it_j}{\sqrt{np_j}} - \frac{t_j^2}{2np_j} + \frac{1}{n}\epsilon(1/n) \tag{36.270}$$

et ensuite le logarithme selon la formule

$$\ln(1 + x) = x - \frac{x^2}{2} + x^2\alpha(x^2). \tag{36.271}$$

Nous avons

$$\begin{aligned}
\ln(A) &= n \ln\left[1 + \sum_j p_j\Big(\frac{it_j}{\sqrt{np_j}} - \frac{t_j^2}{2np_j} + \frac{1}{n}\epsilon(\frac{1}{n})\Big)\right] && \text{(36.272a)}\\
&= n \ln\left[1 + \sum_j p_j\Big(it_j\sqrt{\frac{p_j}{n}} - \frac{t_j^2}{2n} + \frac{1}{n}\epsilon(\frac{1}{n})\Big)\right] && \text{(36.272b)}\\
&= n \underbrace{\sum_{j=1}^{k}\left(it_j\sqrt{\frac{p_j}{n}} - \frac{t_j^2}{2n} + \frac{1}{n}\epsilon(1/n)\right)}_{K} && \text{(36.272c)}\\
&\quad - n\frac{1}{2}\left[\sum_j\left(it_j\sqrt{\frac{p_j}{n}} - \frac{t_j^2}{2n} + \frac{1}{n}\epsilon(1/n)\right)^2\right] && \text{(36.272d)}\\
&\quad + nK^2\alpha(K) && \text{(36.272e)}
\end{aligned}$$

Nous introduisons dans ϵ tous les termes en $1/n^2$ et nous trouvons

$$\ln(A) = \sum_j\left(it_j\sqrt{p_j n} - \frac{t_j^2}{2}\right) - \frac{1}{2}\left(\sum_j it_j\sqrt{p_j}\right)^2 + \epsilon(1/n) + K^2\alpha(K). \tag{36.273}$$

En remplaçant dans (36.267) et en passant à la limite pour $n \to \infty$,

$$\Phi_{(\alpha_1,\dots,\alpha_k)}(t_1,\dots,t_k) = \exp\left(-\frac{1}{2}\sum_j t_j^2 + \frac{1}{2}\big(\sum_j t_j\sqrt{p_j}\big)^2\right). \tag{36.274}$$

Nous reconnaissons des lois gaussiennes dans le premier terme de l'exponentielle. Nous allons maintenant nous atteler à identifier le second terme.

Soit C une matrice orthogonale dont la dernière ligne est $(\sqrt{p_1},\dots,\sqrt{p_k})$. Nous considérons les vecteurs

$$\alpha = \begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_k\end{pmatrix}, \quad t = \begin{pmatrix}t_1\\ \vdots\\ t_k\end{pmatrix}. \tag{36.275}$$

et ensuite nous notons

$$U = Ct = \begin{pmatrix} u_1 \\ \vdots \\ u_k \end{pmatrix}. \tag{36.276}$$

Étant donné que C est orthogonale, nous avons $\sum_{i=1}^k \alpha_i^2 = \sum_{i=1}^k \beta_i^2$ et

$$\Phi_{(\beta_1,\dots,\beta_k)} = E\big(e^{i\langle u,\beta\rangle}\big) = E\big(e^{i\langle t,\alpha\rangle}\big) = \Phi_{(\alpha_1,\dots,\alpha_k)}(t_1,\dots,t_k). \tag{36.277}$$

Nous pouvons récrire l'argument de l'exponentielle (36.274) de la façon suivante :

$$\sum_{i=1}^k t_j^2 = \sum_j u_j^2 \tag{36.278a}$$

$$\sum_{j=1}^k t_j\sqrt{p_j} = (Ct)_k, \tag{36.278b}$$

Nous avons alors

$$\lim_{n\to\infty} \Phi_{(\beta_1,\dots,\beta_k)}(t_1,\dots,t_k) = \lim_{n\to\infty} \Phi_{(\alpha_1,\dots,\alpha_k)}(u_1,\dots,u_k) \tag{36.279a}$$

$$= \exp\left(\frac{1}{2}\sum_{j=1}^{k-1} u_j^2\right) \tag{36.279b}$$

$$= \Phi_{(Z_1,\dots,Z_{k-1},0)}(u_1,\dots,u_k) \tag{36.279c}$$

où les Z_i sont des variables aléatoires indépendantes et identiquement distribuées de distribution normale centrée réduite. Nous avons donc montré que

$$(\beta_1,\dots,\beta_k) \xrightarrow{\mathscr{L}} (Z_1,\dots,Z_{k-1},)). \tag{36.280}$$

Étant donné que l'application $x \mapsto \|x\|^2$ est continue, nous avons aussi

$$\|(\beta_1,\dots,\beta_k)\|^2 \xrightarrow{\mathscr{L}} \|(Z_1,\dots,Z_{k-1},0)\|^2, \tag{36.281}$$

et par conséquent

$$\|(\alpha_1,\dots,\alpha_k)\|^2 \xrightarrow{\mathscr{L}} \sum_{j=1}^{k-1} Z_j^2 \sim \chi^2(k-1). \tag{36.282}$$

D'après la définition (36.263) nous avions

$$\|(\alpha_1,\dots,\alpha_k)\|^2 = \sum_{j=1}^k \frac{(N_j - p_j)^2}{np_j}. \tag{36.283}$$

□

Chapitre 37

Chaînes de Markov à temps discret

Mets tes deux pieds en canard, c'est la chaine de Markov qui se prépare.

37.1 Généralités

Les chaines de Markov interviennent pour la description des systèmes dont l'évolution future ne dépend que de l'état présent.

Définition 37.1.
Soit E un ensemble au plus dénombrable[1] *et $(\Omega, \mathcal{F}, P)$ un espace de probabilité. Une* ***chaine de Markov*** *à valeurs dans E est une famille $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires telles que pour tout $x_0, \dots, x_{n+1} \in E$,*

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0) = P(X_{n+1} = x_{n+1} | X_n = x_n). \tag{37.1}$$

Pour une chaine de Markov, il n'est pas important de savoir l'historique pour prédire la futur : X_{n+1} est seulement déterminé par X_n.

Remarque 37.2.
Il existe une théorie des chaines de Markov à temps continu ou avec E non dénombrable, mais ce n'est pas au programme.

37.3.
Vu que l'ensemble E des états est au plus dénombrable, nous rappelons très humblement au lecteur la proposition 10.99 qui nous permet de changer des sommes sur E en des sommes sur $\mathbb{N}$ sans nous soucier de l'ordre sur E. Si f est une fonction sur E, nous nous écrirons

$$\sum_{x\in E} f(x) = \sum_{k=0}^{\infty} f(x_k) \tag{37.2}$$

sans citer 10.99 à chaque fois.

Définition 37.4.
Si (X_n) est une chaine de Markov[2]*, nous notons*

$$p_n(x, y) = P(X_{n+1} = y | X_n = x) \tag{37.3}$$

la ***probabilité de transition*** *de la chaine à l'instant n. Si cette probabilité ne dépend pas de n, nous disons que la chaine de Markov est* ***homogène****, et nous notons $p(x, y)$ au lieu de $p_n(x, y)$.*

1. Une chaine de Markov sur un ensemble indénombrable demanderait plus de technique à cause du lemme 10.96 qui fait que toutes les sommes sur des ensembles indénombrables sont infinies.
2. Définition 37.1.

Définition 37.5 (Matrice de transition)**.**
Nous notons $Q^{(n)}$ la matrice de transition qui est éventuellement infinie :

$$Q^{(n)}_{xy} = p_n(x, y). \tag{37.4}$$

Si l'ensemble des états E est infini, ce n'est pas une matrice à proprement parler.

Le lemme suivant est intuitivement rien d'autre que le fait que la somme des probabilités doit être 1.

Lemme 37.6.
Soit un processus de Markov (X_n) sur l'ensemble E. Pour chaque $x \in E$ pour pour tout $n \in \mathbb{N}$ nous avons

$$\sum_{y\in E} p_n(x, y) = 1. \tag{37.5}$$

Démonstration. Nous avons le calcul

$$\begin{aligned}
\sum_{y\in E} p_x(x, y) &= \sum_{y\in E} P(X_{n+1} = y|X_n = x) && (37.6a)\\
&= P(\Omega|X_n = x) && (37.6b)\\
&= \frac{P(\Omega \cap X_n = x)}{P(X_n = x)} && (37.6c)\\
&= 1. && (37.6d)
\end{aligned}$$

Justifications :

— Pour (37.6b), c'est le lemme 35.36 en observant que $\Omega = \bigcup_{y\in E}\{X_{n+1} = y\}$.

— Pour (37.6d), c'est $\Omega \cap A = A$.

□

Remarque 37.7.
Attention à ce que ce lemme 37.6 ne fonctionne que sur les colonnes de p_n. En effet, la somme $\sum_{x\in E} p(x, y)$ ne vaut pas spécialement 1. Si les états x_1 et x_2 arrivent tous les deux en y de façon certaine, alors nous avons $\sum_x p(x, y) \geqslant 2$. Il n'y a donc pas de limites aux sommes des lignes.

Lemme-définition 37.8 (Produit de matrices de transition)**.**
Soient deux processus de Markov (X_n) et (Y_n) sur le même ensemble E. En notons p et q leurs matrices de transition[3] *alors*

(1) La somme $\sum_{x\in E} p(a, x)q(x, b)$ converge pour tout $a, b \in E$.

(2) Nous notons pq la « matrice de transition »

$$(pq)(a, b) = \sum_{x\in E} p(a, x)q(x, b). \tag{37.7}$$

(3) La matrice pq vérifie

$$\sum_{y\in E}(pq)(x, y) = 1 \tag{37.8}$$

pour tout $x \in E$.

Démonstration. Nous considérons la mesure de comptage sur E (définition 13.230). Cela nous permet d'écrire la somme comme une intégrale.

3. Qui n'est pas spécialement une matrice, voir la définition 37.5.

Pour (1) Nous devons donc démontrer la convergence de

$$\sum_{x\in E} p(a,x)q(x,b) = \int_E p(a,x)q(x,b)dm(x) \tag{37.9}$$

où $dm(x)$ n'est pas du tout la mesure de Lebesgue (qui n'aurait aucun sens), mais bien la mesure de comptage en x.

Nous pouvons majorer $q(x,b)$ par 1 (tous les nombres sont strictement plus grands que zéro) : pour chaque x nous avons

$$p(a,x)q(x,b) \leqslant p(a,x), \tag{37.10}$$

alors que $\int_E p(a,x)dm(x) = 1$ par le lemme 37.6.

Vu que la fonction $x \mapsto p(a,x)q(x,b)$ est dominée par la fonction $x \mapsto p(a,x)$ et que cette dernière est intégrable, le lemme 13.172 conclu à l'intégrabilité de la première. Bref, l'intégrale (37.9) existe et est finie.

Pour (2) Il n'y a rien à prouver, c'est seulement une définition.

Pour (3) Nous devons calculer la valeur de

$$\sum_{b\in E}\Big(\sum_{x\in E} p(a,x)q(x,b)\Big) = \int_E \Big(\int_E p(a,x)q(x,b)dm(x)\Big)dm(b). \tag{37.11}$$

Pour cela nous allons utiliser le théorème de Fubini comme expliqué en 13.263, et nous partons de la somme dans le sens inverse. D'abord nous prouvons que $(x,b) \mapsto p(a,x)q(x,b)$ est dans $L^1(E\times E)$ en étudiant les intégrales en chaine :

$$\int_E \Big(\int_E |p(a,x)q(x,b)|dm(b)\Big)dm(x) = \int_E p(a,x)\Big(\underbrace{\int_E q(x,b)dm(b)}_{=1}\Big)dm(x) \tag{37.12a}$$

$$= \int_E p(a,x)dm(x) \tag{37.12b}$$

$$= 1. \tag{37.12c}$$

Donc la fonction est $L^1(E\times E)$ et nous pouvons fusionner et permuter les intégrales à volonté. Nous avons alors

$$1 = \int_E \Big(\int_E |p(a,x)q(x,b)|dm(b)\Big)dm(x) \tag{37.13a}$$

$$= \int_{E\times E} p(a,x)q(x,b)dm(x,b) \tag{37.13b}$$

$$= \int_E \Big(\int_E p(a,x)q(x,b)dm(x)\Big)dm(b). \tag{37.13c}$$

□

Définition 37.9.
Une matrice dont tous les éléments sont positifs ou nuls et donc la somme de toutes les lignes sont 1 *est une* ***matrice stochastique****.*

Notons que l'ensemble des matrices stochastiques est un fermé dans l'ensemble des matrices.

Lemme 37.10.
Si U est une matrice stochastique[4]*, alors il existe une chaine de Markov dont la matrice de transition est U.*

4. Définition 37.9.

Exemple 37.11.
Nous considérons une fourmi qui se déplace dans un appartement à trois pièces A, B, C. Supposons qu'à chaque minute, elle a une probabilité 1/3 de rester dans la pièce et une probabilité 2/3 de se déplacer. Le plan de l'appartement est

$$A \longrightarrow B \longrightarrow C \tag{37.14}$$

De la pièce A est est donc uniquement possible d'aller vers la pièce B ; de la B il est possible d'aller en A et en C et de la C il est uniquement possible d'aller en B.

La matrice de transition de cette chaine de Markov est

$$Q = \begin{pmatrix} 1/3 & 2/3 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 2/3 & 1/3 \end{pmatrix} \tag{37.15}$$

△

Proposition 37.12.
Si N_t est un processus de Poisson, alors les variables aléatoires $X_n = N_n$ forment une chaine de Markov.

37.2 Chaînes de Markov sur un ensemble fini

Définition 37.13.
*Une chaine de Markov est **finie** si l'ensemble E dans lequel elle prend ses valeurs est fini.*

Proposition 37.14 ([536]).
Si (X_n) est une chaine de Markov irréductible sur un ensemble fini, alors pour tout ensemble $A \subset E$ nous avons

$$P(\tau_A < \infty) = \lim_{n\to\infty} P(\tau_A \leqslant n) = 1 \tag{37.16}$$

où $\tau_A = \min\{k \text{ tel que } X_k \in A\}$.

Les propositions à venir vont montrer que

(1) Toute matrice stochastique admet un état stationnaire, proposition 37.15.

(2) Si la chaine de Markov est irréductible, alors il y a unicité de l'état stationnaire, proposition 37.16. Mais attention : cela ne veut pas encore dire que la chaine converge effectivement vers cet état.

(3) Si la chaine est irréductible et apériodique, alors il y a convergence en loi vers l'unique loi invariante, théorème 37.19.

Proposition 37.15 ([537]).
Toute matrice stochastique admet un état stationnaire.

Proposition 37.16 ([537]).
Soit une chaine de Markov irréductible finie. Alors il existe une unique loi stationnaire π et de plus nous avons $\pi_i > 0$ pour tout état i de E.

Définition 37.17.
*Une chaine de Markov finie est **régulière** s'il existe un $n \in \mathbb{N}$ tel que P^n a uniquement des éléments strictement positifs.*

Théorème 37.18 ([536]).
Soit P la matrice de transition d'une chaine de Markov régulière sur un ensemble E de cardinal N. Alors il existe des nombres $\pi_1, \dots, \pi_N$ tels que

(1) $\pi_i > 0$ pour tout $i = 1, \dots, N$

(2) $\pi_1 + \cdots + \pi_N = 1$

(3)

$$\lim_{n\to\infty} P^n = \Pi = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_N \\ \vdots & \vdots & & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_N \end{pmatrix} \tag{37.17}$$

De plus le vecteur $\pi = (\pi_1, \dots, \pi_N)$ *est l'unique solution de*

$$\pi P = \pi. \tag{37.18}$$

Démonstration. Si la chaine n'a qu'un seul état c'est évident parce que la probabilité de transition est toujours 1 ; fin de l'histoire.

Hypothèse Sinon nous supposons que P n'a que des éléments positifs, quitte à considérer P^n au lieu de P. Nous notons d le plus petit élément de P ; il vérifie $d \leqslant \frac{1}{2}$ parce que la somme des élément d'une ligne de la matrice P doit être égale à 1.

Les suites min et max Soit x un vecteur quelconque (de composantes positives). Nous notons $m_0 = \min\{x_i\}$ et $M_0 = \max\{x_i\}$. Étant donné que les éléments du vecteur Px sont des moyennes pondérées des éléments de x, si nous posons

$$m_k = \min\{(P^k x)_i\}_{i=1,\dots,N} \tag{37.19a}$$

$$M_k = \max\{(P^k x)_i\}_{i=1,\dots,N}, \tag{37.19b}$$

la suite (m_k) est croissante et la suite (M_k) est décroissante.

Stricte croissance et décroissance Si $M_{k+1} = M_k$, alors toutes les composantes de $P^k x$ sont égales à M_k et le théorème est prouvé. Cela est encore une propriété de la moyenne. Même remarque pour la suite (m_k).

Nous pouvons donc supposer que la suite (m_k) est strictement croissante et que la suite (M_k) est strictement décroissante. Elles sont toutes les deux bornées dans $[m_0, M_0]$. Le lemme 7.137 nous donne la convergence.

Égalité des limites Vu que les éléments de $P^k x$ ne sont pas tous les mêmes et s'étalent de m_k à M_k, pour majorer M_{k+1} nous mettons le plus petit coefficient possible (c'est-à-dire d) devant m_k et nous supposons que toutes les autres composantes sont M_k ; nous avons alors

$$M_{k+1} \leqslant d m_k + (1-d) M_k \tag{37.20}$$

parce que tous les autres coefficients de la ligne contenant le d (dans P^k) sont plus petits ou égaux à $1-d$. De la même façon nous avons la minoration

$$m_{k+1} \geqslant d M_k + (1-d) m_k. \tag{37.21}$$

En faisant la différence, et en nous souvenant que $0 < 1 - 2d < 1$,

$$M_{k+1} - m_k \leqslant (1-2d)(M_k - m_k), \tag{37.22}$$

ce qui signifie que

$$M_{k+1} - m_k \leqslant (1-2d)^k (M_0 - m_0), \tag{37.23}$$

et donc que les deux limites sont égales.

Conclusion pour la limite Pour tout vecteur x, la suite $P^k x$ tend vers un vecteur dont toutes les composantes sont égales. En particulier pour le vecteur e_i de la base canonique,

$$P^k e_i \to \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_1 \end{pmatrix}. \tag{37.24}$$

Mais $P^k e_i$ est la i^{e} colonne de la matrice P^k. Cela prouve la convergence annoncée $P^k \to \Pi$.

Réglons rapidement le cas des deux autres allégations du théorème. D'abord les matrices P^k sont toutes des matrices stochastiques ; et l'ensemble des matrices stochastiques est fermé, donc la convergence se fait à l'intérieur de l'ensemble des matrices stochastiques. Cela prouve que $\pi_1 + \ldots + \pi_N = 1$.

Ensuite la suite (m_k) étant strictement croissante et m_0 étant égal à 0 dans le cas de e_i nous avons toujours $\pi_i > 0$ (strictement). □

Théorème 37.19 ([537]).
Si (X_n) est une chaine de Markov finie, irréductible et apériodique de loi stationnaire π, alors

(1) La suite de matrices stochastiques P^k converge vers la matrice

$$P^k \to \Pi = \begin{pmatrix} \pi \\ \vdots \\ \pi \end{pmatrix}. \tag{37.25}$$

(2) Nous avons convergence des variables aléatoires au sens où

$$P(X_k = \mu P^k) \to \pi. \tag{37.26}$$

37.3 Marche aléatoire sur $\mathbb{Z}$

Soit (Y_n) une suite de variables aléatoires indépendantes et identiquement distribuées valant -1 avec une probabilité p et 1 avec une probabilité $(1-p)$. La loi est

$$Y_n \sim p\delta_{-1} + (1-p)\delta_1. \tag{37.27}$$

Nous considérons la variable aléatoire

$$X_n = X_0 + \sum_{i=1}^{n} Y_i \tag{37.28}$$

où X_0 est une variable aléatoire indépendante des Y_i à valeurs dans $\mathbb{Z}$. Nous vérifions à présent que X_n est une chaine de Markov avec comme espace d'états $E = \mathbb{Z}$. Nous devons montrer que

$$P\big(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0\big) = P\big(X_{n+1} = x_{n+1} | X_n = x_n\big). \tag{37.29}$$

Pour ce faire nous allons exprimer tout cela en termes des Y_i au lieu des X_i. D'abord étant donné que nous avons égalité des événements

$$\{X_{n+1} = x_{n+1}\} \cap \{X_n = x_n, \ldots, X_0 = x_0\} = \{Y_{n+1} = x_{n+1} - x_n\} \cap \{X_n = x_n, \ldots, X_0 = x_0\}, \tag{37.30}$$

nous pouvons, en vertu du principe (35.78), remplacer $X_{n+1} = x_{n+1}$ par $Y_{n+1} = x_{n+1} - x_n$ dans le membre de gauche de (37.29). Nous avons donc déjà

$$P\big(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0\big) = P\big(\underbrace{Y_{n+1} = x_{n+1} - x_n}_{A} \,|\, \underbrace{X_n = x_n, \ldots, X_0 = x_0}_{B}\big). \tag{37.31}$$

L'événement B est égal à l'événement

$$\{X_0 = x_0, Y_1 = x_1 - x_0, Y_2 = x_2 - x_1, \ldots, Y_n = x_n - x_{n-1}\}, \tag{37.32}$$

qui n'est autre que l'ensemble

$$X_0^{-1}(x_0) \cap Y_1^{-1}(x_1 - x_0) \cap \ldots \cap Y_n^{-1}(x_n - x_{n-1}) \tag{37.33}$$

qui est dans la tribu engendrée par les variables aléatoires $X_0, (Y_i)_{i=1,\dots,n}$. Le point délicat du raisonnement est de montrer que les événements A et B donnés par

$$A = \{Y_{n+1} = x_{n+1} - x_n\} \tag{37.34a}$$

$$B = \{X_0 = x_0\} \cap \bigcap_{i=1}^{n} \{Y_i = x_i - x_{i-1}\} \tag{37.34b}$$

sont indépendants. Nous ne pouvons pas montrer directement que $P(A \cap B) = P(A)P(B)$ parce que cela est la formule que nous voulons utiliser pour montrer que la chaine est de Markov. Nous passons donc par les tribus :

$$A \in \sigma(Y_{n+1}) \tag{37.35a}$$

$$B \in \sigma(X_0, Y_1, \dots, Y_n). \tag{37.35b}$$

Nous utilisons maintenant l'hypothèse d'indépendance des variables aléatoires X_0 et Y_i pour conclure que les deux tribus des équations (37.35) sont indépendantes. Les événements A et B sont par conséquent indépendants.

L'événement A est indépendant de l'événement $\{X_n = x_n\}$. Nous avons donc successivement

$$P\big(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0\big) = P\big(Y_{n+1} = x_{n+1} - x_n | X_n = x_n, \dots, X_0 = x_0\big) \tag{37.36a}$$

$$= P\big(Y_{n+1} = x_{n+1} - x_n | Y_i = x_i - x_{i-1}, X_0 = x_0\big) \tag{37.36b}$$

$$= P(Y_{n+1} = x_{n+1} - x_n) \tag{37.36c}$$

$$= P(Y_{n+1} = x_{x+1} - x_n | X_n = x_n) \tag{37.36d}$$

$$= P(Y_{n+1} = x_{n+1} - X_n | X_n = x_n) \tag{37.36e}$$

$$= P(X_{n+1} = x_{n+1} | X_n = x_n). \tag{37.36f}$$

Justifications :

— (37.36c) parce que les tribus $\sigma(Y_{n+1})$ et $\sigma(Y_i, X_0)$ sont indépendantes.

— (37.36d) Nous avons

$$\{X_n = x_n\} \in \sigma(X_0, Y_1, \dots, Y_n) \tag{37.37}$$

tandis que

$$\{Y_{n+1} = x_{n+1} - x_n\} \in \sigma(Y_{n+1}); \tag{37.38}$$

ce sont donc deux événements issus de tribus indépendantes. Donc conditionner ou non l'événement $Y_{n+1} = x_{n+1} - x_n$ à l'événement $X_n = x_n$ ne change rien.

— (37.36e) est encore l'utilisation du fait que $P(A|B) = P(K|B)$ dès que $A \cap B = K \cap B$.

La chaine est par conséquent de Markov.

La matrice de transition de cette chaine de Markov est une matrice infinie « dans tous les sens » :

$$p(x, y) = \begin{cases} p & \text{si } y = x - 1 \\ (1 - p) & \text{si } y = x + 1 \\ 0 & \text{sinon.} \end{cases} \tag{37.39}$$

Remarque 37.20.
La plupart du temps lorsqu'il faut démontrer qu'une chaine est de Markov, il faut suivre la procédure que nous venons de suivre pour la marche aléatoire sur $\mathbb{Z}$.

— Écrire tout en fonction des incréments.

— Dire que les incréments conditionnés sont indépendants des incréments qui conditionnent (via les tribus engendrées).

— Écrire que la probabilité cherchée est égale à l'événement conditionné dans lequel on a juste remplacé l'incrément par sa valeur.

— Conditionner à nouveau par rapport au dernier incrément qui est indépendant.

— Changer la valeur du dernier incrément par la variable aléatoire.

Dans ce raisonnement nous utilisons deux fois le fait que $P(A|B) = P(K|B)$ si $A \cap B = K \cap B$.

37.3.1 Chaînes de Markov homogènes

Proposition 37.21.
Voici quelques propriétés des chaines de Markov homogènes[5].

(1) La probabilité d'une trajectoire donnée est

$$P(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = p(x_{n-1}, x_n) \dots p(x_0, x_1) P(X_0 = x_0) \tag{37.40}$$

où les $p(x, y)$ sont les probabilités de transitions introduits dans la définition 37.4.

(2) La probabilité de transition « en n coups » est donnée par la puissance n^e de la matrice de transition :

$$P(X_n = x_n | X_0 = x_0) = Q^n_{x_0, x_n}. \tag{37.41}$$

(3) Si l'espace des états E est fini, l'espérance d'une fonction bornée[6] *$f\colon E \to \mathbb{R}$ de l'état est donnée par*

$$E\big(f(X_{n+1}) | X_n = x_n, \dots, X_0 = x_0\big) = E\big(f(X_{n+1}) | X_n = x_n\big) = \sum_{y \in E} f(y) p(x_n, y) \tag{37.42}$$

pour tout $x_0, \dots, x_n \in E$.

Pour être précis, ce que nous notons « $f(X_{n+1})$ » est la composée $f \circ X_{n+1}$.

Démonstration. En plusieurs étapes.

Pour (1) Nous écrivons la formule $P(A \cap B) = P(A|B)P(B)$ pour les événements $A = \{X_n = x_n\}$ et $B = \bigcap_{i=0}^{n-1} \{X_i = x_i\}$:

$$P(X_n = x_n, \dots, X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) P(X_{n-1} = x_{n-1}, \dots, X_0 = x_0). \tag{37.43}$$

Par la propriété de Markov, le premier facteur est

$$P(X_n = x_n | X_{n-1} = x_{n-1}) = p(x_{n-1}, x_n). \tag{37.44}$$

Le reste est une récurrence sur n.

Pour (2) Montrons avec $n = 2$. En utilisant les divers points du théorème 35.34, nous avons

$$\begin{aligned}
P(X_2 = x_2 | X_0 = x_0) &= \sum_{y \in E} P(X_2 = x_2, X_1 = y | X_0 = x_0) && \text{(37.45a)}\\
&= \sum_{y \in E} P(X_2 = x_2 | X_1 = y, X_0 = x_0) P(X_1 = y | X_0 = x_0) && \text{(37.45b)}\\
&= \sum_{y \in E} P(X_2 = x_2 | X_1 = y) P(X_1 = y | X_0 = x_0) && \text{(37.45c)}\\
&= \sum_{y \in E} p(x_2, y) p(y, x_0) && \text{(37.45d)}\\
&= Q^2_{x_2, x_0}. && \text{(37.45e)}
\end{aligned}$$

Nous avons utilisé l'homogénéité de la chaine de Markov au moment d'écrire l'expression (37.45d). En principe nous aurions dû écrire $p_2(y, x_2) p_1(x_0, y)$.

5. Définition 37.4.

6. L'hypothèse de borne sur f n'est pas très chère parce que E est fini. Il suffit que f ne soit infinie en aucun point.

Pour (3) Vu que E est fini, $f(E)$ est fini et nous notons $\{a_k\}_{k=1,\dots,N}$ l'ensemble des valeurs non nulles (dans $\mathbb{R}$) atteintes par f. Nous utilisons le lemme 35.50 pour la variable aléatoire[7] $f\circ X_{n+1}\colon\Omega\to\{0,a_k\}_{k=1,\dots,N}$:

$$E\big(f(X_{n+1})|X_n=x_n,\dots,X_0=x_0\big)=\sum_{k=1}^{N}a_kP\big(f\circ X_{n+1}=a_k|X_n=x_n,\dots,X_0=x_0\big).\tag{37.46}$$

L'événement $f\circ X_{n+1}=a_k$ signifie $X_{n+1}\in f^{-1}(a_k)$ ou encore

$$\{\omega\in\Omega\text{ tel que }X_{n+1}(\omega)\in f^{-1}(a_k)\}.\tag{37.47}$$

Vu que E est fini, l'ensemble $f^{-1}(a_k)$ est fini et nous écrivons

$$f^{-1}(a_k)=\{y_{k,1},\dots,y_{k,N_k}\}.\tag{37.48}$$

Nous avons

$$\{\omega\in\Omega\text{ tel que }X_{n+1}(\omega)\in f^{-1}(a_k)\}=\bigcup_{i=1}^{N_k}\{\omega\in\Omega\text{ tel que }X_{n+1}(\omega)=y_{k,i}\}.\tag{37.49}$$

Nous pouvons utiliser le lemme 35.36 pour décomposer

$$P(f\circ X_{n+1}=a_k|\cdots)=\sum_{i=1}^{N_k}P(X_{n+1}=y_{k,i}|\cdots)\tag{37.50}$$

et continuer (37.46) pas

$$E\big(f(X_{n+1})|X_n=x_n,\dots,X_0=x_0\big)=\sum_{k=1}^{N}a_k\sum_{i=1}^{N_k}P(X_{n+1}=y_{k,i}|X_n=x_n,\dots,X_0=x_0)\tag{37.51a}$$

$$=\sum_{k=1}^{N}a_k\sum_{i=1}^{N_k}P(X_{n+1}=y_{k,i}|X_n=x_n)\tag{37.51b}$$

$$=\sum_{k=1}^{N}a_kP\big(f\circ X_{n+1}=a_k|X_n=x_n\big)\tag{37.51c}$$

$$=E\big(f(X_{n+1})|X_n=x_n\big).\tag{37.51d}$$

Pour (37.51b), nous avons utilisé la propriété de Markov.

□

37.3.2 Graphe de transition

Le **graphe de transition** d'une chaine de Markov est le graphe dont les sommets sont les éléments de l'espace des états de la chaine et dont les sommets sont reliés par des arrêtes pondérées par la probabilité de transition correspondante.

Définition 37.22.
*Une chaine de Markov est **irréductible** si pour tout $x,y\in E$, il existe n tel que $p^n(x,y)>0$ où*

$$p^n(x,y)=P(X_n=y|X_0=x).\tag{37.52}$$

Le nombre n peut dépendre de x et y.

7. La composée de fonctions mesurables est mesurable, proposition 13.43. De plus f est mesurable parce que E étant dénombrable, nous y mettons la tribu de toutes les parties.

Lemme 37.23.
Une chaine de Markov homogène est irréductible si et seulement si son graphe de transition est connexe.

Démonstration. Pour chaque couple $(x, y) \in E^2$ nous avons

$$\begin{aligned} p^n(x,y) &= \sum_{z_i \in E} P(X_n = y, X_{n-1} = z_{n-1}, \dots, X_1 = z_1, X_0 = x) \\ &= \sum_{z_i} p(z_{n-1}, y) p(z_{n-2}, z_{n-1}) \dots p(z_1, z_2) p(x, z_1). \end{aligned} \tag{37.53}$$

La positivité d'un des termes de la somme signifie que le graphe est connexe tandis que la positivité de $p^n(x, y)$ signifie que la chaine est irréductible. □

37.3.3 Chaîne de Markov définie par récurrence

37.3.3.1 Le cas général

Proposition 37.24.
Soit X_0 une variable aléatoire à valeurs dans E, un ensemble au plus dénombrable. Soit (Y_n) une suite de variables aléatoires réelles indépendantes et identiquement distribuées indépendantes de X_0.

Soit (X_n) la suite de variables aléatoires à valeurs dans E définie par récurrence selon la formule

$$X_{n+1} = G(X_n, Y_{n+1}) \tag{37.54}$$

où $G\colon E \times \mathbb{R} \to E$ est une fonction mesurable. Alors (X_n) est une chaine de Markov.

Démonstration. Soient $x_0, \dots, x_{n+1}$ des éléments de E. Nous devons calculer la valeur de

$$P(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_0 = x_0). \tag{37.55}$$

Commençons par préciser les espaces sur lesquels nos variable aléatoires sont définies. Nous avons

$$X_0\colon \Omega_0 \to E \tag{37.56}$$

et

$$Y_i\colon \Omega \to \mathbb{R}. \tag{37.57}$$

La variable aléatoire X_1 est donnée par

$$\begin{aligned} X_1\colon \Omega_0 \times \Omega &\to E \\ (\omega_0, \omega_1) &\mapsto G\big(X_0(\omega_0), Y_1(\omega_1)\big). \end{aligned} \tag{37.58}$$

La variable aléatoire X_2 est

$$\begin{aligned} X_2\colon \Omega_0 \times \Omega^2 &\to E \\ (\omega_0, \omega_1, \omega_2) &\mapsto G\big(X_1(\omega_0, \omega_1), Y_2(\omega_2)\big) \\ &= G\Big(G\big(X_0(\omega_0), \omega_1\big), Y_2(\omega_2)\Big) \end{aligned} \tag{37.59}$$

et ainsi de suite.

Considérons maintenant l'événement

$$\{X_1 = x_1, X_0 = x_0\} \subset \Omega_0 \times \Omega. \tag{37.60}$$

Il est donné explicitement par

$$\{X_1 = x_1, X_0 = x_0\} = \{(\omega_0, \omega_1) \text{ tel que } G\big(X_0(\omega_0), Y_1(\omega_1) = x_1, X_0(\omega_0) = x_0\big)\} \tag{37.61a}$$

$$= \{(\omega_0, \omega_1) \text{ tel que } G\big(x_0, Y_1(\omega_1) = x_1, X_0(\omega_0) = x_0\big)\} \tag{37.61b}$$

$$= \{\omega_0 \in \Omega_0 \text{ tel que } X_0(\omega_0) = x_0\} \times \{\omega_1 \in \Omega \text{ tel que } G\big(x_0, Y_1(\omega_1) = x_1\big)\}. \tag{37.61c}$$

Le premier terme du produit cartésien est dans $\sigma(X_0)$, tandis que le second est dans $\sigma(Y_1)$. Étant donné la définition des tribus produit (définition 13.112) nous avons

$$\{X_1 = x_1, X_0 = x_0\} \in \sigma(X_0, Y_1). \tag{37.62}$$

Ce raisonnement se généralise immédiatement et nous trouvons que

$$\{X_n = x_n, \dots, X_0 = x_0\} \in \sigma(X_0, Y_1, \dots, Y_n). \tag{37.63}$$

Nous sommes donc à calculer

$$\begin{aligned}
\diamondsuit &= P(X_{n+1} = x_{n+1} | X_n = X_n, \dots, X_0 = X_0) && \text{(37.64a)}\\
&= P\big(\underbrace{G(x_n, Y_{n+1}) = x_{n+1}}_{\in\sigma(Y_{n+1})} \big| \underbrace{X_n = x_n, \dots, X_0 = x_0}_{\in\sigma X_0, Y_1, \dots, Y_n}\big). && \text{(37.64b)}
\end{aligned}$$

Les tribus $\sigma(Y_{n+1})$ et $\sigma(X_0, Y_1, \dots, Y_n)$ étant indépendantes nous avons

$$\begin{aligned}
\diamondsuit &= P\big(G(x_n, Y_{n+1}) = x_{n+1}\big) && \text{(37.65a)}\\
&= P\big(G(x_n, Y_{n+1}) = x_{n+1} | X_n = X_n\big) && \text{(37.65b)}\\
&= P\big(G(X_n, Y_{n+1}) = x_{n+1} | X_n = x_n\big) && \text{(37.65c)}\\
&= P(X_{n+1} = x_{n+1} | X_n = x_n). && \text{(37.65d)}
\end{aligned}$$

Pour (37.65b) nous avons utilisé le fait que $\sigma(Y_{n+1})$ est indépendante de $\sigma(X_n)$. Nous avons prouvé que la chaine était de Markov. □

Les probabilités de transition de la chaine de Markov définie dans la proposition 37.24 sont

$$P(X_1 = y | X_0 = x) = P\big(G(X_0, Y_1) = y | X_0 = x_0\big) = P\big(G(x_0, Y_1) = y\big). \tag{37.66}$$

37.3.3.2 Exemple : la file de réparation de machines à laver

Nous considérons un magasin de réparation d'électroménager. Durant le jour n, un nombre aléatoire Z_n de machines en panne arrivent au magasin. Une machine est réparée chaque jour (aucune si le magasin est vide). Nous supposons que les Z_n soient indépendantes et identiquement distribuées, et nous posons X_n, le nombre de machines en magasin le jour n.

La loi d'avancement de X_n est

$$X_{n+1} = \begin{cases} X_n + Z_n - 1 & \text{si } X_n \neq 0\\ Z_n & \text{si } X_n = 0. \end{cases} \tag{37.67}$$

Cela est une chaine de Markov en vertu de la proposition 37.24. Ici la fonction est

$$G(x, y) = x + y - \mathbb{1}_{x\neq 0}. \tag{37.68}$$

Les probabilités de transitions sont

$$p(x, y) = \begin{cases} 0 & \text{si } x \leqslant y - 2\\ P(Z = 0) & \text{si } x = y - 1\\ P(Z = k) & \text{si } x = y + k - 1 \end{cases} \tag{37.69}$$

pour $x \neq 0$.

Exemple 37.25.
Soit (X_n) une chaine de Markov de matrice de transition

$$P = \begin{pmatrix} 0.2 & 0.5 & 0.3\\ 0.1 & 0.1 & 0.8\\ 0.5 & 0.2 & 0.3 \end{pmatrix}. \tag{37.70}$$

Calculer $P(X_3 = 1|X_0 = 1)$ et $P(X_7 = 0|X_4 = 0)$.

Déterminer, s'il en existe, une loi stationnaire vers laquelle converge la chaine.

Nous avons

$$P^3 = \begin{pmatrix} 0.344 & 0.251 & 0.405 \\ 0.283 & 0.307 & 0.41 \\ 0.287 & 0.248 & 0.465 \end{pmatrix}. \tag{37.71}$$

La probabilité d'aller de l'état 1 à l'état 1 en trois étapes est donc 0.307. La chaine étant de Markov, sans mémoire, les probabilités entre les temps 4 et 7 sont les mêmes qu'entre 0 et 3. Nous avons alors

$$P(X_7 = 0|X_4 = 0) = 0.344. \tag{37.72}$$

La chaine est irréductible et n'a pas d'états absorbants.

△

37.4 Classification des états

Sauf mention expresse du contraire, nous considérons toujours une chaine de Markov homogène.

Définition 37.26.
*Un état $x \in E$ est **absorbant** pour la chaine (X_n) si $p(x, x) = 1$.*

Il n'est pas spécialement impossible d'arriver sur un état absorbant, mais il est impossible d'en sortir.

Si $x \in E$, nous notons

$$T(x) = \inf\{k \geqslant 1 \text{ tel que } X_k = x\}, \tag{37.73}$$

le **premier temps d'atteinte** de l'état x. Si $X_0 = x$, alors $T(x)$ est le **temps de retour** en x. Si $p \in \mathbb{N}$ nous notons

$$T_p(x) = \inf\{k \geqslant 1 \text{ tel que } X_{k+p} = x\}. \tag{37.74}$$

C'est le temps mis pour atteindre x à partir de l'instant p.

Proposition 37.27.
La loi de la variable aléatoire $[T_p(x)|X_p = x]$ est la même que celle de la variable aléatoire $[T(x)|X_0 = x]$.

Démonstration. Nous devons montrer que

$$P(T_p(x) = k|X_p = x) = P(T(x) = k|X_0 = x). \tag{37.75}$$

Cela est intuitivement évident du fait qu'une chaine de Markov soit un processus sans mémoire. Afin de prouver, nous allons sommer sur tous les états intermédiaires possibles :

$$\begin{aligned}
P(T_p(x) = k|X_0 = x) &= P(X_{p+k} = x, X_{p+k-1} \neq x, \dots, X_{p+1} \neq x|X_p = x) && (37.76a)\\
&= \sum_{z_i \neq x} P(X_{p+k} = x, X_{p+k-1} = z_{k-1}, \dots, X_{p+1} = z_1|X_p = x) && (37.76b)\\
&= \sum_{z_i} P(X_{p+k} = x, X_{p+k-i} = z_i|X_{p+1} = z_1, X_p = x) \underbrace{P(X_{p+1} = z_1|X_p = x)}_{=p(x,z_1)} && (37.76c)\\
&= \sum_{z_i} P(X_{p+k} = x, X_{p+k-i} = z_i|X_{p+2} = z_2, X_{p+1} = z_1, X_p = x) && (37.76d)\\
&\quad \underbrace{P(X_{p+2} = z_2|X_{p+1} = z_1, X_p = x)}_{P(X_{p+2}=z_2|X_{p+1}=z_1)=p(z_1,z_2)} p(x, z_1) && (37.76e)\\
&= \dots && (37.76f)\\
&= \sum_{z_i} p(x, z_1)p(z_1, z_2) \dots p(z_{k-1}, z_{k-1})p(z_{k-1}, x). && (37.76g)
\end{aligned}$$

À ce point ci, nous avons éliminé toute référence à p grâce à l'homogénéité de la chaine. Nous pouvons refaire le calcul à l'envers pour reconstituer l'expression de départ sans le p :

$$\sum_{z_i} p(x,z_1)p(z_1,z_2)\dots p(z_{k-1},z_{k-1})p(z_{k-1},x) \tag{37.77a}$$
$$= P(x_k = x, X_{k-1} \neq x, \dots, X_1 \neq x | X_0 = x) \tag{37.77b}$$
$$= P(T(x) = k), \tag{37.77c}$$

ce qu'il fallait obtenir. □

Définition 37.28.
*Un état x est **récurrent** si $P(T(x) = \infty | X_0 = x) = 0$, c'est-à-dire si la probabilité de ne jamais retourner en x lorsqu'on y est passé est nulle. L'état x est **transient** ou **transitoire** dans le cas contraire.*

*Si x est un état récurrent, et si $E\big(T(x)|X_0 = x\big) < \infty$, nous disons que x est **récurrent positif**. Si $E\big(T(x)|X_0 = x\big) = \infty$ alors nous disons que est **récurrent nul**.*

Nous introduisons une variable aléatoire qui compte le nombre de fois que la chaine de Markov passe par l'état x :

$$N_x = \sum_{k=0}^{\infty} \mathbb{1}_{\{X_k = x\}}. \tag{37.78}$$

C'est une variable aléatoire à valeurs dans $\mathbb{N} \cup \{\infty\}$.

Proposition 37.29.
Les deux propriétés suivantes sont équivalentes à dire que x est récurrent :

(1) $P(N_x < \infty | X_0 = x) = 0$

(2) $E(N_x | X_0 = x) = \infty$.

Les deux propriétés suivantes sont équivalentes à dire que x est transient :

(1) $P(N_x < \infty | X_0 = x) = 1$

(2) $E(N_x | X_0 = x) < \infty$.

Démonstration. En tant que événements, nous avons l'égalité

$$N_x < \infty = \bigcup_{n\in\mathbb{N}} \{\underbrace{X_n = x, X_{n+k} \neq x \forall k \geqslant 1}_{F_n}\}. \tag{37.79}$$

Nous avons donc

$$P(N_x < \infty | X_0 = x) = \sum_{n=0}^{\infty} P(F_n | X_0 = x), \tag{37.80}$$

et

$$P(F_n|X_0 = x) = P(X_{n+k} \neq x, \forall k \geqslant 1, X_n = x | X_0 = x) \tag{37.81a}$$
$$= P(X_{n+k} \neq x, k \geqslant 1 | X_n = x, X_0 = x) P(X_n = x | X_0 = x) \tag{37.81b}$$
$$= P(X_{n+k} \neq x, k \geqslant 1 | X_n = x) P(X_n = x | X_0 = x) \tag{37.81c}$$
$$= P(X_k \neq x, k \geqslant 1 | X_0 = x) P(X_n = x | X_0 = x) \tag{37.81d}$$
$$= P(T(x) = \infty | X_0 = x) P(X_n = x | X_0 = x) \tag{37.81e}$$

Justifications :

(1) Pour (37.81c), nous utilisons le fait que la chaine soit « sans mémoire ».

(2) Pour (37.81d), nous utilisons le fait que la chaine soit homogène.

(3) Pour (37.81e), l'événement $X_k \neq x$ pour tout $k \geqslant 1$ est exactement l'événement $T(x) = \infty$.

En nous servant de la proposition 13.266 (théorème de Fubini et mesure de comptage), nous permutons l'espérance et la somme dans l'expression

$$\sum_{n=0}^{\infty} P(X_n = x|X_0 = x) = \sum_{n=0}^{\infty} E(\mathbb{1}_{\{X_n=x\}}|X_0 = x) \tag{37.82a}$$

$$= E\big(\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=x\}}|X_0 = x\big) \tag{37.82b}$$

$$= E(N_x|X_0 = x). \tag{37.82c}$$

Voyons ce passage plus en détail. D'abord, en général nous avons

$$E(Y|X = x_0) = \int_{\{X=x_0\}} Y(\omega)dP(\omega) = \int_{\Omega} \mathbb{1}_{\{X=x_0\}}(\omega)Y(\omega)dP(\omega). \tag{37.83}$$

Dans notre cas,

$$E\big(\mathbb{1}_{\{X_n=x\}}|X_0 = x\big) = \int_{\Omega} \mathbb{1}_{X_0=x}(\omega)\mathbb{1}_{\{X_n=x\}}(\omega)dP(\omega). \tag{37.84}$$

La fonction qui correspond à la proposiiton 13.266 est

$$f(n,\omega) = f_n(\omega) = \delta_{X_0(\omega),x}\delta_{X_n(\omega),x}, \tag{37.85}$$

qui est bien une fonction positive et mesurable.

Nous reprenons à présent le calcul (37.80) en remplaçant les éléments par leurs valeurs que nous avons calculées :

$$P(N_x < \infty|X_0 = x) = P\big(T(x) = \infty|X_0 = x\big)E(N_x|X_0 = x). \tag{37.86}$$

Si x est récurrent, nous avons $P\big(T(x) = \infty|X_0 = x\big) = 0$, mais la relation (37.86) ne permet pas de conclure que le membre de gauche est nul parce qu'il reste la possibilité que $E(N_x|X_0 = x) = \infty$. Nous devons donc faire un pas en arrière et écrire cette espérance comme la limite des sommes partielles :

$$P(N_x < \infty|X_0 = x) = \lim_{N\to\infty} \sum_{n=0}^{N} P\big(T(x) = \infty|X_0 = x\big)P(X_n = x|X_0 = x) = 0 \tag{37.87}$$

parce que tous les termes de la suite des sommes partielles sont nuls. Nous avons donc bien que $P(N_x < \infty|X_0 = x) = 0$. Il s'ensuit immédiatement que $E(N_x|X_0 = x) = 1$.

