

Intégration et Analyse de Fourier

Cours de première année

donné à l'Ecole normale supérieure de Lyon

année universitaire 2005-2006

Cédric Villani

Unité de Mathématiques Pures et Appliquées

École normale supérieure de Lyon

46 allée d'Italie

69364 Lyon Cedex 07

`cvillani@umpa.ens-lyon.fr`

Table des matières

Avant-Propos	5
Bibliographie	9
Notations et conventions	13
Introduction : Motivations de l'intégrale de Lebesgue	15
CHAPITRE I. Théorie abstraite de la mesure	19
I-1. Espaces mesurés	19
I-2. Rappels de topologie	29
I-3. Régularité des espaces mesurés	40
I-4. Prolongement de mesures	47
I-5. Complétion de mesures	59
I-6. Application : Construction de la mesure de Lebesgue	60
I-7. Complément : Recouvrement et remplissage	61
CHAPITRE II. Intégration selon Lebesgue et selon Riesz	67
II-1. Fonctions mesurables	67
II-2. Construction de l'intégrale	74
II-3. L'intégrale est une forme linéaire positive	80
II-4. Le théorème de Riesz	83
II-5. Intégration à valeurs vectorielles	91
CHAPITRE III. Théorèmes fondamentaux d'intégration	95
III-1. Comportement face aux limites	95
III-2. Intégration sur les espaces produits	110
III-3. Changement de variable	121
III-4. Inégalités intégrales élémentaires	124
III-5. Équi-intégrabilité et tension	133
III-6. Produits infinis	137
Appendice : Rappels sur les fonctions convexes	148
CHAPITRE IV. La mesure de Lebesgue	151
IV-1. Construction de la mesure de Lebesgue	151
IV-2. Propriétés fondamentales de la mesure de Lebesgue	154
IV-3. L'intégrale de Lebesgue généralise l'intégrale de Riemann	161
IV-4. Règles de calcul associées à l'intégrale de Lebesgue	164
IV-5. Mesurabilité, non-mesurabilité, et paradoxes de Banach–Tarski	170
CHAPITRE V. Les mesures de Hausdorff	179
V-1. Motivations	179
V-2. Construction des mesures de Hausdorff	182

V-3.	Identification des mesures de Hausdorff	188
V-4.	Dimension	195
V-5.	Utilisation dans les changements de variables	201
CHAPITRE VI.	Espaces de Lebesgue et mesures signées	203
VI-1.	Espaces L^p de Lebesgue	203
VI-2.	Inégalités et relations entre espaces de Lebesgue	212
VI-3.	Espace des fonctions mesurables	224
VI-4.	Espaces de mesures	228
CHAPITRE VII.	Analyse de Banach - À COMPLÉTER	243
VII-1.	La complétude et ses conséquences	243
VII-2.	Régularité des espaces de Banach	254
VII-3.	Projection	270
VII-4.	Dualité	274
VII-5.	Convergence et compacité	286
VII-6.	Intégration au sens faible	314
VII-7.	Convexité	316
VII-8.	Retour sur les espaces de Hilbert	327
CHAPITRE VIII.	Analyse sur les groupes abéliens - À RÉDIGER	331
CHAPITRE IX.	Fonction maximale de Hardy–Littlewood - À RÉDIGER	333
CHAPITRE X.	Désintégration	335
X-1.	Théorème de Lebesgue–Radon–Nikodym	335
X-2.	Désintégration	339
X-3.	Désintégration constructive	351
X-4.	Dérivation des fonctions absolument continues	351
X-5.	Décomposition des mesures d’une infinité de variables	351
CHAPITRE XI.	Analyse de Fourier – À COMPLÉTER	353
XI-1.	Théorie de Fourier	353
XI-2.	Fourier calculateur : deux lois fondamentales des probabilités	361
XI-3.	Fourier physicien : l’Âge de la Terre	369
XI-4.	Fourier analyste : régularité	373

Avant-Propos

Ce cours constitue une introduction à deux outils fondamentaux de l'analyse : d'une part, la théorie de la mesure et de l'intégration, telle qu'elle a été développée au vingtième siècle à partir des idées de Borel et Lebesgue ; d'autre part, l'analyse de Fourier. Ces deux sujets sont très liés, ne serait-ce que pour des raisons historiques.