Nous devons maintenant démontrer l'implication inverse. Supposons que $P(N_x < \infty|X_0 = x) = 0$. Dans ce cas nous avons immédiatement $P(N_x = \infty|X_0 = x) = 1$ et $E(N_x|X_0 = x) = \infty$. L'équation (37.86) nous indique alors que

$$P\big(T(x) = \infty|X_0 = x\big) = 0, \tag{37.88}$$

c'est-à-dire que x est récurrent. □

37.4.1 Chaînes irréductibles

Proposition 37.30.
Soit (X_n) une chaine de Markov irréductible.

(1) Un état x est récurrent si et seulement si tous les états sont récurrents.

(2) Un état x est transient si et seulement si tous les états sont transients.

Démonstration. Soient x et y des états de la chaine de Markov. Nous devons tester la valeur de $P(X_n = y|X_0 = y)$. Afin d'exploiter l'hypothèse d'irréductibilité, nous considérons $r, s \in \mathbb{N}$ tels que

$$p^r(x,y) > 0 \tag{37.89a}$$

$$p^s(y,x) > 0 \tag{37.89b}$$

et nous calculons majorons en passant par quelques intermédiaires :

$$P(X_{n+r+s} = y|X_0 = y) \geqslant P(X_{n+r+s} = y, X_{n+s} = x, X_s = x|X_0 = y) \tag{37.90a}$$
$$= P(X_{n+r+s} = y|X_{n+s} = x, X_s = x, X_0 = y) \tag{37.90b}$$
$$P(X_{n+s} = x|X_s = x, X_0 = y)P(X_s = x|X_0 = y).$$

Les deux premiers facteurs se calculent en utilisant la propriété de Markov et l'homogénéité de la chaine. Pour le premier,

$$P(X_{n+s} = x|X_s = x, X_0 = y) = P(X_{n+s} = x|X_s = x) = P(X_n = x|X_0 = x). \tag{37.91}$$

Nous avons donc

$$\sum_{n\in\mathbb{N}} P(X_{n+r+s} = y|X_0 = y) \geqslant p^r(x, y)p^s(y, x) \sum_{n\in\mathbb{N}} P(X_n = x|X_0 = x). \tag{37.92}$$

En réutilisant Fubini comme dans l'équation (37.82), nous avons

$$\sum_{n\in\mathbb{N}} P(X_{n+r+s} = y|X_0 = y) \geqslant KE(N_x|X_0 = x) \tag{37.93}$$

où K est une constante strictement positive, par hypothèse d'irréductibilité de la chaine de Markov.

Si x est un état récurrent, alors le membre de gauche est infini par la proposition (37.29) et donc

$$\sum_{n\in\mathbb{N}} P(X_{n+r+s} = y|X_0 = y) = \infty. \tag{37.94}$$

Aux $r + s$ premiers termes près (qui ne changent pas la somme), nous avons

$$\sum_{n\in\mathbb{N}} P(X_n = y|X_0 = y) = \infty, \tag{37.95}$$

ce qui signifie que y est récurrent. □

Nous rappelons que $T(x)$ est le temps que première atteinte de l'état x. Nous notons

$$\pi(x) = \frac{1}{E\big(T(x)|X_0 = x\big)}. \tag{37.96}$$

Étant donné que $T(x)$ est un entier positif ou nul nous avons $E\big(T(x)|X_0 = x\big) \in [1, \infty]$ et donc $\pi(x) \in [0, 1]$.

Si x est un état transient, alors $T(x) = \infty$ lorsque $X_0 = x$ et donc $E\big(T(x)|X_0 = x\big) = 0$ et $\pi(x) = 0$. Si x est récurrent par contre, $P\big(T(x) < \infty|X_0 = x\big) = 1$ et il n'y a pas de garanties sur la valeur de $E\big(T(x)|X_0 = x\big)$.

Corolaire 37.31.
Un état récurrent est récurrent positif si et seulement si $\pi(x) > 0$. Un état récurrent est récurrent nul si et seulement si $\pi(x) = 0$.

Démonstration. C'est la formule (37.96). □

Proposition 37.32.
Soit (X_n) est une chaine de Markov irréductible.

(1) Si x est un état récurrent, alors $T(X) < \infty$ presque surement.

(2) Nous avons une égalité entre les lois

$$\mathscr{L}\big(X_{k+T(x)}|T(x) < \infty\big) = \mathscr{L}(X_k|X_0 = x). \tag{37.97}$$

37.4.2 Nombre de visites

La fonction

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{\{X_k=x\}} \tag{37.98}$$

est la **fréquence empirique** de la chaine de Markov.

Soit x un état récurrent, c'est-à-dire que $P\big(T(x)<\infty|X_0=x\big)=1$. Nous classons les visites de la façon suivante :

$$T_1(x)=T(x)=\inf\{k\geqslant 1 \text{ tel que } X_k=x\} \tag{37.99a}$$
$$T_2(x)=\inf\{k\geqslant 1 \text{ tel que } X_{T_1(x)+k}=x\} \tag{37.99b}$$
$$\vdots \tag{37.99c}$$
$$T_n(x)=\inf\{k\geqslant 1 \text{ tel que } X_{T_{n-1}(x)+k}=x\} \tag{37.99d}$$

La variable aléatoire T_i représente le temps entre la visite numéro $i-1$ et la visite numéro i (si $X_0\neq x$, sinon il faut décaler). Nous définissons l'instant la na visite numéro n :

$$S_n=\sum_{k=1}^{n}T_k(x). \tag{37.100}$$

Lemme 37.33.
Les variables aléatoires T_i sont indépendantes.

Démonstration. Nous choisissons n des T_i et nous calculons la probabilité

$$\spadesuit=P(T_{i_1}=k_1,T_{i_2}=k_2,\ldots,T_{i_n}=k_n) \tag{37.101}$$

où nous supposons $i_1>i_2>\ldots>i_n$. Nous décomposons cette probabilité en sommant sur toutes les histoires de la chaine de Markov compatibles avec les nombres k_i donnés :

$$\spadesuit=\sum_{\substack{\{z_j\}\\ \text{compatibles}}}P(X_j=z_j,j=1,\ldots,N). \tag{37.102}$$

Notons qu'ici, le numéro du dernier terme de la somme n'est pas certain parce que tous les T_i ne sont pas fixés. Nous l'avons noté N, mais en réalité il est différent d'un terme à l'autre de la somme. Il est certain que $z_N=x$ et $z_{N-k_1}=x$ et si $N-k_1<j<N$, alors $z_j\neq x$. Cela est simplement le fait que nous demandions aux z_i de respecter les conditions données par les k_i. Nous avons

$$\spadesuit=\sum_{\{z_j\}}P(X_N=x,X_j=z_j,N-k_1<j<N|X_j=z_j,j\leqslant N-k_1)P(X_j=z_j,j<N-k_1) \tag{37.103a}$$
$$=\sum_{\{z_j\}}P(X_N=x,X_j=z_j,N-k_1<j<N|X_{N-k_1}=x)P(X_j=z_j,j<N-k_1) \tag{37.103b}$$
$$\tag{37.103c}$$

Le premier facteur est $P(T_{i_1}=k_1)$ tandis que le second facteur est précisément $P(T_j=k_j,j>1)$. Nous avons donc montré que

$$P(T_{i_1}=k_1,T_{i_2}=k_2,\ldots,T_{i_n}=k_n)=P(T_{i_1}=k_1)P(T_j=k_j,j>1), \tag{37.104}$$

et donc les T_i sont indépendants. □

Proposition 37.34.
Si (X_n) est une chaine de Markov irréductible et si $x \in E$ alors

$$\pi(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k=x\}} \tag{37.105}$$

presque surement.

Démonstration. Étant donné que la chaine est irréductible, les états sont soit tous transient soit tous récurrents par la proposition 37.30. Nous commençons par considérer que x est transient.

En comparant la définition (37.78) de N_x et le membre de droite de (37.105), nous avons pour chaque n l'inégalité

$$\frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k=x\}} \leqslant \frac{1}{n} E(N_x). \tag{37.106}$$

Dans le cas d'un élément transient, nous avons $\pi(x) = 0$, donc il serait bon de montrer que $E(N_x) < \infty$, de sorte que prendre la limite $n \to \infty$ dans (37.106) donne zéro.

Nous décomposons le calcul en deux morceaux :

$$E(N_x) = E\big(N_x|T(x) = \infty\big)P\big(T(x) = \infty\big) + E\big(N_x|T(x) < \infty\big)P\big(T(x) < \infty\big). \tag{37.107}$$

Le fait que le premier terme soit fini découle immédiatement du fait que $T(x) = \infty$ implique $X_k \neq x$ pour tout $k \geqslant 1$. Dans ce cas l'espérance de N_x est évidemment finie.

Pour le second terme nous avons

$$E\big(N_x|T(x) < \infty\big) = E\big(\sum_{k=0}^{\infty} \mathbb{1}_{\{X_k=x\}}|T(x) < \infty\big) \tag{37.108a}$$

$$= \sum_{k=1}^{\infty} E\big(\mathbb{1}_{\{X_k=x\}}|T(x) < \infty\big). \tag{37.108b}$$

Pour inverser la somme et l'espérance, nous avons utilisé le théorème de théorème de Fubini-Tonelli qui est encore valable pour des fonctions qui prennent la valeur ∞. Le fait d'inverser ne signifie pas que ni la somme ni l'intégrale soit finie. D'ailleurs c'est exactement ce que nous sommes en train de déterminer.

Étant donné que nous voulons seulement savoir si cette somme est finie ou non, nous pouvons nous restreindre à la somme depuis $k = 1$ ou oublier le premier terme. D'autre par nous avons

$$\sum_{k=1}^{\infty} \mathbb{1}_{\{X_k=x\}} = \sum_{j=0}^{\infty} \mathbb{1}_{\{X_{j+T(x)}=x\}} \tag{37.109}$$

parce que les $T(x)$ premiers termes sont par définition nuls. Nous regardons donc

$$\sum_{j=0}^{\infty} E\big(\mathbb{1}_{X_{j+T(x)}=x}|T(x) < \infty\big) = \sum_j P\big(X_{j+T(x)} = x|T(x) < \infty\big) \tag{37.110a}$$

$$= \sum_j P(X_j = x|X_0 = x) \tag{37.110b}$$

$$= \sum_j E\big(\mathbb{1}_{\{X_j=x\}}|X_0 = x\big) \tag{37.110c}$$

$$= E\big(\sum_j \mathbb{1}_{X_j=x}|X_0 = x\big) \tag{37.110d}$$

$$= E(N_x|X_0 = x) \tag{37.110e}$$

$$< \infty \qquad \text{parce que } x \text{ est transient.} \tag{37.110f}$$

L'équation (37.110b) provient de la proposition 37.32 et plus précisément de l'égalité entre les lois (37.97). Nous avons terminé la preuve dans le cas où x est transient.

Nous passons maintenant au cas où x est récurrent, c'est-à-dire $P(T(x) < \infty | X_0 = x) = 1$. Les variables aléatoires T_i définies en (37.99) pour $i \geqslant 2$ sont indépendantes et identiquement distribuées et

$$\mathscr{L}\big(T_k(x)\big) \sim \mathscr{L}\big(T(X)|X_0 = x\big). \tag{37.111}$$

La loi des grands nombres nous indique que

$$\frac{S_n}{n} = \frac{T_1(x)}{n} + \frac{1}{n}\sum_{k=2}^{n} T_k(x) \xrightarrow{p.s.} E\big(T_2(x)\big) \tag{37.112a}$$

$$= E\big(T(x)|X_0 = x\big). \tag{37.112b}$$

Remarque 37.35.
La loi des grands nombres est encore vraie sans l'hypothèse de variables aléatoires dans L^1 pourvu qu'elles soient positives. Alors dans la conclusion de la loi nous devons accepter la possibilité que l'espérance soit infinie.

Nous posons pour $m \in \mathbb{N}$

$$n(m) = \sum_{j=1}^{m} \mathbb{1}_{\{X_j = x\}} \tag{37.113}$$

qui est le nombre de visites de x avant l'instant m. Nous avons évidemment $n(m) \leqslant m$. Mais S_n est l'instant de la $n$ième visite, par conséquent $S_{n(m)}$ est l'instant de la dernière visite avant le moment m. Pour tout m nous avons les inégalités

$$S_{n(m)} \leqslant m < S_{n(m)+1}. \tag{37.114}$$

Nous divisons par $n(m)$ et nous effectuons la limite $m \to \infty$:

$$\frac{S_n(m)}{n(m)} \leqslant \frac{m}{n(m)} \leqslant \frac{S_{n(m)} + 1}{n(m)} \tag{37.115}$$

En ce qui concerne la limite de $n(m)$, nous utilisons la définition (37.113) :

$$n(m) \to \sum_{j=1}^{\infty} \mathbb{1}_{\{X_j = x\}} = \tag{37.116}$$

heur...

$$\lim_{m\to\infty} n(m) = \lim_{m\to\infty} \sum_{n=1}^{m} \mathbb{1}_{\{X_j = x\}} \xrightarrow{p.s.} \infty \tag{37.117}$$

par la proposition (37.29). Plus précisément, la limite vaut N_x qui vaut presque surement ∞ dans le cas où x est récurrent. Par ailleurs la loi des grands nombres (37.112) nous enseigne en particulier que

$$\frac{S_{n(m)}}{n(m)} \xrightarrow{p.s.} E\big(T(x)|X_0 = x\big). \tag{37.118}$$

Le terme de droite dans (37.115) se traite de façon usuelle :

$$\frac{S_{n(m)+1}}{n(m)} = \frac{S_{n(m)+1}}{n(m)+1}\frac{n(m)+1}{n(m)}. \tag{37.119}$$

Le dernier facteur tend vers 1 et le tout a pour limite $E\big(T(x)|X_0 = x\big)$. Par conséquent nous avons

$$\frac{m}{n(m)} \xrightarrow{p.s.} E\big(T(x)|X_0 = x\big) \tag{37.120}$$

et

$$\frac{n(m)}{n} = \frac{1}{m}\sum_{j=1}^{m} \mathbb{1}_{\{X_j = x\}} \to \frac{1}{E\big(T(x)|X_0 = x\big)} = \pi(x). \tag{37.121}$$

□

Lemme 37.36.
Soit (X_k) une chaine de Markov dont l'espace des états est noté E. Pour chaque $x \in E$ nous notons

$$T(x) = \inf\{k \geqslant 1 \text{ tel que } X_k = x\} \tag{37.122}$$

et

$$T_p(x) = \inf\{k \geqslant 1 \text{ tel que } X_{k+p} = x\} \tag{37.123}$$

Alors nous avons

$$P(T_p(x) = k | X_p = y) = P(T(x) = k | X_0 = y). \tag{37.124}$$

La proposition suivante nous permet de parler de chaine de Markov **récurrence positive**.

Proposition 37.37.
Soit (x_n) une chaine de Markov irréductible.

(1) Un état x est transient si et seulement si tous les états sont transients.
(2) Un état est récurrent positif si et seulement si tous les états sont récurrents positifs.

Démonstration. Nous rappelons (proposition 37.34) que si la chaine est irréductible

$$\pi(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{[X_k = x]} \tag{37.125}$$

Notons aussi que

$$\sum_{k=1}^{N} \mathbb{1}_{X_k = x} = \begin{cases} 0 & \text{si } N < T(x) \\ \sum_{k=0}^{N-T(x)} \mathbb{1}_{X_{k+T(x)}=x} & \text{si } N > T(x) \end{cases} \tag{37.126}$$

où dans la seconde ligne nous avons effectué le changement de variable de sommation $k' = k + T(x)$. Dans la limite (37.125) nous sommes toujours dans le cas où N est assez grand. Nous pouvons donc écrire

$$\pi(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-T(x)} \mathbb{1}_{X_{k+T(x)}=x}. \tag{37.127}$$

Nous pouvons aussi écrire

$$\frac{1}{N - T(x)} \sum_{k=0}^{N-T(x)} \mathbb{1}_{X_{k+T(x)}=x} = \frac{N}{N - T(x)} \frac{1}{N} \sum_{k=0}^{N-T(x)} \mathbb{1}_{X_{k+T(x)}=x}. \tag{37.128}$$

Dans cette dernière égalité le membre de droite tend vers $\pi(x)$ et nous avons

$$\lim_{N\to\infty} \frac{1}{N - T(x)} \sum_{k=0}^{N-T(x)} \mathbb{1}_{X_{k+T(x)}=x} = \pi(x) \tag{37.129}$$

ou encore

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N} \mathbb{1}_{X_{k+T(x)}=x} = \pi(x) \tag{37.130}$$

Étant donné que $\pi(x)$ est une constante nous avons évidemment $E(\pi(x)) = \pi(x)$. Nous pouvons cependant considérer les variables aléatoires

$$Z_n = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{X_{k+T(x)}=x} \tag{37.131}$$

et remarquer que $Z_n \xrightarrow{p.s.} \pi(x)$ avec $0 \leqslant Z_n \leqslant 1$. Le théorème de la convergence dominée (13.184) nous permet d'inverser la limite et l'espérance et écrire

$$\pi(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} E\big(\mathbb{1}_{X_{k+T(x)}=x}\big) \tag{37.132a}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P\big(X_{k+T(x)} = x\big). \tag{37.132b}$$

Par le lemme 37.36 nous avons

$$P(X_{k+T(x)} = x) = P(X_k = k|X_0 = x) \tag{37.133}$$

et $\pi(x)$ prend la forme

$$\pi(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P(X_k = x|X_0 = x). \tag{37.134}$$

Soit maintenant un état x positif récurrent et y, un autre état. Par définition 37.28 et par corolaire 37.31 nous avons $\pi(x) > 0$. Nous devons prouver que $\pi(y) > 0$.

Étant donné que la chaine est irréductible il existe r et s tels que

$$\begin{cases} p^r(x,y) = P(X_r = y|X_0 = x) > 0 & \text{(37.135a)} \\ p^s(x,y) = P(X_s = x|X_0 = y) > 0 & \text{(37.135b)} \end{cases}$$

Nous reprenons l'équation (37.92) multipliée par $1/N$:

$$\frac{1}{N} \sum_{n=1}^{N} P(X_{r+s+n=y|X_0=y}) \geqslant \underbrace{p^r(x,y)p^s(y,x)}_{\geqslant 0} \underbrace{\frac{1}{N} \sum_{n=1}^{N} P(X_n = x|X_0 = x)}_{\to\pi(x)} \tag{37.136}$$

et nous prenons la limite lorsque $N \to \infty$. À $r+s$ termes près, nous trouvons à gauche l'expression (37.134) de $\pi(y)$. Par conséquent

$$\pi(y) \geqslant \lim_{N\to\infty} \frac{1}{n} \sum_{n=1}^{N} P(X_{r+s+n} = y|X_0 = y) \geqslant \alpha\pi(x) \tag{37.137}$$

où α est une constante positive. Le nombre $\pi(x)$ étant strictement positif par hypothèse nous avons montré que $\pi(y) > 0$, c'est-à-dire que y est récurrent positif. □

37.5 Mesure invariante

Définition 37.38.
Une mesure de probabilité μ sur l'espace des états E d'une chaine de Markov est ***invariante*** *si pour tout $x \in E$*

$$\mu(x) = \sum_{y\in E} p(y,x)\mu(y). \tag{37.138}$$

Remarque 37.39.
Une mesure invariante est une mesure de probabilité et nous noterons par abus $\mu(x)$ pour $\mu(\{x\})$. Si $A \subset E$ nous avons

$$\mu(A) = \sum_{x\in A} \mu(x). \tag{37.139}$$

Remarque 37.40.
Une loi invariante associée à une chaine de Markov est une loi associée à la matrice de transition de la chaine, mais pas à la loi de X_0. Par conséquent nous pouvons tester si μ est une mesure invariante pour une certaine chaine de Markov (X_k) en considérant la chaine (Y_k) avec $Y_k = X_k$ pour $k > 0$ et Y_0 arbitraire.

L'adjectif *invariant* provient du lemme suivant.

Lemme 37.41.
Soit (X_n) une chaine de Markov telle que $X_0 \sim \mu$ où μ est une mesure invariante sur l'espace des états. Alors $X_k \sim \mu$ pour tout k.

Démonstration. Par hypothèse, $P(X_0 = x) = \mu(x)$. Ensuite nous avons

$$P(X_1 = y) = \sum_{x\in E} P(X_1 = y|X_0 = x)P(X_0 = x) \tag{37.140a}$$
$$= \sum_x p(x, y)\mu(x) \tag{37.140b}$$
$$= \mu(y). \tag{37.140c}$$

Par conséquent X_1 suit également la loi μ. Par récurrence tous les états suivent cette même loi. □

Si les états d'une chaine de Markov ont comme loi une mesure invariante, alors nous disons que la chaine est **stationnaire**.

Remarque 37.42.
Pour une chaine de Markov stationnaire de loi invariante μ nous avons

$$\mu(x) = \sum_y p(y, x)\mu(y) \tag{37.141}$$

et si l'ensemble E est fini cette équation signifie

$$\mu = Q\mu \tag{37.142}$$

où Q est la matrice de transition de la chaine de Markov.

Théorème 37.43 (Théorème ergodique)**.**
Une chaine de Markov irréductible est positive récurrente si et seulement si elle accepte une mesure invariante. Cette mesure est invariante est alors unique et vérifie $\mu = Q\mu$ où Q est la matrice de transition.

Démonstration. Nous allons seulement prouver le théorème ergodique dans le cas où E est fini. Soit (X_n) une chaine de Markov récurrente positive ; nous avons $\pi(x) > 0$ pour tout $x \in E$. Nous allons montrer que π est une mesure invariante.

Nous commençons par montrer que

$$\sum_{x\in E} \pi(x) = 1. \tag{37.143}$$

Pour cela nous reprenons la propriété de chaine irréductible pour écrire

$$\pi(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{X_k=x} \tag{37.144}$$

Étant donné que E est fini nous pouvons sommer sur $x \in E$ et permuter la somme avec la limite :

$$\sum_{x\in E} \pi(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \underbrace{\sum_{x\in E} \mathbb{1}_{X_k=x}}_{=1}. \tag{37.145}$$

Nous nous retrouvons donc avec $\lim_{N\to\infty} \frac{1}{N} N = 1$. La fonction π définit donc bien une mesure de probabilité sur E.

Nous montrons à présent que cette mesure est invariante, c'est-à-dire que

$$\pi(x) = \sum_{y\in E} p(y, x)\pi(y). \tag{37.146}$$

Pour cela nous utilisons encore le théorème de la convergence dominée pour permuter la limite et l'intégrale dans

$$\pi(x) = E(\pi(x)) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \underbrace{E\big(\mathbb{1}_{X_k=x}\big)}_{P(X_k=x)} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} P(X_{k+1} = x). \tag{37.147}$$

La dernière égalité découle du fait que en divisant par N et en faisant tendre N vers l'infini, le fait d'enlever un terme à la somme ne change pas la valeur de la limite. Nous pouvons substituer dans (37.147) la valeur

$$P(X_{k+1} = x) = \sum_{y\in E} p(y,x)P(X_k = y). \tag{37.148}$$

Nous avons alors

$$\pi(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \sum_{y\in E} p(y,x)P(X_k = y) \tag{37.149a}$$

$$= \sum_{y\in E} p(y,x) \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} P(X_k = y) \tag{37.149b}$$

$$= \sum_{y\in E} p(y,x)\pi(y), \tag{37.149c}$$

ce qui signifie que π est une mesure invariante. Notons que nous avons encore utilisé le fait que E soit fini pour permuter avec la limite.

Il nous reste à montrer l'unicité de la mesure invariante sur la chaine de Markov. Soit μ une mesure invariante pour la chaine de Markov (X_k). Comme indiqué dans la remarque 37.40 nous pouvons supposer que X_0 suit la loi μ. Par le lemme 37.41 nous avons $P(X_k = x) = \mu(x)$ pour tout k. Par conséquent

$$\pi(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} P(X_k = x) = \mu(x). \tag{37.150}$$

□

Théorème 37.44 (loi des grands nombres pour les chaine de Markov)**.**
Soit (X_n) une chaine de Markov irréductible acceptant une mesure invariante. Soit $f\colon E \to \mathbb{R}$ une fonction dans $L^1(E,\mu)$. Alors nous avons

$$\frac{1}{N} \sum_{k=1}^{N} f(X_k) \xrightarrow{p.s.} \sum_{x\in E} f(x)\mu(x). \tag{37.151}$$

En ce qui concerne les notations, l'hypothèse $f \in L^1(E,\mu)$ signifie

$$\sum_{x\in E} |f(x)|\mu(x) = \int_E |f(x)|d\mu(x) < \infty. \tag{37.152}$$

Démonstration. Nous prouvons le théorème dans le cas où E est fini. Si nous écrivons

$$f(X_k) = \sum_{y\in E} f(y)\mathbb{1}_{X_k=y}, \tag{37.153}$$

alors

$$\frac{1}{N} \sum_{k=1}^{N} f(X_k) = \sum_{y\in E} f(y) \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{X_k=y}. \tag{37.154}$$

Étant donné que E est fini nous pouvons permuter les sommes et prendre la limite $N \to \infty$:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k} f(X_k) = \sum_{y\in E} f(y) \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{X_k=y} = \sum_{y\in E} f(y)\pi(y). \tag{37.155}$$

□

37.6 Convergence vers l'équilibre

Nous voudrions savoir sous quelles conditions la variable aléatoire X_n converge en loi vers quelque chose lorsque $n \to \infty$. Une telle loi limite doit dépendre de la loi initiale[8] comme le

8. Lorsque la loi limite ne dépend pas de la loi initiale, nous disons que la chaine de Markov est ergodique, nous y reviendrons.

montre l'exemple de la chaine de Markov

$$\begin{tikzcd} A \arrow[loop left, "1"] & C \arrow[l, "1/2"'] \arrow[r, "1/2"] & B \arrow[loop right, "1"] \end{tikzcd} \tag{37.156}$$

Si $X_0 = C$, alors la loi limite est

$$\frac{1}{2}(\delta_A + \delta_B). \tag{37.157}$$

Si par contre $X_0 = B$, la loi limite est δ_B. Notons que la chaine de Markov proposée ici est irréductible.

Notons qu'il n'y a pas toujours de lois limite comme le montre l'exemple

$$\begin{tikzcd} A \arrow[r, shift left, "1"] & B \arrow[l, shift left, "1"] \end{tikzcd} \tag{37.158}$$

avec $X_0 = A$. La loi en est

$$X_k = \begin{cases} \delta_A & \text{si } k \text{ est pair} \\ \delta_B & \text{si } k \text{ est impair.} \end{cases} \tag{37.159}$$

Lemme 37.45.
Si nous avons une loi limite

$$P(X_n = x) \to l(x), \tag{37.160}$$

et que la chaine est irréductible, alors nous avons $l = \pi$.

Démonstration. D'après la proposition 37.34 nous avons

$$\frac{1}{n}\sum_{k=1}^{n} P(X_k = x) \to \pi(x). \tag{37.161}$$

Par le lemme 10.74 sur la moyenne de Cesaro et l'hypothèse (37.160), nous avons aussi

$$\frac{1}{n}\sum_{k=1}^{n} P(X_k = x) \to l(x). \tag{37.162}$$

Du coup $\pi(x) = l(x)$. □

Lemme 37.46 ([537])**.**
Si π est une loi stationnaire et si x est un étant transient, alors $\pi(x) = 0$.

Ce lemme (qui peut être prouvé rigoureusement) est principalement dû au fait que la chaine de Markov ne visite un état transitoire qu'un nombre fini de fois par la proposition 37.29(1).

Définition 37.47.
*Un état $x \in E$ est **apériodique** si*

$$\text{pgcd}\{n \geqslant 1 \text{ tel que } p^n(x,x) > 0\} = 1. \tag{37.163}$$

Mettons que tous les n tels que $p^n(x,x) > 0$ ont 2 comme diviseur. L'état n'est alors pas apériodique, mais on voit que si $X_0 = x$, alors les états impairs ne peuvent pas être sur x. Cela est une forme de périodicité.

Si un état est apériodique, il existe p et q premiers entre eux tels que $p^p(x,x)$ et $p^q(x,x)$ sont non nuls. En particulier pour tout $n \in p\mathbb{N}+q\mathbb{N}$, $P(X_n = x) \neq 0$. Par conséquent la proposition 3.15 nous indique qu'à partir d'un certain moment tous les X_k pourraient être x.

L'état C de la chaine de Markov suivante est apériodique :

$$\begin{tikzcd} A \arrow[rr, shift left, "1"] & & B \arrow[ll, shift left, "2/3"] \arrow[dl, "1/3"] \\ & C \arrow[ul, "1"] & \end{tikzcd} \tag{37.164}$$

En effet $p^3(C,C) \neq 0$ par le chemin $C \to A \to B \to C$ tandis que $p^5(C,C) \neq 0$ également par le chemin $C \to A \to B \to A \to B \to C$. Or $\operatorname{pgcd}\{3,5\} = 1$.

Proposition 37.48 ([537])**.**
Soit (X_n), une chaine de Markov irréductible. Un état x est apériodique si et seulement s'il existe N tel que

$$p^k(x,x) = P(X_k = x | X_0 = x) > 0 \tag{37.165}$$

pour tout $k \geqslant N$.

La proposition suivante va nous permettre de parler de **chaine apériodique** .

Proposition 37.49.
Si une chaine de Markov est irréductible, alors un état est apériodique si et seulement si tous les états sont apériodiques.

Démonstration. Soit x un état apériodique de la chaine de Markov $(X_n)_{n\in\mathbb{N}}$. En vertu de la proposition 37.48 il existe N_x tel que $p^k(x,x) \neq 0$ pour tout $k \geqslant N_x$. Soit $y \in E$. Étant donné que la chaine est irréductible, il existe r et s tels que$p^r(x,y) > 0$ et $p^s(y,x) > 0$. Nous avons

$$p^{k+r+s}(y,y) = P(X_{k+r+s} = y | X_0 = y) \geqslant p^s(x,y) P(X_k = x | X_0 = x) p^r(y,x). \tag{37.166}$$

Si k est assez grand, cette quantité est strictement positive. Donc il suffit de prendre $N_y = N_x + r + s$ pour savoir que y est également apériodique. □

Exemple 37.50.
Quelle est la différence entre une chaine irréductible et une chaine apériodique ? Une chaine est irréductible lorsque aucune sous-chaine ne peut piéger le système. Pour toute paire d'états $x, y \in E$, il existe un n tel qu'il soit possible d'aller de x à y en n pas. Une chaine est apériodique lorsqu'après un temps suffisamment long, *tous* les états soient possibles en même temps.

Un exemple de chaine irréductible non apériodique :

$$\begin{tikzcd} A \arrow[r, bend left, "1"] & B \arrow[l, bend left, "1"] \end{tikzcd} \tag{37.167}$$

Cette chaine est irréductible parce que le graphe est connexe, par contre il n'est pas apériodique parce que si $X_0 = A$ il n'est pas possible d'être dans l'état A après un nombre impair de pas.

Plus formellement, $p^n(A,A) = 1$ dès que n est pair ; le PGCD de la définition 37.47 n'est donc certainement pas 1. △

Si E est fini et si la chaine de Markov est irréductible, alors en posant $N = \max_{x\in E} N(x)$, la matrice P^k a des éléments non nuls sur toute la diagonale pour tout $k > N$. Ces éléments diagonaux ne sont autre que les $p^k(x,x)$.

Théorème 37.51 (Convergence en loi des chaine de Markov)**.**
Si (X_n) est

(1) irréductible,

(2) récurrente positive,

(3) apériodique,

alors X_n converge en loi vers l'unique probabilité invariante π vérifiant

$$\pi(x) = \sum_{u\in E} p(y,x)\pi(y) = \frac{1}{E\big(T(x)|X_0 = x\big)}. \tag{37.168}$$

Cette convergence est indépendante de la loi de X_0 et on a

$$P(X_n = x | X_0 = y) \to_{n\to\infty} \pi(x). \tag{37.169}$$

37.7 Processus de Galton-Watson

Nous considérons une maladie et notons Z_n le nombre de malades à l'instant n. Nous posons $Z_0 = 1$ et

$$Z_{n+1} = \begin{cases} 0 & \text{si } Z_n = 0 \\ \sum_{i=1}^{Z_n} \xi_i^{(n)} & \text{sinon} \end{cases} \tag{37.170}$$

où $\xi_i^{(n)}$ est le nombre de personnes contaminées par le malade i à l'instant n. Nous supposons que ces variables aléatoires sont indépendantes et identiquement distribuées et admettent un moment d'ordre 1.

L'équation de propagation 37.170 signifie que nous supposons qu'une personne malade à l'instant n n'est plus malade à l'instant $n+1$. Par ailleurs les hypothèses d'indépendance signifient qu'à chaque instant, le nombre de personnes contaminées par le malade i est indépendant du nombre de personnes contaminées par le malade j. De plus la façon dont la contamination se passe à l'instant n est indépendant de la façon dont la contamination se passe à l'instant m. Ces hypothèses sont raisonnables tant que le nombre de personnes non contaminées est grand. À partir du moment où presque tout le monde est malade, l'approximation de Galton-Watson ne fonctionne plus.

Nous notons ξ la loi parente des $\xi_i^{(n)}$. Ensuite nous considérons

$$G(s) = E(s^\xi) \tag{37.171a}$$

$$m = E(\xi) \tag{37.171b}$$

$$G_n(s) = E(s^{Z_n}). \tag{37.171c}$$

Par le théorème de transfert (proposition 35.68) avec $f(t) = s^t$. Ce que nous avons est

$$G_n(s) = E\big(f(Z_n)\big) = \int_{\mathbb{R}} s^x dP_{Z_n}(x) = \sum_{k=0}^{\infty} s^k P(Z_n = k) \tag{37.172}$$

où l'intégrale s'est transformée en somme parce que la loi de Z_n est discrète : dP_{Z_n} est une somme de masses de Dirac. En particulier nous avons

$$G_n(s) = \sum_{k=0}^{\infty} s^k P(Z_n = k) \tag{37.173a}$$

$$G(0) = P(Z_n = 0) \tag{37.173b}$$

et

$$\eta = \lim_{n\to\infty} P(Z_n = 0) = \lim_{n\to\infty} G_n(0). \tag{37.174}$$

D'où l'intérêt d'étudier G_n.

Lemme 37.52.
Pour tout $n \in \mathbb{N}^*$ et pour tout $s \in [0,1]$, nous avons

$$G_n(s) = \underbrace{G \circ G \circ \ldots \circ G(s)}_{n \text{ fois}}. \tag{37.175}$$

Démonstration. Pour $n = 1$, nous avons $Z_1 = \xi_1^{(1)}$ et donc

$$G_1(s) = E(s^\xi) = G(s), \tag{37.176}$$

comme il se doit.

Si $n \neq 1$ nous écrivons

$$G_n(s) = E(s^{Z_n}) \tag{37.177a}$$

$$= E\left(s^{\sum_{i=1}^{Z_{n-1}} \xi_i^{(n-1)}}\right) \tag{37.177b}$$

$$= E\left(\sum_{k=0}^{\infty} \mathbb{1}_{\{Z_{n-1}=k\}} s^{\sum_{i=1}^{k} \xi_i^{(n-1)}}\right). \tag{37.177c}$$

À ce niveau, nous voulons permuter la somme et l'espérance. Étant donné que le lemme est facile à vérifier pour $s = 1$, nous supposons $s < 1$. Du coup

$$s^{\sum_{i=1}^{k} \xi_i^{(n-1)}} < 1 \tag{37.178}$$

et ce qui se trouve dans l'espérance est majoré par

$$\sum_{k=0}^{\infty} \mathbb{1}_{Z_{n-1}=k} = 1. \tag{37.179}$$

La fonction constante 1 est intégrable sur Ω (ici nous utilisons à fond le fait que l'espace Ω soit un espace de probabilité) et nous pouvons utiliser le théorème de convergence dominée de Lebesgue 14.2 pour permuter la somme et l'intégrale. Nous continuons donc le calcul (37.177) :

$$G_n(s) = \sum_{k=0}^{\infty} E\left(\mathbb{1}_{\{Z_{n-1}=k\}} s^{\sum_{i=1}^{k} \xi_i^{(n-1)}}\right). \tag{37.180}$$

La tribu engendrée par la variable aléatoire $\mathbb{1}_{\{Z_{n-1}=k\}}$ est une fonction des variables aléatoires $\xi_i^{(m)}$ avec $m \leqslant n-2$ tandis que la variable aléatoire $s^{\sum_{i=1}^{k} \xi_i^{(n-1)}}$ est une fonction des variables aléatoires $\xi_i^{(n-1)}$. Par conséquent le lemme de regroupement 35.16 nous dit que ces variables aléatoires sont indépendantes, donc

$$G_n(s) = \sum_{k=0}^{\infty} \underbrace{E\big(\mathbb{1}_{\{Z_{n-1}=k\}}\big)}_{=P(Z_{n-1}=k)} E\big(s^{\sum_{i=1}^{k} \xi_i^{(n-1)}}\big). \tag{37.181}$$

Nous avons utilisé le fait que l'espérance d'une fonction indicatrice est la probabilité de l'événement.

En ce qui concerne la puissance de s, les événements ξ_i^{n-1} sont indépendants et suivent tous la même loi ξ, donc

$$s^{\sum_{i=1}^{k} \xi_i^{(n-1)}} = \prod_{i=1}^{k} s^{\xi_i^{(n-1)}} \tag{37.182}$$

et

$$E\big(\prod_{i=1}^{k} s^{\xi}\big) = E(s^{\xi})^k = G(s)^k. \tag{37.183}$$

En mettant tout bout à bout,

$$G_n(s) = \sum_{k=1}^{\infty} P(Z_{n-1} = k) G(s)^j = G_{n-1}\big(G(s)\big). \tag{37.184}$$

□

Théorème 37.53.
La probabilité d'extinction η est donnée par

$$\eta = P\left(\bigcup_{n\geqslant 1} (Z_n = 0)\right) = \lim_{n\to\infty} P(Z_n = 0). \tag{37.185}$$

Ce nombre est la plus petite solution positive de l'équation $G(s) = s$.

De plus la classification des cas est comme suit.

(1) Si $P(\xi = 0) = 0$ alors $\eta = 0$.

(2) Si $P(\xi = 0) \neq 0$ alors

(2a) si $m \leqslant 1$ alors $\eta = 1$,

(2b) si $m > 1$ alors $\eta \in]0, 1[$.

Le cas $m < 1$ est dit **sous-critique**, le cas $m = 1$ est dit **critique**. Le cas $m > 1$ est dit **sur-critique**.

Démonstration. Commençons par prouver que G est une fonction continue. En utilisant la théorème de transfert comme pour l'équation (37.172) nous trouvons que

$$G(s) = E(s^\xi) = \sum_{k=0}^{\infty} p_k s^k \tag{37.186}$$

où nous avons noté $p_k = P(\xi = k)$. Si $r < 1$, alors la suite $p_k r^k$ est bornée, donc le critère d'Abel (14.10) nous indique que la série (37.186) converge absolument et la théorie générale des séries entières conclut que la fonction G est en particulier dérivable terme à terme pour tout $s \in]-1, 1[$.

Le probabilité d'extinction est un point fixe de G En utilisant la continuité de G en 0 nous passons à la limite dans $G_{n+1}(0) = G\big(G_n(0)\big)$ et nous obtenons

$$\eta = G(\eta), \tag{37.187}$$

ce qui signifie que la probabilité d'extinction est un point fixe de G.

η est le plus petit point fixe de G Nous démontrons maintenant que η est plus précisément le plus petit point fixe de G sur $[0, 1]$. Nous allons effectuer cette partie en décomposant selon les valeurs de p_0 et de p_1.

Au vu de l'écriture (37.186), si $p_1 = 1$ alors $G(s) = s$ pour tout $s \in [0, 1]$. Mais dans ce cas nous savons par ailleurs que l'extinction est impossible. Zéro est bien la plus petite solution de $G(s) = s$.

Supposons maintenant que $p_1 < 1$ et $p_0 + p_1 = 1$. Alors $G(s) = p_0 + p_1 s$ et $s = 1$ est l'unique solution. Mais vu que nous savons que η est solution, c'est que $\eta = 1$ et l'extinction est certaine.

Nous passons au cas général : $p_0 + p_1 < 1$. D'abord nous remarquons que $s = 1$ est solution parce que

$$G(1) = p_0 + p_1 + \cdots = 1. \tag{37.188}$$

Remarquons aussi que dans ce cas $s = 0$ n'est plus solution.

La fonction G est strictement convexe sur $[0, 1]$ (parce que $G'' > 0$). Cela se voir en effectuant deux dérivations termes à termes (le rayon de convergence de la dérivée est le même que celui de la fonction). Cette stricte convexité entraine que l'équation $G(s) = s$ a au maximum une autre solution que $s = 1$. Nous nommons s_0 la plus petite solution dans $[0, 1]$. Étant donné que G est croissante on a

$$G(0) \leqslant G(s_0) = s_0. \tag{37.189}$$

En appliquant G à cette équation nous obtenons $G\big(G(s_0)\big) \leqslant G(s_0) = s_0$ et en appliquant n fois,

$$G_n(0) \leqslant s_0. \tag{37.190}$$

En passant à la limite, $\eta \leqslant s_0$ mais η étant solution, nous avons $\eta = s_0$. Nous avons donc prouvé que la probabilité d'extinction η est la plus petite solution de $G(s) = s$.

Classification des cas Nous devons encore discuter les cas. Si $P(\xi = 0) = 0$, alors $p_0 = 0$ et $G(0) = 0$, ce qui signifie que $s_0 = \eta = 0$ et l'extinction est impossible.

Nous passons au cas $p_0 \neq 0$. Si $p_0 + p_1 = 1$, alors $m = p_1 < 1$ et nous avions déjà vu que dans le cas $p_0 + p_1 = 1$, la probabilité d'extinction est $\eta = 1$.

Il nous reste à traiter le cas $p_0 + p_1 < 1$. Encore une fois, la courbe G est strictement convexe sur $[0, 1]$ et elle est en particulier plus grande que sa tangente en $s = 1$, c'est-à-dire

$$G(s) > G'(1)(s - 1) + G(1). \tag{37.191}$$

Nous savons que $G(1) = 1$. En ce qui concerne $G'(1)$, nous dérivons encore terme à termes :

$$G'(s) = \sum_{k=1}^{\infty} kp_k s^{k-1}, \tag{37.192}$$

donc

$$G'(1) = \sum_{k=1}^{\infty} kp_k = E(\xi) = m. \tag{37.193}$$

Ce que nous avons donc est

$$G(s) > 1 + m(s-1). \tag{37.194}$$

Nous nous particularisons au cas sous-critique ($m \leqslant 1$). En nous rappelant que $s - 1 < 0$,

$$G(s) > 1 + (s-1) = s, \tag{37.195}$$

donc $s = 1$ est la plus petite solution et effectivement nous avons déjà vu que $\eta = 1$ dans ce cas.
Si $m > 1$, alors on a

$$G(s) > 1 + m(s-1). \tag{37.196}$$

Mais dire $m > 1$ revient à dire $G'(1) > 1$ et donc dans un voisinage de $s = 1$ on a

$$\frac{G(s) - G(1)}{s-1} > 1, \tag{37.197}$$

ce qui implique que

$$G(s) < s - 1 + G(1) = s. \tag{37.198}$$

Nous avons donc $G(s) < s$ dans un voisinage de 1. Mais $G(0) - 0 = p_0 > 0$, donc la fonction $f(s) = G(s) - s$ est positive en 0 et négative proche de $s = 1$. Le théorème de la valeur intermédiaire nous indique alors qu'il existe un $s \in]0, 1[$ tel que $f(s) = 0$, c'est-à-dire tel que $G(s) = s$.

□

Chapitre 38

Martingales

38.1 Convergence de martingales

Définition 38.1.
*Si $\mathcal{A}$ est une tribu, une **filtration** de $\mathcal{A}$ est une suite croissante de sous-tribus $\mathcal{B}_i \subseteq \mathcal{B}_{i+1} \subseteq \mathcal{A}$.*

*Nous disons qu'une suite de variables aléatoires (X_n) est **adaptée** à une filtration $(\mathcal{F}_n)$ si X_i est $\mathcal{F}_i$-mesurable pour tout i.*

Ces définitions impliquent immédiatement que si (X_n) est adapté à $(\mathcal{F}_n)$ alors X_n est $\mathcal{F}_k$-mesurable pour $k \geqslant n$.

Définition 38.2.
*Une **martingale** adaptée à la filtration $(\mathcal{B}_n)_{n\in\mathbb{N}}$ est une suite de variables aléatoires M_n telle que*

(1) $M_n \in L^1(\Omega, \mathcal{A}, P)$,

(2) M_n est $\mathcal{B}_n$-mesurable,

(3) $E(M_{n+1}|\mathcal{B}_n) = M_n$.

*Le processus M_n est une **sur-martingale** si $E(M_{n+1}|\mathcal{B}_n) \leqslant M_n$ pour tout n, et c'est une **sous-martingale** si $E(M_n|\mathcal{B}_n) \geqslant M_n$.*

Exemple 38.3.
Si $M \in L^1(\Omega, \mathcal{A}, P)$ et si $(\mathcal{B}_n)_{n\in\mathbb{N}}$ est une filtration, nous pouvons considérer la martingale $M_n = E(M|\mathcal{B}_n)$. △

Exemple 38.4.
Soit $(X_i)_{i\geqslant 1}$ une suite de variables aléatoires indépendantes et centrées. On pose

$$S_n = X_1 + \cdots + X_n \tag{38.1}$$

et la filtration $\mathcal{B}_n = \sigma(X_1, \ldots, X_n)$. Pour montrer que cela est une martingale, nous commençons par remarquer que

$$E(X_{n+1}|\mathcal{B}_n) = E(X_{n+1}) = 0 \tag{38.2}$$

par indépendance des tribus $\mathcal{B}_n$ et $\sigma(X_{n+1})$. Ici c'est le lemme 35.44 qui joue.

Ensuite nous argumentons que $E(X_1 + \cdots + X_n|\mathcal{B}_n) = X_1 + \cdots + X_n$. En effet d'une part $X_1 + \cdots + X_n$ est $\mathcal{B}_n$-mesurable et évidemment la condition intégrale de l'espérance conditionnelle est satisfaite.

Plus généralement si X est une variable aléatoire et si $\sigma(X) \subset \mathcal{B}$ alors $E(X|\mathcal{B}) = X$. △

Lemme 38.5.
Soit (M_n) une martingales adaptée à la filtration $(\mathcal{F}_n)$ et $n \geqslant k$. Alors

$$E(M_n|\mathcal{F}_k) = M_k \tag{38.3a}$$

$$E(M_k|\mathcal{F}_n) = M_k. \tag{38.3b}$$

Démonstration. La seconde relation revient seulement à dire que M_k est $\mathcal{F}_n$-mesurable, ce qui est évident parce que $\mathcal{F}_k \subset \mathcal{F}_n$.

Nous prouvons la première par récurrence (à l'envers) sur k. D'abord si $k = n$, l'égalité $E(M_n|\mathcal{F}_n) = M_n$. Nous supposons maintenant que $E(M_n|\mathcal{F}_k) = M_k$, et nous prouvons que $E(M_n|\mathcal{F}_{k-1}) = M_{k-1}$. Si $B_{k-1} \in \mathcal{F}_{k-1}$, nous avons

$$\int_{B_{k-1}} M_{k-1} = \int_{B_{k-1}} M_k = \int_{B_{k-1}} M_n. \tag{38.4}$$

La première égalité est la définition d'une martingale, et la seconde est l'hypothèse de récurrence. □

Théorème 38.6 ([538, 539]).
Soit $(M_n)_{n\geqslant 0}$ une martingale bornée dans $L^2(\Omega)$, c'est-à-dire telle que

$$\alpha = \sup_{n\geqslant 0} E(M_n^2) < \infty. \tag{38.5}$$

Alors la suite M_n converge dans $L^2(\Omega)$.

Démonstration. Nous écrivons M_n en somme télescopique

$$M_n = M_0 + \sum_{k=1}^{n} \Delta_k \tag{38.6}$$

où $\Delta_k = M_k - M_{k-1}$. Nous commençons par monter que les incréments sont orthogonaux au sens où $E(\Delta_n\Delta_k) = 0$. Pour $n > k$, la variable aléatoire $E\big(\Delta_n\Delta_k|\mathcal{F}_{n-1}\big)$ est la variable aléatoire $\mathcal{F}_{n-1}$-mesurable telle que

$$\int_{B_{n-1}} E\big(\Delta_n\Delta_k|\mathcal{F}_{n-1}\big) = \int_{B_{n-1}} \Delta_n\Delta_k \tag{38.7}$$

pour tout $B_{n-1} \in \mathcal{F}_{n-1}$. En particulier avec $B_{n-1} = \Omega$ nous trouvons

$$E\Big(E\big(\Delta_n\Delta_k|\mathcal{F}_{n-1}\big)\Big) = E(\Delta_n\Delta_k) \tag{38.8}$$

par la définition de l'espérance (35.52). Par conséquent, en utilisant le lemme 38.5 nous avons [1]

$$E(\Delta_n\Delta_k) = E\Big(E(\Delta_n\Delta_k|\mathcal{F}_{n-1})\Big) = E\Big(\Delta_k E(\Delta_n|\mathcal{F}_{n-1})\Big) = 0 \tag{38.9}$$

parce que $E(\Delta_n|\mathcal{F}_{n-1}) = E(M_n|\mathcal{F}_{n-1}) - E(M_{n-1}|\mathcal{F}_{n-1}) = 0$.

Utilisant l'orthogonalité des incréments, nous avons

$$E(M_n^2) = E(M_0^2) + \sum_{k=1}^{n} E(\Delta_k^2). \tag{38.10}$$

En prenant le supremum (par rapport à n des deux côtés),

$$E(M_0^2) + \sum_{k=1}^{\infty} E(\Delta_k^2) = \alpha < \infty. \tag{38.11}$$

Cela prouve que la suite $\sum_{k=1}^{n} \Delta_k$ converge dans $L^2(\Omega)$. Nous en déduisons immédiatement que (M_n) est de Cauchy dans $L^2(\Omega)$ parce que si $k, l > n$, nous avons (en utilisant encore l'orthogonalité des incréments)

$$E\big(|M_k - M_l|^2\big) = \sum_{i=k+1}^{l} E(\Delta_i^2) \leqslant \sum_{i=k+1}^{\infty} E(\Delta_i^2), \tag{38.12}$$

qui tend vers zéro lorsque $n \to \infty$. □

1. À ce niveau je crois qu'il y a une faute dans [539] qui conditionne par rapport à $\mathcal{F}_n$.

Le théorème suivant complète la conclusion du théorème 38.6.

Théorème 38.7 ([539]).
Soit $(M_n)_{n\in\mathbb{N}}$ une martingale bornée dans L^2. Alors (M_n) converge dans $L^2(\Omega)$ et presque surement vers une même variable aléatoire M_∞ qui vérifie

$$M_n = E(M_\infty|\mathcal{F}_n). \tag{38.13}$$

Notons en particulier que la variable aléatoire M_∞ est presque surement finie parce qu'en vertu de (38.13) nous avons

$$\int_\Omega M_\infty = \int_\Omega M_n < \infty. \tag{38.14}$$

Exemple 38.8.
Soient des variables aléatoires indépendantes $V_k \sim \mathscr{E}(2^n\lambda)$ et la variable aléatoire somme

$$S_n = \sum_{k=1}^n V_k. \tag{38.15}$$

Nous allons montrer que $S_n \xrightarrow{p.s.} X$ où X est une variable aléatoire presque surement finie. Nous posons

$$M_n = S_n - \sum_{k=1}^n \frac{1}{2^k\lambda} \tag{38.16}$$

Cela est une martingale adaptée à la filtration $\mathcal{F}_n = \sigma(V_1,\dots,V_n)$ en vertu de l'exemple 38.4. Nous montrons à présent qu'elle est bornée dans $L^2(\Omega)$ au sens où $\sum_{n\geqslant 1} E(M_n^2) < \infty$. Nous avons

$$E(M_n^2) = E\left(\left[S_n - \sum_k \frac{1}{2^k\lambda}\right]^2\right) = E\left(\left[\sum_k (V_k - \frac{1}{2^k\lambda})\right]^2\right). \tag{38.17}$$

La variable aléatoire $V_k - 1/2^k\lambda$ est une variable aléatoire centrée de variance $1/(2^k\lambda)^2$ (voir proposition 35.105). Étant donné que M_n est centrée, $\mathrm{Var}(M_n) = E(M_n^2)$ et nous avons

$$E(M_n^2) = \sum_{k=1}^n \mathrm{Var}\left(V_k - \frac{1}{2^k\lambda}\right) = \sum_{k=1}^n \frac{1}{(2^k\lambda)^2}, \tag{38.18}$$

cette dernière somme étant bornée par $l = \sum_{k=1}^\infty \frac{1}{(2^k\lambda)^2}$, nous avons

$$E(M_n^2) \leqslant l \tag{38.19}$$

avec l indépendant de n. C'est pour cela que $(M_n)_{n\in\mathbb{N}}$ est une martingale bornée dans $L^2(\Omega)$. Par le théorème 38.7 nous avons $M_n \to M_\infty$ et en faisant $n \to \infty$ dans

$$S_n = M_n + \sum_{k=1}^n \frac{1}{2^k\lambda}, \tag{38.20}$$

nous trouvons

$$S_n \to M_\infty + \sum_{k=1}^\infty \frac{1}{2^k\lambda} = M_\infty + \frac{1}{\lambda} \tag{38.21}$$

qui est presque surement finie. △

38.2 Temps d'arrêt et martingale terminée

Définition 38.9.
Soit $(\Omega, \mathcal{F}_n, \mathcal{F}, P)$ un espace de probabilité filtré. Une application $T\colon \Omega \to \bar{\mathbb{N}}$ est un ***temps d'arrêt*** *adapté à la filtration $(\mathcal{F}_n)$ si pour tout $n \in \mathbb{N}$ nous avons $\{T \leqslant n\} \in \mathcal{F}_n$.*

Le temps d'arrêt T est ***borné*** *s'il existe $k \in \mathbb{N}$ tel que $T(\omega) \leqslant k$ pour presque tout $\omega \in \Omega$.*

Lemme 38.10.
Si T est un temps d'arrêt presque surement fini, alors[2]

— $T \wedge n \xrightarrow{p.s.} T$,

— $\lim_{n\to\infty} E(T \wedge n) = E(T)$.

Démonstration. Vu que T est presque surement finie, il suffit de prouver que

$$(T \wedge n)(\omega) \to T(\omega) \tag{38.22}$$

pour tout ω tel que $T(\omega) = k$ pour tout $k \in \mathbb{N}$. Soient donc $\omega \in \Omega$ tel que $T(\omega) = k$ et $n > k$. Nous avons

$$(T \wedge n)(\omega) = T(\omega) \wedge n = k = T(\omega). \tag{38.23}$$

En ce qui concerne la seconde assertion, la suite de variables aléatoires $X_x = T \wedge n$ est croissante et positive, donc le théorème de la convergence monotone 13.160 montre que

$$\lim_{n\to\infty} E(T \wedge n) = E(T). \tag{38.24}$$

□

Remarque 38.11.
Notons la différence subtile entre $S_T(\omega)$ et $(S_T)(\omega)$. La première est la variable aléatoire

$$\omega' \mapsto S_{T(\omega')}(\omega) \tag{38.25}$$

et la seconde est le nombre $S_{T(\omega)}(\omega)$.

Théorème 38.12 (Théorème d'arrêt borné[539])**.**
Soit (X_n) une sur-martingale et $S \leqslant T$, deux temps d'arrêts bornés. Alors

(1) les variables aléatoires X_S et X_T sont intégrables,

(2) $E(X_T|\sigma(S)) \leqslant X_S$ presque surement.

Si par contre (X_n) est une martingale alors X_S et X_T sont bornées, et

$$E(X_T|\sigma(S)) = X_S. \tag{38.26}$$

Remarque 38.13.
Un cas particulier intéressant de ce théorème 38.12 est le cas $S = 0$ qui est un temps d'arrêt vérifiant $\mathcal{F}_0 = \{\Omega, \emptyset\}$. Si X est n'importe quelle variable aléatoire, la tribu engendrée $\sigma(X)$ est toujours indépendante de la tribu $\{\Omega, \emptyset\}$, donc le résultat $E(X_T|\mathcal{F}_S) = X_S$ donne

$$E(X_T) = X_0. \tag{38.27}$$

Théorème 38.14 (Premier théorème d'arrêt de Doob[540])**.**
Soient (X_n) une martingale et T un temps d'arrêt ; tous deux pour la filtration $(\mathcal{F}_n)$. Nous supposons qu'une des trois propriétés suivantes soit vérifiée :

(1) T est presque surement bornée.

2. Dans [54], dans le problème de la ruine du joueur, la seconde assertion est avec une limite sup et non avec une limite normale.

(2) $E(T) < \infty$ et il existe une constante c telle que

$$E\big(|X_{n+1} - X_n|\,|\mathcal{F}_n\big) \leqslant c \tag{38.28}$$

sur l'événement $\{T \geqslant n\}$.

(3) Il existe une constante c telle que $|X_{T\wedge n}| \leqslant c$ presque surement[3].

Alors X_T est une variable aléatoire presque surement bien définie nous avons

$$E(T_T) = E(X_0). \tag{38.29}$$

Si (X_n) est une sur-martingale, alors la conclusion est $E(X_T) \leqslant E(X_0)$ et si (X_n) est une sous-martingale, la conclusion est $E(X_T) \geqslant E(X_0)$.

Remarque 38.15.
Sous l'hypothèse (3), il est possible d'avoir $T = \infty$ sur un ensemble de mesure non nulle. Sur cet ensemble, la variable aléatoire X_T doit être définie de façon plus fine.

Problèmes et choses à faire

D'après la page de discussion de l'article sur Wikipédia, il semblerait que la seconde condition soit mal énoncée. Je n'ai pas vérifié.

Définition 38.16.
*Nous disons que la martingale $(M_n)_{n\geqslant 1}$ est **terminée** s'il existe $M \in L^1(\Omega, \mathcal{A}, P)$ telle que $E(M|\mathcal{A}_n) = M$ pour tout $n > 1$.*

Définition 38.17.
*Un ensemble $H \subset L^1(\Omega, \mu)$ est **équi-intégrable** si*

$$\lim_{a\to\infty} \left(\sup_{f\in H} \int_{|f|>a} |f(x)| d\mu(x) \right) = 0. \tag{38.30}$$

Notons dans cette définition que vu que $f \in L^1$ nous avons toujours

$$\lim_{a\to\infty} \int_{|f|>a} |f(x)| d\mu(x) = 0. \tag{38.31}$$

L'équi-intégrabilité donne une sorte d'uniformité en f de cette limite.

Théorème 38.18.
Si (M_n) est une martingale, nous avons équivalence entre

(1) (M_n) converge dans L^1 ;

(2) (M_n) est terminée ;

(3) l'ensemble $\{M_n\}_{n\geqslant 1}$ est équi-intégrable.

Attention : en vertu de la proposition 26.19 et surtout de l'exemple 26.20, la convergence L^1 n'implique pas la convergence presque partout.

Théorème 38.19 (Théorème de Doob[270])**.**
À propos de convergence de martingales.

(1) Toute martingale terminée converge presque surement et pour la norme L^1.

(2) Toute martingale bornée dans L^2 converge presque surement et pour la norme L^2.

Proposition 38.20 ([541])**.**
Soient (M_n) une martingale et T un temps d'arrêt (pour la même filtration $(\mathcal{B}_n)$). Alors le processus $V_n = M_{n\wedge T}$ est une martingale.

3. Il est d'usage assez classique de noter $a \wedge b$ le minimum de a et b.

Démonstration. Nous décomposons V_n de la façon suivante :

$$V_n = M_{n\wedge T} = M_n \mathbb{1}_{T\geqslant n} + M_T \mathbb{1}_{T<n} = M_n \mathbb{1}_{T\geqslant n} + \sum_{k<n} M_k \mathbb{1}_{T=k}. \tag{38.32}$$

Nous avons, grâce au lemme 13.3,

$$\{T \geqslant n\} = \complement\{T < n\} = \complement\{T \leqslant n-1\} \in \mathcal{B}_{n-1} \tag{38.33}$$

et, si $k \leqslant n$,

$$\{T = k\} = \underbrace{\{T \leqslant k\}}_{\in\mathcal{B}_k} \setminus \underbrace{\{T \leqslant k-1\}}_{\in\mathcal{B}_{k-1}} \in \mathcal{B}_k \subset \mathcal{B}_n. \tag{38.34}$$

La forme (38.32) donne donc manifestement la $\mathcal{B}_n$-mesurabilité de V_n.

En ce qui concerne l'espérance nous devons calculer

$$E(V_{n+1}|\mathcal{B}_n) = E(M_{n+1}\mathbb{1}_{T\geqslant n+1}|\mathcal{B}_n) + \sum_{k<n+1} E(M_k \mathbb{1}_{T=k}|\mathcal{B}_n) \tag{38.35}$$

où nous avons utilisé la proposition 35.29. Étant donné que $\mathbb{1}_{T\geqslant n+1}$ et $\mathbb{1}_{T=k}$ sont des variables aléatoires $\mathcal{B}_n$-mesurables nous pouvons utiliser la proposition 35.47 pour les sortir :

$$E(V_{n+1}|\mathcal{B}_n) = \mathbb{1}_{T\geqslant n+1} M_n + \sum_{k\leqslant n} \mathbb{1}_{T=k} M_k = M_{T\wedge n} = V_n. \tag{38.36}$$

Pour cette ligne, nous avons aussi utilisé les égalités suivantes :

— $E(M_{n+1}|\mathcal{B}_n) = M_n$ parce que (M_n) est une martingale

— $E(M_k|\mathcal{B}_n) = M_k$ parce que M_k est $\mathcal{B}_n$-mesurable.

□

Définition 38.21.
*Si (X_n) est un processus adapté à la filtration $(\mathcal{F}_n)$ et si T est un temps d'arrêt $\mathcal{F}_n$-mesurable alors le **processus arrêté** à l'instant T est le processus $Y_n = X_{n\wedge T}$.*

Nous avons déjà vu par la proposition 38.20 que si (X_n) est une martingale alors son processus arrêté est encore une martingale.

38.3 Décomposition de martingales

Définition 38.22 (Processus croissant prévisible[539])**.**
*Un processus X_n adapté à la filtration $\mathcal{F}_n$ est un processus **croissant prévisible** si*

(1) $A_0 = 0$

(2) $A_n \leqslant A_{n+1}$ *; c'est cette condition qui correspond à « croissant »,*

(3) A_{n+1} *est $\mathcal{F}_n$-mesurable ; c'est cette condition qui correspond à « prévisible ».*

Proposition 38.23 (Décomposition de Doob pour une sous-martingale[539])**.**
Toute sous-martingale (X_n) s'écrit de façon unique sous la forme

$$X_n = M_n + A_n \tag{38.37}$$

où (M_n) est une martingale et (A_n) est un processus croissant prévisible.

Démonstration. Nous considérons le processus

$$\begin{cases} A_0 = 0 & (38.38a) \\ A_{n+1} = A_n + E(X_{n+1} - X_n|\mathcal{F}_n). & (38.38b) \end{cases}$$

Nous vérifions que cela est un processus croissant prévisible. D'abord

$$E(X_{n+1} - X_n|\mathcal{F}_n) = E(X_{n+1}|\mathcal{F}_n) - E(X_n|\mathcal{F}_n). \tag{38.39}$$

Le second terme est égal à X_n parce que cette variable aléatoire est $\mathcal{F}_n$-mesurable tandis que (X_n) étant une sous-martingale nous avons $E(X_{n+1}|\mathcal{F}_n) \geqslant X_n$. Nous avons donc bien $A_{n+1} \geqslant A_n$ et le processus (A_n) est croissant.

En ce qui concerne la prévisibilité nous devons prouver que A_{n+1} est $\mathcal{F}_n$-mesurable. D'une part A_n est $\mathcal{F}_n$-mesurable et d'autre part par définition de l'espérance conditionnelle, la variable aléatoire $E(X_{n+1} - X_n|\mathcal{F}_n)$ est également $\mathcal{F}_n$-mesurable.

Nous posons alors $M_n = X_n - A_n$ et nous devons prouver que cela est une martingale. Nous avons

$$E(M_{n+1} - M_n|\mathcal{F}_n) = E(X_{n+1} - X_n|\mathcal{F}_n) - E(A_{n+1} - A_n|\mathcal{F}_n). \tag{38.40}$$

Le second terme vaut

$$E(A_{n+1} - A_n|\mathcal{F}_n) = E\Big(E(X_{n+1} - X_n|\mathcal{F}_n)|\mathcal{F}_n\Big) = E(X_{n+1} - X_n|\mathcal{F}_n) \tag{38.41}$$

par la proposition 35.42. Le processus (M_n) est donc une martingale. La preuve de l'existence d'une décomposition (38.37) est achevée.

Nous passons maintenant à l'unicité en posant $X_n = M_n + A_n = M'_n + A'_n$. Nous avons $A_0 = A'_0 = 0$ et $A'_n = X_n - M'_n$, donc

$$A'_{n+1} - A'_n = X_{n+1} - X_n + M'_{n+1} - M'_n = X_{n+1} - X_n - (M'_{n+1} - M'_n). \tag{38.42}$$

Nous appliquons $E(.|\mathcal{F}_n)$ des deux côtés de cette égalité :

$$\underbrace{E(A'_{n+1} - A'_n|\mathcal{F}_n)}_{=A'_{n+1}-A'_n} = E(X_{n+1} - X_n|\mathcal{F}_n) - \underbrace{E(M'_{n+1} - M'_n|\mathcal{F}_n)}_{=0}. \tag{38.43}$$

Nous avons utilisé le que que (M_n) étant une martingale, $E(M_{n+1} - M_n\mathcal{F}_n) = 0$, et idem avec (M'_n). Donc

$$A'_{n+1} - A'_n = E(X_{n+1} - X_n|\mathcal{F}_n) = E(M_{n+1} - M_n|\mathcal{F}_n) + E(A_{n+1} - A_n|\mathcal{F}_n) = A_{n+1} - A_n. \tag{38.44}$$

Nous avons donc montré que $A_{n+1} - A_n = A'_{n+1} - A'_n$ et donc que $A_n = A'_n$ pour tout n. Nous en déduisons immédiatement que $M_n = M'_n$ pour tout n et l'unicité de la décomposition. □

Lemme 38.24.
Si (X_n) est une martingale de carré intégrable adaptée à la filtration $(\mathcal{F}_n)$ alors

(1) Le processus (X_n^2) est une sous-martingale.

(2) Si $X_n^2 = M_n + A_n$ est la décomposition de Doob, alors

$$A_n = \sum_{i=1}^{n} \Big(E(X_i^2|\mathcal{A}_{i-1}) - X_{i-1}^2\Big) = \sum_{i=1}^{n} E\Big((X_i - X_{i-1})^2|\mathcal{A}_{i-1}\Big). \tag{38.45}$$

Démonstration. Pour la première assertion, nous utilisons l'inégalité de Jensen 35.60 :

$$E(X_n^2|\mathcal{F}_{n-1}) \geqslant \big(E(X_n|\mathcal{F}_{n-1})\big)^2 = X_n^2 \tag{38.46}$$

parce que $E(X_n|\mathcal{F}_{n-1}) = X_n$ du fait que (X_n) soit une martingale.

En ce qui concerne la seconde assertion, nous nous souvenons que le processus prévisible de la décomposition de Doob d'une sous-martingale est donné par la récurrence (38.38) que nous recopions ici :

$$\begin{cases} A_0 = 0 & \text{(38.47a)} \\ A_{n+1} = A_n + E(X_{n+1}^2 - X_n^2|\mathcal{F}_n) & \text{(38.47b)} \end{cases}$$

Vu que X_n^2 est $\mathcal{F}_n$-mesurable, il peut sortir de l'espérance :

$$A_{n+1} = A_n + E(X_{n+1}^2|\mathcal{F}_n) - X_n^2 \tag{38.48}$$

et donc

$$A_n = \sum_{i=1}^{n} \Big(E(X_i^2|\mathcal{F}_{i-1}) - X_{i-1}^2 \Big). \tag{38.49}$$

Pour obtenir la dernière partie de (38.45) nous travaillons un peu :

$$\begin{aligned}
E\big((X_i - X_{i-1})^2|\mathcal{F}_{i-1}\big) &= E\big(X_i^2 + X_{i-1}^2 - 2X_iX_{i-1}|\mathcal{F}_{i-1}\big) && \text{(38.50a)}\\
&= E(X_i^2|\mathcal{F}_{i-1}) + X_{i-1}^2 - 2E(X_iX_{i-1}|\mathcal{F}_{i-1}) && \text{(38.50b)}\\
&= E(X_i^2|\mathcal{F}_{i-1}) + X_{i-1}^2 - 2X_{i-1}E(X_i|\mathcal{F}_{i-1}) && \text{(38.50c)}\\
&= E(X_i^2|\mathcal{F}_{i-1}) + X_{i-1}^2 - 2X_{i-1}X_i && \text{(38.50d)}
\end{aligned}$$

où nous avons utilisé la proposition 35.47 pour obtenir (38.50c). □

38.4 Problème de la ruine du joueur

Nous considérons un joueur compulsif qui joue à un jeu très simple[4] : il joue à pile ou face contre la banque avec une pièce truquée. Si pile sort, la banque donne 1 au joueur et si c'est face, c'est le joueur qui donne 1 à la banque. Nous nommons a la fortune initiale du joueur, b celle de la banque et p la probabilité d'obtenir pile.

Nous supposons que le jeu se poursuit jusqu'à la ruine du joueur ou de la banque. La modélisation est comme suit : nous considérons (Y_n) une suite de variables aléatoires indépendantes et identiquement distribuées de loi

$$Y_n \sim p\delta_1 + (1-p)\delta_{-1}. \tag{38.51}$$

C'est le résultat financier pour le joueur du n^{e} lancé. La fortune du joueur au bout de n lancés est la variable aléatoire

$$S_n = a + \sum_{j=1}^{n} Y_j. \tag{38.52}$$

Nous notons $Y_0 = a$.

Nous considérons la filtration

$$\mathcal{A}_n = \sigma\big(S_i \text{ tel que } 0 \leqslant i \leqslant n\big) = \sigma\big(Y_i \text{ tel que } 0 \leqslant i \leqslant n\big), \tag{38.53}$$

et le temps d'arrêt du jeu :

$$T = \inf\{n \geqslant 1 \text{ tel que } S_n \in \{0, a+b\}\}; \tag{38.54}$$

c'est le temps qu'il faut pour que tout l'argent appartienne soit au joueur soit à la banque.

Nous voulons étudier les paramètres suivants :

(1) $\rho = P(S_T = a+b)$, c'est-à-dire la probabilité que ce soit le joueur qui gagne contre la banque.

(2) $P(T < \infty)$, c'est-à-dire la probabilité que le jeu se finisse.

(3) $E(T)$, la durée moyenne du jeu.

Lemme 38.25.
Le processus S_n du problème de la ruine du joueur est vérifie

$$E(S_n|\mathcal{A}_{n-1}) = S_{n-1} + p - q. \tag{38.55}$$

De plus le processus S_n est

4. Le gros des choses dites à propos de la ruine du joueur provient de [54].

(1) une martingale si $p = q = \frac{1}{2}$,
(2) une sous-martingale si $p > q$.

Démonstration. Pour $n \geqslant 1$ nous avons

$$E(S_n|\mathcal{A}_{n-1}) = a + \sum_{j=1}^{n} E(Y_j|\mathcal{A}_{n-1}) = a + \sum_{j=1}^{n-1} E(Y_j|\mathcal{A}_{n-1}) + E(Y_n|\mathcal{A}_{n-1}). \tag{38.56}$$

Si $j \leqslant n-1$ alors $Y_j \in m(\mathcal{A}_{n-1})$. Mais nous savons que si X est $\mathcal{F}$-mesurable, alors $E(X|\mathcal{F}) = X$ (c'est la définition de l'espérance conditionnelle), donc $\sum_{j=1}^{n-1} E(Y_j|\mathcal{A}_{n-1}) = \sum_{j=1}^{n-1} Y_j$.

En ce qui concerne le terme $j = n$ nous utilisons le fait que $\sigma(Y_n)$ soit une tribu indépendante de $\mathcal{A}_{n-1}$; nous avons donc au final pour tout j que $E(Y_j|\mathcal{A}_{n-1}E(Y_j) = p - q$. Nous avons donc

$$E(S_n|\mathcal{A}_{n-1}) = S_{n-1} + p - q. \tag{38.57}$$

Si $p = q = \frac{1}{2}$ alors c'est une martingale, et si $p > q$ c'est une sous-martingale. □

Lemme 38.26.
La variable aléatoire T est un temps d'arrêt.

Démonstration. Par définition $T = \inf\{n \geqslant 1 \text{ tel que } S_n \in \{0, a+b\}\}$. Vu que les variables aléatoires S_i avec $i \leqslant n$ sont $\mathcal{F}_n$-mesurables, les ensembles $\big\{S_k \notin \{0, a+b\}\big\}$ avec $k \leqslant n$ sont $\mathcal{F}_n$-mesurables. Donc les ensembles

$$\{T = n\} = \bigcap_{k \leqslant n} \big\{S_k \notin \{0, a+b\}\big\} \cap \big\{S_n \in \{0, a+b\}\big\} \tag{38.58}$$

sont $\mathcal{F}_n$-mesurables. Nous en concluons que l'ensemble $\{T \leqslant n\}$ est également mesurable. □

38.4.1 Le cas où la pièce est truquée

Nous supposons être dans le cas $p > q$.

38.4.1.1 Introduction d'une martingale

Considérons le processus

$$\begin{cases} A_0 = 0 & \text{(38.59a)} \\ A_n = A_{n-1} + E(S_n - S_{n-1}|\mathcal{A}_{n-1}). & \text{(38.59b)} \end{cases}$$

Vu que $E(S_n|\mathcal{A}_{n-1}) = S_{n-1} + p - q$ (lemme 38.25) et que $E(S_{n-1}|\mathcal{A}_{n-1}) = S_{n-1}$ (parce que S_{n-1} est dans la tribu de $\mathcal{A}_{n-1}$), nous avons $A_n = A_{n-1} + (p - q)$ et donc

$$A_n = n(p - q). \tag{38.60}$$

Ce processus (A_n) est croissant et prévisible. Nous introduisons le processus

$$M_n = S_n - A_n \tag{38.61}$$

et nous montrons que c'est une martingale[5]. Nous conditionnons la définition (38.61) par rapport à $\mathcal{A}_{n-1}$:

$$\begin{aligned} E(M_n\mathcal{A}_{n-1}) &= E(S_n|\mathcal{A}_{n-1}) - \underbrace{E(A_n|\mathcal{A}_{n-1})}_{=A_n} && \text{(38.62a)} \\ &= A_n - A_{n-1} + E(S_{n-1}|\mathcal{A}_{n-1}) - A_n && \text{(38.62b)} \\ &= E(S_{n-1}|\mathcal{A}_{n-1}) - A_{n-1}. && \text{(38.62c)} \end{aligned}$$

Mais S_{n-1} est $\mathcal{A}_{n-1}$-mesurable, donc $E(S_{n-1}|\mathcal{A}_{n-1}) = S_{n-1}$ et

$$E(M_n|\mathcal{A}_{n-1}) = S_{n-1} - A_{n-1} = M_{n-1}, \tag{38.63}$$

ce qui signifie que (M_n) est une martingale.

5. Ceci est un peu le contraire de la décomposition de Doob.

38.4.1.2 Finitude du temps d'arrêt

Nous montrons maintenant, en étudiant $M_{T\wedge n}$ que T est intégrable et nous prouvons que $P(T=\infty)=0$.

Proposition 38.27.
La variable aléatoire T est intégrable, et $P(T=\infty)=0$, c'est-à-dire que le jeu se termine presque certainement après un temps fini.

Démonstration. Nous rappelons que le lemme 38.26 nous indique que T est un temps d'arrêt. Le temps d'arrêt $T\wedge n$ est borné (par n évidemment) et nous pouvons donc lui appliquer le théorème d'arrêt 38.14 pour dire que

$$E(M_{T\wedge n})=E(M_0). \tag{38.64}$$

Le membre de droite est simple parce que $M_0=S_0-A_0=S_0=a$ parce que c'est l'argent de départ du joueur. Pour l'autre :

$$E(M_{T\wedge n})=E(S_{T\wedge n})-E(A_{T\wedge n}). \tag{38.65}$$

D'une part, $E(A_{T\wedge n})=E\big((T\wedge n)(p-q)\big)$ et d'autre part, $E(S_{T\wedge n})\leqslant a+b$ parce que S_T vaut zéro ou $a+b$ (avec des probabilités encore inconnues [6]). En combinant avec ce qui était dit juste au dessus et remarquant que $(p-q)E(T\wedge n)\geqslant 0$ nous pouvons écrire

$$0\leqslant (p-q)E(T\wedge n)\leqslant b. \tag{38.66}$$

La suite de variables aléatoires $T\wedge n$ est donc croissante, positive et intégrable [7] et donc nous avons du travail pour le théorème de la convergence monotone 13.160. La variable aléatoire T est alors mesurable et [8]

$$\lim_{n\to\infty}E(T\wedge n)=E(T). \tag{38.67}$$

Notons que nous n'avons pas encore prouvé que $E(T)<\infty$, mais en passant à la limite dans (38.66) nous écrivons

$$0\leqslant (p-q)E(T)\leqslant b. \tag{38.68}$$

Maintenant nous avons prouvé que T est intégrable et même L^1. Par conséquent

$$P(T=\infty)=0. \tag{38.69}$$

Le jeu se termine donc presque certainement après un temps fini. □

38.4.1.3 Temps moyen de jeu

Le lemme 38.10 nous indique que $S_{T\wedge n}\xrightarrow{p.s.}S_T$.

Nous avons les bornes $0\leqslant S_{T\wedge n}\leqslant a+b$ et comme $a+b$ est intégrable, $S_{T\wedge n}$ l'est aussi et nous pouvons parler de $E(S_{T\wedge n})$. Repartons de (38.65) :

$$a=E(M_0)=E(M_{T\wedge n})=E(S_{T\wedge n})-E(A_{T\wedge n})=E(S_{T\wedge n})-(p-q)E(T\wedge n). \tag{38.70}$$

La variable aléatoire $S_{T\wedge n}$ est majorée par $a+b$ indépendamment de n ; donc le théorème de la convergence dominée 13.184 donne $\lim_{n\to\infty}E(S_{T\wedge n})=E(S_T)$. En ce qui concerne le second terme, la convergence dominée ne fonctionne pas parce que $T\wedge n$ n'est pas à priori majoré par quelque chose d'indépendant de n. Heureusement, le théorème de la convergence monotone donne $\lim_{n\to\infty}E(T\wedge n)=E(T)$. Au final en passant à la limite dans (38.70) nous avons

$$a=E(S_T)-(p-q)E(T). \tag{38.71}$$

6. Mais on y travaille.
7. Je rappelle que les constantes sont des fonctions intégrables sur Ω. Oui, je sais, quand on est habitué à faire de l'analyse sur $\mathbb{R}^n$ c'est un truc qu'on perd toujours un peu de vue.
8. Dans [54], l'équation (38.67) vient avec une $\limsup$ et non une limite normale. Je ne comprends pas pourquoi.

Étant donné que $T > 0$ et $p - q > 0$ nous pouvons récrire cela sous la forme

$$0 \leqslant (p-q)E(T) = E(S_T) - a. \tag{38.72}$$

Par définition de T nous avons aussi

$$E(S_T) = (a+b)P(S_T = a+b) + 0 \cdot P(S_T = 0) = \rho(a+b). \tag{38.73}$$

Nous déduisons

$$E(T) = \frac{(a+b)\rho - a}{p-q}. \tag{38.74}$$

Ne crions pas victoire trop vite : nous n'avons pas encore d'expression de $\rho = P(S_T = a+b)$. Le temps moyen de jeu n'est donc pas encore tout à fait connu.

38.4.1.4 Probabilité de victoire du joueur

Nous avons besoin d'exprimer ρ en termes de a, b et p. Pour cela nous introduisons la variable aléatoire [9]

$$U_n = \left(\frac{p}{q}\right)^{S_n}. \tag{38.75}$$

Nous commençons par prouver que c'est une martingale en calculant

$$E(U_n|\mathcal{A}_{n-1}) = E\left(\left(\frac{q}{p}\right)^{S_n-1}\left(\frac{q}{p}\right)^{Y_n}|\mathcal{A}_{n-1}\right) \tag{38.76}$$

Nous utilisons la proposition 35.47. Dans notre cas, S_{n-1} et Y_n sont des variables aléatoires $\mathcal{A}_n$-mesurables ; la variable aléatoire Y_n est même $\mathcal{A}_{n-1}$-mesurable et sort donc du conditionnement ; nous avons donc

$$E(U_n|\mathcal{A}_{n-1}) = \left(\frac{q}{p}\right)^{S_{n-1}} E\left(\left(\frac{q}{p}\right)^{Y_n}\right) \tag{38.77}$$

Nous allons utiliser le théorème de transfert 35.68 :

$$E(s^{Y_n}) = \int_\Omega s^{Y_n(\omega)} DP(\omega) = \int_{Y_n=1} s dP(\omega) + \int_{Y_n=-1} \frac{1}{s} dP(\omega). \tag{38.78}$$

Mais nous savons que $P(Y_n = 1) = p$ et $P(Y_n = -1) = 1 - p = q$, donc

$$E(s^{Y_n}) = ps + \frac{1-p}{s} \tag{38.79}$$

et

$$E\left(\left(\frac{p}{q}\right)^{Y_n}\right) = p + q = 1. \tag{38.80}$$

Donc

$$E(U_n|\mathcal{A}_{n-1}) = \left(\frac{q}{p}\right)^{S_n-1} = U_{n-1}, \tag{38.81}$$

ce qui prouve que (U_n) est une martingale.

Par définition nous avons toujours $S_n \geqslant 0$ tant que $n \leqslant T$ [10], donc $U_{T\wedge n} \in [0,1]$. Il est donc évident que si $a \geqslant 1$ nous avons

$$\int_{|U_{T\wedge n}|>a} |U_{T\wedge n}| dP = 0 \tag{38.82}$$

9. Nous dirons un mot sur ce choix dans le « petit complément » plus bas
10. Pour $n > T$ le jeu est terminé, donc on ne se pose pas la question.

parce que le domaine d'intégration est vide. Donc les variables aléatoires $V_n = U_{T\wedge n}$ sont équi-intégrables[11] et le théorème 38.18 montre que la martingale (V_n) est terminée ; par ricochet[12] le théorème de Doob 38.19 montre qu'il existe une variable aléatoire X telle que $V_n \xrightarrow{p.s.} X$. Nous allons prouver que $X = U_T$ presque partout. Nous savions déjà (voir l'équation (38.22) et ses alentours) que

$$S_{n\wedge T} \xrightarrow{p.s.} S_T. \tag{38.83}$$

Nous avons alors (au sens du presque surement) :

$$\lim_{n\to\infty} V_n = \lim_{n\to\infty} U_{T\wedge n} = \lim_{n\to\infty} \left(\frac{q}{p}\right)^{S_{T\wedge n}} = \left(\frac{q}{p}\right)^{S_T} = U_T. \tag{38.84}$$

Donc par unicité de la limite presque partout nous avons $X = U_T$ presque partout. Par le théorème de transfert 35.68 nous évaluons

$$E(U_T) = \left(\frac{q}{p}\right)^0 P(S_T = 0) + \left(\frac{q}{p}\right)^{a+b} P(S_T = a+b) = (1-\rho) + \left(\frac{q}{p}\right)^{a+b} \rho. \tag{38.85}$$

La remarque 38.13 nous permet de dire que

$$E(U_{T\wedge n}) = U_0. \tag{38.86}$$

Mais par définition

$$U_0 = \left(\frac{q}{p}\right)^{S_0} = \left(\frac{q}{p}\right)^a, \tag{38.87}$$

donc nous avons

$$E(U_{T\wedge n}) = \left(\frac{q}{p}\right)^a. \tag{38.88}$$

Nous voudrions passer à la limite $n \to \infty$ dans cette équation. Pour permuter la limite et l'espérance, il faut utiliser le théorème de la convergence dominée 13.184. Vu que nous avons choisi $q > p$, nous avons $q/p > 1$ et donc $U_{T\wedge n} \leqslant (q/p)^{a+b}$, ce qui montre que la fonction $\omega \mapsto (U_{T\wedge n})(\omega)$ est majorée par une constante (qui est une fonction intégrable). Nous pouvons donc permuter la limite et l'espérance :

$$\lim_{n\to\infty} E(U_{T\wedge n}) = E\big(\lim_{n\to\infty} U_{T\wedge n}\big). \tag{38.89}$$

Mais nous avion déjà montré que $U_{T\wedge n} \xrightarrow{p.s.} U_T$. Donc

$$E(U_T) = \left(\frac{q}{p}\right)^a. \tag{38.90}$$

En égalisant avec l'expression (38.85) de $E(U_T)$ nous trouvons

$$\rho = \frac{\left(\frac{q}{p}\right)^a - 1}{\left(\frac{q}{p}\right)^{a+b} - 1} \tag{38.91}$$

et ensuite nous trouvons $E(T)$ en remettant ce ρ dans l'expression (38.74) donnée plus haut.

38.4.2 Le cas où la pièce est non truquée

Maintenant $p = q = 1/2$.

11. Définition 38.17.
12. Nous rappellons que la convergence L^1 n'implique pas la convergence presque partout.

38.4.2.1 Probabilité de gagner

Le lemme 38.25 nous indique alors que (S_n) est une martingale et le lemme 38.24 nous permet de dire que (S_n^2) est alors une sous-martingale. Le processus croissant prévisible de (S_n^2) est donné par (38.38) qui en adaptant les notations est

$$\begin{cases} B_0 = 0 & (38.92a) \\ B_n = B_{n-1} + E\Big((S_n - S_{n-1})^2 | \mathcal{A}_{n-1}\Big). & (38.92b) \end{cases}$$

Nous avons toujours $S_n - S_{n-1} = \pm 1$ parce que soit le joueur gagne soit le joueur perd, mais de toutes façons sa fortune varie de 1 à chaque étape du jeu. Donc (38.92b) nous donne $B_n = B_{n-1} + 1$ et

$$B_n = n. \tag{38.93}$$

Cela nous dit que la variable aléatoire

$$S_n^2 - B_n = S_n^2 - n \tag{38.94}$$

est une martingale (une sur-martingale moins son processus prévisible croissant). Nous lui appliquons le théorème d'arrêt 38.12 avec les temps d'arrêt 0 et $T \wedge n$:

$$E(S_{T\wedge n}^2 - T \wedge n | \mathcal{F}_0) = S_0^2 - 0 \tag{38.95}$$

où $\mathcal{F}_0$ est la tribu engendrée par la variable aléatoire 0, c'est-à-dire $\{\Omega, \emptyset\}$. Cette tribu est indépendante de toute autre tribu et nous pouvons donc supprimer le conditionnement dans (38.95). Nous avons aussi $S_0 = a$ par définition. Avec tout ça nous avons la majoration

$$E(T \wedge n) = E(S_{T\wedge n}^2) - a^2 \leqslant (a+b)^2 - a^2 \tag{38.96}$$

parce que S_k est toujours positif et entre 0 et $a + b$. En utilisant le lemme 38.10 et en passant à la limite,

$$E(T) \leqslant (a+b)^2 - a^2. \tag{38.97}$$

En particulier, $T \in L^1(\Omega)$ et $P(T < \infty) = 1$.

En suivant exactement les mêmes étapes que dans le lemme 38.1038.10 nous avons aussi

$$\lim_{n\to\infty} S_{T\wedge n} = S_T \tag{38.98}$$

presque partout. De plus nous savons que

$$0 \leqslant S_{T\wedge n}^2 \leqslant (a+b)^2, \tag{38.99}$$

et nous pouvons donc utiliser le théorème de la convergence dominée 13.184 pour dire que

$$\lim_{n\to\infty} E(S_{T\wedge n}^2) = E(S_T^2). \tag{38.100}$$

Nous montrons à présent que $S_{T\wedge n} \xrightarrow{L^2} S_T$. Pour cela nous devons évaluer la limite

$$\lim_{n\to\infty} \int_\Omega |S_{T\wedge n} - S_T|^2. \tag{38.101}$$

La fonction $|S_{T\wedge n} - S_T|^2$ est majorée par $(a+b)^2$ et nous pouvons à nouveau appliquer la convergence dominée :

$$\lim_{n\to\infty} E(|S_{T\wedge n} - S_T|^2) = \lim_{n\to\infty} \int_\Omega |S_{T(\omega)\wedge n}(\omega) - S_{T(\omega)}(\omega)|^2 dP(\omega) = \int_\Omega \lim_{n\to\infty} |S_{T\wedge n} - S_T|^2 = 0. \tag{38.102}$$

La même chose en n'écrivant pas les carrés montre que l'on a aussi $S_{T\wedge n} \xrightarrow{L^1} S_T$.

Il n'y a pas que $n \mapsto S_n^2 - n$ qui est une martingale. Il y a aussi (S_n) lui-même (lemme 38.25). Nous pouvons lui appliquer le théorème d'arrêt 38.12 pour les temps d'arrêts $T \wedge n$ et 0 :

$$E(S_{T\wedge n}) = E(S_0) = a. \tag{38.103}$$

En passant à la limite, $E(S_T) = a$. L'espérance $E(S_T)$ peut par ailleurs être calculée comme

$$E(S_T) = 0 \cdot P(S_T = 0) + (a+b)P(S_T = a+b). \tag{38.104}$$

En égalisant les valeurs (38.103) et (38.104) de $E(S_T)$ nous trouvons

$$\rho = \frac{a}{a+b}. \tag{38.105}$$

Cette formule est assez logique : la probabilité que le joueur gagne est égale à la proportion d'argent en jeu qu'il a amené.

38.4.2.2 Temps moyen de jeu

Nous calculons maintenant l'espérance $E(T)$ du temps de jeu (sans compter les pauses ni les jours de fermeture du casino[13]).

Nous recopions la première égalité de (38.96) sous la forme

$$a^2 = E(S_{T\wedge n}^2 - T \wedge n) \tag{38.106}$$

et nous passons à la limite[14] en sachant que $E(S_T^2) = \rho(a+b)^2$:

$$a^2 = \rho(a+b)^2 - E(T). \tag{38.107}$$

En reprenant la valeur (38.105) de ρ,

$$E(T) = ab. \tag{38.108}$$

Et là, on voit que si le joueur amène 1000 euros contre une banque qui en a un million, et si ils jouent toutes les secondes , on en a pour 32 ans de jeu en moyenne.

Voilà. C'est fini pour la ruine du joueur.

38.4.3 Un petit complément

Nous avons introduit lors de l'équation (38.75) la variable aléatoire $U_n = (p/q)^{S_n}$. Sans aller jusqu'à motiver complètement ce choix, nous nous proposons maintenant de voir que parmi les variables aléatoires $U_n = s^{S_n}$, le choix $s = p/q$ est le seul qui donne une martingale.

Soit donc $U_n = s^{S_n}$ et exprimons le fait que ce soit une martingale. Nous avons

$$E(U_n|\mathcal{A}_{n-1}) = E(s^{S_{n-1}}s^{Y_n}|\mathcal{A}_{n-1}) \tag{38.109a}$$
$$= s^{S_{n-1}}E(s^{Y_n|\mathcal{A}_{n-1}}) \tag{38.109b}$$
$$= s^{S_{n-1}}E(s^{Y_n}). \tag{38.109c}$$

Le passage à (38.109b) se justifie en disant que $s^{S_{n-1}}$ est une variable aléatoire bornée et $\mathcal{A}_{n-1}$-mesurable, et en invoquant proposition 35.47. La variable aléatoire Y_n vaut 1 avec probabilité p et -1 avec probabilité q ; donc l'espérance est vite vue :

$$E(s^{Y_n}) = ps + q\frac{1}{s} \tag{38.110}$$

et nous avons

$$E(U_n|\mathcal{A}_{n-1}) = \left(ps + q\frac{1}{s}\right)s^{S_{n-1}} = (ps + \frac{q}{s})U_{n-1}. \tag{38.111}$$

13. Le joueur est un *vrai* joueur compulsif.
14. Comme il est dit dans La Grande Illusion, à quoi sert un n ? À passer à la limite.

Pour que (U_n) soit une martingale il faut (et il suffit) que

$$ps + \frac{q}{s} = 1. \tag{38.112}$$

Les solutions de cette équation sont $s \in \{1, \frac{p}{q}\}$. C'est évidemment $s = p/q$ qui donne une martingale non triviale. Attention pour être complet, il faut se demander ce qu'il se passe si $s = 0$ séparément parce que manifestement l'équation (38.112) ne traite pas ce cas. Encore une fois, en repartant du début, $s = 0$ ne se révèle pas être une martingale très excitante.

Bref, nous devons poser

$$U_n = \left(\frac{p}{q}\right)^{S_n} \tag{38.113}$$

pour avoir une martingale.

Chapitre 39

Processus de Poisson

39.1 Processus de Poisson

Définition 39.1.
Une famille de variables aléatoires $(N_t)_{t\geqslant 0}$ est un ***processus de Poisson*** *d'intensité λ s'il existe une suite de variables aléatoires indépendantes et identiquement distribuées $(T_k)_{k\in\mathbb{N}}$ de loi $\mathscr{E}(\lambda)$ telles que*

$$N_t = \sup\{n \geqslant 0 \text{ tel que } \sum_{k=1}^{n} T_k \leqslant t\}. \tag{39.1}$$

Si nous posons $S_n = \sum_{k=1}^{n} T_k$, alors nous avons une expression plus pratique pour N_t :

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{\{S_n \leqslant t\}}. \tag{39.2}$$

Nous avons par la proposition 35.111 vu que $N_t \sim \mathscr{P}(\lambda t)$.

Pour chaque $\omega \in \Omega$, la fonction $t \mapsto N_t(\omega)$ est une fonction (pas du tout strictement) croissante à valeurs dans $\mathbb{N}$. Cette fonction part de 0 et fait un saut de taille 1 après des intervalles de temps $T_1(\infty)$, $T_2(\omega)$, etc. Elle est continue à droite.