En rédigeant ces notes, j'ai adopté les principes suivants.

Ce cours n'est pas un ouvrage de référence. La bibliographie est volontairement réduite à un petit nombre d'ouvrages — ceux que j'utilise moi-même régulièrement.

Pour autant, les perspectives historiques ne seront pas complètement négligées, et seront souvent brièvement évoquées au moment de l'introduction des nouveaux concepts. J'ai également tâché de motiver l'introduction de l'intégrale de Lebesgue, et expliqué en détail pourquoi elle généralise celle de Riemann. Outre l'intérêt historique de cette question, il m'a semblé qu'elle était importante pour les étudiants, qui ont presque tous commencé la théorie de l'intégration par l'intégrale de Riemann ; et pour les applications pratiques, puisque c'est presque toujours l'intégrale de Riemann ou ses variantes que l'on utilise pour effectuer des calculs numériques d'intégrales.

Les liens entre théorie de la mesure et théorie des probabilités d'une part, logique axiomatique d'autre part, seront esquissés, mais ces sujets ne seront pas développés. En particulier, ce cours ne contient pas d'introduction à la théorie des probabilités. Dans mon opinion, une telle introduction devrait se concentrer sur le développement de l'intuition probabiliste, au détriment de l'outil technique que constitue la théorie de la mesure ; en corollaire, il m'a semblé naturel d'incorporer à ce cours les principaux résultats de théorie de la mesure qui sont utiles en probabilités (y compris des résultats assez subtils tels que les théorèmes d'existence de Kolmogorov et de Ionescu Tulcea, ou la loi du 0-1 de Hewitt et Savage).

Le choix du cadre général dans lequel développer la théorie de la mesure abstraite n'est pas anodin. Traditionnellement, les ouvrages à tendance probabiliste (comme ceux de Billingsley) insistent sur la théorie des probabilités dans les espaces polonais (métriques séparables complets), tandis que ceux qui sont plus centrés sur l'analyse fonctionnelle, tout en recherchant une grande généralité, préfèrent le cadre localement compact, non nécessairement métrique (c'est le cas des ouvrages de Rudin, Halmos, Bourbaki, etc.). Ce dernier point de vue paraît aujourd'hui difficile à soutenir, étant donné le très faible nombre de personnes intéressées à faire de l'intégration dans des espaces localement compacts non Polonais, en regard de la quantité prodigieuse de personnes (dont beaucoup de non-mathématiciens) qui utilisent la théorie de la mesure et de l'intégration dans des espaces polonais non localement compacts tels que l'espace de Wiener.

C'est donc le point de vue des espaces polonais qui est ici développé prioritairement, et en particulier toutes les preuves seront effectuées dans un cadre métrique. J'ai cependant conservé les hypothèses de compacité locale, non nécessairement métrique, dans les énoncés où elles semblent naturelles, en particulier le théorème de Riesz. Les rappels nécessaires de topologie générale (lemme d'Urysohn, théorème de Tychonov) sont donc énoncés sans démonstration dans leur version générale, et démontrés dans le cadre métrique. J'ai tenté par là de satisfaire à la fois les lecteurs novices qui auront ainsi accès à des démonstrations complètes dans un cadre métrique séparable, sans avoir jamais à manipuler de topologie abstraite ; et les lecteurs amateurs de topologie, qui pourront facilement reconstituer les preuves d'énoncés plus abstraits.