Nous avons les égalités d'événements suivantes qui sont pratiques :

$$\{s < S_n \leqslant t\} = \{N_t \geqslant n > N_s\} \tag{39.3a}$$

$$\{N_t = n\} = \{S_n \leqslant t \leqslant S_{n+1}\}. \tag{39.3b}$$

Théorème 39.2.
Les variables aléatoires $(N_t)_{t\geqslant 0}$ est un processus de Poisson d'intensité λ si et seulement si elles vérifient les trois propriétés suivantes.

Accroissements indépendants *Pour tout choix $0 < t_0 < t_1 < \dots t_n$, les variables aléatoires $N_{t_{i+1}} - N_{t_i}$ sont indépendantes.*

Accroissements stationnaires *Si $0 < s < t$ et $h > 0$ alors*

$$N_{t+h} - N_{s+h} \overset{\mathcal{L}}{=} N_t - N_s, \tag{39.4}$$

c'est-à-dire que les accroissements décalés suivent les mêmes lois.

Poisson *Pour tout t nous avons $N_t \sim \mathscr{P}(\lambda t)$.*

Une conséquence des accroissements stationnaires est que $N_t - N_s \overset{\mathcal{L}}{=} N_{t-s} - N_0 = N_{t-s}$ parce que $N_0 = 0$.

Proposition 39.3.
Si (N_t) est un processus de Poisson d'intensité λ, alors

$$\lim_{t\to\infty} N_t = +\infty \tag{39.5}$$

presque surement. De plus

$$\lim_{t\to\infty} \frac{N_t}{t} = \lambda \tag{39.6}$$

presque surement.

La relation (39.6) est appelée **loi des grands nombres**.

Démonstration. Par définition nous savons que

$$N_t = \sup\{n \geqslant 0 \text{ tel que } S_n \leqslant t\}. \tag{39.7}$$

Évidemment la fonction $t \mapsto N_t$ est croissante, donc la limite

$$\lim_{t\to\infty} N_t(\omega) \tag{39.8}$$

existe dans $[0, \infty]$. Nous pouvons nous restreindre à $t \in \mathbb{N}$ et considérer $L(\omega) = \lim_{n\to\infty} N_n(\omega)$. Par somme télescopique avec $N_0 = 0$,

$$\frac{N_n}{n} = \frac{\sum_{k=1}^{n}(N_k - N_{k-1})}{n}. \tag{39.9}$$

Étant donné que le processus est de Poisson, les variables aléatoires $(N_k - N_{k-1})_{k=1,\dots,n}$ sont indépendantes et suivent toutes la loi de $N_1 - N_0$, c'est-à-dire la loi de N_1. Encore par le fait que N_t soit de Poisson nous savons que $N_1 \sim \mathscr{P}(\lambda)$. La loi des grands nombres (35.91) appliquée aux variables aléatoires $N_k - N_{k-1}$ nous dit que

$$\frac{N_n}{n} \xrightarrow{p.s.} E(N_1) = \lambda > 0. \tag{39.10}$$

Du coup $N_n \to \infty$ et $L(\omega) = \infty$.

Nous démontrons maintenant la loi des grands nombres pour les processus de Poisson. Étant donné que pour les entiers $N_n/n \to \lambda$, pour les réels, si la limite existe, ça ne peut pas être autre chose. Si nous notons $\bar{t}$ la partie entière de $t \in \mathbb{R}^+$,

$$\frac{N_t}{t} = \frac{N_t - N_{\bar{t}}}{t} + \frac{N_{\bar{t}}}{t}. \tag{39.11}$$

Le second terme est relativement simple à traiter :

$$\frac{N_{\bar{t}}}{t} = \underbrace{\frac{N_{\bar{t}}}{\bar{t}}}_{\to\lambda} \cdot \underbrace{\frac{\bar{t}}{t}}_{\to 1}. \tag{39.12}$$

où nous avons utilisé le premier point, $\bar{t}$ étant entier. Pour le premier terme nous savons que $t \mapsto N_t$ est croissante et donc que

$$\frac{N_t - N_{\bar{t}}}{t} \leqslant \frac{N_{\bar{t}+1} - N_{\bar{t}}}{t} = \frac{N_{\bar{t}+1} - N_{\bar{t}}}{\bar{t}+1}\frac{\bar{t}+1}{t}. \tag{39.13}$$

Le second facteur tend vers 1 lorsque $t \to \infty$. Le premier s'écrit

$$\frac{N_n - N_{n-1}}{n} \tag{39.14}$$

et tend vers zéro en tant que terme général de la série (39.9) qui converge. □

Proposition 39.4.
La variable aléatoire N_t/t est un estimateur sans biais de λ. De plus il converge vers λ en moyenne quadratique.

Démonstration. Vu que $N_t/t \to \lambda$ presque surement, la variable aléatoire N_t/t est un estimateur de λ. Le fait qu'il soit sans biais a été fait dans l'exemple 36.25.

D'autre part nous avons (voir théorème 35.104)

$$\operatorname{Var}\left(\frac{N_t}{t}\right) = \frac{1}{t^2}\operatorname{Var}(N_t) = \frac{\lambda}{t}. \tag{39.15}$$

En appliquant la formule $\operatorname{Var}(X) = E(X^2) - E(X)^2$ à $X = N_t/t$ nous trouvons

$$E\left(\frac{N_t^2}{t^2}\right) = \frac{\lambda}{t} + \lambda^2. \tag{39.16}$$

Cela montre que $\frac{N_t}{t} \xrightarrow{L^2} \lambda$. □

Pour le théorème central limite d'un processus de Poisson, nous visons un résultat du style de

$$\frac{\frac{1}{n}\sum_i X_i - mn}{\sigma\sqrt{n}} \xrightarrow{\mathscr{L}} \mathscr{N}(0,1). \tag{39.17}$$

Nous écrivons le théorème central limite pour le nombre de sauts que le processus de Poisson a connu en un temps t. Le rôle de la moyenne empirique est joué par N_t. Nous considérons avoir fait *une seule expérience* qui a duré un temps t. Donc le rôle de n est joué par 1 (et non t comme on pourrait le croire). Pour le reste, le nombre de succès en un temps t d'une variable aléatoire exponentielle de paramètre λ est une variable aléatoire de Poisson de paramètre λt, en vertu de ce qui est raconté au point 35.5.8. C'est cela qui motive l'énoncé suivant.

Théorème 39.5 (Théorème central limite pour les processus de Poisson)**.**
Si $(N_t)_{t>0}$ est un processus de Poisson de paramètre λ, alors nous avons

$$\frac{N_t - \lambda t}{\sqrt{\lambda t}} \xrightarrow{\mathscr{L}} \mathscr{N}(0,1). \tag{39.18}$$

Remarque 39.6.
Avant de nous lancer dans la démonstration, remarquons que si nous nous limitons à $t \in \mathbb{N}$, alors nous avons

$$\frac{N_n - \lambda n}{\sqrt{\lambda n}} = \frac{\sum_{k=1}^n (N_k - N_{k-1}) - \lambda n}{\sqrt{\lambda n}} \tag{39.19}$$

or par définition nous avons les égalités de lois

$$N_k - N_{k-1} \sim N_1 \sim \mathscr{P}(\lambda), \tag{39.20}$$

donc

$$\frac{S_n - \lambda n}{\sqrt{\lambda n}} = \frac{\frac{1}{n}S_n - \lambda}{\frac{\sqrt{\lambda n}}{n}} = \frac{\frac{1}{n}S_n - \lambda}{\sqrt{\lambda}/\sqrt{n}}, \tag{39.21}$$

ce qui est exactement le théorème central limite pour une suite de lois de Poisson [1].

Démonstration. Nous écrivons $\bar{t}$ la partie entière de $\bar{t}$ et nous décomposons :

$$\frac{N_t - \lambda t}{\sqrt{\lambda t}} = \underbrace{\frac{N_t - N_{\bar{t}}}{\sqrt{\lambda t}}}_{A} + \underbrace{\frac{N_{\bar{t}} - \lambda \bar{t}}{\sqrt{\lambda t}}}_{B} + \underbrace{\frac{\lambda \bar{t} - \lambda t}{\sqrt{\lambda t}}}_{C}. \tag{39.22}$$

En ce qui concerne le terme B, nous avons

$$B = \sqrt{\frac{\bar{t}}{t}}\frac{N_{\bar{t}} - \lambda \bar{t}}{\sqrt{\lambda \bar{t}}} \to \mathscr{N}(0,1). \tag{39.23}$$

1. Au fait près que nous devrions encore montrer que S_n est de carré intégrable.

Notons que nous utilisons le fait que si $a_n \to 1$ (en tant que suite de nombre) et si $X_n \to \mathscr{N}(0,1)$ (limite en loi), alors $a_n X_n \to \mathscr{N}(0,1)$ en loi.

Le terme C est également facile parce que $\lambda\bar{t} - \lambda t$ est majoré en norme par λ. Du coup

$$-\frac{\lambda}{\sqrt{\lambda t}} \leqslant C \leqslant \frac{\lambda}{\sqrt{\lambda t}}. \tag{39.24}$$

Donc $\lim_{t\to\infty} C = 0$.

Reste à travailler sur A. Vu que $t \mapsto N_t$ est croissante, la différence $N_t - N_{\bar{t}}$ est positive. Soit $\eta > 0$, nous avons

$$P(|A| > \eta) = P(N_t - N_{\bar{t}} > \sqrt{\lambda t}\eta) \leqslant P\big(N_{\bar{t}+1} - N_{\bar{t}} \geqslant \sqrt{\lambda t}\eta\big) = P(N_1 \geqslant \eta\sqrt{\lambda t}) \tag{39.25}$$

parce que nous savons que $N_{\bar{t}+1} - N_{\bar{t}} \sim N_1 \sim \mathscr{P}(\lambda)$. En vertu des propriétés de la loi de Poisson,

$$\lim_{t\to\infty} P(N_1 \geqslant \eta\sqrt{\lambda t}) = 0. \tag{39.26}$$

En effet si Z est une variable aléatoire de Poisson de paramètre λ nous avons

$$P(Z > l) = \sum_{k=l}^{\infty} P(Z = k) = e^{-\lambda} \sum_{k=l}^{\infty} \frac{\lambda^k}{k!}. \tag{39.27}$$

Nous reconnaissons la queue de série de e^{λ}, qui tend donc vers zéro lorsque $l \to \infty$. Nous avons donc prouvé que

$$\lim_{t\to\infty} P\big(|A| > \eta\big) = 0, \tag{39.28}$$

c'est-à-dire la convergence en probabilité de A vers zéro.

Nous avons montré que

$$B + C \xrightarrow{\mathscr{L}} U \sim \mathscr{N}(0,1) \tag{39.29a}$$

$$A \xrightarrow{P} 0. \tag{39.29b}$$

Le lemme de Slutsky (35.82) nous avons une convergence du couple

$$(A, B + C) \xrightarrow{\mathscr{L}} (0, U). \tag{39.30}$$

Utilisant le corolaire 35.84, nous trouvons la convergence en loi

$$A + (B + C) \xrightarrow{\mathscr{L}} 0 + U, \tag{39.31}$$

ce qu'il fallait. □

39.2 Quelques trucs sur la simulation

Le théorème ergodique dit que

$$\pi(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{X_k = x}. \tag{39.32}$$

C'est avec cela qu'on calcule $\pi(x)$ à partir d'une simulation de chaine de Markov.

39.2.1 Le théorème central limite pour Markov

Théorème 39.7 (Version allégée)**.**
Si (X_n) est irréductible et positive récurrente, alors pour toute fonction f,

$$\frac{1}{\sqrt{N}}\left[\sum_{k=1}^{N} -N\int f d\pi\right] \xrightarrow{\mathscr{L}} \mathscr{N}(0,\sigma^2) \tag{39.33}$$

où σ^2 dépend de la fonction f et de la chaine de Markov.
Ici, $\int f d\pi = \sum_{x\in E} f(x)\pi(x)$.

Nous allons simuler la variable aléatoire

$$Z = \frac{1}{\sqrt{N}}\left[\sum f(X_k) - N\sum_{x\in E} f(x)\pi(x)\right] \tag{39.34}$$

et puis on va mettre sa réalisation dans un histogramme. Dans le cas où on prend $f(i) = \mathbb{1}_{i=i_0}$, il y a de la simplification dans l'intégrale qui devient

$$Z = \frac{1}{\sqrt{N}}\left[\sum_{i=1}^{N} \mathbb{1}_{X_k=i_0} - N\pi(i_0)\right]. \tag{39.35}$$

39.2.2 Feuille 5

On pose

$$D_n = \sqrt{n}\sup_{x\in\mathbb{R}} |F_n(x) - F(x)|. \tag{39.36}$$

On en génère un millier de fois D_n, on note $D_n^{(k)}$ ces réalisations, et on regarde ce que vaut

$$\frac{1}{1000}\sum_{k=1}^{1000} \mathbb{1}_{D_n^{(k)\geqslant c}}. \tag{39.37}$$

Cela nous donne une approximation de

$$P\big(\sqrt{n}\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \geqslant c\big). \tag{39.38}$$

Note que chacun des $D_n^{(k)}$ demande de créer un nouveau vecteur Y_i de lois qu'on veut regarder. Par exemple de loi exponentielle.

39.2.3 Feuille 6

Pour créer une fonction qui renvoie i avec probabilité p_i pour $i = 1, 2, 3$, on peut faire

$$U \sim \mathscr{U}[0,1] \tag{39.39}$$

et puis on a

$$P(U < p_0) = p_0 \tag{39.40a}$$
$$P(p_0 < U < p_0 + p_1) = p_1 \tag{39.40b}$$
$$P(p_0 + p_1 < U < p_2) = p_2. \tag{39.40c}$$

Une façon de faire une loi uniforme $[0,1]$ est de faire `rand`

39.2.4 Feuille 7

L'échantillon est $(Y_1, \dots, Y_n)$ et nous écrivons le vecteur

$$Y = X\beta + \epsilon \tag{39.41}$$

où $Y \sim \mathscr{N}(X\beta, \sigma^2 \operatorname{Id})$ et $\epsilon \sim \mathscr{N}(0, \sigma^2 \operatorname{Id})$. Nous utilisons le principe de maximum de vraisemblance. Soit $(y_1, \dots, y_n)$ un échantillon et

$$P_\theta(y_1, \dots, y_n) = \prod_i \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y_i - X_i^t\beta}{\sigma}\right)^2\right]. \tag{39.42}$$

L'astuce est de faire que $y_i - X_i^t\beta$ est la $i$ième composante du vecteur $Y - X\beta$ et donc la somme qui est dans l'exponentielle devient la norme de $Y - X\beta$:

$$f_\theta(y_1, \dots, y_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}\|Y - X\beta\|^2\right]. \tag{39.43}$$

On passe au logarithme et on dérive par rapport à σ^2. Attention : la variable est σ^2, donc la dérivée de σ^2 est 1 et non 2σ. Bref, on trouve

$$\sigma^2 = \frac{1}{2n}\|U + X\beta\|. \tag{39.44}$$

39.2.5 Simuler des lois conditionnelles

Nous voulons générer des couples (X, Y) tels que Y prend les valeurs 0 ou 1 et tels que

$$\begin{cases} P(X|Y = 0) \sim \mathscr{E}(\lambda_0) & \text{(39.45a)} \\ P(X|Y = 1) \sim \mathscr{E}(\lambda_1). & \text{(39.45b)} \end{cases}$$

Le plus simple est de générer une liste

$$(X_1, 0) \qquad (X_4, 1) \tag{39.46a}$$
$$(X_2, 0) \qquad (X_5, 1) \tag{39.46b}$$
$$(X_3, 0) \qquad (X_6, 1) \tag{39.46c}$$

avec $X_1, X_2, X_3 \sim \mathscr{E}(\lambda_0)$ et $X_4, X_5, X_6 \sim \mathscr{E}(\lambda_1)$.

Avec cette méthode cependant la liste est triée et en plus on a autant de 1 que de 0. On peut faire un peu plus technologique pour corriger cela. Pour créer un couple, on commence par $Y \sim \mathscr{B}(p)$ et puis suivant que $Y = 0$ ou $y = 1$, on génère $X \sim \mathscr{E}(\lambda_0)$ ou $X \sim \mathscr{E}(\lambda_1)$.

Chapitre 40

Langages

40.1 Langages

40.1.1 Alphabets et mots

Définition 40.1.
*Un **alphabet** est un ensemble fini de symboles appelés **lettres**.*

On utilise aussi parfois le terme **vocabulaire** pour désigner un alphabet.

Définition 40.2.
*Un **mot** sur l'alphabet Σ est une suite finie et ordonnée, éventuellement vide, de lettres de Σ. Le **mot vide** est toujours noté ε.*

Définition 40.3.
*La **longueur d'un mot** w, noté $|w|$, est le nombre de lettres constituant le mot w. Le mot vide a une longueur de 0.*

Soit w un mot de longueur k, on peut désormais noter $w = w_1 \cdots w_k$, où chacun des $w_i, 1 \leqslant i \leqslant k$ représente une lettre de w. Par convention, si $k = 0$, alors le mot w est le mot vide.

Définition 40.4.
*Soient w un mot sur l'alphabet Σ et $a \in \Sigma$ une lettre, le **nombre d'occurrences** de la lettre a dans le mot w, noté $|w|_a$, est le nombre de fois où apparaît la lettre a dans le mot w, c'est-à-dire le cardinal de l'ensemble $\{i \mid w_i = a, 1 \leqslant i \leqslant |w|\}$.*

Définition 40.5.
*Soit Σ un alphabet, l'**ensemble des mots non-vides** sur l'alphabet Σ, noté Σ^+, est l'ensemble :*

$$\Sigma^+ = \{w = w_1 \dots w_n, n > 0\} \tag{40.1}$$

Définition 40.6.
*Soit Σ un alphabet, l'**ensemble des mots** sur l'alphabet Σ, noté Σ^*, est l'ensemble :*

$$\Sigma^* = \{w = w_1 \dots w_n, n \geqslant 0\} \tag{40.2}$$

Des deux définitions précédentes, on tire l'égalité suivante :

$$\Sigma^* = \Sigma^+ \cup \{\varepsilon\} \tag{40.3}$$

Définition 40.7.
*Soient Σ un alphabet et $x, y \in \Sigma^*$ deux mots sur l'alphabet Σ de longueur respective n et m, le **produit** w de x et y, noté $x \cdot y$ est défini par $w = x_1 \dots x_n y_1 \dots y_m$.*

Le produit est également appelé **concaténation**.

Proposition 40.8 (Longueur du produit de deux mots)**.**
La longueur du produit de deux mots x et y est la somme des longueurs des mots x et y.

$$|x \cdot y| = |x| + |y| \tag{40.4}$$

Proposition 40.9 (Monoïde $(\Sigma^*, \cdot, \varepsilon)$)**.**
L'ensemble Σ^* munie de l'opération produit d'élément neutre ε est un monoïde[1]*.*

Démonstration. Soient $x, y, z \in \Sigma^*$, avec les définitions précédentes, on peut vérifier facilement que :

— le produit est une loi interne : $x \cdot y \in \Sigma^*$;

— le produit est associatif : $x \cdot (y \cdot z) = (x \cdot y) \cdot z$;

— ε est l'élément neutre du produit : $x \cdot \varepsilon = \varepsilon \cdot x = x$.

□

Le produit n'est pas commutatif.

Définition 40.10.
*Soient Σ un alphabet et $w \in \Sigma^*$, la **puissance** n^ed'un mot w, notée w^n, est définie par :*

$$w^n = \begin{cases} \varepsilon & si\ n = 0 \\ w \cdot w^{n-1} & si\ n > 0 \end{cases}$$

40.1.2 Langage

Définition 40.11.
*Un **langage** sur un alphabet Σ est un sous-ensemble de Σ^*. C'est un ensemble de mots sur l'alphabet Σ.*

Un langage étant défini comme un ensemble, on peut appliquer toutes les notions de la théorie des ensembles aux langages.

Définition 40.12.
*Le **langage vide**, noté $\varnothing$ est le langage qui ne contient aucun mot.*

Définition 40.13.
*Le **langage unité** est le langage qui contient uniquement le mot vide : $\{\varepsilon\}$.*

Définition 40.14.
*Soient Σ un alphabet et $L_1, L_2 \subseteq \Sigma^*$ deux langages sur l'alphabet Σ, on définit le **produit** L de L_1 et L_2, noté $L_1.L_2$ par :*

$$L = L_1.L_2 = \{u_1 \cdot u_2, u_1 \in L_1, u_2 \in L_2\} \tag{40.5}$$

Le produit de langages est également appelé **concaténation**. Il ne faut pas confondre le produit de langage avec le produit cartésien de deux ensembles. Le langage unité est l'élément neutre du produit de langages.

Proposition 40.15 (Distributivité du produit de langage par rapport à l'union)**.**
Le produit de langage est distributif par rapport à l'union. Soient Σ un alphabet et $L_1, L_2, L_3 \subseteq \Sigma^*$, alors :

$$L_1.(L_2 \cup L_3) = (L_1.L_2) \cup (L_1.L_3) \ et\ (L_1 \cup L_2).L_3 = (L_1.L_3) \cup (L_2.L_3) \tag{40.6}$$

Démonstration. Soit $w \in L_1.(L_2 \cup L_3)$, montrons que $w \in (L_1.L_2) \cup (L_1.L_3)$. $\exists w_1 \in L_1, w' \in L_2 \cup L_3, w = w_1 \cdot w'$. Donc $w' \in L_2$ ou $w' \in L_3$. Si $w' \in L_2$, alors $w = w_1 \cdot w' \in L_1.L_2$. Si $w' \in L_3$, alors $w = w_1 \cdot w' \in L_1.L_3$. Donc, $w \in (L_1.L_2) \cup (L_1.L_3)$. Donc $L_1.(L_2 \cup L_3) \subseteq (L_1.L_2) \cup (L_1.L_3)$.

1. Définition 6.19.

Soit $w \in (L_1.L_2) \cup (L_1.L_3)$, montrons que $w \in L_1.(L_2 \cup L_3)$. $w \in L_1.L_2$ ou $w \in L_1.L_3$. Si $w \in L_1.L_2, i$ avec $i \in \{2,3\}$ alors $\exists w_1 \in L_1, w_i \in L_i, w = w_1 \cdot w_i$. Donc $w \in L_1.(L_2 \cup L_3)$. Donc, $(L_1.L_2) \cup (L_1.L_3) \subseteq L_1.(L_2 \cup L_3)$

Donc $(L_1.L_2) \cup (L_1.L_3) = L_1.(L_2 \cup L_3)$

L'autre partie de la proposition se montre de manière analogue. □

Définition 40.16.
*Soient Σ un alphabet et $L \subseteq \Sigma^*$, la **puissance** n^edu langage L, notée L^n est définie par :*

$$L^n = \begin{cases} \{\varepsilon\} & si\ n = 0 \\ L.L^{n-1} & si\ n > 0 \end{cases}$$

Définition 40.17 ([542])**.**
*L'**étoile de Kleene** est un opérateur unaire noté $*$. L'**itéré** d'un langage L, noté L^*, est l'application de l'étoile de Kleene à un langage L et est défini par :*

$$L^* = \bigcup_{i \geqslant 0} L^i \tag{40.7}$$

En particulier, on remarque que le mot vide fait toujours partie de l'itéré d'un langage, y compris quand ce même langage ne contient pas le mot vide.

Définition 40.18.
*L'**itéré strict** d'un langage L, noté L^+, est défini par :*

$$L^+ = \bigcup_{i > 0} L^i \tag{40.8}$$

Proposition 40.19 (Relations entre itéré et itéré strict)**.**
Soit L un langage, alors on a :

$$L^* = L^+ \cup \{\varepsilon\} \tag{40.9}$$

$$L^+ = L.L^+ = L^+.L \tag{40.10}$$

Chapitre 41

Utilisation dans les autres sciences

Dans ce chapitre nous donnons des applications de divers théorèmes dans les autres sciences que la mathématique.

41.1 Démystification du MRUA

41.1.1 Preuve de la formule

Nous sommes maintenant en mesure de donner une démonstration complète de la formule du MRUA :

$$x(t) = \frac{at^2}{2} + v_0 t + x_0. \tag{41.1}$$

Au niveau de la physique, nous considérons un mobile qui se déplace avec une accélération constante a. Nous notons par v_0 sa vitesse initiale et par x_0 sa position initiale.

Nous savons que, pour tout mouvement, si $x(t)$ est la position en fonction du temps, et si $v(t)$ et $a(t)$ représentent la vitesse et l'accélération en fonction du temps, alors

$$v(t) = x'(t) \quad \text{et} \quad a(t) = v'(t) = x''(t). \tag{41.2}$$

Afin de trouver $x(t)$ en connaissant $a(t)$, il « suffit » donc de prendre deux fois la primitive. Essayons ça dans le cas facile du MRUA où $a(t) = a$ est constante.

La vitesse $v(t)$ doit être une primitive de la constante a. Il est facile de voir que $v(t) = at$ est une primitive de a. Par le corolaire 11.137(bis),

$$v(t) = at + C_1 \tag{41.3}$$

pour une certaine constante C_1. Afin de fixer C_1, il faut faire appel à la physique : d'après la formule (41.3), la vitesse initiale est $v(0) = C_1$. Donc il faut identifier C_1 à la vitesse initiale : $C_1 = v_0$. Nous avons donc déjà obtenu que

$$v(t) = at + v_0. \tag{41.4}$$

Afin de trouver $x(t)$, il faut trouver une primitive de $v(t)$. Il n'est pas très difficile de voir que $at^2/2 + v_0 t$ fonctionne, donc il existe une constante C_2 telle que

$$x(t) = \frac{at^2}{2} + v_0 t + C_2. \tag{41.5}$$

Encore une fois, regardons la condition initiale : la formule donne comme position initiale $x(0) = C_2$, et donc nous devons identifier C_2 avec la position initiale x_0. En définitive, nous avons bien

$$x(t) = \frac{at^2}{2} + v_0 t + x_0. \tag{41.6}$$

Cette formule est donc maintenant *démontrée* à partir de la seule définition de la vitesse comme dérivée de la position et de l'accélération comme dérivée de la vitesse.

Remarquons cependant que la preuve complète fut *très* longue. En effet, nous avons utilisé les règles de dérivation de la proposition 11.115, pour la démonstration desquels, les résultats 11.19 et 11.18 ont étés utiles. Mais nous avons surtout utilisé le corolaire 11.137(bis) qui repose sur le théorème de Rolle 11.127, qui lui-même demande le théorème de Borel-Lebesgue 7.124 dans lequel la notion d'ensemble compact a été cruciale.

41.1.2 Interprétation graphique

La distance parcourue $x(t)$ en un temps t est la primitive de la vitesse. Nous avons, par ailleurs, que l'opération inverse de la dérivée donnait la surface. Pour reprendre les mêmes notations, nous notons $S_v(t)$ la surface contenue en dessous de la fonction v entre 0 et x. Nous ne serions donc pas étonné que

$$S_v(t) = \frac{at^2}{2} + v_0 t + x_0 \tag{41.7}$$

soit la surface en dessous de la fonction $v(t) = at + v_0$. Nous voyons que la surface totale sous la fonction $v(t) = at + v_0$ est exactement

$$S_v(t) = \frac{at^2}{2} + v_0 t. \tag{41.8}$$

Cela est un bon début, mais hélas nous ne retrouvons pas le terme « $+x_0$ » de la formule (41.7). Cela n'est pas tout à fait étonnant parce que nous savons que la surface sous une fonction était *une* primitive de la fonction, mais nous n'avons pas dit *laquelle*. D'après le fameux corolaire 11.137(bis), la primitive n'est définie qu'à une constante près. Ici, c'est la constante x_0 qu'on a perdue en chemin.

Nous parlerons plus en détail du lien entre les surfaces et les primitives dans la section dédié à l'intégration.

41.2 Relativité en mécanique newtonienne

41.2.1 Relativité du mouvement

Prenons quelqu'un qui cours le cent mètres en onze secondes. Par rapport à un spectateur dans les gradins, il se sera déplacé de cent mètres. Mais si je cours à côté de lui de telle façon à avoir parcouru 80 mètres le temps qu'il en fasse cent, alors par rapport à moi l'athlète ne se sera déplacé que de 20 mètres. Par contre, par rapport à mon chronomètre, il aura également mit onze secondes : ce n'est pas parce que je cours que mon chronomètre s'affole !

Entre moi et les spectateurs, on a donc une loi de transformation

$$x' = x - vt \qquad\qquad t' = t. \tag{41.9}$$

C'est-à-dire que la distance x' qu'aura parcouru l'athlète par rapport à moi vaut la distance x parcourue par le spectateur moins la vitesse que j'ai courue moi-même, c'est-à-dire moins vt.

41.2.2 Bob et Alice

Formalisons le concept de changement de repères. Pour cela, prenons deux amoureux, Bob et Alice[1]. Mettons que Bob reste assis sur un banc pendant qu'Alice cours en ligne droite à une vitesse v. Tout deux déclenchent leur chronomètre quand Alice passe devant Bob. À tout moment, Bob et Alice ont leur repères de temps et d'espace. Par exemple si après un temps t, Alice voir une peau de banane à 1 mètre devant elle, elle va dire « Il y a une peau de banane à un mètre. », tandis que Bob va dire « Il y a une peau de banane à $(1 + vt)$ mètres ».

Plus généralement, s'il se passe quelque chose à la position x au temps t pour Bob, ce quelque chose se passera au temps $t' = t$ à l'endroit $x' = x - vt$ pour Alice parce qu'en un temps t, elle aura déjà avancé d'une distance vt.

Ça c'est ce dont tout le monde était persuadé depuis Galilée jusqu'au début du vingtième siècle.

1. C'est plus poétique que dire « soient A et B deux observateurs ».

41.3 Invariance de la vitesse de la lumière

41.3.1 Champ de gravitation et électrique

Nous savons que que la force de gravitation s'écrit :

$$F_{grav} = G\frac{mm'}{r^2},$$

tandis que la force électrique entre deux charges q et q' est donnée par

$$F_{elec} = k\frac{qq'}{r^2}. \tag{41.10}$$

Nous avons aussi fait remarquer que dans le cas de la gravitation, la force a l'air d'être instantanée, et que cela posait quelques problèmes conceptuels. La force électrique a apparemment le même problème. Une différence entre les deux est qu'une charge électrique c'est tout petit et qu'on peut expérimenter à souhait, tandis que pour avoir une masse dont on peut mesurer le champ de gravitation correctement, il faut quelque chose grand comme la Terre [2].

41.3.1.1 Finitude de la vitesse de propagation de la force électrique

Si un micro est placé juste à côté de ton oreille, et qu'il commence à faire biiiiip, tu l'entends directement. Quand il s'arrête, tu ne l'entends plus. Si le micro est placé à 600 m de toi, tu ne commenceras à l'entendre que deux secondes après le commencement du son, et tu continueras à l'entendre deux secondes après qu'il ait fini.

Eh bien, pour la force électrique, on a pu mesurer que c'est la même chose (sauf que ça va beaucoup plus vite). Si on place une charge quelque part, on ne ressent la force (41.10) qu'après qu'elle ait eut le temps d'arriver. Si on déplace la charge électrique, on continue à ressentir la même force pendant un certain temps : il faut que la modification du champ électrique ait le temps d'arriver. Exactement comme quand on fait des remous quelque part dans un étang : il faut du temps que les remous arrivent plus loin.

On a pu faire des dizaines d'expériences de ce type avec l'électricité, le magnétisme et la lumière ; et les résultats sont clairs : il faut du temps pour que ça se déplace. Tout cela provoque des ondes électromagnétiques qui se déplacent à une vitesse finie. On peut produire de telles ondes avec n'importe quel courant électrique alternatif.

41.3.1.2 Pourquoi pas la gravitation ?

La gravitation telle que donnée par Newton pose le même problème de vitesse de propagation que l'électricité. Est-ce qu'en réalité la gravitation se propage également à une vitesse finie ?

Avec la gravitation c'est beaucoup plus compliqué parce qu'elle est beaucoup plus faible, et donc c'est beaucoup plus difficile à détecter. D'après la théorie d'Einstein de la gravitation, la gravitation devrait également produire des ondes gravitationnelles. Seulement, si un simple courant électrique suffit pour mesurer une onde électromagnétique, affin de mesurer une onde gravitationnelle, il faudrait un déplacement de masse de l'ampleur d'une étoile qui explose. Or ça, on ne sait pas produire dans un laboratoire. Les physiciens sont donc pour l'instant (2009) en train d'attendre qu'une étoile explose pas trop loin d'ici affin d'être capable de mesurer une onde gravitationnelle.

L'existence de ces ondes de gravitation ne fait aucun doute dans la tête d'aucun physicien parce qu'elles sont une conséquence logique (et mathématique) de la théorie de la relativité générale, laquelle a déjà eut beaucoup de confirmations expérimentales. Mais comme on est dans le cadre d'une science expérimentale, il faut être patient et attendre d'en avoir effectivement observée une avant de dire avec certitude que ça existe.

2. Une autre différence fondamentale est qu'il existe des charges électriques négatives, mais pas de masses négatives ; de ce fait on ne peut pas construire d'isolant gravitationnel, contrairement aux isolants électriques qui existent. Cela augmente encore la difficulté de faire des expériences avec la gravitation.

41.3.2 Support du champ : pas d'éther

Nous avons dit qu'une onde électromagnétique se propage comme une onde sonore (quoique beaucoup plus vite). Une question se pose alors. En effet, une onde sonore est matérialisé par de l'air qui vibre. Qu'est-ce qui vibre pour une onde électromagnétique ?

Étant donné que les ondes électromagnétiques se propagent dans le vide (c'est pour ça que la radio fonctionne dans l'espace), la question est problématique. Les physiciens ont donc supposé que tout l'univers était rempli d'un fluide invisible appelé **l'éther**. L'électromagnétisme consiste en une vibration de l'éther, exactement comme l'acoustique consiste en une vibration de l'air.

En fait, vérifier cette hypothèse n'est pas très compliqué. En effet il n'y a aucune raison que l'éther suive la Terre dans son mouvement. Or la Terre se déplace à environ 30 km/s autour du Soleil. Donc les ondes électromagnétiques doivent se propager plus vite dans le sens du mouvement de la Terre que dans le sens perpendiculaire. Tout comme le son se propage plus vite dans le sens du vent.

La célèbre expérience de Michelson-Morley a mesuré cet effet ... et ce fut la consternation : il n'y a aucun effet ! Or, la lumière se déplace à 300.000 km/s ; une variation de 30 km/s devrait être détectable !

Mais rien ! On a recommencé les expériences dans tout les sens, à tous les mois de l'année, à tous les endroits de la Terre. On n'a pas observé un poil de variation de la vitesse de la lumière. Et ça, ça pose un gros problème à la physique.

41.3.3 Le problème

Si je joue au football dans un train qui avance à 100 km/h et que je lance une balle à 20 km/h, quelqu'un au sol mesura la vitesse de la balle soit à 120 km/h soit à 80 km/h d'après que l'on ait shooté vers l'avant ou l'arrière du train. Cela paraît logique. Mais ce qu'on vient de voir c'est que ça ne marche pas avec la lumière.

Si un train avance à 100.000 km/s et qu'on y allume une lampe de poche, la lumière avancera à 300.000 km/s par rapport au train et 400.000 km/s par rapport au sol. Non ! Justement pas ! La lumière avancera quand même à 300.000 km/s par rapport au sol.

Là encore, on a fait des dizaines d'expériences partout, sur Terre, dans des avions, dans l'espace avec des atomes, des lampes de poche et des horloges atomique, et dans tous les sens, le sens de déplacement de la Terre, le sens inverse, le sens perpendiculaire, vers le haut, vers le bas : rien ! Personne n'a jamais observé un rayon de lumière se déplacer à une autre vitesse que 300.000 km/s.

Le problème est que le principe d'addition des vitesses est faux pour la lumière. Puisque l'expérience nous force, nous devons faire avec.

Loi numéro 1.
La réalité est que la vitesse de la lumière est la même dans tous les référentiels. On note c cette vitesse. C'est une constante fondamentale de la Nature.

Étant donné que c'est une loi expérimentale, nous n'en pouvons rien. C'est la nature qui est comme ça. En particulier tu ne peux pas en vouloir à ton prof de physique d'avoir inventé une théorie compliquée. Ce n'est pas lui qui l'a inventée et ce n'est pas de sa faute.

41.4 Conséquences

C'est maintenant que les choses vraiment graves commencent (cela soit dit sans vouloir te faire peur). Affin d'un peu simplifier les choses, nous n'allons étudier que les mouvements en une dimension, c'est-à-dire sur une droite.

41.4.1 Ligne d'univers

Un événement a une coordonnée (t, x). Si je pose un objet juste à mes pieds (disons en $x = 0$), ses coordonnées seront à tout moment $(t, 0)$. Il est bon de voir cette coordonnée comme l'équation

paramétrique d'une droite horizontale dans le plan des coordonnées t et x. Plus généralement quand un mobile effectue un mouvement $x(t)$, on appelle la **ligne d'univers** du mobile la ligne (pas forcément droite) $(t, x(t))$. Dans le premier exemple, on avait $x(t) = 0$ pour tout t.

Le cas d'un mobile se déplaçant à vitesse constante v donne comme ligne d'univers la droite[3] $(t, x_0 + vt)$, et un objet qui se déplace selon un MRUA a comme ligne d'univers

$$\big(t, x_0 + v_0 t + \frac{at^2}{2}\big).$$

41.4.2 Transformations de Lorentz

Reprenons les amours scientifiques de Bob et Alice, mais cette-fois, analysons celles-ci en tenant compte du fait que la vitesse de la lumière soit invariante. Maintenant, si Bob voit se passer quelque chose au temps t à l'endroit x, on va dire qu'Alice voit cette chose au temps t' à la position x', et on va chercher (t', x') en fonction de (t, x).

Posé en termes mathématiques, le problème s'énonce ainsi : trouver les fonctions f et g telles que les formules

$$t' = f(t, x) \tag{41.11a}$$
$$x' = g(t, x) \tag{41.11b}$$

donnent les coordonnées vues par Alice pour un événement vu par Bob à l'instant t au point x. Une première étape importante est franchie par la proposition suivante[4].

Proposition 41.1.
Les fonctions f et g contenues dans les transformations (41.11) sont nécessairement linéaires (affines), c'est-à-dire qu'elles doivent s'écrire sous la forme

$$t' = \alpha t + \beta x + p$$
$$x' = \gamma t + \delta x + q$$

pour certaines fonctions α, β, γ, δ, p et q de la vitesse d'Alice relativement à Bob.

Démonstration. Pendant qu'Alice court et que Bob la regarde, Ève tente de lancer une pierre sur Alice (Ève est jalouse). Bob et Alice regardent deux événements. Le premier est la pierre qui quitte la main de Ève, et le second est la pierre qui percute le sol. Pour Bob, le jet s'est passée au temps t_0 au point x_0, et la pierre touche le sol un petit peu plus tard, au temps $t_0 + \Delta t$ et un peu plus loin, au point $x_0 + \Delta x$. Bob écrit donc ceci sur sa feuille de papier :

$$E_1 = (t_0, x_0)$$
$$E_2 = (t_0 + \Delta t, x_0 + \Delta x),$$

tandis qu'Alice, en observant les mêmes deux événements, aura noté

$$E_1' = \big(f(t_0, x_0), g(t_0, x_0)\big)$$
$$E_2' = \big(f(t_0 + \Delta t, x_0 + \Delta x), g(t_0 + \Delta t, x_0 + \Delta x)\big).$$

Bob et Alice se demandent combien de temps la pierre est restée en l'air et quelle distance elle a parcourue. Par le principe général d'homogénéité, les deux seules quantités pertinentes (qui ont un sens physique) pour Bob sont $(t_0 + \Delta t) - t_0$ et $(x_0 + \Delta x) - x_0$, c'est-à-dire Δt et Δx. En effet, si Bob avait choisi de s'asseoir autre part et si Alice avait commencé à courir un peu plus tard, ça n'aurait rien changé à la longueur du jet de Ève.

D'une façon ou d'une autre, il doit exister une façon de déduire les mesures de Alice en connaissant celles de Bob ; je ne connais pas avec quelles formules, mais ces formules ne peuvent contenir que Δt, Δx et v parce que ce sont les seules quantités qui définissent tous les événements.

3. bon exercice de révision de ton cours de math de vérifier que c'est une droite.
4. dont je te suggère fortement de ne pas lire la preuve si tu ne veux pas que ton cerveau éclate.

Cela dit, Alice va caractériser le mouvement de la pierre avec la différence des coordonnée entre le jet et la chute sur le sol mesurées par elle-même. En d'autres termes, pour Alice ce qui compte c'est la différence entre E_1' et E_2', soit

$$\big(f(t_0+\Delta t, x_0+\Delta x), g(t_0+\Delta t, x_0+\Delta x)\big) - \big(f(t_0,x_0), g(t_0,x_0)\big). \tag{41.12}$$

Mais nous venons de signaler que ce qu'Alice mesurait devait pouvoir être exprimé en termes de Δt et Δx. Nous concluons que la différence (41.12) ne dépend en fait pas de x et t mais seulement de Δt et Δt.

Prenons maintenant une notation plus compacte et notons $X = (t,x)$, $\Delta X = (\Delta t, \Delta x)$ puis $F = (f,g)$. Avec ça, l'expression (41.12) se note $F(X+\Delta X) - F(X)$. Comme mentionné, cette expression ne dépend que de Δx. En particulier, elle ne dépend pas de X.

Maintenant tu vas comprendre pourquoi on apprend les dérivées dans ton cours de math. Comme $F(X+\Delta X)-F(X)$ ne dépend pas de X, le rapport $\big(F(X+\Delta X)-F(X)\big)/\Delta X$ non plus. La limite de ce rapport quand ΔX tend vers zéro non plus :

$$\lim_{\Delta X\to 0} \frac{F(X+\Delta X)-F(X)}{\Delta X} \tag{41.13}$$

ne dépend pas de X. Tu reconnais là la dérivée de F au point X. En d'autres termes, $F'(X)$ est constante, elle ne dépend pas de X. Disons donc que $F'(X) = a$. Tu connais beaucoup de fonctions dont la dérivée est constante ? Non ? En effet, il n'y en a pas beaucoup. Les fonctions qui vérifient $F'(X) = a$ signifie sont toutes de la forme

$$F(X) = aX + b.$$

À ce niveau, il convient de re-déballer les notations compactes : si $a = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ et $b = (p,q)$ on trouve

$$f(t,x) = \alpha t + \beta x + p \tag{41.14a}$$
$$g(t,x) = \gamma t + \delta x + q, \tag{41.14b}$$

comme annoncé.

□

Nous savons que lorsque $(t,x) = (0,0)$, alors $(t',x') = (0,0)$. En effet, Bob et Alice ont lancés leurs chronos en même temps au moment où ils étaient au même endroit. En mettant $(t,x) = (0,0)$ dans les équations (41.14), on trouve $(t',x') = (p,q)$, et donc $p = q = 0$. Ça fait une chose de réglée ; on se retrouve avec

$$\begin{cases} t' = \alpha t + \beta x & (41.15a) \\ x' = \gamma t + \delta x. & (41.15b) \end{cases}$$

Quelles sont les contraintes à vérifier pour que ces transformations décrivent correctement la physique que l'on cherche à écrire ?

(1) Il faut que les transformations décrivent correctement que Alice avance à une vitesse v par rapport à Bob,

(2) dans le même ordre d'idée, il faut que l'on trouve que Bob avance à la vitesse $-v$ par rapport à Alice,

(3) il faut que si Alice et Bob observent un rayon lumineux, ce rayon aille à la vitesse c par rapport à Alice et à la *même* vitesse c par rapport à Bob,

(4) enfin, il faut avoir le principe de relativité, c'est-à-dire que comme les équations (41.15) disent ce que Alice voit en fonction de ce que Bob voit, on demande que les équations qui disent ce que Bob voit en fonction de ce que Alice voit soient les mêmes. En d'autres termes, il faut que les transformations et les transformations inverses soient les mêmes au changement près du signe de v.

Étudions une à une ce que chacune de ses contraintes impose. Rappelons que (t,x) et (t',x') sont les coordonnées que Bob et Alice mettent sur le même événement. Par exemple sur l'événement qui consiste à ce que Ève, par jalousie envers Bob, jette une peau de banane sous les pieds d'Alice. Cet événement a lieu à un certain moment, à un certain endroit. C'est ce moment et cet endroit qui sont notés (t,x) et (t',x').

(1) Les coordonnées (t,x) et (t',x') peuvent décrire n'importe quoi. Regardons les coordonnées de Alice qui cours. Pour Alice, cela correspond à $(t',x') = (t',0)$ parce que si x' désigne la position de Alice par rapport à Alice, alors x' est toujours nul. Pour Bob par contre, Alice ne reste pas en place, mais se déplace à une vitesse v. C'est-à-dire que si (t,x) sont les coordonnées de Alice pour Bob, alors $x/t = v$. Écrivons les équations (41.15) en tenant compte de tout ça : avec $x' = 0$, la seconde équation donne

$$0 = \gamma t + \delta x, \tag{41.16}$$

d'où on déduit que $x/t = -\gamma/\delta$. En imposant que cela soit v, on trouve $\gamma = -v\delta$, et on ré-écrit les transformations en tenant compte de ça :

$$\begin{cases} t' = \alpha t + \beta x & (41.17a) \\ x' = -v\delta t + \delta x. & (41.17b) \end{cases}$$

Nous voilà débarrassé d'un paramètre.

(2) Maintenant, on regarde ce qu'il se passe quand (t,x) et (t',x') décrivent les positions de Bob. On a $(t,x) = (t,0)$ parce que selon Bob, Bob est au repos. Les équations deviennent :

$$t' = \alpha t \qquad\qquad x' = -v\delta t. \tag{41.18}$$

La vitesse de Bob par rapport à Alice est $-v$, donc on exige que $x'/t' = -v$, c'est-à-dire que

$$\frac{-v\delta t}{\alpha t} = -v,$$

ce qui implique que $\delta = \alpha$. On avance encore un peu. Écrivons à nouveau les lois de transformation en en tenant compte :

$$\begin{cases} t' = \alpha t + \beta x & (41.19a) \\ x' = -v\alpha t + \alpha x. & (41.19b) \end{cases}$$

(3) Si maintenant Bob et Alice regardent un même rayon de lumière (comme c'est romanesque !), alors (t,x) et (t',x') expriment les coordonnées d'un rayon lumineux expriment les coordonnées d'un rayon lumineux. Le fait que Bob regarde un rayon lumineux fait que $x = ct$, et donc que les coordonnées du rayon lumineux, observé par Alice sont :

$$t' = \alpha t + \beta ct \qquad\qquad x' = -\alpha vt + \alpha ct. \tag{41.20}$$

L'invariance de la vitesse de la lumière exige que Alice mesure une vitesse c pour le rayon de lumière, c'est-à-dire $x' = ct'$. On exige donc que

$$-\alpha vt + \alpha ct = c\alpha t + \beta c^2 t,$$

ce qui implique que

$$\beta = -\frac{\alpha v}{c^2}.$$

Une fois de plus, l'avant-dernière, on ré-écrit les lois de transformations en tenant compte de ce fait ; mais cette fois, on fait l'effort d'écrire aussi les transformations inverses :

$$t' = \alpha t - \frac{\alpha v}{c^2}x \qquad\qquad t = \frac{1}{\Delta}\big(\alpha t' + \frac{\alpha v}{c^2}x'\big) \tag{41.21}$$

$$x' = -\alpha vt + \alpha x \qquad\qquad x = \frac{1}{\Delta}(\alpha vt' + \alpha x') \tag{41.22}$$

où $\Delta = \alpha^2 - \frac{\alpha^2 v^2}{c^2}$ que tu noteras au passage être toujours positif, et nul uniquement quand $v = c$.

(4) Maintenant il reste à imposer le principe de relativité. Les transformations (41.21) montrent comment Alice voit le monde (c'est-à-dire (t', x')) en fonction de la façon dont Bob voit le monde (c'est-à-dire (t, x)). On se demande donc quelle seraient, pour Bob, les coordonnées (t, x) d'un point vu en (t', x') par Alice. Cela signifie que l'on impose que les deux systèmes (41.21) soient en réalité les mêmes, à un changement de signe près.

Attention : il *à priori* faux de dire qu'en changeant le signe de v dans $\alpha v/c^2$, j'obtiens $-\alpha v/c^2$ parce que α est une fonction de v. En réalité, il faut noter $\alpha(v)v/c^2$ et donc le changement de signe de v donne $-\alpha(-v)v/c^2$. Ceci étant clair, on peut un petit peu calculer.

Commençons par égaliser le coefficient de x dans t' à celui de x' dans t, en changeant le signe de v :

$$\frac{\alpha(-v)v}{c^2} = \frac{\alpha(v)v}{c^2},$$

et donc $\alpha(v) = \alpha(-v)$. Ça c'est une bonne nouvelle. Égalisons maintenant le coefficient de t dans t' à celui de t' dans t en changeant le signe de v :

$$\alpha(-v) = \frac{\alpha(v)}{\Delta(v)} = \frac{\alpha(v)}{\alpha(v)^2\big(1 - \frac{v^2}{c^2}\big)}.$$

Comme $\alpha(-v) = \alpha(v)$, on en déduit que

$$\alpha(v) = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{41.23}$$

Maintenant qu'on a tout, on peut écrire les transformations de Lorentz. On met donc l'expression (41.23) dans les lois de transformations (41.21) :

$$\begin{aligned} t' &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}\Big(t - \frac{v}{c^2}x\Big) & t &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}\Big(t' + \frac{v}{c^2}x'\Big) \\ x' &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}(x - vt) & x &= \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}(vt' + x'). \end{aligned} \tag{41.24}$$

Tu remarqueras que $\Delta = 1$; si tu ne sais pas ce qu'est le déterminant d'une application linéaire, ça n'a pas d'importance. Mais si tu sais ce qu'est le déterminant d'une application linéaire, alors ce $\Delta = 1$ est crucial !

Affin d'avoir des équations un peu plus courtes, à partir de maintenant nous allons noter

$$\gamma(v) = \sqrt{1 - \frac{v^2}{c^2}}.$$

41.4.3 Conditions d'existence

Comme tu vois une racine carrée et un dénominateur dans ces formules, tu dois te demander quelles sont les conditions d'existence. Étant donné que $v < c$, on a $v^2/c^2 \leqslant 1$ et en particulier, $v^2/c^2 = 1$ si et seulement si $v = c$.

Ce qui se trouve dans la racine carrée ne pose donc jamais de problèmes parce que ce n'est jamais négatif.

Le dénominateur est par contre plus problématique : quand $v = c$ il n'y a plus rien qui fonctionne. Quelle est la physique de ce problème ? Pour le comprendre, il faut se souvenir ce que représente v. Nous avons dit que v est la vitesse à laquelle Alice court. Ce que la condition d'existence nous enseigne, c'est que personne ne peut courir à la vitesse de la lumière. C'est une vitesse que l'on ne peut pas atteindre.

Dit en termes plus savants, on ne peut pas choisir un repère qui se déplace à la vitesse de la lumière.

La question qui se pose alors est « ah bon, on ne peut pas atteindre la vitesse de la lumière ! Et la lumière, comment elle fait ? ». Bonne question, merci de l'avoir posée. Hélas la réponse sort du cadre de ce cours.

Loi numéro 2.
Aucun objet ne peut atteindre la vitesse de la lumière.

Loi numéro 3.
Tu ne dois pas te demander pourquoi la lumière elle-même se déplace à la vitesse de la lumière malgré la loi numéro 2.

41.4.4 La notion d'intervalle

Un **événement** est quelque chose qui se passe à un endroit à un certain moment. C'est donc caractérisé par le moment et le lieu. Comme on travaille à une dimension, c'est un couple de réels (t, x).

Regardons un rayon de lumière. Un événement est le fait d'allumer une lampe de poche, et un autre est le fait que le lumière arrive sur l'objet qu'on éclaire. Appelons-les (t_1, x_1) et (t_2, x_2). Comme d'habitude, on note $\Delta t = t_2 - t_1$ et $\Delta x = x_2 - x_1$. Comme le rayon de lumière va à la vitesse c, on a $c = \Delta x/\Delta t$, ou encore

$$c^2\Delta t^2 - \Delta x^2 = 0.$$

Pour cette raison, on va dire que l'**intervalle** entre deux événements (t_1, x_1) et (t_2, x_2) vaut en général

$$s^2 = c^2(t_2 - t_1)^2 - (x_2 - x_1)^2. \tag{41.25}$$

Par invariance de la vitesse de la lumière, si un intervalle est nul pour un observateur, il sera nul pour tous les observateurs.

41.4.4.1 En mécanique newtonienne

Affin de voir un peu mieux l'enjeu de l'invariance de l'intervalle, regardons un exemple chiffré. Si par exemple je me déplace de 10 m en 5 s, mon intervalle mesuré par une personne extérieure est

$$c^2\Delta t^2 - \Delta x^2 = (300.000.000)^2 \cdot (5)^2 - (10)^2 = 2,25 \cdot 10^{18}\,\text{m}.$$

Si je fais le calcul pour moi, j'ai que $\Delta x' = 0$ parce que je ne me déplace pas, et $\Delta t' = 5$ parce que je me suis déplacé en 5 secondes. Le truc est que à côté de $(300.000.000)^2$, l'intervalle spatial Δx ne pèse pas grand chose. Ça ne change presque rien qu'il soit de 5 mètres ou de zéro. Ça ne change pas grand chose, mais ça change quand même ! Entre moi qui calcule ou une personne extérieure, l'intervalle change de 100 sur un nombre de la grandeur de 200.000.000.000.000.000.0000 !

Reprenons plus clairement le raisonnement. D'après la mécanique classique, l'intervalle mesuré par deux personnes est différent, mais très peu différent. Inutile de dire que du temps de Newton, on n'avait pas les moyens techniques de mesurer si cet intervalle est effectivement différent ou bien s'il est en réalité égal. C'est un peu comme si on te mettait un spot dans les yeux et qu'on te demandait si c'est un spot de 1000 W ou de 1001 W. Bonne chance pour le dire !

41.4.4.2 En mécanique relativiste

Maintenant qu'on a des moyens techniques nettement plus poussés que Newton, on a pu mesurer que l'intervalle est égal. L'intervalle est un invariant. Cela n'est pas un nouveau principe physique parce qu'il découle des transformations de Lorentz.

41.4.5 Le cône de lumière d'un point

Il est intéressant de dessiner dans le plan (t, x) l'ensemble des points atteints par le rayon lumineux. Le point (t, x) est atteint si $c^2t^2 - x^2 = 0$, ou encore si $x = \pm ct$. Cela forme deux droites dans le plan tracé par les coordonnées t et x. Ces deux droites forment ce qu'on appelle le **cône de lumière** du point $(0, 0)$.

41.4.6 Contraction des longueurs

Bob prend un morceau de bois qu'il mesure de longueur l et le dépose devant lui. À l'instant t (de Bob), les deux extrémités sont aux coordonnées $e_1 = (t, 0)$ et $e_2 = (t, l)$.

Afin de savoir quelle est la longueur de ce même morceau de bois pour Alice, il faut qu'elle mesure les deux extrémités en même temps (pour elle), et qu'elle fasse la différence. Comme Bob et Alice déclenchent leurs chronomètres en même temps, le plus simple est de faire la mesure à cet instant.

Pour Bob, c'est clair : les coordonnées des deux extrémités sont $e_1 = (0, 0)$ et $e_2 = (0, l)$. La longueur du bois est l. Pour savoir quelle est la longueur mesurée par Alice, on utilise les transformations de Lorentz qui donnent les coordonnées e'_1 et e'_2 relatives à Alice. On trouve $e'_1 = (0, 0)$ et

$$e'_2 = \left(\frac{-vl/c^2}{\gamma(v)}, \frac{l}{\gamma(v)}\right). \tag{41.26}$$

En d'autres termes, on a $x_1 = 0$ et $x_2 = l/\gamma(v)$, ce qui fait que la longueur observée par Alice est $l' = x_2 - x_1 = l/\gamma(v)$.

Eh bien ce résultat est faux. Si tu vois pourquoi sans lire la suite, tu es très fort.

Pour mesurer la longueur d'un objet, il faut mesurer la position des deux bouts *en même temps* puis faire la différence entre les deux. Effectivement, e_1 et e_2 sont en même temps pour Bob, et donc Bob peut mesurer la longueur de son bout de bois en faisant la différence $x_2 - x_1$. Mais comme le montre les coordonnées (41.26), les événements e'_1 et e'_2 ne se passent pas en même temps pour Alice ! Eh oui : $t'_1 = 0$ et $t'_2 = -vl/c^2\gamma(v)$; c'est pas la même chose.

Il faut donc trouver un événement qui pour Alice correspond à l'extrémité du bout de bois au temps $t' = 0$. Comme l'événement général qui correspond au bout du bois pour Bob est (t, l), l'événement général est pour Alice

$$t' = \frac{t - \frac{v}{c^2}l}{\gamma(v)} \qquad x' = \frac{l - vt}{\gamma(v)}. \tag{41.27}$$

Affin d'avoir $t' = 0$, il faut $t = vl/c^2$. En mettant cette valeur de t dans x', on trouve

$$x' = \frac{l - v\left(\frac{vl}{c^2}\right)}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{l\left(1 - \frac{v^2}{c^2}\right)}{\sqrt{1 - \frac{v^2}{c^2}}} = l\gamma(v).$$

Et là, c'est la bonne formule. Si un objet a une longueur l dans le référentiel où il est au repos, il aura une longueur

$$l' = l\sqrt{1 - \frac{v^2}{c^2}} \tag{41.28}$$

dans un référentiel qui se déplace à la vitesse v par rapport à l'objet.

41.4.7 Dilatation des intervalles de temps

Encore un petit effort et promit, je te donne une application concrète que tu connais des bizarreries de la relativité. Mais en attendant, regarde bien ta montre, tu ne va pas en croire tes yeux !

Reprenons Bob et Alice. On se rappelle que Bob et Alice avaient déclenchés leurs chronomètres en même temps quand Alice était passée devant Bob. Un peu plus tard, Alice regarde sa montre qui indique un temps t. Et elle se demande si Bob a aussi à ce moment une montre qui indique un temps t.

Ce serait dingue que non hein ! ? ! En effet, si je synchronise ma montre avec quelqu'un et que je pars faire un tour, ma montre ne sera pas tout d'un coup désynchronisée. Oui, mais Alice, elle cours presque à la vitesse de la lumière ... et à ces vitesses-là, on a déjà vu des choses incroyables. Calculons donc pour en avoir le cœur net.

Le fait qu'Alice regarde sa montre est un événement qui se passe pour Alice aux coordonnées $(t', 0)$ (le zéro c'est parce que par rapport à elle-même, Alice est toujours au repos). À quelles coordonnées (t, x) pour Bob correspond cet événement ?

L'équation de t en fonction de t' et x' dans les transformations de Lorentz (41.24) prise avec $x' = 0$ donnent

$$t = \frac{t'}{\gamma(v)}.$$

Et si, juste pour le plaisir, on faisait l'inverse ? Bob regarde sa montre, il voir un temps t et sa coordonnée spatiale est $x = 0$. À quel temps d'Alice cela correspond ? Mettons $x = 0$ dans la transformation de Lorentz de t' en fonction de t et x. Ce qu'on obtient c'est

$$t' = \frac{t}{\gamma(v)}.$$

N'est-ce pas génial ? C'est la même ! Évidemment, ça ne pouvait pas être autre chose : le principe de relativité demande qu'on ne puisse pas faire la différence entre Alice qui cours vers la droite avec Bob assis et Alice assise avec Bob qui cours vers la gauche. C'est exactement pour ça que dans une gare, quand le train d'à côté démarre, il t'arrive de croire que c'est ton train qui démarre : tu ne peux pas faire la différence, c'est un principe physique.

41.4.8 Invariance de l'intervalle

Dans deux secondes, je vais te montrer comment une utilisation intelligente des exponentielles permet de trouver un résultat très fort en relativité. Quoi ? Les exponentielles, les mêmes qu'au cours de math ? Eh oui : la même exponentielle que celle qu'on t'a introduit avec des populations de bactéries qui se multiplient, cette même exponentielle qui la la miraculeuse propriété d'être égale à sa propre dérivée.

Mais n'anticipons pas.

Nous avons déjà signalé que si la quantité $\Delta s^2 = c^2\Delta t^2 - \Delta x^2$ était nulle pour un observateur, alors elle était nulle pour tous les observateurs. Supposons deux événements A et B observés par Alice et Bob. Bob les note aux coordonnées (t_a, x_a), et (t_b, x_b) tandis qu'Alice les note en (t'_a, x'_a) et (t'_b, x'_b).

L'intervalle entre les deux événements mesuré par Bob sera

$$s^2 = c^2(t_b - t_a)^2 - (x_b - x_a)^2,$$

tandis que ce même intervalle mesuré par Alice sera

$$s'^2 = c^2(t'_b - t'_a)^2 - (x'_b - x'_a)^2.$$

On peut bien entendu remplacer dans la première équation les t_a, t_b, x_a et x_b par leurs valeurs en termes de t'_a, t'_b, x'_a et x'_b données par les transformations de Lorentz. Tu paries que les trois quart des termes dans le calcul se simplifient et qu'il restera exactement s'^2 ? Je te dis que oui, et je te conseille de me croire sur parole, sinon tu vas devoir lire le calcul suivant :

$$\begin{aligned} s^2 = c^2(t_b - t_a)^2 - (x_b - x_a)^2 = c^2 &\left(\frac{1}{\gamma(v)}(t'_b + \frac{v}{c^2}x'_b) - \frac{1}{\gamma(v)}(t'_a + \frac{v}{c^2}x'_a)\right)^2 \\ &- \left(\frac{1}{\gamma(v)}(vt'_b + x'_b) - \frac{1}{\gamma(v)}(vt'_a + x'_a)\right)^2. \end{aligned}$$

Jusqu'ici, on n'a fait que remplacer les choses par leurs valeurs données par les transformations de Lorentz. Maintenant on regroupe à l'intérieur de chaque parenthèse les termes de façon à faire

apparaitre $\Delta x'$ et $\Delta t'$:

$$\begin{aligned} s^2 &= \frac{c^2}{\gamma(v)^2}\left((t'_b - t'_a) + \frac{v}{c^2}(x'_b - x'_a)\right)^2 \\ &\quad - \frac{1}{\gamma(v)^2}\left((x'_b - x'_a) + v(t'_b - t'_a)\right)^2 \\ &= \frac{c^2}{\gamma(v)^2}\left((\Delta t')^2 + 2\frac{v}{c^2}\Delta t'\Delta x' + \frac{v^2}{c^4}(\Delta x')^2\right) \\ &\quad - \frac{1}{\gamma(v)^2}\left((\Delta x')^2 + 2v\Delta x'\Delta t' + v^2(\Delta t')^2\right). \end{aligned}$$

Là, on a utilisé le produit remarquable $(a+b)^2 = a^2+2ab+b^2$, et on a systématiquement renommé tous les intervalles avec la notation Δ pour être plus compact. Maintenant, on va regrouper tous les termes contentant $(\Delta t')^2$ ensemble, tous ceux qui contiennent $\Delta t'\Delta x'$ ensemble et ceux qui contiennent $(\Delta x')^2$ ensemble. Autre manière de le dire, on met les Δ en évidence comme on peut. On trouve ceci :

$$\begin{aligned} (\Delta t')^2\left(\frac{c^2}{\gamma(v)^2} - \frac{v^2}{\gamma(v)^2}\right) + \Delta t'\Delta x'\left(\frac{2vc^2}{\gamma(v)^2c^2} - \frac{2v}{\gamma(v)^2}\right) \\ + (\Delta x')^2\left(\frac{c^2v^2}{c^4\gamma(v)^2} - \frac{1}{\gamma(v)^2}\right). \end{aligned}$$

À partir de là, je te laisse vérifier (en utilisant le fait que $\gamma(v)^2 = 1 - v^2/c^2$) que les coefficients se simplifient beaucoup et valent finalement respectivement c^2, 0 et -1 comme il se doit. Avec tout ça, nous avons montré le résultat très important suivant :

> L'intervalle entre deux événements est invariant sous les changements de repères d'inertie, c'est-à-dire que la valeur mesurée par n'importe qui qui se déplace en MRU sera toujours la même.

Pourquoi cela est tellement important ? À cause de Pythagore et d'une petite démonstration à coups d'exponentielles [5].

41.4.8.1 Rappel de trigonométrie hyperbolique

Les fonctions de trigonométrie hyperboliques sont :

$$\cosh(x) = \frac{e^x + e^{-x}}{2} \qquad \sinh = \frac{e^x - e^{-x}}{2}. \tag{41.29}$$

Elles ont pas mal de propriétés en commun avec les sinus cosinus et normaux. D'abord, leurs dérivées sont faciles à calculer :

$$\begin{aligned} \cosh'(x) &= \sinh(x) \\ \sinh'(x) &= \cosh(x) \end{aligned}$$

où tu noteras qu'il n'y a pas de signe moins qui apparaît, contrairement au cas de la trigonométrie normale. Une autre propriété qui ressemble fort à une propriété de la trigonométrie est :

Proposition 41.2.
Pour tout $x \in \mathbb{R}$,

$$\cosh^2(x) - \sinh^2(x) = 1 \tag{41.30}$$

avec un signe moins comme différence avec la trigonométrie.

5. oui oui tout ton cours de math va finir par y passer.

Démonstration. La preuve revient simplement à calculer en utilisant le produit remarquable de $(a+b)^2$. D'abord, on a :

$$\cosh^2(x) = \frac{1}{4}\big(e^x + e^{-x}\big)^2 = \frac{1}{4}\big(e^{2x} + 2e^x e^{-x} + e^{-2x}\big) = \frac{1}{4}(e^{2x} + 2 + e^{-2x})$$

où l'on a utilisé le fait que $(e^x)^2 = e^{2x}$ et que $e^x e^{-x} = 1$. Il te reste à faire la même chose pour $\sinh^2(x)$, la réponse est :

$$\sinh^2(x) = \frac{1}{4}\big(e^{2x} - 2 + e^{-2x}\big).$$

En faisant la différence entre les deux, il reste 1. □

Une propriété qui est par contre très différente entre la trigonométrie plane et la trigonométrie hyperbolique, c'est la périodicité. Les fonctions usuelles cos et sin sont périodiques. Pas les fonctions hyperboliques.

Proposition 41.3.
La fonction $\sinh\colon \mathbb{R} \to \mathbb{R}$ *est bijective.*

Démonstration. Il faut démontrer que sinus hyperbolique est injective et surjective. Calculons d'abord les limites. Comme tu sais que $\lim_{x\to\infty} e^x = \infty$ et $\lim_{x\to-\infty} e^x = 0$, tu vois facilement que

$$\lim_{x\to-\infty} \sinh(x) = -\infty \qquad \lim_{x\to\infty} \sinh(x) = \infty. \tag{41.31}$$

Par ailleurs, la fonction sinus hyperbolique est continue et respecte donc le théorème de la valeur intermédiaire[6] 11.50. Soit $y \in \mathbb{R}$. Voyons s'il existe un $x \in \mathbb{R}$ tel que $\sinh(x) = y$. Les deux limites indiquent qu'il existe $x_1 \in \mathbb{R}$ tel que $\sinh(x_1) < y$ et $x_2 \in \mathbb{R}$ tel que $\sinh(x_2) > y$. Le théorème de la valeur intermédiaire conclu qu'il existe un x entre x_1 et x_2 tel que $\sinh(x) = y$. Cela prouve la surjectivité.

Pour l'injectivité, on va utiliser le théorème de Rolle et une petite preuve par l'absurde. Supposons que $\sinh(x_1) = \sinh(x_2)$ avec $x_1 \neq x_2$. Dans ce cas, le théorème de Rolle nous dit qu'il existe un x entre x_1 et x_2 tel que $\sinh'(x) = 0$. La dérivée de sinus hyperbolique étant cosinus hyperbolique, il faut se demander il existe un x tel que $\cosh(x) = 0$. Étant donné que $e^x > 0$ pour tout x, en fait le cosinus hyperbolique ne s'annule jamais. □

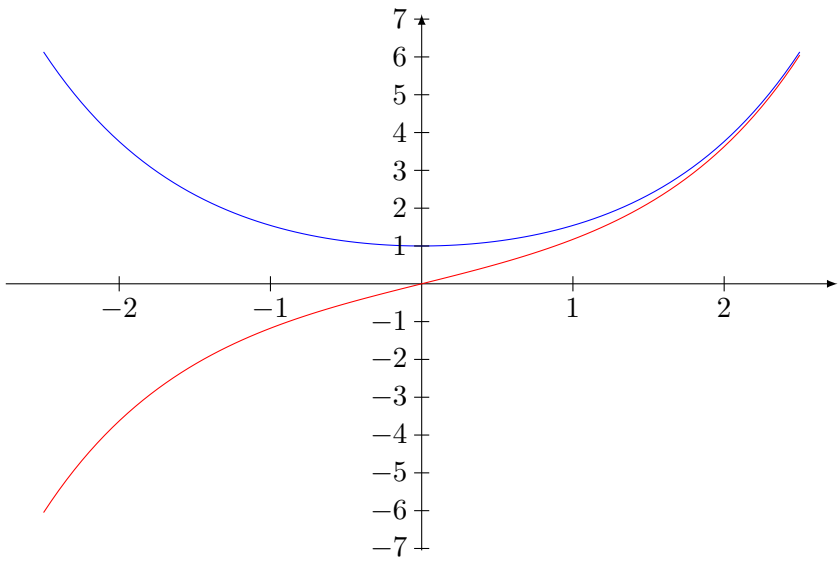


FIGURE 41.1 – En rouge, la fonction $x \mapsto \sinh(x)$ et en bleu, la fonction $x \mapsto \cosh(x)$.

6. Je t'avais dit que tout tons cours de math allait y passer hein.

Un très bon exercice serait de faire un étude complète des fonctions cosinus et sinus hyperbolique. Leur graphes sont donnés à la figure 41.1

Un corolaire de la surjectivité de sinh sur $\mathbb{R}$ est que si je prends n'importe quel deux nombres dont la différence des carrés vaut 1, alors ces carrés sont représentables avec des fonctions hyperboliques :

$$\forall x, y \in \mathbb{R} \text{ tels que } x^2 - y^2 = 1, \exists \xi \in \mathbb{R} \text{ tel que } x^2 = \cosh(\xi) \text{ et } y^2 = \sinh(\xi).$$

La **tangente hyperbolique** est définie par

$$\tanh(x) = \frac{\sinh(x)}{\cosh(x)}. \tag{41.32}$$

Un bon exercice est de prouver les deux relations suivantes :

$$\sinh(x) = \frac{\tanh(x)}{\sqrt{1 - \tanh^2(x)}} \qquad \cosh(x) = \frac{1}{\sqrt{1 - \tanh^2(x)}}. \tag{41.33}$$

41.4.8.2 Les transformations de Lorentz (bis)

Nous avons prouvé qu'en relativité, l'intervalle est un invariant. Pour cela, nous avons utilisé les transformations de Lorentz démontrées à partir de l'hypothèse d'invariance de la vitesse de la lumière. Eh bien, oublions un instant que la vitesse de la lumière soit invariante, et posons à la place comme hypothèse que l'intervalle soit invariant. C'est-à-dire que si Bob mesure un événement aux coordonnées (t, x) et Alice en (t', x'), alors $c^2t^2 - x^2 = c^2(t')^2 - (x')^2$.

Théorème 41.4.
Les transformations de Lorentz sont les seules qui laissent l'intervalle invariant.

Démonstration. Toute la partie comme quoi les transformations doivent êtres linéaires reste parce que cette partie ne demandait pas l'invariance de la vitesse de la lumière.

Nous cherchons donc les transformations entre Alice et Bob sous la forme

$$\begin{aligned} t' &= \alpha t + \beta x \\ x' &= \gamma t + \delta x \end{aligned}$$

telles que $c^2(t')^2 - (x')^2 = c^2t^2 - x^2$. Lorsque Alice passe devant Bob, ils déclenchent tout deux leurs chronomètre et leurs axes. C'est-à-dire que si à ce moment un événement se trouve à droite pour Alice, il est aussi à droite pour Bob. On doit donc avoir, quand $t = t' = 0$, que $x > 0$ implique $x' > 0$. Cela donne la contrainte que $\delta > 0$. D'autre part, comme leurs chronomètres vont dans le même sens (ils choisissent tout les deux de *compter* le temps et non *décompter*), on a $\alpha > 0$.

En développant l'expression de $(s')^2$ en termes de t et x, on trouve la condition d'invariance de l'intervalle sous la forme :

$$c^2(\alpha^2t^2 + 2\alpha\beta tx + \beta^2x^2) - (\gamma^2t^2 + 2\gamma\delta tx + \delta^2x^2) = c^2t^2 - x^2, \tag{41.34}$$

qui doit être valable pour tout t et pour tout x. En $t = 0$ on trouve la condition

$$\delta^2 - c^2\beta^2 = 1. \tag{41.35}$$

Cela implique qu'il existe un $\xi \in \mathbb{R}$ tel que $\delta^2 = \cosh^2(\xi)$ et $c^2\beta^2 = \sinh(\xi)$. La première équation donne $\delta = \cosh(\xi)$ (il faut rejeter $\delta = -\cosh(\xi)$ parce qu'on a demandé que $\delta > 0$). Pour la seconde, on trouve $c\beta = \sinh(\xi)$ où l'on peut oublier la possibilité $c\beta = -\sinh(\xi)$ parce que cela revient juste à renommer $\xi \to -\xi$ (la fonction sinus hyperbolique est impaire). Bref, il existe un ξ tel que

$$\begin{aligned} \delta &= \cosh(\xi) \\ \beta &= \frac{\sinh(\xi)}{c} \end{aligned} \tag{41.36}$$

En mettant maintenant $x = 0$ dans la condition (41.34), on trouve la condition

$$\alpha^2 - \frac{\gamma^2}{c^2} = 1.$$

Pour les mêmes raisons qu'avant, il existe un $\eta \in \mathbb{R}$ tel que

$$\begin{aligned} \alpha &= \cosh(\eta) \\ \gamma &= c\sinh(\eta). \end{aligned} \tag{41.37}$$

Rien qu'en regardant deux cas très particuliers, on a déjà bien avancé, non ? Remettons maintenant les valeurs (41.36) et (41.37) dans la condition (41.34). En utilisant l'identité $\cosh^2(x) - \sinh^2(x) = 1$, et en séparant les termes en t^2, x^2 et tx pour satisfaire la condition, il faut

$$\cosh(\eta)\sinh(\xi) = \sinh(\eta)\cosh(\xi) \tag{41.38}$$

parce que les termes en t^2 et x^2 donnent exactement $c^2t^2 - x^2$ et qu'il faut que le terme en tx s'annule. Mettons la condition (41.38) au carré, et substituons $\cosh^2(\eta) = 1 + \sinh^2(\eta)$ et $\cosh^2(\xi) = 1 + \sinh^2(\xi)$, il reste

$$\sinh^2\xi = \sinh^2\eta,$$

ce qui signifie $\sinh\xi = \pm\sinh\eta$, ou encore $\xi = \pm\eta$. On voit que $\xi = -\eta$ ne fonctionne pas dans (41.38), donc on reste avec $\xi = \eta$ et les transformations prennent la forme

$$\begin{aligned} t' &= \cosh(\xi)t + \frac{\sinh(\xi)}{c}x \\ x' &= c\sinh(\xi)t + \cosh(\xi)x. \end{aligned} \tag{41.39}$$

Ce que nous avons prouvé, c'est qu'il existe un $\xi \in \mathbb{R}$ tel que les transformations entre Alice et Bob aient cette forme. Il faut trouver ce que vaut ξ en fonction de la vitesse v à laquelle Alice court.

Pour ce faire, étudions le mouvement d'Alice. Bob la voit aux coordonnées (t, vt), ce qui correspond à

$$x' = c\sinh(\xi)t + \cosh(\xi)vt$$

pour Alice. Mais ces coordonnées sont celles de Alice elle-même, donc $x' = 0$, ce qui donne[7] $vt = -c\sinh(\xi)t/\cosh(\xi)$, ou encore

$$\tanh(\xi) = -\frac{v}{c} \tag{41.40}$$

En utilisant les relations (41.33), on trouve

$$\cosh(\xi) = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \qquad \sinh(\xi) = \frac{-v/c}{\sqrt{1 - \frac{v^2}{c^2}}}. \tag{41.41}$$

En remettant ces valeurs dans les transformations (41.39), on trouve

$$t' = \frac{t - \frac{v}{c^2}x}{\gamma(v)} \tag{41.42}$$

$$x' = \frac{x - vt}{\gamma(v)}, \tag{41.43}$$

exactement les transformations de Lorentz !

□

Ce résultat est important pour une raison assez simple : maintenant, la théorie de la relativité est indépendante de toute considérations sur la lumière. En effet, ce que nous venons de prouver, c'est que s'il existe une vitesse c telle que $c^2t^2 - x^2 = c^2(t')^2 - (x')^2$, alors (t, x) et (t', x') sont liés par les transformations de Lorentz.

7. Conditions d'existence : $\cosh(\xi) \neq 0$; heureusement, nous avons vu que le cosinus hyperbolique ne s'annule jamais.

41.4.9 Vitesse limite

Affin de nous passer de l'hypothèse d'invariance de la vitesse de la lumière, nous avons prouvé que l'hypothèse d'invariance de l'intervalle était suffisante. Mais il faut avouer que cette hypothèse n'est pas très intuitive. Nous allons montrer maintenant que l'existence d'une vitesse limite est une troisième hypothèse qui peut être utilisée comme alternative aux deux premières.

41.5 Applications

Une première application sympa est le logiciel[8] *lightspeed*. Si tu es sous Ubuntu-Linux, installe juste le paquet nommé lightspeed, et régales-toi ! Tu verras c'est marrant. Si tu utilise des fenêtres, je laisse faire l'adage « Windows c'est facile ».

41.5.1 Le GPS

Pour qu'un système GPS puisse te localiser, en gros, il t'envoie un signal, tu lui réponds et il mesure le temps qu'il a fallu à la lumière pour faire l'aller-retour. Déjà, tu remarques que cela n'est possible que grâce au fait que la vitesse de la lumière soit finie. Sinon, le GPS ne fonctionnerait pas. Mais il y a mieux.

Comme pour te localiser il faut plusieurs satellites en plus de ton appareil, il faut que les horloges internes de tout ce petit monde soient bien synchronisées, sinon pour mesurer des intervalles de temps et calculer des distances, c'est mal parti. Eh mais un satellite, ça bouge assez vite (surtout que les mesures doivent être très précises), et en plus ça ne fait même pas un MRU, vu que ça tourne en rond. Comme tu vois tout le travail qu'il a fallu faire pour trouver les transformations de Lorentz d'un MRU, tu t'imagines le travail pour un mouvement circulaire ! Eh bien ce travail a été fait, et le résultat est que si on en tient pas compte, les contractions temporelles liées à la relativité sont suffisamment grandes pour complètement dérégler le GPS.

41.5.2 Les ondes électromagnétiques

Tu te souviens qu'au début du chapitre, nous avons dit que le problème qui a amené la relativité était la propagation du champ électrique. Maintenant que nous avons déjà vu une partie des conséquences du problème, il est temps de se rendre compte que les champs électriques et magnétiques sont les objets les plus soumis aux bizarreries relativistes du monde : elles se propagent à la vitesse de la lumière. Regarde un coup autour de toi ; tout ce qui est champ électromagnétique a besoin de la relativité pour être bien compris : GSM, lumière, four à micro-onde, radio, wifi, fibre optique, ...

Si un jour un ingénieur te dit qu'il n'y a pas besoin de connaitre la relativité pour inventer la radio (c'est vrai : la radio a été inventée avant la relativité), ni pour construire une fibre optique, dis lui en pensant à moi qu'il utilise tout le temps les équations de Maxwell[9], et que ces équations sont relativistes.

Bref, soit convaincu que tu vis dans un monde relativiste et que les transformations de Lorentz te suivent à chacun de tes pas.

41.6 Mécanique relativiste

Cela est bien beau, mais la dilatation du temps, et les contractions de longueurs doivent bien avoir des répercussions sur la cinématique et la dynamique des objets. Est-ce que le théorème de l'énergie cinétique est encore valable ? est-ce que les lois de Newton tiennent encore la route ?

8. jeu de mot sur « application » ! ah ah !
9. C'est sous ce nom là qu'on nomme l'ensemble des équations de l'électromagnétisme comme la loi de l'induction.

41.6.1 Des problèmes, toujours des problèmes

Attardons-nous un peu pour faire quelques commentaires sur cette citation du chevalier pégase dans les chevaliers du zodiaque :

> Ses coups vont à la vitesse de la lumière et pourtant je les vois distinctement arriver.

Est-ce possible ? Nous avons vu qu'il y avait des dénominateurs qui s'annulent quand des objets se déplacent plus vite que la lumière ; or pour voir venir un rayon de lumière qui vient vers soi, il faudrait que le rayon émette de la lumière devant elle. Ça semble un peu mal parti pour respecter les lois de la relativité, non ?

Cela pose en tout cas une question qu'il faudra résoudre. On *entend* venir une ambulance parce qu'elle émet du son qui avance plus vite qu'elle. Pas de problèmes avec ça. Mais quid de la *voir* venir ?

On peut voir venir un tram parce qu'il émet de la lumière ; cette lumière allant plus vite que le tram, elle arrive à nos yeux avant le tram lui-même. Cela est très bien. Mettons que le tram avance à 50 km/h ; pour le conducteur, la lumière de son phare avant avance devant lui à la vitesse c. Par conséquent pour un observateur au sol, cette même lumière devrait avancer à la vitesse $c + 50$. Encore une fois, on a un problème d'invariance de la vitesse de la lumière ; mais comme c'est de la lumière, on est habitué à ce que des trucs bizarres arrivent. On ne sera pas étonné que $c + 50 = c$ d'une manière ou d'une autre[10]. Pire. Si un vaisseau spatial avance à la vitesse 200000 km/s et qu'il envoie en reconnaissance un vaisseau devant lui à la vitesse de 150000 km/s, le vaisseau de reconnaissance ira à la vitesse 150000 km/s par rapport au vaisseau principal. Et par rapport au sol, il ira à la vitesse $150000 + 200000 = 350000$ km/s, ce qui est impossible. Il faudra trouver quelque chose pour que ça se passe bien.

Un autre problème maintenant.

Prenons une masse m que l'on soumet à une force constante F. Par la loi de Newton, $a = F/m$ est constante et la vitesse après un temps t vaut $v = Ft/m$. Pas de bol, ça devient plus grand que la vitesse de la lumière à partir du temps $t = cm/F$. Ça est un problème hein ? Il faut trouver un truc pour qu'avec une force constante, l'accélération diminue.

41.6.2 Loi d'addition des vitesses

Si Bob observe un objet se déplacer à la vitesse V, alors Alice devrait l'observer bouger à la vitesse $V - v$. Tout comme si une vache voit passer un train à 90 km/h, alors le vélo qui avance à 25 km/h le voit passer à 65 km/h.

Maintenant, tu es habitué à ce que rien ne se passe comme d'habitude, donc tu te doutes bien qu'en réalité la bonne formule ne va pas être $V - v$.

Bob observe l'objet aux coordonnées (t, Vt), ce qui fait pour Alice :

$$\left(\frac{t - \frac{v}{c^2}Vt}{\gamma(v)}, \frac{Vt - vt}{\gamma(v)}\right).$$

En divisant le x' d'Alice par le t' d'Alice, on trouve la vitesse mesurée par Alice :

$$V' = \frac{(V - v)t}{\gamma(v)} \frac{\gamma(v)}{t\left(1 - \frac{vV}{c^2}\right)} = \frac{V - v}{1 - \frac{vV}{c^2}}.$$

La loi de transformation des vitesses relativiste est donc

$$V' = \frac{V - v}{1 - \frac{vV}{c^2}}. \tag{41.44}$$

Qu'en est-il de notre $c + 50 = c$? Disons que Bob lance un bisou à Alice pendant qu'elle arrive vers lui. Le bisou arrive à la vitesse de la lumière (càd $V = c$) tandis que Alice s'approche de Bob

10. et je ne te cache pas que c'est ce qui va arriver.

à la vitesse 50 m/s (càd $v = 50$). Donc la vitesse à laquelle Alice devrait voir arriver le bisou est bien $c + 50$. En utilisant la formule d'addition relativiste des vitesses (41.44), nous trouvons

$$V' = \frac{c - 50}{1 - \frac{50c}{c^2}} = \frac{c - 50}{1 - \frac{50}{c}} = \frac{c(c - 50)}{c - 50} = c.$$

Donc effectivement en relativité quand on additionne des vitesses il faut penser à la règle du « $c + 50 = c$ ».

41.6.3 L'action d'une force

L'équation fondamentale de la mécanique classique est

$$F = ma.$$

Or tu n'es pas sans savoir que l'accélération est la dérivée seconde de la position par rapport au temps. Nous noterions donc $F = mx''(t)$. Le problème est évidemment que si F est constante, on trouve $v = Ft/m$ qui dépasse toujours la vitesse c quand t est assez grand. Il faudra donc modifier la loi $F = ma$. Pour cela, posons-nous des questions sur la dérivée $x'(t)$. On dérive par rapport au temps ; oui mais nous avons vu que le temps n'est pas le même pour tout le monde. Introduisons donc la notation

$$v = \frac{dx}{dt} \tag{41.45}$$

qui ne signifie rien d'autre que nous dérivons x par rapport à t et non par rapport au temps t' de quelqu'un d'autre. Dans le cadre de la relativité, ce que signifie l'équation (41.45) est que v est la dérivée de x par rapport à t. Dans le cas où x et t sont les coordonnées de la position d'Alice mesurées par Bob, cela signifie qu'on dérive la position *mesurée par Bob* par rapport au temps *mesuré par Bob*.

Ce que dit la relativité est que cette quantité v ne peut pas varier proportionnellement à la force sous peine de dépasser la vitesse de la lumière. La subtilité est de modifier la loi de Newton en disant que la quantité qui varie sous l'action d'une force n'est plus $dx/dt = v$, mais

$$\frac{dx}{dt'} = \frac{v}{\sqrt{1 - \frac{v^2}{c^2}}},$$

c'est-à-dire la dérivée de la position *mesurée par Bob* par rapport au temps *mesuré par Alice* ! La loi de Newton $v = Ft/m$ devient donc

$$\frac{v}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{Ft}{m}. \tag{41.46}$$

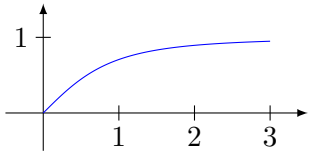
Est-ce que cela résous le problème ? Pour le savoir, regardons la vitesse acquise par le mobile de masse m soumit à la force F pendant un temps t Il faut résoudre l'équation (41.46) par rapport à v et voir si cela reste bien toujours inférieur à c. On commence par mettre la racine à droite et à élever toute l'équation au carré :

$$\begin{aligned} v^2 &= \frac{F^2t^2}{m^2}\left(1 - \frac{v^2}{c^2}\right) \\ v^2\left(1 + \frac{F^2t^2}{c^2m^2}\right) &= \frac{F^2t^2}{m^2} \\ v &= \frac{\sqrt{F^2t^2/m^2}}{\sqrt{1 + \frac{F^2t^2}{c^2m^2}}}, \end{aligned}$$

et donc finalement

$$v(t) = \frac{Ft}{m\sqrt{1 + \frac{F^2t^2}{c^2m^2}}}. \tag{41.47}$$

Tu dois remarquer que si F et t ne sont pas trop grands, l'expression F^2t^2/c^2m^2 est minuscule parce que c est énorme. Si on fait l'approximation $F^2t^2/c^2m^2 = 0$ dans cette expression, on retrouve $v = Ft/m$. Cela montre qu'à moins de faire des expérience avec de très grandes forces pendant énormément de temps, on ne peut pas voir la différence entre la mécanique de Newton et la mécanique relativiste.



Sur le graphe suivant, la vitesse en fonction du temps lorsqu'une particule de masse $m = 1$ est soumise à une force constante. Pour les besoins du graphique, nous avons mit à 1 la vitesse c. Tu vois que quand la vitesse n'est pas très grande, le graphiques est presque celui d'une droite ; et à partir d'un certain moment, la courbe s'infléchit pour tendre vers 1 sans l'atteindre.

Remarque que si on maintient une accélération constante égale à celle de la gravité terrestre pendant deux heures, on arrive déjà sur la Lune, à une vitesse de 75 km/s, c'est-à-dire encore rien par rapport à la vitesse de la lumière ! Cela pour te dire que la formule (41.47) a l'air d'être très différente de la formule classique $v = Ft/m$, mais en réalité tant qu'on n'atteint pas des forces énormes, elle ressemble très fort.

Vérifions maintenant que la formule (41.47) n'est pas en contradiction avec l'impossibilité de dépasser la vitesse de la lumière. Pour cela, regardons ce qu'il se passe si on applique une force constante F sur un objet de masse m pendant un temps très long. C'est-à-dire : calculons la limite

$$\lim_{t\to\infty} v(t).$$

Tu vois tout de suite qu'on est sur un cas $\frac{\infty}{\infty}$, ce qui t'oblige à utiliser la règle de l'Hospital. On peut cependant un peu réfléchir et deviner la réponse sans passer par des math trop compliquées.

En effet, quand t est vraiment énorme, l'expression $\frac{F^2t^2}{m^2c^2}$ devient très grande, et le 1 qui se trouve à côté ne vaut plus grand chose, on peut le négliger.

$$\begin{aligned}\lim_{t\to\infty} \frac{Ft}{m\sqrt{1+\frac{F^2t^2}{c^2m^2}}} &= \lim_{t\to\infty} \frac{Ft}{m\sqrt{\frac{F^2t^2}{m^2c^2}}} \\ &= \lim_{v\to\infty} \frac{Ft}{m\frac{Ft}{cm}} \\ &= c.\end{aligned} \tag{41.48}$$

Tout est bien : on arrive au maximum à la vitesse de la lumière, mais il faut un temps infini pour y parvenir. Conclusion : il n'est pas possible d'accélérer un objet jusqu'à atteindre la vitesse de la lumière.

41.6.4 Équivalence entre la masse et l'énergie

Le moment est venu de montrer ce que signifie la fameuse formule $E = mc^2$.

41.7 Principe de correspondance

Nous ne sommes pas parvenu à démontrer la formule (41.46) de la mécanique relativiste qui montre comme un objet accélère sous l'effet d'une force constante. Nous avons juste montré qu'il fallait modifier la loi $v = Ft/m$ et nous avons prit la première modification qui nous soit tombée sous la main, à savoir qu'il faut dériver la position par rapport au temps de l'objet qu'on observe plutôt que par rapport au temps de l'observateur.

En fait, il est possible de prouver rigoureusement [11] la formule

$$\frac{Ft}{m} = \frac{\alpha v}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Mais il n'y a pas moyen de trouver la valeur de la constante α. Tout ce qu'il y a moyen de trouver avec l'hypothèse de l'invariance de la vitesse de la lumière est l'existence d'une constante telle que cette formule soit vraie.

Affin de fixer la constante α, il faut faire intervenir un principe physique supplémentaire, le **principe de correspondance**

Loi numéro 4.
Lorsque la vitesse d'une particule est faible, les équations doivent être en première approximation les mêmes que celles de la mécanique classique.

Que signifie *en première approximation* ? Tu sais qu'une fonction $x \mapsto f(x)$ peut être approximée (pour des petits x) par la formule

$$f(x) \simeq f(0) + xf'(0).$$

Nous voudrions donc que Ft/m soit en première approximation égal à v. Nous devons étudier la fonction

$$f(v) = \frac{\alpha v}{\sqrt{1 - \frac{v^2}{c^2}}}.$$

Voir ce que vaut cette fonction en première approximation lorsque v est petit est un exercice de dérivation. En utilisant la règle de dérivation des fractions, on trouve que

$$f'(v) = \frac{\alpha}{\sqrt{1 - \frac{v^2}{c^2}}} + \frac{\alpha v^2}{c^2 \left(1 - \frac{v^2}{c^2}\right)^{3/2}},$$

et donc que $f'(0) = \alpha$. Bien entendu, $f(0) = 0$. En première approximation, nous trouvons donc

$$f(v) \simeq \alpha v \tag{41.49}$$

qui doit être égal à la quantité non relativiste v. Nous en déduisons qu'il faut fixer $\alpha = 1$, et on tombe sur la formule relativiste proposée plus haut

$$\frac{Ft}{m} = \frac{v}{\sqrt{1 - \frac{v^2}{c^2}}}$$

L'utilisation cruciale du principe de correspondance a une répercussion énorme sur notre vision de la physique. En effet, la relativité d'Einstein ne parvient pas à *remplacer* la mécanique de Newton. On a besoin d'invoquer la mécanique de Newton pour fixer la théorie. On peut écrire l'axiome suivant :

$$\lim_{v \to 0} \text{Einstein} = \text{Newton}. \tag{41.50}$$

Cela n'est pas une propriété de la théorie d'Einstein, mais un de ses axiomes !

La relativité ne fait donc pas table rase des principes physiques de la mécanique newtonienne : elle les complète et les contient.

11. Mais il n'existe pas de démonstrations simples à ma connaissance.

Chapitre 42

Exemples avec Sage

Ce chapitre est un foure-tout de choses que l'on peut faire avec Sage.

42.1 Graphiques

Pour afficher le graphe d'une fonction, vous pouvez faire

```
+----------------------------------------------------------------------+
| SageMath version 8.1, Release Date: 2017-12-07                       |
| Type "notebook()" for the browser-based notebook interface.          |
| Type "help()" for help.                                              |
+----------------------------------------------------------------------+
sage: plot(cos(x),0,5)
Launched png viewer for Graphics object consisting of 1 graphics primitive
sage: f(x)=sin(x)
sage: f.plot(-pi,pi)
Launched png viewer for Graphics object consisting of 1 graphics primitive
```

Un programme externe se lance automatiquement pour afficher le graphique que vous avez demandé.

Il se peut qu'aucun programme ne se lance et vous ayez, au lieu de `Launched png viewer for Graphics object ...` uniquement `Created graphics object ...`. Disons pour faire court que Sage a produit un `png` et qu'il ne sait pas quel programme externe utiliser pour l'afficher.

La solution est à l'adresse http://doc.sagemath.org/html/en/reference/misc/sage/misc/viewer.html

42.1.1 Autres

Dans le but d'automatiser certaines tâches, j'ai écrit ce module, nomé `outilsINGE.sage`, dans le cadre d'un cours de première année donné à des ingénieurs. Certaines des fonctions définies ici sont utilisée dans les exemples qui suivent.

```
# -*- coding: utf8 -*-
from sage.all import *

"""
This module provides _pragmatic_ tools for solving exercise for
a first year in general mathematics.
"""

# TODO : trouver une bonne traduction pour "point de selle."

```

```
def automatedVar(symbol,n):
  """ If symbol = "x" and n=4, return the string 'x1,x2,x3,x4'  ↩
      """
  s = ",".join([ symbol+str(i) for i in range(1,n+1)])
  return s

class SolveLinearSystem(object):
  """
  Solve Ax=v and print it in a nice way

  Example :

  A=matrix([ [1,-2,3,-2,0],[3,-7,-2,4,0],[4,3,5,2,0] ])
  v=vector((0,0,0,0,0))
  print SolveLinearSystem(A,v)
  """
  def __init__(self,A,v):
    self.matrix = A
    self.vector = v
    self.nvars = A.ncols()
    s = automatedVar("x",self.nvars)
    self.xx=var(s)
  def equations(self):
    """Return the equations corresponding to the
                self.matrix and self.vector as a list of equations↩
                   """
    X=matrix( [self.xx[i] for i in range(0,self.nvars) ]).↩
       transpose()
    eqs=[]
    for i in range(0,self.matrix.nrows()):
      exp = (self.matrix*X)[i][0]==self.vector[i]
      eqs.append(exp)
    return eqs
  def solutions(self):
    return solve(self.equations(),self.xx)
  def latex(self):
    """"Return the LaTeX's code of the system."""
    a=[]
    a.append(r"""

        \item
          $
          \left\{
          \begin{array}{ll}
    """)
    for eq in self.equations():
      a.append("          "+str(eq).replace("x","x_").replace("*"↩
         ,"").replace("==","=")+"\\\\ \n")
    a.append(r"""
          \end{array}
          \right.
          $
          """)
```

```
    return "".join(a)
  def __str__(self):
    a = []
    a.append("The given matrix corresponds to the system")
    for eq in self.equations():
      a.append(str(eq))
    a.append("And the solutions are")
    a.append(str(self.solutions()))
    return "\n".join(a)

def QuadraticMap(A,v):
  """
  Return the result of the quadratic form associated
        with A applied on the vector v, that is the number
  A_ij v^iv^j
  using the summation convention.
  """
  n = A.nrows()
  if not A.is_symmetric():
    print "Warning : Given matrix is not symmetric"
  if not A.is_square():
    raise TypeError,"Error : The matrix A is not square"
  if not v.degree()==n :
    raise TypeError,"The size do not agree"
  return sum([ A[i,j]*v[i]*v[j] for i in range(n) for j in range(n)  ]).simplify_full()

class SymmetricMatrix(object):
  """
  Provide informations about the matrix A assuming it is symmetric.
  """
  def __init__(self,A):
    if not A.is_square():
      print "Error : A symmetric matrix must be square"
      raise TypeError
    self.matrix = A
    self.degree = A.nrows()
    self.matrix.set_immutable()
  def primary_principal_submatrix(self,n):
    """
    Return the primary principal submatrix of order n, that is the matrix obtained
                  by removing the n last lines and columns from self.
    """
    taille=self.degree-n
    v=[]
    for i in range(0,taille):
      v.append(self.matrix[i][0:taille])
    return matrix(v)
  def principal_minors(self):
    """
```

```
    Return the list of principal minors. The principal minor of order k is
                the determinant of the primary principal matrix of order k.
    """
    a=[]
    for i in range(self.degree):
      a.append(self.primary_principal_submatrix(i).determinant())
    return a
  def genre_list(self):
    """
    Return the genius of the matrix as a list of booleans in the order
    positive defined, negative defined;
                semidefinite positive, semidefinite negative, indefinite.

    """
    defpos = True
    defneg = True
    semidefpos = True
    semidefneg = True
    indefinie=True
    mineurs = self.principal_minors()
    for i in range(len(mineurs)):
      m = mineurs[i]
      if m == 0:
        defneg=False
        defpos=False
      if m < 0:
        defpos=False
        semidefpos=False
        if i%2==0:
          defneg=False
      if m > 0:
        semidefneg=False
        if i%2==1:
          defneg=False
    if 0 not in mineurs:
      semidefneg=False
      semidefpos=False
    if (defpos==True)or(defneg==True)or(semidefpos==True)or(semidefneg==True):indefinie=False
    return [defpos,defneg,semidefpos,semidefneg,indefinie]
  def __str__(self):
    return str(self.matrix)

class QuadraticForm(SymmetricMatrix):
  """
  From a symmetric matrix A, provide informations concerning the associated quadratic form.
  """
  def __init__(self,A):
```

```
    SymmetricMatrix.__init__(self,A)
    if not A.is_symmetric():
      print "Warning : matrix is not symmetric"
  def evaluate(self,v):
    """
    Return the value of the quadratic form on the vector v.
    """
    return QuadraticMap(self.matrix,v)
  def diagonalizing_martrix(self):
    """
    Return the matrix B such that B^tAB is diagonal.
    """
    # The transposition is because, in the matrix B, the ↩
       eigenvectors have
    # to be read as column while Sage's matrix constructor takes ↩
       rows.
    return matrix(self.orthonormal_basis()).transpose()
  def new_variables(self):
    """
    Give the change of variables needed to put the quadratic form ↩
       under its normal form
    X=BY
    where X are the "old" variables
    """
    variables = var(automatedVar("y",self.degree))
    Y = vector(variables)
    return self.diagonalizing_martrix()*Y
  def eigenmatrix_left(self):
    return self.matrix.eigenmatrix_left()
  def eigenvectors(self):
    """
    Return a list of eigenvectors of the matrix.

    As the matrix is symmetric, that list has to be a basis.
    """
    D,P = self.eigenmatrix_left()
    return [P[i] for i in range(P.nrows())]
  def eigenvalues(self):
    """
    Return a list of eigenvalues of the matrix in the same order ↩
       as the list of eigenvectors given in
      self.eigenvectors()
    """
    D,P = self.eigenmatrix_left()
    return [ D[i,i] for i in range(D.nrows()) ]
  def orthonormal_basis(self):
    """
    Return a basis of eigenvectors normalised to 1 as a list.
    """
    M,mu = matrix(self.eigenvectors()).gram_schmidt()
    return [ v/v.norm() for v in M ]
  def verification(self):
    """
```

```
    return the value of the quadratic form on the vector new_variables()
    """
    return self.evaluate(self.new_variables())

  def __str__(self):
    a = []
    a.append("Hi guy; I'm the quadratic form associated with the matix")
    a.append(str(self.matrix))
    a.append("My eigenvalues and eigenvectors are : ")
    veps = self.eigenvectors()
    vaps = self.eigenvalues()
    for i in range(len(veps)):
      a.append("%s -> %s"%(str(vaps[i]),str(veps[i])))
    a.append("I've the following orthonormal basis of eigenvectors:")
    for v in self.orthonormal_basis():
      a.append(str(v))
    a.append("A matrix B such that B^tAB is diagonal is ")
    a.append(str(self.diagonalizing_martrix()))
    a.append("I'm quite pretty in the following variables ...")
    for i in range(self.degree):
      a.append("x%s = %s"%(str(i+1),str(self.new_variables()[i]))
      )
    a.append("Look at me when I wear my cool variables")
    a.append(str(self.verification()))
    return "\n".join(a)

class Extrema(object):
  """
  From a function f, provides the informations for the study of the extrema :
  partial derivative
  critical points
  Hessian matrix at the critical points
  Genius of the Hessian and conclusion as local min/max

  Dear student : remember that this class does not furnish any informations
      concerning *global* extrema. The latter have to be found
    among the critical points OR on the border of the domain.
  """
  def __init__(self,f):
    var('x,y')
    self.fun = f
    self.gx=self.fun.diff(x).full_simplify()
    self.gy=self.fun.diff(y).full_simplify()
    self.gxx=self.gx.diff(x).simplify_full()
    self.gxy=self.gx.diff(y).full_simplify()
    self.gyy=self.gy.diff(y).full_simplify()
    self.cp = solve( [self.gx(x,y)==0,self.gy(x,y)==0],[x,y] )
  def critical_points(self):
```

```
    """
    Return the critical points as a list of tuples (x,y)
    """
    a = []
    for pt in self.cp :
      try :
        px = SR(pt[0].rhs())
        py = SR(pt[1].rhs())
        a.append((px,py))
      except TypeError :
        a.append("  I'm not able to solve these equations.")
    return a
  def hessienne(self,a,b):
    return matrix(SR,2,2,[self.gxx(a,b),self.gxy(a,b),self.gxy(a,↩
       b),self.gyy(a,b)])
  def __str__(self):
    a = []
    a.append ("The function :")
    a.append(str(self.fun))
    a.append ("Derivative x and y :")
    a.append(str(self.gx))
    a.append(str(self.gy))
    a.append ("Hessian matrix :")
    a.append(str(self.hessienne(x,y)))
    a.append ("Critical points :")
    for pt in self.critical_points() :
      a.append(str(pt))
    for pt in self.critical_points():
      try :
        px = pt[0]
        py = pt[1]
        a.append("At (%s,%s), the Hessian is"%(str(px),str(py)))
        try :
          Hess = SymmetricMatrix(self.hessienne(px,py))
          for l in Hess.matrix:
            a.append("  "+str(l))
          a.append("  Primary principal minors are %s"%str(Hess.↩
             principal_minors()))
          l = Hess.genre_list()
          if l[0]==True:
            a.append("  Hessian positive defined")
            a.append("  local minimum")
          if l[1]==True:
            a.append("  Hessian negative defined")
            a.append("  local maximum")
          if l[2]==True:
            a.append("  Hessian positive semidéfinite")
            a.append("  I don't conclude")
          if l[3]==True:
            a.append("  Hessian negative semidefinite")
            a.append("  I don't conclude")
          if l[4]==True:
            a.append("  Undefinite Hessian")
```

```
            a.append("  «selle» point")
        except RuntimeError,data :
          a.append("  "+str(data))
      except TypeError :
        a.append("  I'm not able to solve these equations.")
    return "\n".join(a)
```

tex/sage/outilsINGE.sage

Exemple 42.1.
Calculer la limite

$$\lim_{x\to\infty} \frac{\sin(x)\cos(x)}{x} \tag{42.1}$$

```
var('x')
f(x)=sin(x)*cos(x)/x
limit(f(x),x=oo)
```

La première ligne déclare que la lettre `x` désignera une variable. Pour la troisième ligne, notez que l'infini est écrit par deux petits « o ». △

Exemple 42.2.
Quelques limites et graphes avec Sage.

(1) $\lim_{x\to 0} \frac{\sin(\alpha x)}{\sin(\beta x)}$.
Pour effectuer cet exercice avec Sage, il faut taper les lignes suivantes :

```
sage: var('x,a,b')
(x, a, b)
sage: f(x)=sin(a*x)/sin(b*x)
sage: limit( f(x),x=0  )
a/b
```

Noter qu'il faut déclarer les variables `x`, `a` et `b`.

(2) $\lim_{x\to\pm\infty} \frac{\sqrt{x^2+1}-x}{x-2}$

```
sage: f(x)=(sqrt(x**2+1))/(x-2)
sage: limit(f(x),x=oo)
1
sage: limit(f(x),x=-oo)
-1
```

Noter la commande pour la racine carré : `sqrt`. Étant donné que cette fonction diverge en $x = 2$, si nous voulons la tracer, il faut procéder en deux fois :

```
sage: plot(f,(-100,1.9))
Launched png viewer for Graphics object consisting of 1 graphics primitive
sage: plot(f,(2.1,100))
Launched png viewer for Graphics object consisting of 1 graphics primitive
```

La première ligne trace de -100 à 1.9 et la seconde de 2.1 à 100. Ces graphiques vous permettent déjà de voir les limites. Attention : ils ne sont pas des *preuves* ! Mais ils sont de sérieux indices qui peuvent vous inspirer dans vos calculs.

△

Exemple 42.3.
Calculer les dérivées partielles $\partial_x f$, $\partial_y f$, $\partial^2_x f$, $\partial^2_{xy} f$, $\partial^2_{yx} f$ et $\partial^2_y f$ des fonctions suivantes.

(1) $2x^3 + 3x^2y - 2y^2$

(2) $\ln(xy^2)$

(3) $\tan(x/y)$

(4) $\frac{xy^2}{x+y}$

Le script Sage suivant (`exoDV002.sage`) résout l'exercice :

```
# -*- coding: utf8 -*-

def LesCalculs(f):
  print "Pour la fonction %s"%str(f)
  print "d_x",f.diff(x).simplify_full()
  print "d_y",f.diff(y).simplify_full()
  print "d^2_x",f.diff(x).diff(x).simplify_full()
  print "d_xd_y",f.diff(x).diff(y).simplify_full()
  print "d_yd_x",f.diff(y).diff(x).simplify_full()
  print "d^2_y",f.diff(y).diff(y).simplify_full()
  print ""

def exercise_DV002():
  var('x,y')
  fa(x,y)=2*x**3+3*x**2*y-2*y**2
  fb(x,y)=ln(x*y**2)
  fc(x,y)=tan(x/y)
  fd(x,y)=x*y**2/(x+y)
  LesCalculs(fa)
  LesCalculs(fb)
  LesCalculs(fc)
  LesCalculs(fd)
```

tex/sage/exoDV002.sage

La sortie est :

```
Pour la fonction (x, y) |--> 2*x^3 + 3*x^2*y - 2*y^2
d_x (x, y) |--> 6*x^2 + 6*x*y
d_y (x, y) |--> 3*x^2 - 4*y
d^2_x (x, y) |--> 12*x + 6*y
d_xd_y (x, y) |--> 6*x
d_yd_x (x, y) |--> 6*x
d^2_y (x, y) |--> -4

Pour la fonction (x, y) |--> log(x*y^2)
d_x (x, y) |--> 1/x
d_y (x, y) |--> 2/y
d^2_x (x, y) |--> -1/x^2
d_xd_y (x, y) |--> 0
d_yd_x (x, y) |--> 0
d^2_y (x, y) |--> -2/y^2

Pour la fonction (x, y) |--> tan(x/y)
d_x (x, y) |--> 1/(y*cos(x/y)^2)
d_y (x, y) |--> -x/(y^2*cos(x/y)^2)
d^2_x (x, y) |--> 2*sin(x/y)/(y^2*cos(x/y)^3)
d_xd_y (x, y) |--> -(2*x*sin(x/y) + y*cos(x/y))/(y^3*cos(x/y)^3)
d_yd_x (x, y) |--> -(2*x*sin(x/y) + y*cos(x/y))/(y^3*cos(x/y)^3)
```

```
d^2_y (x, y) |--> 2*(x^2*sin(x/y) + x*y*cos(x/y))/(y^4*cos(x/y)^3)

Pour la fonction (x, y) |--> x*y^2/(x + y)
d_x (x, y) |--> y^3/(x^2 + 2*x*y + y^2)
d_y (x, y) |--> (2*x^2*y + x*y^2)/(x^2 + 2*x*y + y^2)
d^2_x (x, y) |--> -2*y^3/(x^3 + 3*x^2*y + 3*x*y^2 + y^3)
d_xd_y (x, y) |--> (3*x*y^2 + y^3)/(x^3 + 3*x^2*y + 3*x*y^2 + y^3)
d_yd_x (x, y) |--> (3*x*y^2 + y^3)/(x^3 + 3*x^2*y + 3*x*y^2 + y^3)
d^2_y (x, y) |--> 2*x^3/(x^3 + 3*x^2*y + 3*x*y^2 + y^3)
```

△

Exemple 42.4.
Résoudre les systèmes suivants.

(1) $\begin{cases} x_1 - 2x_2 + 3x_3 - 2x_4 = 0 \\ 3x_1 - 7x_2 - 2x_3 + 4x_4 = 0 \\ 4x_1 + 3x_2 + 5x_3 + 2x_4 = 0 \end{cases}$

(2) $\begin{cases} 2x_1 + x_2 - 2x_3 + 3x_4 = 0 \\ 3x_1 + 2x_2 - x_3 + 3x_4 = 4 \\ 3x_1 + 3x_2 + 3x_3 - 3x_4 = 9 \end{cases}$

(3) $\begin{cases} x_1 + 2x_2 - 3x_3 = 0 \\ 2x_1 + 5x_2 + 2x_3 = 0 \\ 3x_1 - x_2 - 4x_3 = 0 \end{cases}$

(4) $\begin{cases} x_1 + 2x_2 - x_3 = 0 \\ 2x_1 + 5x_2 + 2x_3 = 0 \\ x_1 + 4x_2 + 7x_3 = 0 \\ x_1 + 3x_2 + 3x_3 = 0 \end{cases}$

(5) $\begin{cases} x_1 + x_2 + x_3 + x_4 = 0 \\ x_1 + x_2 + x_3 - x_4 = 4 \\ x_1 + x_2 - x_3 + x_4 = -4 \\ x_1 - x_2 + x_3 + x_4 = 2 \end{cases}$

(6) $\begin{cases} x_1 + 3x_2 + 3x_3 = 1 \\ x_1 + 3x_2 + 4x_3 = 0 \\ x_1 + 4x_2 + 3x_3 = 3 \end{cases}$

(7) $\begin{cases} x_1 - 3x_2 + 2x_3 = -6 \\ -3x_1 + 3x_2 - x_3 = 17 \\ 2x_1 - x_2 = 3 \end{cases}$

(8) $\begin{cases} x_1 - 2x_2 + 3x_3 - 2x_4 = 0 \\ 3x_1 - 7x_2 - 2x_3 + 4x_4 = 0 \\ 4x_1 + 3x_2 + 5x_3 + 2x_4 = 0 \end{cases}$

(9) $\begin{cases} 2x_1 + x_2 - 2x_3 + 3x_4 = 0 \\ 3x_1 + 2x_2 - x_3 + 3x_4 = 4 \\ 3x_1 + 3x_2 + 3x_3 - 3x_4 = 9 \end{cases}$

(10) $\begin{cases} x_1 + 2x_2 - 3x_3 = 0 \\ 2x_1 + 5x_2 + 2x_3 = 0 \\ 3x_1 - x_2 - 4x_3 = 0 \end{cases}$

(11) $\begin{cases} x_1 + 2x_2 - x_3 = 0 \\ 2x_1 + 5x_2 + 2x_3 = 0 \\ x_1 + 4x_2 + 7x_3 = 0 \\ x_1 + 3x_2 + 3x_3 = 0 \end{cases}$

(12) $\begin{cases} x_1 + x_2 + x_3 + x_4 = 0 \\ x_1 + x_2 + x_3 - x_4 = 4 \\ x_1 + x_2 - x_3 + x_4 = -4 \\ x_1 - x_2 + x_3 + x_4 = 2 \end{cases}$

(13) $\begin{cases} x_1 + 3x_2 + 3x_3 = 1 \\ x_1 + 3x_2 + 4x_3 = 0 \\ x_1 + 4x_2 + 3x_3 = 3 \end{cases}$

(14) $\begin{cases} x_1 - 3x_2 + 2x_3 = -6 \\ -3x_1 + 3x_2 - x_3 = 17 \\ 2x_1 - x_2 = 3 \end{cases}$

Nous résolvons les systèmes en utilisant Sage avec le script suivant.

```
# -*- coding: utf8 -*-
"""
Ce script Sage résout un certain nombre
de systèmes d'équations linéaires du cours INGE1121
"""

import outilsINGE

def exercise_1_1_bcdefhi():
  # Exercice 1.1.b (INGE1121)
  A=matrix([ [1,-2,3,-2,0],[3,-7,-2,4,0],[4,3,5,2,0] ])
```

```
v=vector((0,0,0,0,0))
print outilsINGE.SolveLinearSystem(A,v)
# Exercice 1.1.c (INGE1121)
A=matrix([ [2,1,-2,3],[3,2,-1,3],[3,3,3,-3] ])
v=vector((0,4,9))
print outilsINGE.SolveLinearSystem(A,v)
# Exercice 1.1.d (INGE1121)
A=matrix([ [1,2,-3],[2,5,2],[3,-1,-4] ])
v=vector((0,0,0))
print outilsINGE.SolveLinearSystem(A,v)
# Exercice 1.1.e (INGE1121)
A=matrix([ [1,2,-1],[2,5,2],[1,4,7],[1,3,3] ])
v=vector((0,0,0,0))
print outilsINGE.SolveLinearSystem(A,v)
# Exercice 1.1.f (INGE1121)
A=matrix([ [1,1,1,1],[1,1,1,-1],[1,1,-1,1],[1,-1,1,1] ])
v=vector((0,4,-4,2))
print outilsINGE.SolveLinearSystem(A,v)
# Exercice 1.1.h (INGE1121)
A=matrix([ [1,3,3],[1,3,4],[1,4,3] ])
v=vector((1,0,3))
print outilsINGE.SolveLinearSystem(A,v)
# Exercice 1.1.i (INGE1121)
A=matrix([ [1,-3,2],[-3,3,-1],[2,-1,0] ])
v=vector((-6,17,3))
print outilsINGE.SolveLinearSystem(A,v)
```

tex/sage/exo11.sage

Le résultat est le suivant :

```
The given matrix corresponds to the system
x1 - 2*x2 + 3*x3 - 2*x4 == 0
3*x1 - 7*x2 - 2*x3 + 4*x4 == 0
4*x1 + 3*x2 + 5*x3 + 2*x4 == 0
And the solutions are
[
[x1 == -23/16*r19, x2 == -5/16*r19, x3 == 15/16*r19, x4 == r19, x5 == r18]
]
The given matrix corresponds to the system
2*x1 + x2 - 2*x3 + 3*x4 == 0
3*x1 + 2*x2 - x3 + 3*x4 == 4
3*x1 + 3*x2 + 3*x3 - 3*x4 == 9
And the solutions are
[
[x1 == 3*r20 - 7, x2 == -4*r20 + 11, x3 == r20, x4 == 1]
]
The given matrix corresponds to the system
x1 + 2*x2 - 3*x3 == 0
2*x1 + 5*x2 + 2*x3 == 0
3*x1 - x2 - 4*x3 == 0
And the solutions are
[
[x1 == 0, x2 == 0, x3 == 0]
]
```

```
The given matrix corresponds to the system
x1 + 2*x2 - x3 == 0
2*x1 + 5*x2 + 2*x3 == 0
x1 + 4*x2 + 7*x3 == 0
x1 + 3*x2 + 3*x3 == 0
And the solutions are
[
[x1 == 9*r21, x2 == -4*r21, x3 == r21]
]
The given matrix corresponds to the system
x1 + x2 + x3 + x4 == 0
x1 + x2 + x3 - x4 == 4
x1 + x2 - x3 + x4 == -4
x1 - x2 + x3 + x4 == 2
And the solutions are
[
[x1 == 1, x2 == -1, x3 == 2, x4 == -2]
]
The given matrix corresponds to the system
x1 + 3*x2 + 3*x3 == 1
x1 + 3*x2 + 4*x3 == 0
x1 + 4*x2 + 3*x3 == 3
And the solutions are
[
[x1 == -2, x2 == 2, x3 == -1]
]
The given matrix corresponds to the system
x1 - 3*x2 + 2*x3 == -6
-3*x1 + 3*x2 - x3 == 17
2*x1 - x2 == 3
And the solutions are
[
[x1 == 37, x2 == 71, x3 == 85]
]
```

△

Exemple 42.5.
Pour chacun des systèmes suivants $A \cdot X = B$,

(1) Résoudre le système par échelonnement,

(2) Calculer A^{-1},

(3) Vérifier votre réponse en calculant $A^{-1}B$. Qu'êtes-vous censé obtenir ?

Les énoncés sont

(1)

$$A = \begin{pmatrix} 2 & 1 & -2 \\ 3 & 2 & 2 \\ 5 & 4 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 10 \\ 1 \\ 4 \end{pmatrix} \tag{42.2}$$

Nous utilisons Sage pour fournir la réponse. Le code suivant résout le système et donne l'inverse de la matrice :

```
# -*- coding: utf8 -*-

```

```
import outilsINGE

def exercise_1_3():
  A=matrix([[2,1,-2],[3,2,2],[5,4,3]])
  v=vector((10,1,4))
  print outilsINGE.SolveLinearSystem(A,v)
  print "Matrice inverse :"
  print A.inverse()
```

tex/sage/exo13.sage

La sortie est ici :

```
The given matrix corresponds to the system
2*x1 + x2 - 2*x3 == 10
3*x1 + 2*x2 + 2*x3 == 1
5*x1 + 4*x2 + 3*x3 == 4
And the solutions are
[
[x1 == 1, x2 == 2, x3 == -3]
]
Matrice inverse :
[  2/7  11/7  -6/7]
[ -1/7 -16/7  10/7]
[ -2/7   3/7  -1/7]
```

△

Exemple 42.6.
Sachant que $(-1,0,1,0)$ est un vecteur propre de la matrice

$$A=\begin{pmatrix}2&1&-1&1\\1&0&1&1\\-1&1&2&1\\1&1&1&0\end{pmatrix} \tag{42.3}$$

(1) Diagonaliser A au moyen d'une matrice orthogonale
(2) Écrire la forme quadratique X^tAX sous forme d'une somme pondérée de carrés.

Calculons Av afin de savoir la valeur propre associée au vecteur donné :

$$\begin{pmatrix}2&1&-1&1\\1&0&1&1\\-1&1&2&1\\1&1&1&0\end{pmatrix}\begin{pmatrix}-1\\0\\1\\0\end{pmatrix}=\begin{pmatrix}-3\\0\\3\\0\end{pmatrix}. \tag{42.4}$$

La valeur propre est donc 3. Nous savons donc que $(\lambda-3)$ pourra être factorisé dans le polynôme caractéristique.

Pour le reste de l'exercice c'est standard et c'est résolu de la façon suivante :

```
# -*- coding: utf8 -*-

import outilsINGE

def exercise_6_5():
  A=matrix(QQ,4,4,[2,1,-1,1,1,0,1,1,-1,1,2,1,1,1,1,0])
```

```
x=outilsINGE.QuadraticForm(A)
print x
```

tex/sage/exo65.sage

qui retourne

```
Hi guy; I'm the quadratic form associated with the matix
[ 2  1 -1  1]
[ 1  0  1  1]
[-1  1  2  1]
[ 1  1  1  0]
My eigenvalues and eigenvectors are :
3 -> (1, 0, -1, 0)
3 -> (0, 1, 2, 1)
-1 -> (1, 0, 1, -2)
-1 -> (0, 1, 0, -1)
I've the following orthonormal basis of eigenvectors :
(1/2*sqrt(2), 0, -1/2*sqrt(2), 0)
(1/2, 1/2, 1/2, 1/2)
(1/6*sqrt(6), 0, 1/6*sqrt(6), -1/3*sqrt(6))
(-1/2*sqrt(1/3), 3/2*sqrt(1/3), -1/2*sqrt(1/3), -1/2*sqrt(1/3))
A matrix B such that B^tAB is diagonal is
[   1/2*sqrt(2)            1/2    1/6*sqrt(6) -1/2*sqrt(1/3)]
[             0            1/2              0  3/2*sqrt(1/3)]
[  -1/2*sqrt(2)            1/2    1/6*sqrt(6) -1/2*sqrt(1/3)]
[             0            1/2   -1/3*sqrt(6) -1/2*sqrt(1/3)]
I'm quite pretty in the following variables ...
x1 = -1/2*sqrt(1/3)*y4 + 1/2*sqrt(2)*y1 + 1/6*sqrt(6)*y3 + 1/2*y2
x2 = 3/2*sqrt(1/3)*y4 + 1/2*y2
x3 = -1/2*sqrt(1/3)*y4 - 1/2*sqrt(2)*y1 + 1/6*sqrt(6)*y3 + 1/2*y2
x4 = -1/2*sqrt(1/3)*y4 - 1/3*sqrt(6)*y3 + 1/2*y2
Look at me when I wear my cool variables
3*y1^2 + 3*y2^2 - y3^2 - y4^2
```

△

Exemple 42.7.
Rechercher les extrémums des fonctions suivantes

(1) $f(x,y) = 2 - \sqrt{x^2+y^2}$

(2) $f(x,y) = x^3 + 3xy^2 - 15x - 12y$

(3) $f(x,y) = \frac{x^3}{3} + \frac{4y^3}{3} - x^2 - 3x - 4y - 3$

Les corrigés sont créés par le script Sage `exo101.sage`

```
# -*- coding: utf8 -*-

import outilsINGE

def exercise_10_1_A():
   var('x,y')
   f(x,y)=2-sqrt(x**2+y**2)
   print outilsINGE.Extrema(f)
def exercise_10_1_B():
   var('x,y')
```

```
   f(x,y)=x**3+3*x*y**2-15*x-12*y
   print outilsINGE.Extrema(f)
def exercise_10_1_C():
   var('x,y')
   f(x,y)=x**3/3+4*y**3/3-x**2-3*x-4*y-3
   print outilsINGE.Extrema(f)
```

Des réponses :

(1)
```
The function :
(x, y) |--> -sqrt(x^2 + y^2) + 2
Derivative x and y :
(x, y) |--> -x/sqrt(x^2 + y^2)
(x, y) |--> -y/sqrt(x^2 + y^2)
Hessian matrix :
[-sqrt(x^2 + y^2)*y^2/(x^4 + 2*x^2*y^2 + y^4)                          x*y/(x^2 + y^2)^(3
[                        x*y/(x^2 + y^2)^(3/2) -sqrt(x^2 + y^2)*x^2/(x^4 + 2*x^2*y^2 + y
Critical points :
(0, 0)
At (0,0), the Hessian is
   power::eval(): division by zero
```

Ici nous voyons que Sage a du mal à calculer la matrice hessienne en $(0,0)$. En effet, nous tombons sur une division par zéro. Pour résoudre l'exercice, il faut se rendre compte que la fonction $(x,y) \mapsto \sqrt{x^2+y^2}$ est toujours positive et est nulle seulement au point $(0,0)$. Donc f est toujours plus petite ou égale à deux tandis que $f(0,0) = 2$. Le point est donc un maximum global.

(2)
```
The function :
(x, y) |--> x^3 + 3*x*y^2 - 15*x - 12*y
Derivative x and y :
(x, y) |--> 3*x^2 + 3*y^2 - 15
(x, y) |--> 6*x*y - 12
Hessian matrix :
[6*x 6*y]
[6*y 6*x]
Critical points :
(2, 1)
(1, 2)
(-1, -2)
(-2, -1)
At (2,1), the Hessian is
   (12, 6)
   (6, 12)
   Primary principal minors are [108, 12]
   Hessian positive defined
   local minimum
At (1,2), the Hessian is
   (6, 12)
   (12, 6)
   Primary principal minors are [-108, 6]
   Undefinite Hessian
   «selle» point
At (-1,-2), the Hessian is
   (-6, -12)
   (-12, -6)
```

```
   Primary principal minors are [-108, -6]
   Undefinite Hessian
   «selle» point
At (-2,-1), the Hessian is
   (-12, -6)
   (-6, -12)
   Primary principal minors are [108, -12]
   Hessian negative defined
   local maximum
```

(3)
```
The function :
(x, y) |--> 1/3*x^3 - x^2 + 4/3*y^3 - 3*x - 4*y - 3
Derivative x and y :
(x, y) |--> x^2 - 2*x - 3
(x, y) |--> 4*y^2 - 4
Hessian matrix :
[2*x - 2       0]
[      0     8*y]
Critical points :
(3, 1)
(-1, 1)
(3, -1)
(-1, -1)
At (3,1), the Hessian is
   (4, 0)
   (0, 8)
   Primary principal minors are [32, 4]
   Hessian positive defined
   local minimum
At (-1,1), the Hessian is
   (-4, 0)
   (0, 8)
   Primary principal minors are [-32, -4]
   Undefinite Hessian
   «selle» point
At (3,-1), the Hessian is
   (4, 0)
   (0, -8)
   Primary principal minors are [-32, 4]
   Undefinite Hessian
   «selle» point
At (-1,-1), the Hessian is
   (-4, 0)
   (0, -8)
   Primary principal minors are [32, -4]
   Hessian negative defined
   local maximum
```

△

Exemple 42.8.
Déterminer les valeurs extrêmes et les points de selle des fonctions suivantes.

(1) $f(x,y) = x^2 + 4x + y^2 - 2y$.
(2) $f(x,y) = e^{x^2+xy}$.
(3) $f(x,y) = e^x \sin(y)$.

Certains corrigés de cet exercice ont étés réalisés par Sage. Le script utilisé est `exo103.sage`

```
# -*- coding: utf8 -*-

import outilsINGE

def exercise_10_3_A():
  var('x,y')
  f(x,y)=x**2+4*x+y**2-2*y
  print outilsINGE.Extrema(f)

def exercise_10_3_H():
  var('x,y')
  f(x,y)=exp(x**2+x*y)
  print outilsINGE.Extrema(f)

def exercise_10_3_Q():
  var('x,y')
  f(x,y)=exp(x)*sin(y)
  print outilsINGE.Extrema(f)
```

tex/sage/exo103.sage

Des réponses :

(1)
```
The function :
(x, y) |--> x^2 + y^2 + 4*x - 2*y
Derivative x and y :
(x, y) |--> 2*x + 4
(x, y) |--> 2*y - 2
Hessian matrix :
[2 0]
[0 2]
Critical points :
(-2, 1)
At (-2,1), the Hessian is
   (2, 0)
   (0, 2)
   Primary principal minors are [4, 2]
   Hessian positive defined
   local minimum
```

(2)
```
The function :
(x, y) |--> e^(x^2 + x*y)
Derivative x and y :
(x, y) |--> (2*x*e^(x^2) + y*e^(x^2))*e^(x*y)
(x, y) |--> x*e^(x^2 + x*y)
Hessian matrix :
[(4*x*y*e^(x^2) + y^2*e^(x^2) + 2*(2*x^2 + 1)*e^(x^2))*e^(x*y)                    (x*y*e
[                  (x*y*e^(x^2) + (2*x^2 + 1)*e^(x^2))*e^(x*y)
Critical points :
(0, 0)
```

```
At (0,0), the Hessian is
   (2, 1)
   (1, 0)
   Primary principal minors are [-1, 2]
   Undefinite Hessian
   «selle» point
```

(3)
```
The function :
(x, y) |--> e^x*sin(y)
Derivative x and y :
(x, y) |--> e^x*sin(y)
(x, y) |--> e^x*cos(y)
Hessian matrix :
[ e^x*sin(y)  e^x*cos(y)]
[ e^x*cos(y) -e^x*sin(y)]
Critical points :
   I'm not able to solve these equations.
   I'm not able to solve these equations.
At (   ,I), the Hessian is
   I'm not able to solve these equations.
At (   ,I), the Hessian is
   I'm not able to solve these equations.
```

Ici, Sage n'est pas capable de résoudre les équations qui annulent le jacobien. Les équations à résoudre sont pourtant faciles :

$$\begin{cases} e^x \cos(y) = 0 & (42.5a) \\ e^x \sin(y) = 0 & (42.5b) \end{cases}$$

Étant donné que l'exponentielle ne s'annule jamais, il faudrait avoir en même temps $\cos(y) = 0$ et $\sin(y) = 0$, ce qui est impossible. La fonction n'a donc aucun extrémums local.

△

Exemple 42.9.
Considérons la fonction

$$f(x, y) = xy^2 e^{-(x^2+y^2)/4}. \tag{42.6}$$

(1) Montrer qu'il y a une infinité de points critiques.

(2) Déterminer leur nature.

Voici la fonction Sage qui fournit les informations :

```
# -*- coding: utf8 -*-

import outilsINGE

def exercise_10_4():
  var('x,y')
  f(x,y)=x*y**2*exp(-(x**2+y**2)/4)
  print outilsINGE.Extrema(f)
```

tex/sage/exo104.sage

La sortie est

```
The function :
(x, y) |--> x*y^2*e^(-1/4*x^2 - 1/4*y^2)
```

```
Derivative x and y :
(x, y) |--> -1/2*(x^2 - 2)*y^2*e^(-1/4*x^2 - 1/4*y^2)
(x, y) |--> -1/2*(x*y^3 - 4*x*y)*e^(-1/4*x^2 - 1/4*y^2)
Hessian matrix :
[                1/4*(x^3 - 6*x)*y^2*e^(-1/4*x^2 - 1/4*y^2) 1/4*((x^2 - 2)*y^3 - 4*(x^2 - 2
[1/4*((x^2 - 2)*y^3 - 4*(x^2 - 2)*y)*e^(-1/4*x^2 - 1/4*y^2)        1/4*(x*y^4 - 10*x*y^2 +
Critical points :
(r17, 0)
(-sqrt(2), -2)
(sqrt(2), -2)
(-sqrt(2), 2)
(sqrt(2), 2)
At (r17,0), the Hessian is
   (0, 0)
   (0, 2*r17*e^(-1/4*r17^2))
   Primary principal minors are [0, 0]
   Hessian positive semidéfinite
   I don't conclude
   Hessian negative semidefinite
   I don't conclude
At (-sqrt(2),-2), the Hessian is
   (4*sqrt(2)*e^(-3/2), 0)
   (0, 4*sqrt(2)*e^(-3/2))
   Primary principal minors are [32*e^(-3), 4*sqrt(2)*e^(-3/2)]
   Hessian positive defined
   local minimum
At (sqrt(2),-2), the Hessian is
   (-4*sqrt(2)*e^(-3/2), 0)
   (0, -4*sqrt(2)*e^(-3/2))
   Primary principal minors are [32*e^(-3), -4*sqrt(2)*e^(-3/2)]
   Hessian negative defined
   local maximum
At (-sqrt(2),2), the Hessian is
   (4*sqrt(2)*e^(-3/2), 0)
   (0, 4*sqrt(2)*e^(-3/2))
   Primary principal minors are [32*e^(-3), 4*sqrt(2)*e^(-3/2)]
   Hessian positive defined
   local minimum
At (sqrt(2),2), the Hessian is
   (-4*sqrt(2)*e^(-3/2), 0)
   (0, -4*sqrt(2)*e^(-3/2))
   Primary principal minors are [32*e^(-3), -4*sqrt(2)*e^(-3/2)]
   Hessian negative defined
   local maximum
```

Notez la présence de `r1` comme paramètres dans les solutions. Tous les points avec $y = 0$ sont des points critiques. Cependant, Sage [1] ne parvient pas à conclure la nature de ces points $(x, 0)$.

Notons que le nombre $f(x, y)$ a toujours le signe de x parce que y^2 et l'exponentielle sont positives. Toujours ? En tout cas lorsque $x \neq 0$. Prenons un point $(a, 0)$ avec $a > 0$. Dans un voisinage de ce point, nous avons $f(x, y) > 0$ parce que si $a > 0$, alors $x > 0$ dans un voisinage de a. Le point $(a, 0)$ est un minimum local parce que $0 = f(a, 0) \leqslant f(x, y)$ pour tout (x, y) dans un voisinage de $(a, 0)$.

1. ou, plus précisément, le programme que j'ai écrit avec Sage.

De la même façon, les points $(a, 0)$ avec $a < 0$ sont des maximums locaux parce que dans un voisinage, la fonction est négative.

Le point $(0, 0)$ n'est ni maximum ni minimum local. C'est un point de selle.

△

Exemple 42.10.
Dériver les fonctions suivantes.

(1) $\sin\big(\ln(x)\big)$

(2) $\dfrac{\sin x}{x}$;

(3) e^{x^2}

(4) $\cos(x)^{\sin(x)}$

Le programme suivant par Sage résout l'exercice :

```
#! /usr/bin/sage -python
# -*- coding: utf8 -*-

from sage.all import *

var('x')
f=sin(ln(x))
print f.diff(x)
f=sin(x)/x
print f.diff(x)
f=exp(x**2)
print f.diff(x)
f=cos(x)**(sin(x))
print f.diff(x)
```

tex/sage/corrDerive_0002.sage

Le résultat est :

```
cos(log(x))/x
cos(x)/x - sin(x)/x^2
2*x*e^(x^2)
(log(cos(x))*cos(x) - sin(x)^2/cos(x))*cos(x)^sin(x)
```

△

Exemple 42.11.
Donner une approximation de $\ln(1.0001)$.

```
----------------------------------------------------------------------
| Sage Version 4.5.3, Release Date: 2010-09-04                       |
| Type notebook() for the GUI, and license() for information.        |
----------------------------------------------------------------------
sage: numerical_approx(ln(1.0001))
0.0000999950003332973
```

△

Chapitre 43

Épilogue : la constante de Weiner

Nous voici à la fin du Frido. Nous avons étudié beaucoup de math, et beaucoup reste à voir. En guise de conclusion, je voudrais vous parler de la constante de Weiner, introduite dans [543]. Il s'agit d'une constante, qui comme π ou e intervient dans à peu près tous les domaines de la mathématique.

Comme toujours, il existe énormément de définitions équivalentes différentes ; nous choisissons celle-ci, motivée par le le théorème de Weinersmith 26.42.

Définition 43.1.
*La **constante de Weiner** W_c est l'unique entier p tel que l'espace L^p soit un espace de Hilbert.*

Cette constante intervient de façon centrale dans de nombreux résutats dans tous les domaines ; nous en citons quelques-uns.

(1) La moyenne de tout couple de réels peut être calculée en divisant leur somme par la constante de Weiner [1].
(2) La constante de Weiner donne l'indice du groupe alterné dans le groupe symétrique pour tous les ordres, théorème 5.30.
(3) La constante de Weiner donne une borne inférieure optimale pour l'ensemble des nombres premiers.
(4) L'unique point fixe non trivial de la fonction factorielle est la constante de Weiner.
(5) Tout automorphisme d'anneau a un polynôme minimal dont le degré est donné par la constante de Weiner.
(6) L'égalité $ab = 0$ dans un anneau n'implique pas spécialement $a = 0$ ou $b = 0$ lorsque la caractéristique de l'anneau est égale à la constante de Weiner, et seulement dans ce cas.
(7) Pour l'anecdote, la constante de Weiner donne le rapport τ/π ; elle est aussi la partie entière de e.

Il reste encore de nombreuses conjectures mettant en valeur la constante de Weiner :

(1) La fameuse droite critique de la conjecture de Riemann est donnée par l'inverse de la constante de Weiner.
(2) Soit $\mathcal{P}$ l'ensemble de nombres premiers. Est-ce que l'ensemble
$$\{(p,q) \in \mathcal{P} \times \mathcal{P} \text{ tel que } 0 < |p-q| = W_c\} \tag{43.1}$$
est fini ?

D'aucuns pourraient objecter que tout cela n'est que fantaisie et trivialité. Il n'en est rien. La preuve que la constante de Weiner est centrale en mathématique est précisément qu'elle avait déjà un nom et un symbole réservé bien avant le début de l'histoire des mathématiques.

Le fait est que toutes les mathématiques que vous connaissez se basent sur les nombres entiers ; cela n'est pas du tout une trivialité.

1. C'est historiquement la première propriété énoncée de la constante de Weiner ; elle suggère également une notion de constante de Weiner généralisée pour moyenner un nombre arbitraire de nombres. La construction des nombres de Weiner généralisés est en projet dans la section 1.2.

Chapitre 44

GNU Free Documentation License

Version 1.3, 3 November 2008
Copyright © 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc.

http://fsf.org/

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom : to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation : a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals ; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "**Document**", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "**you**". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "**Modified Version**" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "**Secondary Section**" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical

connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "**Invariant Sections**" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "**Cover Texts**" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "**Transparent**" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "**Opaque**".

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "**Title Page**" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "**publisher**" means any person or entity that distributes copies of the Document to the public.

A section "**Entitled XYZ**" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "**Acknowledgements**", "**Dedications**", "**Endorsements**", or "**History**".) To "**Preserve the Title**" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties : any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts : Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version :

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another ; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of

it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/ .

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM : How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page :

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation ; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with ... Texts." line with this :

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.

Bibliographie

[1] L'auteur de ces lignes. Invention personnelle de l'auteur de ces lignes, lire avec prudence et merci de me dire si c'est correct. Attention : je ne prétends pas avoir découvert ce résultat ; un résultat qui a cette référence signifie seulement que j'ai ré-inventé une partie de l'énoncé ou de la preuve. Je serais étonné qu'il y ait ici des résultats vraiment présents nulle part ailleurs.

[2] A. Casamayou, N Cohen, G. Connan, T. Dumont, L. Fousse, F. Maltey, M. Meulien, M. Mezzarobba, C. Pernet, M. N. Thiéry, and Zimmermann P. Calcul mathématique avec Sage. 2013. URL http://sagebook.gforge.inria.fr/.

[3] Wikipedia contributors. Nth root algorithm — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Nth_root_algorithm&oldid=873962836, 2018. [Online ; accessed 31-January-2019].

[4] Sylvie Benzoni. Distributions tempérées. 13 novembre 2013. URL http://math.univ-lyon1.fr/~benzoni/Agreg/Distributions.pdf.

[5] Wikipédia. Axiome du choix — wikipédia, l'encyclopédie libre, 2015. URL http://fr.wikipedia.org/w/index.php?title=Axiome_du_choix&oldid=116151439. [En ligne ; Page disponible le 9-septembre-2015].

[6] O. Teytaud, C. Antonini, J.-B. Bardet, J.-F. Quint, M. De Crisenoy, P. Borgnat, J. Bérard, E. Lebeau, E. Souche, and A. Chateau. Les mathématiques pour l'Agrégation. 21 mai 2002. http://les.mathematiques.free.fr/pdf/agreg.zip.

[7] Wikipédia. Lemme de zorn — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=Lemme_de_Zorn&oldid=164572297. [En ligne ; Page disponible le 16-novembre-2019].

[8] Christine Laurent-Thiébaut. Axiomatique des nombres. URL http://ljk.imag.fr/membres/Bernard.Ycart/mel/ax/ax.pdf.

[9] Wikipédia. Équipotence — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=%C3%89quipotence&oldid=157942537. [En ligne ; Page disponible le 28-mars-2019].

[10] Jérôme Von Buhren. Cardinal d'un ensemble et théorème de cantor-bernstein. URL http://vonbuhren.free.fr/Prepa/Problemes/Cardinal_ensemble_theoreme_Cantor_Bernstein.pdf.

[11] Ilie Grigorescu. Lecture 4 - finite and infinite sets. URL http://www.math.miami.edu/~igrigore/teaching/mth533/lecture4_mth433.pdf.

[12] Wikipédia. Ensemble dénombrable — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Ensemble_d%C3%A9nombrable&oldid=150759981. [En ligne ; Page disponible le 28-juillet-2018].

[13] Alexander Bogomolny. Cantor-bernstein-schroeder theorem. URL https://www.cut-the-knot.org/WhatIs/Infinity/Bernstein.shtml.

[14] Wikipédia. Théorème de comparabilité cardinale — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_comparabilit%C3%A9_cardinale&oldid=158933871. [En ligne ; Page disponible le 2-mai-2019].

[15] Ramez Labib-Sami. Théorie de la cardinalité -2. ensembles infini non dénombrables. 2015. URL http://www.logique.jussieu.fr/~sami/Ensembles_L3/5_Cardinaux_non_denombrables.pdf.

[16] Wikipédia. Théorème de cantor — wikipédia, l'encyclopédie libre, 2020. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Cantor&oldid=166939268. [En ligne ; Page disponible le 31-janvier-2020].

[17] Pedro Tamaroff. Prove that if A is an infinite set then Ax2 is equipotent to A. 2014. URL https://math.stackexchange.com/questions/1041731/prove-that-if-a-is-an-infinite-set-then-a-times-2-is-equipotent-to-a.

[18] cdrcprds. Ensembles équipotents : ajouter ou soustraire des cardinalités. 2020. URL https://github.com/LaurentClaessens/mazhe/issues/146#issuecomment-588513588.

[19] MikeMathMan. Cardinality of the cartesian product of two equinumerous infinite sets. 2017. URL https://math.stackexchange.com/questions/1383755/cardinality-of-the-cartesian-product-of-two-equinumerous-infinite-sets.

[20] Patrice Tauvel. *Cours d'algèbre*. Dunod, 1999. ISBN 2-10-005490-3.

[21] Wikipédia. Morphisme d'anneaux — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Morphisme_d%27anneaux&oldid=129892066. [En ligne ; Page disponible le 23-septembre-2016].

[22] Wikipédia. Diviseur de zéro — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Diviseur_de_z%C3%A9ro&oldid=143858271. [En ligne ; Page disponible le 26-décembre-2017].

[23] Wikipédia. Corps commutatif — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Corps_commutatif&oldid=152289565. [En ligne ; Page disponible le 18-septembre-2018].

[24] Wikipédia. Corps des fractions — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Corps_des_fractions&oldid=130674526. [En ligne ; Page disponible le 14-octobre-2016].

[25] Frédéric Paulin. Topologie, analyse et calcul différentiel. 2008-2009. http://www.math-question-center.com/publications-pdf/cours_d'analyse+topologie_et_calcul_differntiel_Frederic_Paulin.pdf.

[26] Wikipédia. Construction des nombres réels — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Construction_des_nombres_r%C3%A9els&oldid=145402164. [En ligne ; Page disponible le 12-février-2018].

[27] Wikipédia. Nombre positif — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Nombre_positif&oldid=146043202. [En ligne ; Page disponible le 3-mars-2018].

[28] Wikipedia contributors. Least-upper-bound property — Wikipedia, the free encyclopedia, 2019. URL https://en.wikipedia.org/w/index.php?title=Least-upper-bound_property&oldid=893001075. [Online ; accessed 30-April-2019].

[29] Peter Hendrikus Kropholler. Groups rings and fields. 2004-2005. http://www.maths.gla.ac.uk/~phk/ulearning.pdf.

[30] Jean-Louis Rouget. Compléments d'algèbre. 2017. URL https://www.maths-france.fr/MathSpe/Cours/01-structures.pdf.

[31] Anonyme. Théorie des groupes. 2008. Lien original mort : http://theoriedesgroupes.perso.sfr.fr/cours/theoriePDF.pdf autre lien, qui fonctionne pour l'instant : http://ekladata.com/CKoLuTDRT9S2sx5BPX9DkCDAcdo.pdf et https://web.archive.org/web/20150701233527/http://theoriedesgroupes.perso.sfr.fr/cours/theoriePDF.pdf.

[32] Jean Luc W. groupe-quotient et création de nouveaux sous-groupes. 2007. URL https://fr.wikipedia.org/wiki/Discussion:Groupe_quotient.

[33] Pierre Lissy. Théorème de Jordan-Hölder. 6 avril 2010. http://www.ljll.math.upmc.fr/~lissy/Agreg/developpements/JordanHolder.pdf.

[34] Muriel Galley. Groupes résolubles. 24 mai 2012. URL http://matthieu.gendulphe.com/Galley.pdf.

[35] François Combes. *Algèbre et géométrie*. Bréal, 2003. ISBN 2-84291-202-0.

[36] Fabrice Castel. Groupes finis. 2009-2012. http://agreg-maths.univ-rennes1.fr/documentation/docs/Groupes_finis.pdf.

[37] Wikipédia. Groupe symétrique — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Groupe_sym%C3%A9trique&oldid=146656747. [En ligne ; Page disponible le 21-mars-2018].

[38] Jean Delcourt. Groupes-permutations. URL http://delcourt.u-cergy.fr/StrucAlg/chap_5.pdf.

[39] Wikipédia. Permutation — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Permutation&oldid=89805226. [En ligne ; Page disponible le 24-juin-2013].

[40] Patrick Polo. Idéaux premiers et maximaux, localisation, théorème des zéros de Hilbert. 23 octobre 2005. URL https://webusers.imj-prg.fr/~patrick.polo/M1Galois/ATG05ch3.pdf.

[41] cdrcprds. Questions (9), (15) et (21) d'algèbre. 2017. URL https://github.com/LaurentClaessens/mazhe/issues/60.

[42] Wikiversité. Arithmétique/divisibilité et congruences dans z — wikiversité,, 2017. URL https://fr.wikiversity.org/w/index.php?title=Arithm%C3%A9tique/Divisibilit%C3%A9_et_congruences_dans_Z&oldid=650617. [En ligne ; accédé le 25-mai-2017].

[43] Mongi Amorri. Théorème de bezout, théorème de gauss. http://lux.lyceefrancais-brasilia.net/documents/cours/ts/bezout_gauss.pdf.

[44] Gilles Costantini. pgcd et ppcm dans $\mathbb{Z}$, théorème de Bezout, applications. . http://gilles.costantini.pagesperso-orange.fr/agreg_fichiers/bezout.pdf.

[45] Congrugences dans $\mathbb{Z}$, anneaux $\mathbb{Z}/n\mathbb{Z}$. . URL http://math.univ-lyon1.fr/capes/IMG/pdf/new.congruence.pdf.

[46] Keith Conrad. Cyclicity of $(Z/(p))^{\times}$. . URL https://kconrad.math.uconn.edu/blurbs/grouptheory/cyclicmodp.pdf.

[47] minutephysics. How quantum computing breaks encryption, shor's algorithm explained. 2019. URL https://www.youtube.com/watch?v=lvTqbM5Dq4Q.

[48] Wikipédia. Lemme d'euclide — wikipédia, l'encyclopédie libre, 2015. URL http://fr.wikipedia.org/w/index.php?title=Lemme_d%27Euclide&oldid=112046247. [En ligne ; Page disponible le 28-juin-2015].

[49] Wikipédia. Théorème fondamental de l'arithmétique — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_fondamental_de_l%27arithm%C3%A9tique&oldid=110036012. [En ligne ; Page disponible le 26-juin-2015].

[50] Wikipédia. Théorème de cauchy (groupes) — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Cauchy_(groupes)&oldid=140167231. [En ligne ; Page disponible le 17-septembre-2017].

[51] Alexei Pantchichkine. Algèbre2. 2004-2005. http://www-fourier.ujf-grenoble.fr/~panchish/05ensl.pdf.

[52] Niels Borne. Fiche numéro 2 : morphismes, sous-groupe distingué, quotient. 2009-2012. http://math.univ-lille1.fr/~borne/Enseignement/TD2_M308.pdf.

[53] Georges Skandalis. Algèbre générale et algèbre linéaire. 6 mars 2015. URL http://www.math.univ-paris-diderot.fr/_media/formations/prepa/agreginterne/polycopiealgebre.pdf.

[54] Arnaud Girand. Développements pour l'agrégation externe de mathématiques. 18 août 2012. URL http://perso.univ-rennes1.fr/arnaud.girand/pdf/dvp_agreg/dvp.pdf.

[55] Wikipédia. Formule du binôme de newton — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Formule_du_bin%C3%B4me_de_Newton&oldid=144161875. [En ligne ; Page disponible le 4-janvier-2018].

[56] Wikipédia. Caractéristique d'un anneau — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Caract%C3%A9ristique_d%27un_anneau&oldid=136316021. [En ligne ; Page disponible le 22-mai-2017].

[57] Wikipédia. Module sur un anneau — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Module_sur_un_anneau&oldid=141582982. [En ligne ; Page disponible le 16-octobre-2017].

[58] Patrick Polo. Algèbre et théorie de galois. 2007-2008. URL https://webusers.imj-prg.fr/~patrick.polo/M1Galois/ATG07chI-III.pdf.

[59] David Harari. Algèbre 1 - anneaux et modules. . URL https://www.math.u-psud.fr/~harari/exposes/rings.pdf.

[60] Wikipédia. Algèbre sur un corps — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Alg%C3%A8bre_sur_un_corps&oldid=98992040. [En ligne ; Page disponible le 6-mars-2014].

[61] Patrick Polo. Groupes et polynômes symétriques, résolution d'équations. URL https://webusers.imj-prg.fr/~patrick.polo/M1Galois/ATG07chIXcorr.pdf.

[62] Wikipédia. Primalité dans un anneau — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Primalit%C3%A9_dans_un_anneau&oldid=127919489. [En ligne ; Page disponible le 30-mai-2017].

[63] Wikipédia. Idéal premier — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Id%C3%A9al_premier&oldid=125964642. [En ligne ; Page disponible le 7-mai-2016].

[64] Anonyme. sans titre. http://www.ac-grenoble.fr/champo/IMG/arithmetique.pdf.

[65] Wikipédia. Primalité dans un anneau — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Primalit%C3%A9_dans_un_anneau&oldid=145407116. [En ligne ; Page disponible le 12-février-2018].

[66] Michel Cretin. Anneaux principaux et factoriels. URL http://math.univ-lyon1.fr/~cretin/OralAlgebre/annprincfact.pdf.

[67] Wikipédia. Idéal premier — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Id%C3%A9al_premier&oldid=145977206. [En ligne ; Page disponible le 1-mars-2018].

[68] Danny-Jack Mercier. Anneaux factoriels. 5 octobre 2003. http://megamaths.perso.neuf.fr/cours/ari/cann0001.pdf.

[69] Wikipédia. Anneau euclidien — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Anneau_euclidien&oldid=126145344. [En ligne ; Page disponible le 1-juin-2017].

[70] Wikiversité. Anneau (mathématiques)/exercices/exercices — wikiversité,, 2017. URL https://fr.wikiversity.org/w/index.php?title=Anneau_(math%C3%A9matiques)/Exercices/Exercices&oldid=676468. [En ligne ; accédé le 26-octobre-2017].

[71] Sébastien Pellerin. Développements d'algèbre pour l'oral de l'agrégation. . http://pellerin.xyz/doc/agreg/algebre.pdf.

[72] Wikipédia. Triplet pythagoricien — wikipédia, l'encyclopédie libre, 2015. URL http://fr.wikipedia.org/w/index.php?title=Triplet_pythagoricien&oldid=114628048. [En ligne ; Page disponible le 24-juin-2015].

[73] Wikipédia. Primalité dans un anneau — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Primalit%C3%A9_dans_un_anneau&oldid=127919489. [En ligne ; Page disponible le 1-juin-2017].

[74] Wikipédia. Lemme de gauss (polynômes) — wikipédia, l'encyclopédie libre, 2015. URL http://fr.wikipedia.org/w/index.php?title=Lemme_de_Gauss_(polyn%C3%B4mes)&oldid=112056131. [En ligne ; Page disponible le 1-juin-2017].

[75] Wikipédia. Espace vectoriel — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Espace_vectoriel&oldid=143370323. [En ligne ; Page disponible le 10-décembre-2017].

[76] Espaces vectoriels de dimension infinie. . URL http://www.les-mathematiques.net/b/e/e/node8.php.

[77] Wikipedia contributors. Tensor product — Wikipedia, the free encyclopedia, 2018. URL https://en.wikipedia.org/w/index.php?title=Tensor_product&oldid=842778935. [Online ; accessed 18-June-2018].

[78] Marie-Pierre Lebaux. Rappels sur les applications linéaires. URL https://perso.univ-rennes1.fr/marie-pierre.lebaud/agint/ecrit/algebre-lineaire/applications-lineaires/V-appli-lin.pdf.

[79] Jean-Louis Rouget. Dimension d'un espace vectoriel. 2018. URL https://www.maths-france.fr/MathSup/Cours/22-dimensions.pdf.

[80] Daniel Ferrand. Étendre le corps. Juillet 2007. URL http://webusers.imj-prg.fr/~daniel.ferrand/ExtCorps.pdf.

[81] Ycart Bernard and Luc Rozoy. Déterminants. URL https://ljk.imag.fr/membres/Bernard.Ycart/mel/de/de.pdf.

[82] G. Donald Allen. Matrices and linear algebra. 2004. URL http://www.math.tamu.edu/~dallen/m640_03c/lectures/chapter2.pdf.

[83] Wikipédia. Mineur (algèbre linéaire) — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Mineur_(alg%C3%A8bre_lin%C3%A9aire)&oldid=133960288. [En ligne ; Page disponible le 25-janvier-2017].

[84] Khalid Koufany. Formes linéaires et dualité. Janvier 2018. URL http://www.iecl.univ-lorraine.fr/~Khalid.Koufany/Alg-Bil/Chap1-Formes-Lineaires.pdf.

[85] Marc Sage. Dualité en dimension finie. 24 octobre 2005. http://www.normalesup.org/~sage/Enseignement/Cours/DualDimFinie.pdf.

[86] Wikipédia. Théorème de cauchy (groupes) — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Cauchy_(groupes)&oldid=122684298. [En ligne ; Page disponible le 1-août-2017].

[87] Wikiversité. Groupe (mathématiques)/exercices/premiers résultats sur les groupes simples — wikiversité,, 2016. URL https://fr.wikiversity.org/w/index.php?title=Groupe_(math%C3%A9matiques)/Exercices/Premiers_r%C3%A9sultats_sur_les_groupes_simples&oldid=597195. [En ligne ; accédé le 31-mai-2017].

[88] Wikiversité. Groupe (mathématiques)/automorphismes d'un groupe cyclique — wikiversité,, 2016. URL https://fr.wikiversity.org/w/index.php?title=Groupe_(math%C3%A9matiques)/Automorphismes_d%27un_groupe_cyclique&oldid=595608. [En ligne ; accédé le 3-juin-2017].

[89] Structure des groupes d'ordre pq. URL http://agregmaths.free.fr/doc/docs_nicolas/developpementAlgebre/grouped'ordrepq.pdf.

[90] Olivier Debarre. Td4 : produit semi-direct. 2015-2016. URL http://www.math.ens.fr/~debarre/TDC4.pdf.

[91] Christophe Mourougane. Théorie des groupes et géométrie. 2009-2010. http://perso.univ-rennes1.fr/christophe.mourougane/enseignements/2009-10/THGG/poly.groupes.pdf.

[92] Michel Emsalem and Pierre Dèbes. Théorème de sylow. http://exo7.emath.fr/ficpdf/fic00023.pdf.

[93] N.G.J. Pagnon. Le groupe symétrique et le groupe alterné. URL http://amatheux.com/IMG/pdf/groupe_symetrique.pdf.

[94] Wikipédia. Corps ordonné — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Corps_ordonn%C3%A9&oldid=125185412. [En ligne ; Page disponible le 22-mai-2017].

[95] Automorphisme de R et continuité. 2008. URL http://forums.futura-sciences.com/mathematiques-superieur/261451-automorphisme-de-r-continuite.html.

[96] Wikipédia. Monoïde — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Mono%C3%AFde&oldid=148198215. [En ligne ; Page disponible le 6-mai-2018].

[97] Sylvain Duchet. Congrugence dans $\mathbb{Z}$, anneaux $\mathbb{Z}/n\mathbb{Z}$, applications. http://epsilon.2000.free.fr/Csup/congruences.pdf.

[98] Wikipédia. Polynôme irréductible — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Polyn%C3%B4me_irr%C3%A9ductible&oldid=132607764. [En ligne ; Page disponible le 12-décembre-2016].

[99] Patrick Polo. Extension de corps, caractéristique, corps de rupture, corps de décomposition, clôtures algébriques. 2007-2008. URL https://webusers.imj-prg.fr/~patrick.polo/M1Galois/ATG07chVII.pdf.

[100] C. Antonini, JF. Quint, P. Borgnat, J. Bérard, E. Lebeau, E. Souche, A. Chateau, and O. Teytaud. Le théorème de wedderburn. . URL http://www.les-mathematiques.net/d/a/w/node5.php.

[101] cdrcprds. Lemme 5.43 sur les pgcd de polynômes. 2017. URL https://github.com/LaurentClaessens/mazhe/issues/52#issuecomment-333251728.

[102] Wikipédia. Extension de corps — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Extension_de_corps&oldid=133373097. [En ligne ; Page disponible le 5-janvier-2017].

[103] Wikipédia. Extension de corps — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Extension_de_corps&oldid=133373097. [En ligne ; Page disponible le 5-janvier-2017].

[104] David Harari. Algèbre 1 - notions de théorie des corps. . URL https://www.math.u-psud.fr/~harari/exposes/corps.pdf.

[105] Pierre Bernard. Entiers algébriques et racines de l'unité. 7 janvier 2012. URL http://allken-bernard.org/pierre/weblog/?p=2061.

[106] Mortajine Abdellatif. Les extensions de corps. 2012-2013. URL http://www.iecn.u-nancy.fr/~mortajin/chap3-L3-S5.pdf.

[107] Wikipédia. Extension algébrique — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Extension_alg%C3%A9brique&oldid=148162879. [En ligne ; Page disponible le 5-mai-2018].

[108] Olivier Debarre. Algèbre 2. 2012-2013. URL https://www.math.ens.fr/~debarre/Algebre2.pdf.

[109] Wikipédia. Clôture algébrique — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=Cl%C3%B4ture_alg%C3%A9brique&oldid=159475673. [En ligne ; Page disponible le 22-mai-2019].

[110] Herscovich Estanislao. Existence et « unicité » d'une clôture algébrique. URL https://www-fourier.ujf-grenoble.fr/~eherscov/MAT4111/ThmSteinitz.pdf.

[111] Wikipédia. Corps de décomposition — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Corps_de_d%C3%A9composition&oldid=139881008. [En ligne ; Page disponible le 19-août-2017].

[112] Patrick Polo. Extensions normales, séparables, galoisiennes. corps fini. 2007-2008. http://www.math.jussieu.fr/~polo/M1/ATG07chVIII.pdf.

[113] Olivier Dodane. Le théorème de zéros de Hilbert. 8 octobre 2007. http://www.math.ens.fr/~debarre/nullstellensatz.pdf.

[114] Patrick Polo. Polynômes symétriques et résolutions d'équations. 13 décembre 2004. http://www.math.jussieu.fr/~polo/M1/ATGch9.pdf.

[115] Vincent Beck, Jérôme Malick, and Gabriel Peyré. Sur c tout est connexe ! URL http://objagr.gforge.inria.fr/documents/files/connexite-polynome.pdf.

[116] El Hage. Équation générale de degré n. http://les.mathematiques.free.fr/pdf/gal10.pdf.

[117] Florian Morel Chevillet. Résolubilité par radicaux des équations algébriques. 31 mai 2012. URL http://matthieu.gendulphe.com/MorelChevillet.pdf.

[118] Wikipédia. Espace topologique — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=Espace_topologique&oldid=164610481. [En ligne ; Page disponible le 17-novembre-2019].

[119] Erwan Aubry. URL https://math.unice.fr/~eaubry/Enseignement/L3/rappelstopo.pdf.

[120] Wikipédia. Prébase — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=Pr%C3%A9base&oldid=162334814. [En ligne ; Page disponible le 3-septembre-2019].

[121] Wikipédia. Espace séparé — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Espace_s%C3%A9par%C3%A9&oldid=97013072. [En ligne ; Page disponible le 27-septembre-2013].

[122] Wikipédia. Équivalent — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=%C3%89quivalent&oldid=149344743. [En ligne ; Page disponible le 8-juin-2018].

[123] Wikipédia. Connexité (mathématiques) — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=Connexit%C3%A9_(math%C3%A9matiques)&oldid=164885091. [En ligne ; Page disponible le 26-novembre-2019].

[124] Wikipédia. Base (topologie) — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Base_(topologie)&oldid=140151583. [En ligne ; Page disponible le 29-août-2017].

[125] C. Antonini, JF. Quint, P. Borgnat, J. Bérard, E. Lebeau, E. Souche, A. Chateau, and O. Teytaud. Généralités. . URL http://www.les-mathematiques.net/a/a/b/node19.php.

[126] Emmanuel Vieillard Baron. Sous-espaces compacts. 2001. URL http://www.les-mathematiques.net/a/t/c/node5.php.

[127] Wikipédia. Compactifié d'alexandrov — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Compactifi%C3%A9_d%27Alexandrov&oldid=146811167. [En ligne ; Page disponible le 26-mars-2018].

[128] Wikipedia contributors. Filtration (mathematics) — Wikipedia, the free encyclopedia, 2018. URL https://en.wikipedia.org/w/index.php?title=Filtration_(mathematics). [Online ; accessed 14-October-2018].

[129] C. Antonini, JF. Quint, P. Borgnat, J. Bérard, E. Lebeau, E. Souche, A. Chateau, and O. Teytaud. Espaces métriques compacts. . http://www.les-mathematiques.net/a/a/b/node22.php.

[130] Éric Brunelle. Norme matricielle. http://www.dms.umontreal.ca/~math1600/6Supplement/Normematricielle-1.pdf.

[131] Wikipédia. Théorème de borel-lebesgue — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Borel-Lebesgue&oldid=96560796. [En ligne ; Page disponible le 11-mars-2014].

[132] Wikipédia. Théorème des suites adjacentes — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_des_suites_adjacentes&oldid=148318442. [En ligne ; Page disponible le 10-mai-2018].

[133] Mathieu Mansuy. Autour des séries alternées. URL http://www.mathieu-mansuy.fr/pdf/ECS2-complément1.pdf.

[134] M. Quéffelec and A. Bodin. Topologie générale. URL http://exo7.emath.fr/ficpdf/fic00037.pdf.

[135] Ambigraphe. Demande de référence. 2008. URL https://fr.wikipedia.org/wiki/Discussion:Ouvert_(topologie).

[136] Wikipédia. Espace topologique — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Espace_topologique&oldid=152161195. [En ligne ; Page disponible le 13-septembre-2018].

[137] beaucoup de monde. Annonce frido 2019. . https://linuxfr.org/news/le-frido-livre-collaboratif-de-mathematique-de-niveau-agregat

[138] beaucoup de monde. Annonce frido 2018. . https://linuxfr.org/news/le-frido-2018-livre-libre-de-mathematique-pour-l-agregation.

[139] Daniel Perrin. À propos de la définition de la limite d'une fonction en un point. . URL https://www.math.u-psud.fr/~perrin/CAPES/analyse/fonctions/definitiondelimite.pdf.

[140] Usual limit from filter. 2019. URL https://math.stackexchange.com/questions/3364039/usual-limit-from-filter.

[141] Wikipédia. Application bilinéaire — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Application_bilin%C3%A9aire&oldid=144184234. [En ligne ; Page disponible le 5-janvier-2018].

[142] Michel Coste. Formes bilinéaires symétriques, formes quadratiques. URL https://perso.univ-rennes1.fr/michel.coste/Bil.pdf.

[143] Wikipédia. Hermitien — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Hermitien&oldid=124476255. [En ligne ; Page disponible le 18-mars-2016].

[144] Daniel Li. Espaces de hilbert. . http://www.editions-ellipses.fr/product_info.php?products_id=11183 http://www.editions-ellipses.fr/product_info.php?products_id=9387.

[145] user127.0.0.1. Parallelogram law in normed vector space without an inner product. 2014. URL https://math.stackexchange.com/questions/641077/parallelogram-law-in-normed-vector-space-without-an-inner-product.

[146] Frank Jones. Chapter 7 : cross product. 2004. URL http://www.owlnet.rice.edu/~fjones/chap7.pdf.

[147] William F. Trench. Introduction to real analysis. 2010. http://ramanujan.math.trinity.edu/wtrench/texts/TRENCH_REAL_ANALYSIS.PDF.

[148] David Wilkins. Normed vector spaces and functional analysis. 1997-2001. URL http://www.maths.tcd.ie/~dwilkins/Courses/212/212PtIII.pdf.

[149] Wikipédia. Connexité (mathématiques) — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Connexit%C3%A9_(math%C3%A9matiques)&oldid=123744973. [En ligne ; Page disponible le 26-avril-2016].

[150] C. Antonini, JF. Quint, P. Borgnat, J. Bérard, E. Lebeau, E. Souche, A. Chateau, and O. Teytaud. Connexité. . URL http://www.les-mathematiques.net/a/a/b/node23.php.

[151] Huiqiang Jiang. Functional analysis. URL http://www.math.pitt.edu/~hqjiang/2303/functional.pdf.

[152] Nicolas Bourbaki. General topology 2 chapters 5 - 10. URL https://books.google.be/books?id=bQwhdmL6IjUC. Oui je sais c'est une honte de citer Bourbaki en anglais en pointant vers une version partielle disponible sur Googlebooks. Mais c'est surtout une honte que ce livre ne soit pas disponible gratuitement en version électronique.

[153] Guillaume Carlier. Analyse fonctionnelle. 2009-20010. URL https://www.ceremade.dauphine.fr/~carlier/poly2010.pdf.

[154] Raz Kupferman. Topological vector spaces. URL http://www.ma.huji.ac.il/~razk/iWeb/My_Site/Teaching_files/TVS.pdf.

[155] Ron Freiwald. Chapter iv completeness and compactness. URL http://www.math.wustl.edu/~freiwald/ch4.pdf.

[156] Wikipédia. Suite de cauchy — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Suite_de_Cauchy&oldid=98046451. [En ligne ; Page disponible le 11-mars-2014].

[157] Wikipédia. Continuité — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Continuit%C3%A9&oldid=96548980. [En ligne ; Page disponible le 17-septembre-2013].

[158] Wikipédia. Théorème des fermés emboités — Wikipédia, l'encyclopédie libre, 2012. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_des_ferm%C3%A9s_emboit%C3%A9s&oldid=85987870. [En ligne ; Page disponible le 1-avril-2013].

[159] C. Antonini, JF. Quint, P. Borgnat, J. Bérard, E. Lebeau, E. Souche, A. Chateau, and O. Teytaud. Le théorème de tykhonov. . URL http://www.les-mathematiques.net/a/a/b/node20.php.

[160] Wikipédia. Équicontinuité — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=%C3%89quicontinuit%C3%A9&oldid=137910699. [En ligne ; Page disponible le 4-juin-2017].

[161] Wikipédia. Continuité uniforme — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Continuit%C3%A9_uniforme&oldid=138238518. [En ligne ; Page disponible le 16-juin-2017].

[162] Zied Ammari. Analyse fonctionnelle : Pré-requis. http://perso.univ-rennes1.fr/zied.ammari/other/pdf/chapitre1.pdf.

[163] Jean Saint Raymond. Semi-normes. URL http://www.math.jussieu.fr/~raymond/preprints/seminormes.pdf.

[164] Stéphane Mischler. Semi-norme et introduction aux evtlcs. Février 2007. URL https://www.ceremade.dauphine.fr/~mischler/Enseignements/AFAENS/Chap1evtlcs.pdf.

[165] Wikipédia. Semi-norme — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Semi-norme&oldid=89981857. [En ligne ; Page disponible le 22-septembre-2013].

[166] Wikipédia. Théorème de baire — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Baire&oldid=95077621. [En ligne ; Page disponible le 4-août-2013].

[167] Marie-Claure David, Frédéric Haglund, and Daniel Perrin. Géométrie affine. 8 décembre 2003. http://webens.math.u-psud.fr/~mcld/GAEL/Gael/bar.pdf.

[168] Marie-Claude David, Frédéric Haglund, and Daniel Perrin. Géometrie affine. Document de travail pour la préparation au CAPES. Deuxième partit : barycentre. 8 décembre 2003. URL http://omphale.math.u-psud.fr/~mcld/GAEL/Gael/bar.pdf.

[169] Marie-Claude David, Frédéric Haglund, and Daniel Perrin. Géometrie affine. Document de travail pour la préparation au CAPES. Troisième partie : convexité. 8 décembre 2003. URL http://omphale.math.u-psud.fr/~mcld/GAEL/Gael/conv.pdf.

[170] Jean-Marc Decauwert. Géométrie affine. 8 novembre 2011. URL http://ljk.imag.fr/membres/Bernard.Ycart/mel/ga/ga.pdf.

[171] Wikiversité. Barycentre/théorème de l'associativité du barycentre — wikiversité,, 2016. URL https://fr.wikiversity.org/w/index.php?title=Barycentre/Th%C3%A9or%C3%A8me_de_l%27associativit%C3%A9_du_barycentre&oldid=582092. [En ligne ; accédé le 16-janvier-2017].

[172] Emmanuel Vieillard Baron. Déterminant d'une matrice, d'une application linéaire. 2001. http://www.les-mathematiques.net/b/e/d/node5.php.

[173] Robert Rolland. Le théorème de Müntz-szász. . http://megamaths.perso.neuf.fr/rr/fichexo_201.pdf.

[174] Wikipédia. Résultant — wikipédia, l'encyclopédie libre, Mai 2013. URL http://fr.wikipedia.org/w/index.php?title=R%C3%A9sultant&oldid=90109864. [En ligne ; Page disponible le 24-mai-2013].

[175] Emmanuel Pedon. Cours de géométrie affine et euclidienne pour la licence de mathématiques. 23 mars 2015. URL http://pedon.perso.math.cnrs.fr/fichiers/enseignement/CoursGeoLicence.pdf.

[176] Jean-Etienne Rombaldi. Polynômes d'endomorphismes en dimension finie. applications. . http://www-fourier.ujf-grenoble.fr/~rombaldi/AgregInterne/Oral1/110.pdf.

[177] Matthieu Romagny. Endomorphismes cycliques. 2012. URL https://perso.univ-rennes1.fr/matthieu.romagny/agreg/dvt/endom_cycliques.pdf.

[178] Callus. URL http://math.stackexchange.com/questions/1829332/prove-that-if-v-is-finite-dimensional-then-v-is-even-dimensional.

[179] La taverne de l'Irlandais. De la réduction des endomorphismes. 2001. URL http://tanopah.jo.free.fr/epilogues/reduction.pdf.

[180] Wikipédia. Trigonalisation — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Trigonalisation&oldid=109085753. [En ligne ; Page disponible le 22-décembre-2014].

[181] Henry C. King. Unitary diagonalization of matrices. http://www-users.math.umd.edu/~hck/Normal.pdf.

[182] David Delaunay. Nilpotence. 10 juillet 2014. URL http://mp.cpgedupuydelome.fr/pdf/Réductiondesendomorphismes-Nilpotence.pdf.

[183] G. Donald Allen. Lectures on linear algebras. 2004. http://www.math.tamu.edu/~dallen/m640_03c/readings.htm.

[184] Wikipedia. Spectral theorem — wikipedia, the free encyclopedia, 2013. URL http://en.wikipedia.org/w/index.php?title=Spectral_theorem&oldid=575488135. [Online ; accessed 21-October-2013].

[185] theorem for normal triangular matrices. URL http://planetmath.org/theoremfornormaltriangularmatrices.

[186] Michel Granget. Normes matricielles, conditionnement. URL http://math.univ-angers.fr/~granger/ananum/Chapitre_II.pdf.

[187] Wikipédia. Forme quadratique — wikipédia, l'encyclopédie libre, 2020. URL http://fr.wikipedia.org/w/index.php?title=Forme_quadratique&oldid=172697192. [En ligne ; Page disponible le 6-juillet-2020].

[188] Wikipédia. Réduction de gauss — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=R%C3%A9duction_de_Gauss&oldid=155723411. [En ligne ; Page disponible le 11-janvier-2019].

[189] Marie-Pierre Lebaud. Formes quadratiques. URL https://perso.univ-rennes1.fr/marie-pierre.lebaud/agint/ecrit/algebre-lineaire/formes-quadratiques/V-formes-quadratiques.pdf.

[190] Wikipédia. Loi d'inertie de sylvester — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=Loi_d%27inertie_de_Sylvester&oldid=156712630. [En ligne ; Page disponible le 13-février-2019].

[191] Emmanuel Vieillard Baron. Formes quadratiques réelles. 2001. URL http://www.les-mathematiques.net/b/c/h/node12.php.

[192] Robert Rollan. Produit tensorile d'espaces vectoriels. URL http://robert.rolland.acrypta.com/telechargements/algebre/tensor.pdf.

[193] Marc Sage. Algèbre multilinéaire. . URL http://www.normalesup.org/~sage/Cours/ProdTens.pdf.

[194] Ministère de l'éducation nationale. Rapport de jury de concours – agrégation de mathématiques, concours externe. 2011. http://agreg.org/Rapports/rapport2011provisoire.pdf.

[195] Hervé Carrieu, Maurice Fadel, Etienne Fieux, Patrice Lassère, and Frédéric Rodriguez. Autour des matrices de Frobenius ou compagnon. 2007. http://www.math.univ-toulouse.fr/~lassere/pdf/vfcomp.pdf.

[196] Grégory Vial. Autour du théorème des invariants de similitude. Octobre 2005. http://w3.bretagne.ens-cachan.fr/math/people/gregory.vial/files/cplts/ivs.pdf.

[197] Arnaud Moncet. Invariants de similitude. http://blogperso.univ-rennes1.fr/arnaud.moncet/public/IVS.pdf.

[198] Commutant d'un endomorphisme. . URL http://minerve.bretagne.ens-cachan.fr/images/Commutant.pdf.

[199] Bernard Alken. Dans un commentaire de « dimension du commutant d'une matrice ». 2009. URL http://www.mathoman.com/index.php/1538-dimension-du-commutant-d-une-matrice.

[200] Wikipédia. Forme linéaire — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Forme_lin%C3%A9aire&oldid=154647982. [En ligne ; Page disponible le 9-décembre-2018].

[201] Wikipédia. Théorème de burnside (problème de 1902) — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Burnside_(probl%C3%A8me_de_1902)&oldid=98510166. [En ligne ; Page disponible le 2-novembre-2015].

[202] Pierre Monmarché. Développements. 18 mai 2011. URL http://laurent.claessens-donadello.eu/pdf/1-total.pdf.

[203] Transformation laissant invariante une forme quadratique. URL http://www.ilemaths.net/forum-sujet-500814.html.

[204] Éric Jourdain. Angles. URL https://perso.univ-rennes1.fr/eric.jourdain/GEEU/Cours/Angles.pdf.

[205] Maxime Pouvreau. Pseudo-réduction simultanée. 2013-2014. URL http://minerve.bretagne.ens-cachan.fr/images/Pseudored.pdf.

[206] Christian Squarcini. Dégénérescence. 2005. URL http://christian-squarcini.pagesperso-orange.fr/AgregInterne/Algebrelineaire/2_2.pdf.

[207] Wikipédia. Norme d'opérateur — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Norme_d%27op%C3%A9rateur&oldid=125963366. [En ligne ; Page disponible le 26-avril-2017].

[208] Wikipedia. Spectral radius — wikipedia, the free encyclopedia, 2017. URL https://en.wikipedia.org/w/index.php?title=Spectral_radius&oldid=758072091. [Online ; accessed 26-April-2017].

[209] Pierre-Emmanuel Jabin. Analyse numérique, correction du TD 8. 2008-2009. URL http://math.unice.fr/~jabin/CTD3-8.pdf.

[210] Christian Squarcini. Applications linéaires. . URL http://christian-squarcini.pagesperso-orange.fr/AgregInterne/Topo/4_3.pdf.

[211] Espaces métriques et espaces normés. URL http://www.les-mathematiques.net/a/a/b/node4.php.

[212] Gallouët !Thierry. Norme et conditionnement d'une matrice. URL https://www.i2m.univ-amu.fr/~gallouet/licence.d/anum.d/anum-tg2.pdf.

[213] Gilles Costantini. Espaces vectoriels de dimension infinie, normes usuelles, équivalence de normes. . URL http://gilles.costantini.pagesperso-orange.fr/agreg_fichiers/evn.pdf.

[214] Gabriel Nagy. Operator theory in Hilbert spaces. URL https://www.math.ksu.edu/~nagy/real-an/2-07-op-th.pdf.

[215] H. Jerome Keisler. Elementaty calculus, an infinitesimal approach. 2010. http://www.math.wisc.edu/~keisler/keislercalc-810.pdf.

[216] Wikipédia. Suite arithmético-géométrique — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Suite_arithm%C3%A9tico-g%C3%A9om%C3%A9trique&oldid=88168841. [En ligne ; Page disponible le 16-juin-2013].

[217] Prime.mover. Sum of sequence of products of consecutive reciprocals. URL https://proofwiki.org/wiki/Sum_of_Sequence_of_Products_of_Consecutive_Reciprocals.

[218] Daniel Daigle. Cardinalité. URL http://mysite.science.uottawa.ca/asavag2/mat2762/cardinal.pdf.

[219] Alan D. Sokal. A really simple elementary proof of the uniform boundedness theorem. 2010. URL https://arxiv.org/pdf/1005.1585.pdf.

[220] Wikipédia. Théorème de banach-steinhaus — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Banach-Steinhaus&oldid=92377069. [En ligne ; Page disponible le 31-août-2013].

[221] Wikipedia contributors. Strong operator topology — Wikipedia, the free encyclopedia, 2018. URL https://en.wikipedia.org/w/index.php?title=Strong_operator_topology&oldid=849989429. [Online ; accessed 16-February-2019].

[222] Wikipedia contributors. Direct sum of modules — Wikipedia, the free encyclopedia, 2018. URL https://en.wikipedia.org/w/index.php?title=Direct_sum_of_modules&oldid=835636392. [Online ; accessed 22-June-2018].

[223] John Douglas Moore. Othogonal complement. 2010. URL http://web.math.ucsb.edu/~moore/orthogonalcomplements.pdf.

[224] Wikipedia contributors. Tensor product — Wikipedia, the free encyclopedia, 2018. URL https://en.wikipedia.org/w/index.php?title=Tensor_product&oldid=846586202. [Online ; accessed 21-June-2018].

[225] Arjeh Cohen. Tensor product. 2004. URL https://www.win.tue.nl/~amc/ow/lba/lba3.pdf.

[226] Jerome E. Marsden and Tudor S. Ratiu. Manifolds, tensor analysis and applications. 2002. URL http://www.springer.com/gp/book/9780387967905.

[227] F. Laudenbach. *Calcul différentiel et intégral*. les Éd. de l'École polytechnique, 2000. ISBN 9782730207249. URL http://books.google.fr/books?id=Ws9A7ZoRJNEC.

[228] Pierre Lairez. Théorème de dérivation d'une limite. Communication personnelle.

[229] Wikipédia. Droite réelle achevée — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Droite_r%C3%A9elle_achev%C3%A9e&oldid=144800302. [En ligne ; Page disponible le 24-janvier-2018].

[230] Ud779. Limits of functions and left hand right hand limit. 2015. URL https://math.stackexchange.com/questions/1418673/limits-of-functions-and-left-hand-right-hand-limit.

[231] Wikipédia. Fonction cauchy-continue — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=Fonction_Cauchy-continue&oldid=157229246. [En ligne ; Page disponible le 3-mars-2019].

[232] Provatiscus. Uniforme continuité utilisee sans justification page 656. 2019. URL https://github.com/LaurentClaessens/mazhe/issues/124.

[233] Wikiversité. Topologie générale/complétude — wikiversité,, 2019. URL https://fr.wikiversity.org/w/index.php?title=Topologie_g%C3%A9n%C3%A9rale/Compl%C3%A9tude&oldid=753513. [En ligne ; accédé le 24-juillet-2019].

[234] Adrien Fontaine. Développement : théorème de Sarkiowski. 8 octobre 2013. URL http://perso.eleves.bretagne.ens-cachan.fr/~afontain/dvptthmdesarkowski.pdf.

[235] Wikipédia. Théorème de heine — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Heine&oldid=145940281. [En ligne ; Page disponible le 28-février-2018].

[236] J.L. Littlewoor. Every polynomyal has a root. 1941. URL http://www.math.univ-toulouse.fr/~bauval/Littlewood-95-8.pdf.

[237] Wikipédia. Théorème fondamental de l'algèbre — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_fondamental_de_l%27alg%C3%A8bre&oldid=139327371. [En ligne ; Page disponible le 20-août-2017].

[238] URL https://www.math.hmc.edu/calculus/tutorials/quotient_rule/proof.pdf.

[239] Wikiversité. Fonctions d'une variable réelle/dérivabilité — wikiversité,, 2018. URL https://fr.wikiversity.org/w/index.php?title=Fonctions_d%27une_variable_r%C3%A9elle/D%C3%A9rivabilit%C3%A9&oldid=718842. [En ligne ; accédé le 24-mai-2018].

[240] Wikiversité. Fonction dérivée/dérivée d'un quotient — wikiversité,, 2017. URL https://fr.wikiversity.org/w/index.php?title=Fonction_d%C3%A9riv%C3%A9e/D%C3%A9riv%C3%A9e_d%27un_quotient&oldid=675129. [En ligne ; accédé le 24-mai-2018].

[241] Wikipédia. Opérations sur les dérivées — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Op%C3%A9rations_sur_les_d%C3%A9riv%C3%A9es&oldid=99099598. [En ligne ; Page disponible le 28-juillet-2014].

[242] Wikiversité. Fonctions d'une variable réelle/dérivabilité — wikiversité,, 2019. URL https://fr.wikiversity.org/w/index.php?title=Fonctions_d%27une_variable_r%C3%A9elle/D%C3%A9rivabilit%C3%A9&oldid=753518. [En ligne ; accédé le 19-juillet-2019].

[243] Wikipédia. Théorème de rolle — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Rolle&oldid=132483699. [En ligne ; Page disponible le 28-décembre-2016].

[244] Guillaume Connan. Dérivation. URL http://gconnan.free.fr/les%20pdf/Deriv.pdf.

[245] Keith Konrad. The remainder in Taylor series. URL http://www.math.uconn.edu/~kconrad/blurbs/analysis/TaylorRemainder.pdf.

[246] Wikipédia. Règle de l'hôpital — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=R%C3%A8gle_de_L%27H%C3%B4pital&oldid=131491717. [En ligne ; Page disponible le 29-décembre-2016].

[247] Wikipedia contributors. Squeeze theorem — Wikipedia, the free encyclopedia, 2019. URL https://en.wikipedia.org/w/index.php?title=Squeeze_theorem&oldid=903141752. [Online ; accessed 20-July-2019].

[248] Livio Flaminio. Éléments de géométrie différentielle. 29 octobre 2009. URL http://www-gat.univ-lille1.fr/~flaminio/M403/2009-2010/cours_geo_top.pdf.

[249] Wikipédia. Application multilinéaire — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=Application_multilin%C3%A9aire&oldid=161449382. [En ligne ; Page disponible le 1-août-2019].

[250] Wikipédia. Convergence uniforme — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Convergence_uniforme&oldid=95493463. [En ligne ; Page disponible le 15-octobre-2013].

[251] Suites et séries de fonctions. . URL http://blog.psi945.fr/public/maths-psi/cours-psi-suites-series-fonctions.pdf.

[252] F. Poupaud. Analyse fonctionnelle pour la licence. URL http://math.unice.fr/~rascle/pdffiles/coursanapp/ana-fonc.pdf.

[253] Michael Gechele. Théorème de stone-weierstrass. http://michael.gechele.perso.neuf.fr/Agregation/Theoreme_Stone_Weierstrass.pdf.

[254] Primitives et intégrales. URL http://math.univ-lyon1.fr/capes/IMG/pdf/new.primitive.pdf.

[255] Nicole Bopp. Un complément à la leçon sur l'équation fonctionnelle de la fonction exponentielle. 2008. URL http://irma.math.unistra.fr/~bopp/CAPES/cours/equation-felle-exp.pdf.

[256] John K. Hunter. Chapter 13 : Metric, normed and topological spaces. URL https://www.math.ucdavis.edu/~hunter/intro_analysis_pdf/ch13.pdf.

[257] G.B. Folland. Taylor's formula. URL https://sites.math.washington.edu/~folland/Math425/taylor.pdf.

[258] Wikipedia. Développement de taylor, 2016. URL http://fr.wikipedia.org/wiki/Développement_de_Taylor.

[259] Wikiversité. Calcul différentiel/théorèmes utiles — wikiversité,, 2017. URL https://fr.wikiversity.org/w/index.php?title=Calcul_diff%C3%A9rentiel/Th%C3%A9or%C3%A8mes_utiles&oldid=657239. [En ligne ; accédé le 10-juillet-2017].

[260] Rached Mneimé and Frédéric Testard. *Groupes de Lie classiques*. Hermann, 1986. ISBN 2-7056-6040-2.

[261] Yoann Gelineau. Générateurs du groupe linéaire. . URL http://math.univ-lyon1.fr/~gelineau/devagreg/Generateurs_Groupe_Lineaire.pdf.

[262] Sandrine Caruso. Générateurs de $\mathrm{GL}_n(K)$ et $\mathrm{SL}_n(K)$. URL http://sandrine.toonywood.org/pageperso/agreg/geneSL.pdf.

[263] Wikipédia. Théorème de cayley-hamilton — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Cayley-Hamilton&oldid=104470625. [En ligne ; Page disponible le 10-juin-2014].

[264] David Monniaux. 9 juin 2014. URL http://david.monniaux.free.fr/dotclear/index.php/post/2014/06/09/Mémoire-à-effaçage-rapide. David Monniaux esquisse la preuve dans un commentaire à cette note.

[265] Pierre Lissy. Décomposition polaire. 30 avril 2010. URL http://www.ljll.math.upmc.fr/~lissy/Agreg/developpements/DecPol.pdf.

[266] Wikipédia. Décomposition polaire — wikipédia, l'encyclopédie libre, avril 2013. URL http://fr.wikipedia.org/w/index.php?title=D%C3%A9composition_polaire&oldid=90240040. [En ligne ; Page disponible le 24-avril-2013].

[267] Jean-Etienne Rombaldi. Endomorphismes symétriques d'un espace vectoriel. applications. . http://www-fourier.ujf-grenoble.fr/~rombaldi/AgregInterne/Oral1/120.pdf.

[268] Brice Loustau. Développements d'algèbre pour l'agrégation. 2008. URL http://myismail.net/docs/divers/agreg/Sortie/LessonsAlgLoustau.pdf.

[269] Michel Coste. Sous-groupes compacts du groupe linéaire. 29 septembre 2006. URL http://agreg-maths.univ-rennes1.fr/documentation/docs/ssgrpecompact.pdf.

[270] Daniel Li. Notions fondamentales de la théorie des probabilités. . http://labomathlens.free.fr/Liens/ProbaM1/PROBA01.pdf.

[271] Daniel Saada. Tribu de borel et tribu de baire d'un espace topologique. 2009. URL http://www.daniel-saada.eu/fichiers/16-Tribus_de_Baire.pdf.

[272] Wikipédia. Tribu-trace — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Tribu-trace&oldid=155238311. [En ligne ; Page disponible le 28-décembre-2018].

[273] Daniel Li. Construction de la mesure de lebesgue. 28 javier 2008. http://www.editions-ellipses.fr/product_info.php?products_id=11183 http://www.editions-ellipses.fr/product_info.php?products_id=9387.

[274] Wikipédia. Mesure (mathématiques) — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Mesure_(math%C3%A9matiques)&oldid=145203435. [En ligne ; Page disponible le 5-février-2018].

[275] ProofWiki. Outer measure of limit of increasing sequence of sets — proofwiki,, 2012. URL http://www.proofwiki.org/w/index.php?title=Outer_Measure_of_Limit_of_Increasing_Sequence_of_Sets&oldid=85952. [Online ; accessed 2-February-2014].

[276] C. Antonini, JF. Quint, P. Borgnat, J. Bérard, E. Lebeau, E. Souche, A. Chateau, and O. Teytaud. Quelques résultats d'unicité. . URL http://www.les-mathematiques.net/a/i/d/node1.php#T42.

[277] Daniel Li. Tribus et mesures. 24 mars 2011. http://www.editions-ellipses.fr/product_info.php?products_id=11183 http://www.editions-ellipses.fr/product_info.php?products_id=9387.

[278] C. Antonini, JF. Quint, P. Borgnat, J. Bérard, E. Lebeau, E. Souche, A. Chateau, and O. Teytaud. Ensembles négligeables et complétion de tribus. . URL http://www.les-mathematiques.net/a/i/c/node1.php.

[279] Wikipédia. Complétion d'une mesure — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Compl%C3%A9tion_d%27une_mesure&oldid=89819650. [En ligne ; Page disponible le 27-mars-2014].

[280] Gerald Teschl. Topics in real and functional analysis. 2013. URL http://www.mat.univie.ac.at/~gerald/ftp/book-fa/fa.pdf.

[281] Emmanuel Vieillard Baron. Les fonctions mesurables. 2001. URL http://www.les-mathematiques.net/a/i/b/node1.php.

[282] Wikipédia. Fonction étagée — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Fonction_%C3%A9tag%C3%A9e&oldid=104836104. [En ligne ; Page disponible le 6-septembre-2015].

[283] B. Mauray. Intégration jusqu'au théorème de Lebesgue. 2006. URL https://webusers.imj-prg.fr/~bernard.maurey/agreg/Textes/Lebesgue.pdf.

[284] John K. Hunter. Measure theory. 2011. URL https://www.math.ucdavis.edu/~hunter/measure_theory/measure_notes.pdf.

[285] Daniel Li. Intégration sur un espace produit. 29 mars 2011. http://www.editions-ellipses.fr/product_info.php?products_id=11183 http://www.editions-ellipses.fr/product_info.php?products_id=9387.

[286] Rémi Peyre and Frédéric Simon. Travaux dirigés de probabilités. 2009. URL http://www.normalesup.org/~rpeyre/pro/enseignement/td09pbas-e.pdf.

[287] Wikipédia. Ensemble de vitali — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Ensemble_de_Vitali&oldid=100925109. [En ligne ; Page disponible le 9-septembre-2015].

[288] christophe c. Re : mesure non nulle et existence d'un ouvert. URL http://www.les-mathematiques.net/phorum/read.php?4,608327,608430#msg-608430.

[289] Daniel Choï. Mathématiques pour la mécanique. 2008. http://www.meca.unicaen.fr/Enseignement/Document/CoursChoi/mathmeca.pdf.

[290] Gilles Dubois. URL http://dubois.gilles.pagesperso-orange.fr/analyse_reelle/intlimites.html.

[291] Prove the countable additivity of lebesgue integral. . URL https://math.stackexchange.com/questions/668958/prove-the-countable-additivity-of-lebesgue-integral.

[292] Daniel Li. Construction de l'intégrale de lebesgue. 10 février 2011. http://www.editions-ellipses.fr/product_info.php?products_id=11183 http://www.editions-ellipses.fr/product_info.php?products_id=9387.

[293] David Madore. Sur la rédaction des maths et la recherche de l'inambiguïé. 24 octobre 2018. URL http://www.madore.org/~david/weblog/d.2018-10-24.2562.html.

[294] Yao Jianfeng. Statistique et logiciels. 2010. http://perso.univ-rennes1.fr/jian-feng.yao/pedago/statlog/.

[295] Daniel Li. Théorème de radon-nikodým et applications. . http://www.editions-ellipses.fr/product_info.php?products_id=11183 http://www.editions-ellipses.fr/product_info.php?products_id=9387.

[296] Bruce Driver. Analysis tools with applications. April 10, 2003. URL http://www.math.ucsd.edu/~bdriver/240-01-02/Lecture_Notes/anal.pdf.

[297] B. Mauray. Préparation à l'agrégation, analyse. 2011. http://www.math.jussieu.fr/~maurey/agreg/index.html.

[298] Thierry Gallouët and Raphaèle Herbin. Mesure, intégration, probabilités. 2012. http://www.cmi.univ-mrs.fr/~gallouet/licence.d/mes-int-pro.pdf.

[299] Bruno Demange. Théorème de changement de variables. 2012. URL https://www-fourier.ujf-grenoble.fr/~demange/integration/2012/integration-chap5.pdf.

[300] Daniel Li. Changement de variables dans les intégrales sur un ouvert de $\mathbb{R}^n$. 8 juin 2010. URL http://li.perso.math.cnrs.fr/textes/Integration/change-v.pdf.

[301] Wikipédia. Intégrale impropre — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Int%C3%A9grale_impropre&oldid=103902526. [En ligne ; Page disponible le 5-août-2014].

[302] Romain Boillaud. Séries de fonctions. 2010-2011. URL http://www.rblld.fr/cours/chap6.pdf.

[303] Gérard Eguether. Ex-séries entières. URL http://iecl.univ-lorraine.fr/~Gerard.Eguether/zARTICLE/EX.pdf.

[304] Georges Skandalis. Analyse, résumés et exercices. 6 mars 2015. URL https://www.math.univ-paris-diderot.fr/_media/formations/prepa/agreginterne/polycopieanalyse.pdf.

[305] Rahul Krishna. Approximations by taylor polynomials. URL http://math.columbia.edu/~krishna/08092011.pdf.

[306] Vogel Thomas. Analytic functions. 2012. URL http://www.math.tamu.edu/~tvogel/410/sect74a.pdf.

[307] Dan Klain. The matrix exponential (with exercices). 2018. URL http://faculty.uml.edu/dklain/exponential.pdf.

[308] ENS Cachan. Théorème de stabilité de Lyapunov. URL http://minerve.bretagne.ens-cachan.fr/images/Lyapunov.pdf.

[309] Rached Mneimé. *Réduction des endomorphismes*. Calvage et Mounet, 2006. ISBN 2-916352-01-5.

[310] Matthieu Romagny. Représentations linéaires des groupes finis. 2010-2011. http://www.math.jussieu.fr/~romagny/agreg/theme/rlgf_paysage.pdf.

[311] Sébastien Pellerin. Développements d'analyse pour l'oral de l'agrégation. . http://pellerin.xyz/doc/agreg/analyse.pdf.

[312] Kevin Quirin. Théorème de Müntz. http://kevin.quirin.free.fr/Trucs/agreg/Dvt_Muntz.pdf.

[313] Pierre Lissy. Déterminant de cauchy et application au théorème de Müntz. 5 mai 2010. http://www.ljll.math.upmc.fr/~lissy/Agreg/developpements/CauchyMuntz.pdf.

[314] Benjamin Dadoun. Théorème de Müntz. URL http://benjamin.dadoun.free.fr/muntz.pdf.

[315] Jean-Louis Rouget. Intégrales dépendant d'une paramètre. 2017. URL http://www.maths-france.fr/MathSpe/Cours/09-integrales-a-parametres.pdf.

[316] Wikipédia. Lemme de hadamard — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Lemme_de_Hadamard&oldid=90659396. [En ligne ; Page disponible le 14-juillet-2014].

[317] Jean-Pierre Demailly. Analyse numérique et équations différentielles. 1991.

[318] Clémence Minazzo and Kelsey Rider. Théorèmes du point fixe et applications aux équations différentielles. 2006-2007. http://math.unice.fr/~eaubry/Enseignement/M1/memoire.pdf.

[319] Ivan Nourdin. *Leçons d'analyse, probabilités algèbre et géométrie*. Dunod, 2001. ISBN 2-10-005668-9.

[320] Sandrine Caruso. Théorème de Cauchy-Lipschitz. 2009. http://boumbo.toonywood.org/sandrine/pageperso/agreg/cauchy-lipschitz.pdf Voir aussi la page toute pleine de développement http://boumbo.toonywood.org/sandrine/pageperso/agreg.html.

[321] Franck Boyer. Équations différentielles ordinaires. 2012. URL http://www.latp.univ-mrs.fr/~fboyer/Enseignement/Agreg/cours_EDO_Agreg_FBoyer.pdf.

[322] Théo Pierron. Théorème de Cauchy-Lipschitz global. 5 janvier 2014. URL http://perso.eleves.ens-rennes.fr/~tpier758/agreg/dvpt/maths/cauchy_lipschitz.pdf.

[323] Thierry Audibert. Le problème de Cauchy, résultats fondamentaux. URL http://www.univenligne.fr/Documents/PolycopiesAgreg/EDO1.pdf.

[324] Thomas Budzinski. Théorème de Cauchy-Lipschitz. 2014. URL https://www.eleves.ens.fr/home/budzinsk/polys/Nonolympique/2014_cauchylipschitz.pdf.

[325] Wikipédia. Théorème d'inversion locale — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_d%27inversion_locale&oldid=100975433. [En ligne ; Page disponible le 19-février-2014].

[326] Antoine Chambert-Loir. Introduction aux groupes et algèbres de Lie. 2004-2005. URL http://www.math.u-psud.fr/~chambert/enseignement/2004-05/lie/lie.pdf.

[327] E.P. van den Ban. Lie groups. 2003. http://www.math.uu.nl/people/ban/lecnot.html.

[328] Joachim Stubbe. URL http://mathaa.epfl.ch/prob/enseignement/analyse2/series/AnalysisE.pdf.

[329] Laurent Guillopé. Optimisation sous contrainte. 2015-2016. URL https://www.math.sciences.univ-nantes.fr/~guillope/l3-osc/osc.pdf.

[330] Bruno Galerne. Optimisation, algorithmique. 2016-2017. URL http://www.math-info.univ-paris5.fr/~bgalerne/m1_opti_algo/poly_opti_algo.pdf.

[331] Wikipédia. Fonction convexe — wikipédia, l'encyclopédie libre, 2019. URL http://fr.wikipedia.org/w/index.php?title=Fonction_convexe&oldid=157832047. [En ligne ; Page disponible le 24-mars-2019].

[332] Wikiversité. Fonctions convexes/définition et premières propriétés — wikiversité,, 2013. URL http://fr.wikiversity.org/w/index.php?title=Fonctions_convexes/D%C3%A9finition_et_premi%C3%A8res_propri%C3%A9t%C3%A9s&oldid=381775. [En ligne ; accédé le 9-février-2014].

[333] Wikiversité. Fonctions convexes/fonctions convexes dérivables — wikiversité,, 2013. URL http://fr.wikiversity.org/w/index.php?title=Fonctions_convexes/Fonctions_convexes_d%C3%A9rivables&oldid=379779. [En ligne ; accédé le 9-février-2014].

[334] Stephen Boyd and Lieven Vandenberghe. Convex optimization. 2004. URL http://www.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf.

[335] Document 33 : fonctions convexes. . URL http://math.univ-lyon1.fr/capes/IMG/pdf/new.convexe.pdf.

[336] Wikipédia. Inégalité arithmético-géométrique — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=In%C3%A9galit%C3%A9_arithm%C3%A9tico-g%C3%A9om%C3%A9trique&oldid=99682278. [En ligne ; Page disponible le 1-juillet-2014].

[337] Wikipédia. Inégalité de kantorovitch — wikipédia, l'encyclopédie libre, juillet 2014. URL http://fr.wikipedia.org/w/index.php?title=In%C3%A9galit%C3%A9_de_Kantorovitch&oldid=102417007. [En ligne ; Page disponible le 1-juillet-2014].

[338] Jean-François Burnol. Normes lp. URL http://jf.burnol.free.fr/agregnormeslp.pdf.

[339] Georges Comte. Lemme de Morse. 2009-2010. URL http://gc83.perso.sfr.fr/Agregation/LemmedeMorse.pdf.

[340] Wikipédia. Calcul du volume de l'hypersphère — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Calcul_du_volume_de_l%27hypersph%C3%A8re&oldid=147651804. [En ligne ; Page disponible le 18-avril-2018].

[341] Florian Bouguet. Ellipsoïde de John-Lowner. . URL http://florian.bouguet.free.fr/doc/developpements/john-loewner.pdf.

[342] Martin Henk. Löewnr-john ellipsoids. URL https://www.math.uni-bielefeld.de/documenta/vol-ismp/24_henk-martin.pdf.

[343] Shuzhong Zhang. Supplement : the löwner-john ellipsoids. URL http://www.isye.umn.edu/courses/ie8534/pdf/Loewner-John_ellipsoid.pdf.

[344] B. Mauray. Analyse fonctionnelle et théorie spectrale (version longue). 2001–2002. http://www.math.jussieu.fr/~maurey/ts012/poly/lths.pdf.

[345] Michel Rascle. Analyse fonctionnelle pour la licence. http://math.unice.fr/~rascle/pdffiles/coursanapp/ana-fonc.pdf.

[346] Mourad Besbes. Chapitre 5 : espaces métriques complets, espaces de banach. http://besbes.mourad.free.fr/Enseignement/Topologie/Chapitre5.pdf.

[347] C. Antonini, JF. Quint, P. Borgnat, J. Bérard, E. Lebeau, E. Souche, A. Chateau, and O. Teytaud. Complété d'un espace métrique. . URL http://www.les-mathematiques.net/a/a/b/node26.php.

[348] Wikipédia. Espace complet — Wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Espace_complet&oldid=91125353. [En ligne ; Page disponible le 1-avril-2013].

[349] B. Maurey. Holomorphie. 7 février 2005. http://www.math.jussieu.fr/~maurey/agreg/Textes/Holomorphe045.pdf.

[350] Romain Boillaud. Séries entières. http://www.rblld.fr/cours/chap8.pdf.

[351] Philippe Picart. Définition de cos et sin par les séries entières. URL http://trucsmaths.free.fr/telech/exp_cos_sin.pdf.

[352] Alfred Gray. Modern differential geometry of curves and surfaces with mathematica. 1998. URL http://webmath2.unito.it/paginepersonali/sergio.console/CurveSuperfici/.

[353] Wikipédia. Pi — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Pi&oldid=133197907. [En ligne ; Page disponible le 3-janvier-2017].

[354] David Cimasoni. Cours de géométrie I, semestre de printemps. Printemps 2014. URL http://www.unige.ch/math/folks/cimasoni/GeometrieI.pdf.

[355] Sylvain. Générateurs de $\simeq (e)$. URL https://agreg-maths.fr/developpements/147.

[356] Christophe Mourougane. Théorie des groupes et géométrie. 2008-2009. http://perso.univ-rennes1.fr/christophe.mourougane/enseignements/2008-9/THGG/poly.groupes.pdf.

[357] Groupes diédraux. http://theoriedesgroupes.perso.sfr.fr/cours/Diedraux.pdf.

[358] Pierre Lissy. 6 avril 2010. URL http://www.ljll.math.upmc.fr/~lissy/Agreg/developpements/Died.pdf.

[359] Pierre Renfer. Roulettes et colliers. http://les.mathematiques.free.fr/pdf/collier.pdf.

[360] Daniel Perrin. Isométries du plan. . URL https://www.math.u-psud.fr/~perrin/CAPES/geometrie/isometries11-12.pdf.

[361] Wikipédia. Angle — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Angle&oldid=130370952. [En ligne ; Page disponible le 15-novembre-2016].

[362] B. Guesmi. Théorème de l'angle inscrit. URL http://s999cdfd874dd6835.jimcontent.com/download/version/1427398317/module/5509391568/name/theoremedel'arccapable.pdf1.pdf.

[363] Nicole Bopp. Un complément à la leçon sur les isométries d'un polygone régulier. Octobre 2008. URL http://irma.math.unistra.fr/~bopp/CAPES/cours/groupe-fini-isometrie.pdf.

[364] Théo Pierron. Groupes de pavage du plan. 4 juin 2014. URL http://perso.eleves.ens-rennes.fr/~tpier758/agreg/dvpt/useless/gpe_pavage.pdf.

[365] Frédéric Hélein. Géométrie euclidienne. 2005-2006. URL https://webusers.imj-prg.fr/~frederic.helein/cours/euclide.pdf.

[366] Laurent Dietrich. Les sous-groupes finis de so(3). 16 mai 2011. URL http://www.math.cmu.edu/~ldietric/doc/ssgrso3.pdf.

[367] Sous groupes finis de so(3,r). 2013-2014. URL http://minerve.bretagne.ens-cachan.fr/images/Sous-groupes_finis_SO3.pdf.

[368] Marc Sage. Sons, fréquences, harmoniques, tons : le compromis du piano. . http://www.normalesup.org/~sage/Musique/Harmonie.pdf.

[369] Wikipédia. Chiffrement rsa — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=Chiffrement_RSA&oldid=146774926. [En ligne ; Page disponible le 25-mars-2018].

[370] URL https://www.torproject.org/about/torusers.html.en.

[371] URL https://www.torproject.org/projects/torbrowser.html.en.

[372] Wikipédia. Polynôme cyclotomique — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Polyn%C3%B4me_cyclotomique&oldid=93751495. [En ligne ; Page disponible le 25-juin-2013].

[373] Adam Parusinski. Théorème de Dirichlet (version faible). 2010-2011. http://math.unice.fr/~parus/AGREG/Dev/dirichlet.pdf.

[374] Michel Merle. corps finis. 2005-2006. http://www-math.unice.fr/~merle/Complements/corps_finis.pdf.

[375] Gabriel Peyré. Corps finis. 2007. https://objectifagregation.github.io/.

[376] François Rodier. Corps finis. http://iml.univ-mrs.fr/~rodier/Cours/RappelCorpsfinis.pdf.

[377] Stephane Vento. La loi de réciprocité quadratique. 2005. http://www.proba.jussieu.fr/pageperso/nourdin/LeSiteDeLAgregatif/vento2.pdf.

[378] Wikipedia contributors. Chevalley–warning theorem — Wikipedia, the free encyclopedia, 2017. URL https://en.wikipedia.org/w/index.php?title=Chevalley%E2%80%93Warning_theorem&oldid=792119700. [Online ; accessed 21-May-2018].

[379] Christian Squarcini. Anneaux factoriels. . URL http://christian-squarcini.pagesperso-orange.fr/AgregInterne/Anneauxcorps/3_2.pdf.

[380] Wikipédia. Fonction de möbius — wikipédia, l'encyclopédie libre, mai 2013. URL http://fr.wikipedia.org/w/index.php?title=Fonction_de_M%C3%B6bius&oldid=91236321. [En ligne ; Page disponible le 1-mai-2013].

[381] Wikipédia. Corps fini — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Corps_fini&oldid=92489466. [En ligne ; Page disponible le 1-mai-2013].

[382] Wikipédia. Nombre constructible — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Nombre_constructible&oldid=103130793. [En ligne ; Page disponible le 13-mai-2014].

[383] Tracés usuels « à la règle et au compas ». 2012-2013. URL http://dokeos3.u-cergy.fr/dokeos20/courses/EEMEM2CRPE/scorm/2012_2013SCORMEEMEUE3EC2M2P1_SCORM/page_04.htm.

[384] Wikipédia. Théorème de wantzel — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Wantzel&oldid=103220511. [En ligne ; Page disponible le 13-mai-2014].

[385] Wikipédia. Nombre de fermat — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Nombre_de_Fermat&oldid=103080861. [En ligne ; Page disponible le 18-juin-2014].

[386] Philippe Spindel. Éléments de géométrie différentielle pour la mécanique analytique et le gravitation. Novembre 2010.
https://portail.umons.ac.be/EN2/universite/facultes/fs/services/institut_physique/mecanique_et_gravitation/Documents/Textespédagogiques/geodiff.pdf.

[387] William K Allard. The Brouwer fixed point theorem. 2004.
http://www.math.duke.edu/~wka/math204/fixed.pdf
see also http://www.math.duke.edu/~wka/math204/.

[388] Kenneth Kuttler. Topics in analysis. Janvier 2012.
http://www.math.byu.edu/~klkuttle/SobolevSpacesB.pdf
Juste un indice : ce document fait 2086 pages . . . j'y arriverais, courage !

[389] Hervé Le Dret. Notes de cours M2 Équations aux dérivées partielles élliptiques. 4 mars 2010. URL https://www.ljll.math.upmc.fr/~ledret/M2Elliptique/chapitre4.pdf.

[390] Bachir Bekka. Théorème du point fixe de Markov-Kakutani : applications à l'existence de mesures de haar et aux chaînes de Markov. 2005.
http://agreg-maths.univ-rennes1.fr/documentation/docs/PointFixe.pdf.

[391] Olivier Casterá. Différentielles totales exactes.
http://o.castera.free.fr/pdf/Differentielle_totale_exacte.pdf.

[392] Intégration de fractions rationnelles. URL http://blascheck.franck.free.fr/IMG/pdf/IntegFracRat.pdf.

[393] Ophélie Rouby. Théorème de Rothstein-Trager. URL http://perso.eleves.bretagne.ens-cachan.fr/~oroub842/dvlpts/Rothstein_Trager.pdf.

[394] Frédéric Chyzak. Intégration symbolique des fractions rationnelles. Décembre 2004. URL http://www.enseignement.polytechnique.fr/informatique/profs/Frederic.Chyzak/CF-2004/notes-IntRat.pdf.

[395] Sylvain Golénia. Approximation de ln(2). Novembre 2018. URL http://mathaoutils.blogspot.com/2018/11/approximation-de-ln2.html.

[396] Giuseppe De Marco. *Analisi due*. 1999.

[397] Louis Fauchier-Magnan. Introduction à la théorie de Morse. URL http://cqfd.epfl.ch/webdav/site/cqfd/shared/projets/igat/IntroductionàlathéoriedeMorse-LouisFauchier-Magnan.pdf.

[398] Reduction formula for integral of power of sine. . URL https://proofwiki.org/wiki/Reduction_Formula_for_Integral_of_Power_of_Sine.

[399] Michel Stainer. Compléments sur les séries numériques. URL http://michel.stainer.pagesperso-orange.fr/PSIx/Cours/Ch03-ComplSeriesNum.pdf.

[400] Wallis's product. . URL https://proofwiki.org/wiki/Wallis%27s_Product.

[401] Jacek Cichoń. Stirling approximation formula. URL http://cs.pwr.edu.pl/cichon/Math/StirlingApp.pdf.

[402] Wikipédia. Formule de stirling — wikipédia, l'encyclopédie libre, Mai 2013. URL http://fr.wikipedia.org/w/index.php?title=Formule_de_Stirling&oldid=93224272. [En ligne ; Page disponible le 20-mai-2013].

[403] Jean-Louis Rouget. Formule de stirling. 2008. URL https://www.maths-france.fr/MathSpe/GrandsClassiquesDeConcours/SeriesNumeriques/FormuleDeStirling.pdf.

[404] Stirling's formula. . URL https://proofwiki.org/wiki/Stirling%27s_Formula.

[405] Keith Conrad. Differentiating under the integral sign. . URL https://kconrad.math.uconn.edu/blurbs/analysis/diffunderint.pdf.

[406] gb. Cercle circonscrit à une courbe. URL http://www.les-mathematiques.net/phorum/read.php?8,750556,750586.

[407] Mohammad Ghomi. Lecture note on differential geometry. 2007. URL http://people.math.gatech.edu/~ghomi/LectureNotes/LectureNotes5U.pdf.

[408] Wikipedia. Convex curve — wikipedia, the free encyclopedia, 2015. URL https://en.wikipedia.org/w/index.php?title=Convex_curve&oldid=685739052. [Online ; accessed 13-April-2016].

[409] Wikipédia. Théorème de jordan — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Jordan&oldid=124990679. [En ligne ; Page disponible le 6-avril-2016].

[410] Wikipédia. Théorème de jordan — wikipédia, l'encyclopédie libre, Mai 2013. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Jordan&oldid=92525205. [En ligne ; Page disponible le 25-mai-2013].

[411] Les quadrilatères. URL http://ddata.over-blog.com/xxxyyy/4/39/08/01/cinquieme/5eme.parallelogramme.jeu-set-et-maths.mathematiques.pdf.

[412] Robert Rolland. Géométrie projective. . http://robert.rolland.acrypta.com/telechargements/geometrie/pr0.pdf.

[413] Daniel Bertrand. Géométrie projective. 2009-2010. http://www.math.jussieu.fr/~bertrand/Enseignement/epdf/Algeo09_chap2.pdf.

[414] Nicolas Jacon. Géométrie projective. URL http://njacon.perso.math.cnrs.fr/jacon_Geometrieproj.pdf.

[415] François Labourie. Géométrie affine et projective. 9 février 2019. URL http://math.unice.fr/~labourie/preprints/pdf/geomproj.pdf.

[416] Matthieu Romagny. Droites projectives, homographies. 2005-2006. URL https://perso.univ-rennes1.fr/matthieu.romagny/agreg/droite_projective_Halberstadt.pdf.

[417] Arnaud Bodin. L'inversion. 2012. URL http://math.univ-lille1.fr/~bodin/geometrie/ch_inversion.pdf.

[418] Laurent Dietrich. Le groupe circulaire. 21 mai 2011. URL http://www.math.cmu.edu/~ldietric/doc/groupecirc.pdf.

[419] Florian Bouget. Le groupe circulaire. URL http://florian.bouguet.free.fr/doc/developpements/Groupe_circulaire.pdf.

[420] Florian Bouguet. Le groupe circulaire. . URL http://minerve.bretagne.ens-cachan.fr/images/Groupe_circulaire.pdf.

[421] 182 : utilisation des nombres complexes en géométrie. Homographies. 2015-2016. URL http://minerve.bretagne.ens-cachan.fr/images/182_2015-2016.pdf.

[422] Sébastien Pellerin. Action du groupe modulaire sur le demi-plan de Poincaréé. . http://dynamaths.free.fr/docs/lecons/developpement_algebre_6.pdf.

[423] Alexandre Alessandri. Action du groupe modulaire sur le demi-plan de Poincaré. 2011-2012. http://minerve.bretagne.ens-cachan.fr/images/Dvt_psl2.pdf.

[424] Aline Kurtzmann. Espérance conditionnelle. http://www.iecn.u-nancy.fr/~kurtzman/cours/agreg/esperance-conditionnelle.pdf.

[425] Wikipédia. Théorème de projection sur un convexe fermé — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_projection_sur_un_convexe_ferm%C3%A9&oldid=90246160. [En ligne ; Page disponible le 3-mai-2013].

[426] Michel Merle. Espaces de Hilbert. 2004-2005. URL http://www-math.unice.fr/~merle/Algebre_et_geometrie/hilbert.pdf.

[427] Douglas N. Arnold. Functional analysis. 1997. URL http://www.math.umn.edu/~arnold/502.s97/functional.pdf.

[428] Wikipédia. Base de hilbert. . URL https://fr.wikipedia.org/wiki/Base_de_Hilbert?oldid=128533376.

[429] Christiane Schomblond. *Théorie quantique des champs, QED, QCD*. Notes de cours, en français. http://homepages.ulb.ac.be/~cschomb/intfonc+QED+QCD.pdf.

[430] Remsirems. Re : Réponses à quelques questions. 2016. URL http://linuxfr.org/nodes/110155/comments/1675813.

[431] Jonathan Bain. Kochen-Specker and measurement. 2011. http://faculty.poly.edu/~jbain/philqm/lectureslides/08.KS&Measurement.pdf.

[432] Michel Kern. Introduction à la méthode des éléments finie. 2004-2005. URL http://mms2.ensmp.fr/ef_paris/formulation/polycop/f_coursEF.pdf.

[433] Antoine Bensalah. Communication électronique privée. 2017.

[434] Pascal Dingoyan. Chapitre 5. URL http://www.math.jussieu.fr/~dingoyan/lescours/lm367-pageweb/lm367-chapitre5.pdf.

[435] Wikipédia. Anneau principal — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Anneau_principal&oldid=141994549. [En ligne ; Page disponible le 27-octobre-2017].

[436] Corentin. Re : formule de cauchy et dérivation. URL http://www.les-mathematiques.net/phorum/read.php?4,385272,385289.

[437] Philippe Charpentier. Analyse complexe. Septembre 2010. http://www.math.u-bordeaux1.fr/~pcharpen/enseignement/fichiers-master1/Analyse_Complexe.pdf.

[438] Christian Kkein. Séries de laurent, résidus. URL http://math.u-bourgogne.fr/IMB/klein/Complement_dAnalyse_files/chap607.pdf.

[439] Wikipédia. Intégrale de fresnel — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Int%C3%A9grale_de_Fresnel&oldid=132170602. [En ligne ; Page disponible le 26-novembre-2016].

[440] David Sirajuddin. Fresnel integrals. 2008. http://itcanbeshown.com/integrals/FresnelIntegrals/fresnel_integrals.pdf.

[441] Noach Dana-Picard. The logarithm of a complex number. URL http://ndp.jct.ac.il/tutorials/complex/node26.html.

[442] Guillaume Carlier. Analyse complexe. 2012-2013. URL https://www.ceremade.dauphine.fr/~carlier/analysecomplexe.pdf.

[443] Wikiversité. Fonctions d'une variable complexe/le logarithme complexe — wikiversité,, 2016. URL https://fr.wikiversity.org/w/index.php?title=Fonctions_d%27une_variable_complexe/Le_logarithme_complexe&oldid=617412. [En ligne ; accédé le 16-décembre-2016].

[444] Jean-Marie Lion. Fonctions holomorphes d'une variable. 5 mai 2004. http://perso.univ-rennes1.fr/jean-marie.lion/coursholo.pdf.

[445] Wikipédia. Théorème d'Ascoli — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_d%27Ascoli&oldid=92958118. [En ligne ; Page disponible le 11-mai-2013].

[446] Keith Conrad. l^p spaces for $0 < p < 1$. . URL https://www.math.uconn.edu/~kconrad/blurbs/analysis/lpspace.pdf.

[447] B. Maurey. 9 mars 2011. http://www.math.jussieu.fr/~maurey/IntegProba/Cours/CoursIP101-14.pdf.

[448] Danarmk. 2016. URL http://linuxfr.org/nodes/110155/comments/1675589.

[449] B. Mauray. Convolution, inégalités, approximations et régularisation. 2007. https://webusers.imj-prg.fr/~bernard.maurey/agreg/Textes/Convolution.pdf https://webusers.imj-prg.fr/~bernard.maurey/agreg/ag045/Convolution045.pdf.

[450] Schilling. Minkowsky's inequality for integrals. URL http://math.ucsd.edu/~lni/math240/suppl-1.pdf.

[451] Charles Suquet. Intégration, analyse de fourier,probabilités. 2003-2004. http://math.univ-lille1.fr/~suquet/ens/IFP/Cours/cours04/CoursIFP04.html.

[452] Wikipédia. Théorème de Riesz-Fischer — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Riesz-Fischer&oldid=90074115. [En ligne ; Page disponible le 3-mai-2013].

[453] JanTinbergen1991. Lp -space is a hilbert space if and only if p=2. Novembre 2018. URL https://math.stackexchange.com/questions/3017814/lp-space-is-a-hilbert-space-if-and-only-if-p-2.

[454] Thierry Gallouët. Les espaces lp. URL https://www.i2m.univ-amu.fr/perso/thierry.gallouet/licence.d/int/part4.pdf.

[455] Gilles Leborgne. Introduction à la théorie des distributions. 16 mai 2013. URL http://www.isima.fr/~leborgne/IsimathDistributions/distrib.pdf.

[456] Hart Smith. Lecture 2 : convolution. 2013. URL https://sites.math.washington.edu/~hart/m526/Lecture2.pdf.

[457] R. Fortet. Remarques sur les espaces uniformément convexes. 1941. URL http://www.numdam.org/article/BSMF_1941__69__23_0.pdf.

[458] Kai-Seng Chou and Tianwen Luo. Chapter 4 : The lebesgue spaces. 2014. URL https://www.math.cuhk.edu.hk/course_builder/1415/math5011/MATH5011_Chapter_4.2014.pdf.

[459] Olof Hanner. On the uniform convexity of L^p and l^p. 1955. URL https://projecteuclid.org/download/pdf_1/euclid.afm.1485893271.pdf.

[460] Assaf Naor. Proof of the uniform convexity lemma. February 2004. URL https://web.math.princeton.edu/~naor/homepagefiles/inequality.pdf.

[461] David Giraudo. 2011. URL https://math.stackexchange.com/questions/95307/showing-left-fracab2-rightp-left-fraca-b2-rightp-leq-frac12/95312#95312.

[462] gerw. How to prove clarkson's inequality ? 2017. URL https://math.stackexchange.com/questions/1607683/how-to-prove-clarksons-inequality.

[463] M .A Usman, F.A. Hammed, and M.O Olayiwola. Uniformly convex spaces. 2009. URL http://www.sciencepub.net/researcher/0101/10_0506_UNIFORMLY_CONVEX.pdf.

[464] Jean-François Burnol. Le théorème de dualité des espaces l^p. 2006. URL http://jf.burnol.free.fr/0506L312annexeDualiteLp.pdf.

[465] Wikipédia. Espace Lp — wikipédia, l'encyclopédie libre, 2013. URL http://fr.wikipedia.org/w/index.php?title=Espace_Lp&oldid=92618226. [En ligne ; Page disponible le 10-mai-2013].

[466] Didier Smets. Base d'analyse fonctionnelle. 2011-2012. URL https://www.ljll.math.upmc.fr/smets/MM005/MM005_Chapitre_6.pdf.

[467] Haïm Brézis. *Analyse fonctionnelle*. Masson, Paris, 1983. ISBN 2-225-77198-7. Théorie et applications.

[468] MoebiusCorzer Giraudo, Davide. Show smooth functions of compact support are dense in the Schwartz space. 2012. URL http://math.stackexchange.com/questions/229751/show-smooth-functions-of-compact-support-are-dense-in-the-schwartz-space.

[469] Richard Gomez. Séries de Fourier. 10 octobre 2010. URL http://megamaths.perso.neuf.fr/rg/fourier.pdf.

[470] David Delaunay. Séries de Fourier. 8 juin 2012. http://mp.cpgedupuydelome.fr/pdf/SériesdeFourier.pdf.

[471] B. Maurey. Notes de cours : analyse hilbertienne et de Fourier. 2008-2009. http://www.math.jussieu.fr/~maurey/FourHilb/PolyFH.pdf.

[472] Richard Zekri. Analyse de fourier. 15 septembre 2009. URL http://lumimath.univ-mrs.fr/infoetudiant/ANALYSEDEFOURIERMASTER2009.pdf.

[473] Michel Crouzeix. Compléments d'analyse. URL https://perso.univ-rennes1.fr/michel.crouzeix/publis/anep.pdf.

[474] Bachir Bekka. Le théorème d'inversion de Fourier du point de vue des distributions. URL http://agreg-maths.univ-rennes1.fr/documentation/docs/Theoreme-Inversion-Fourier.pdf.

[475] Erdős László. Fourier transform. 2008. URL https://www.mathematik.uni-muenchen.de/~lerdos/WS08/QM/four08.pdf.

[476] Isabelle Gallagher. Chapitre b. distributions. URL https://www.math.ens.fr/~gallagher/ChapitreB2017-2018.pdf.

[477] Daniel Li. Notions sur les distributions. . URL http://labomathlens.free.fr/Liens/AF/distrib.pdf.

[478] Claude Zuily. *Éléments de distributions et d'équations aux dérivées partielles.* Dunod, 2002.

[479] Sabin Lessard. Cours de processus stochastiques. 2013. URL http://www.dms.umontreal.ca/~lessards/ProcessusStochastiquesLessard2014.pdf.

[480] Peter Young. Singular Fourier transforms and the integral representation of the Dirac delta function. November, 2009.

[481] Daniel Choï. Méthode des éléments-finis par l'exemple. Avril 2010. URL http://meca.unicaen.fr/~choi/pdf/cours-mef.pdf.

[482] Michael Reiter and Arthur Schuste. Fourier transform and Sobolev spaces. 2008. URL http://www.mat.univie.ac.at/~stein/lehre/SoSem08/sobolev_fourier.pdf.

[483] Norbert Heuer. Ltcc module computation methods. URL http://www.ltcc.ac.uk/courses/AppliedComputationalMethods/ACM2007Chap1.pdf\etautresfichiersdelamêmesérie.

[484] Arkhnor. Densité des fonctions Cinfty à support compact. 2014. URL http://www.ilemaths.net/sujet-densite-des-fonctions-c-infty-a-support-compact-587937.html.

[485] Gilles Leborgne. Complément : espaces de Sobolev fractionnaires. 2007. URL http://ws.isima.fr/~leborgne/IsimathDistributions/sobolev_frac.pdf.

[486] Richard B. Melrose. Lecture notes for 18.155, fall 2002. 2002. URL https://ocw.mit.edu/courses/mathematics/18-155-differential-analysis-fall-2004/lecture-notes/section10.pdf.

[487] Julien Vovelle. Équations différentielles, cours no 2. URL http://math.univ-lyon1.fr/~vovelle/2Cours.pdf.

[488] A. Munnier. Théorie des équations différentielles ordinaires. 2006-2007. URL http://www.iecn.u-nancy.fr/~munnier/files/cours_edo.pdf.

[489] Olivier Debarre. Équations différentielles. URL http://www.math.ens.fr/~debarre/EquaDiff.pdf.

[490] Sylvie Benzoni-Gavage. Propagation d'ondes. URL http://math.univ-lyon1.fr/~benzoni/Ondes.pdf.

[491] Sonia Fliss. L'équation de transport à coefficients variables. URL http://perso.ensta-paristech.fr/~fliss/Sonia_Fliss_web_page/Enseignement_files/Amphi3_2014.pdf.

[492] Sylvie Benzoni-Gavage. Équations différentielles ordinaires. 11 mai 2007. URL http://math.univ-lyon1.fr/~benzoni/EDO-M1.pdf.

[493] Wikipédia. Flot (mathématiques) — wikipédia, l'encyclopédie libre, 2017. URL http://fr.wikipedia.org/w/index.php?title=Flot_(math%C3%A9matiques)&oldid=134130419. [En ligne ; Page disponible le 1-février-2017].

[494] Alexander Grigor'yan. Ordinary differential equation. 2009. URL https://www.math.uni-bielefeld.de/~grigor/odelec2009.pdf.

[495] Christopher P. Grant. Theory of ordinary differential equation. 2007. URL http://www.math.pitt.edu/~bard/bardware/classes/2920/Grant_4july2007.pdf.

[496] John B. Picard-lindelöf theorem with parameter : what regularity ? 2017. URL http://math.stackexchange.com/questions/2126539/picard-lindelöf-theorem-with-parameter-what-regularity.

[497] Kevin Quirin. Développements. 2011-2012. URL http://kevin.quirin.free.fr/Trucs/agreg/developpements.pdf.

[498] Wikipédia. Équations de lotka-volterra — wikipédia, l'encyclopédie libre, 2015. URL http://fr.wikipedia.org/w/index.php?title=%C3%89quations_de_Lotka-Volterra&oldid=113756354. [En ligne ; Page disponible le 5-juillet-2015].

[499] Predrag Cvitanović. Partial differential equations. 2012. URL http://www.cns.gatech.edu/~predrag/courses/PHYS-6124-12/StGoChap6.pdf.

[500] Ralph Chill. Introduction aux équations aux dérivées partielles. 2010-2011. URL http://www.math.univ-metz.fr/~chill/edp.pdf.

[501] Gianluca Bontempi. Un cours d'analyse numérique et de matlab. 2009. http://www.ulb.ac.be/di/map/gbonte/calcul/ https://web.archive.org/web/20090203224115/http://www.ulb.ac.be/di/map/gbonte/calcul/.

[502] Wikipedia. Two's complement — wikipedia, the free encyclopedia, 2016. URL https://en.wikipedia.org/w/index.php?title=Two%27s_complement&oldid=706335033. [Online ; accessed 6-April-2016].

[503] Wikipédia. Ieee 754 — wikipédia, l'encyclopédie libre, 2015. URL http://fr.wikipedia.org/w/index.php?title=IEEE_754&oldid=120234746. [En ligne ; Page disponible le 12-avril-2016].

[504] Comprendre les nombres à virgules flottantes. URL http://blog.netinfluence.ch/2009/09/24/comprendre-les-nombres-a-virgule-flottante/.

[505] Harald Schmidt. IEEE754 converter. URL http://www.h-schmidt.net/FloatConverter/IEEE754.html.

[506] David Goldberg. What every computer scientist should know about floating-point arithmetic. 1991. URL http://docs.oracle.com/cd/E19957-01/806-3568/ncg_goldberg.html#11655.

[507] Emmanuel Frénod. Calcul scientifique. 2011-2012. URL http://web.univ-ubs.fr/lmam/frenod/IMG/DocEtudiant/MTH1504/calculscientifique.pdf.

[508] Franck Boyer. Agrégation externe de mathématiques. Analyse numérique. 14 octobre 2014. URL http://www.math.univ-toulouse.fr/~fboyer/_media/enseignements/agreg/cours_an_agreg_fboyer_2014.pdf.

[509] Antoine Chambert-Loir. Autour de la méthode de newton. http://perso.univ-rennes1.fr/antoine.chambert-loir/2005-06/agreg/newton.pdf.

[510] Wikipédia. Vitesse de convergence — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=Vitesse_de_convergence&oldid=110252913. [En ligne ; Page disponible le 22-mai-2016].

[511] Martin Pinto Campos. Approximations polynomiales de fonctions et et des intégrales. 21 décembre 2006. D'après le cours de Laurence Halpern http://documents.lamacs.fr/cours/macs1/cours-macs1-pinto1.pdf.

[512] Ming Yang. Matrix decomposition. URL http://users.eecs.northwestern.edu/~mya671/files/Matrix_YM_.pdf.

[513] Benjamin Ambrosio. Chapitre 3 : Méthodes directes de résolution du système $Ax = b$. URL http://lmah.univ-lehavre.fr/~ambrosio/CoursMN/Chapitre3.pdf.

[514] Emmanuel Frénod. Méthodes directes de résolution des systèmes linéaires. URL http://web.univ-ubs.fr/lmam/frenod/IMG/DocEtudiant/MTH1504/methodes-directes.pdf.

[515] Christian Lécot. Analyse numérique matricielle. 2009-2010. URL https://www.lama.univ-savoie.fr/~lecot/data/M1_ANM.pdf.

[516] Wikipédia. Matrice à diagonale dominante — wikipédia, l'encyclopédie libre, 2016. URL http://fr.wikipedia.org/w/index.php?title=Matrice_%C3%A0_diagonale_dominante&oldid=132268176. [En ligne ; Page disponible le 4-mai-2017].

[517] Habib Joulak. Méthodes des différences finies en élasticité. 1974. URL https://ori-nuxeo.univ-lille1.fr/nuxeo/site/esupversions/0e05f3d2-fe12-4d67-b74f-f622869539f2.

[518] Philippe Briand. Probabilités de base. 2005-2006. http://www.lama.univ-savoie.fr/~briand/proba/g12_cours.pdf.

[519] Vincent Bansaye. Variables aléatoires, espérance, indépendance. http://www.cmap.polytechnique.fr/~bansaye/CoursTD2.pdf.

[520] Alfio Marazzi. There are 22 chapters in 22 files. http://www.iumsp.ch/Unites/us/Alfio/polybiostat/ch06.pdf.

[521] Pierre-Loïc Méliot. Notion de conditionnement. URL https://www.math.u-psud.fr/~meliot/agreg/condition.pdf.

[522] Cédric Boutillier. 2016. URL https://github.com/LaurentClaessens/mazhe/issues/16#issue-180279120.

[523] Automaths. Paradoxe des deux enfants – episode 2 ! 25 août 2018. URL https://automaths.blog/2018/08/25/paradoxe-des-deux-enfants-episode-2/.

[524] Wikipédia. Inégalité de jensen — wikipédia, l'encyclopédie libre, août 2013. URL http://fr.wikipedia.org/w/index.php?title=In%C3%A9galit%C3%A9_de_Jensen&oldid=90235295. [En ligne ; Page disponible le 9-août-2013].

[525] Gérard Letac. Calcul des probabilités, Deug 2ième année. juin 2001. http://les.mathematiques.free.fr/pdf/proba.zip.

[526] Daniel Saada. Note sur le théorème de transfert pour les variables aléatoires réelles. 18 octobre 2012. http://www.les-mathematiques.net/phorum/file.php?12,file=17435,filename=th_du_tranfert.pdf.

[527] Christophe Sabot. Différents types de convergence de suites de v.a. URL http://math.univ-lyon1.fr/~sabot/Convergence.pdf.

[528] Antonin Macé and Thomas Pradeau. Exposé de maîtrise. Estimations de densités de probabilité. 2007. Encadré par Gilles Stoltz. http://www.eleves.ens.fr/home/amace/travaux/exposemaitrise.pdf.

[529] Wikipedia. Continuous mapping theorem — wikipedia, the free encyclopedia, 2013. URL http://en.wikipedia.org/w/index.php?title=Continuous_mapping_theorem&oldid=545523937. [Online ; accessed 15-June-2013].

[530] Wikipédia. Inégalité de markov — wikipédia, l'encyclopédie libre, 2014. URL http://fr.wikipedia.org/w/index.php?title=In%C3%A9galit%C3%A9_de_Markov&oldid=100208179. [En ligne ; Page disponible le 11-février-2014].

[531] Charles Suquet. Simulation. 2005-2006. http://math.univ-lille1.fr/~suquet/ens/Agr/simul06.pdf.

[532] Yoann Gelineau. Vecteurs gaussiens. . http://math.univ-lyon1.fr/~gelineau/files/vecteurs_gaussiens.pdf.

[533] Wikipédia. Nombre normal — wikipédia, l'encyclopédie libre, avril 2013. URL http://fr.wikipedia.org/w/index.php?title=Nombre_normal&oldid=89848922. [En ligne ; Page disponible le 28-avril-2013].

[534] Wikipédia. Théorème de glivenko-cantelli — wikipédia, l'encyclopédie libre, . URL http://fr.wikipedia.org/w/index.php?title=Th%C3%A9or%C3%A8me_de_Glivenko-Cantelli&oldid=90841113. [En ligne ; Page disponible le 15-juin-2013].

[535] Vincent Rivoirard and Gilles Stoltz. Estimation de densité de probabilités. URL http://www.math.ens.fr/statenaction/PDF/Densite-Principal.pdf.

[536] Nils Berglund. Chaînes de Markov. Décembre 2007. URL http://www.univ-orleans.fr/mapmo/membres/berglund/markov.pdf.

[537] Arnaud Guyader. Espérance conditionnelle et chaines de Markov. 200-2008. http://www.sites.univ-rennes2.fr/laboratoire-statistique/AGUYADER/doc/proba/poly.pdf https://perso.univ-rennes2.fr/system/files/users/jegou_n/EsperanceConditionnelle.pdf.

[538] Massimiliano Gubinelli. Comportement asymptotique des martingales. 2011-2012. http://www.ceremade.dauphine.fr/~mgubi/e1112/pd3.pdf.

[539] Jean Lacroix and Pierre Priouret. Probabilités approfondies. 2005-2006. http://www.proba.jussieu.fr/cours/processus.pdf.

[540] Wikipedia. Optional stopping theorem — wikipedia, the free encyclopedia, 2013. URL http://en.wikipedia.org/w/index.php?title=Optional_stopping_theorem&oldid=547547174. [Online ; accessed 14-August-2013].

[541] Nizar Touzi. Martingales en temps discret et chaine de Markov. Septembre 2009. URL http://www.cmap.polytechnique.fr/~touzi/MAP432-Poly.pdf.

[542] Wikipédia. Étoile de kleene — wikipédia, l'encyclopédie libre, 2018. URL http://fr.wikipedia.org/w/index.php?title=%C3%89toile_de_Kleene&oldid=149113546. [En ligne ; Page disponible le 1-juin-2018].

[543] Zach Weinersmith. 2008. URL http://www.smbc-comics.com/comic/2008-02-21.