On présente assez souvent la complétion de la tribu borélienne comme une opération agréable et peu coûteuse. Il ne me semble pas clair que cette opération permette de gagner beaucoup, sinon quelques mots de justification supplémentaires ici ou là ; en revanche je suis convaincu qu'elle apporte des complications inutiles dans l'étude de certaines questions naturelles, par exemple sur des sujets tels que le réarrangement de fonctions. Il est significatif que dans le récent ouvrage d'analyse de Lieb et Loss, seule la tribu borélienne soit utilisée, et ce, sans que les démonstrations en souffrent aucunement. J'ai donc présenté au lecteur la complétion comme une opération dont il faut se méfier – pour ma part, je ne l'utilise pas.

L'étude des espaces de Lebesgue est l'occasion de développer une introduction élémentaire mais assez complète à l'analyse fonctionnelle (chapitre VII). En écrivant cette partie, par souci de rendre les preuves plus intuitives et plus constructives, j'ai soigneusement évité tout recours à des structures topologiques abstraites (on y considère uniquement des espaces de Banach), et j'ai systématiquement utilisé des hypothèses de séparabilité ou de dénombrabilité, quand le besoin s'en fait sentir, pour éviter l'axiome du choix. C'est ainsi que le théorème de Hahn-Banach n'est démontré que dans le cadre "restreint" des espaces vectoriels normés séparables. On pourra voir dans ce choix une régression : par endroits, la présentation dans ce chapitre a plus de points communs avec le traitement originel de Banach, dans les années 1930, qu'avec la théorie des espaces vectoriels topologiques développée dans les années 1950 et 1960.

De même, pour ce qui est de l'analyse harmonique sur des groupes localement compacts, j'ai choisi de présenter des démonstrations complètes des énoncés les plus importants (Théorème de Haar, dualité de Pontryagin, etc.) mais uniquement dans le cas de groupes qui, en plus d'être localement compacts, sont métrisables et σ -compacts. Ce choix diminue bien sûr la généralité des énoncés démontrés, mais il est suffisant pour traiter la quasi-totalité des groupes que l'on rencontre en pratique, avec des preuves parfois nettement simplifiées.

En ce qui concerne le plan des notes, voici les quelques autres points qui me paraissent notables.

Les questions de régularité des mesures sont abordées dès le premier chapitre, et plus généralement diverses propriétés faisant intervenir topologie et théorie de la mesure en même temps. Des théorèmes d'extension à la Carathéodory sont présentés dans la foulée. J'ai pris soin d'énoncer une version du théorème de Carathéodory qui soit suffisamment générale pour pouvoir être utilisée dans le théorème d'existence de

la mesure produit, dans celui de l'existence de la mesure de Lebesgue, mais aussi dans le théorème de représentation de Riesz (dont la démonstration "classique" reprend certains éléments de la démonstration du théorème de Carathéodory).

La linéarité de l'intégrale de Lebesgue est d'habitude établie comme corollaire du théorème de convergence monotone ; cette approche est économique, mais a l'inconvénient pédagogique de commencer à traiter des propriétés de l'intégrale par passage à la limite, avant de parler de la propriété plus fondamentale (au cahier des charges de toute notion d'intégrale !) de linéarité. J'ai donc dans un premier temps établi la linéarité par un argument qui copie la preuve du théorème de convergence monotone, et juste après j'ai fait le lien avec l'approche de Riesz, basée sur les formes linéaires. La discussion du théorème de convergence monotone est reprise par la suite, dans un chapitre indépendant consacré aux propriétés de l'intégrale.

Une place importante a été accordée au théorème d'Egorov, qui en pratique se révèle souvent plus maniable que le théorème de convergence dominée. Pour faire sentir la puissance de ce théorème, j'ai indiqué comment on peut l'utiliser pour démontrer le théorème de convergence dominée. En fait on pourrait choisir le théorème d'Egorov comme point de départ de la théorie des passages à la limite ; mais il est sans doute plus naturel de réserver ce rôle au théorème de convergence monotone, pour son analogie formelle avec la propriété d'additivité dénombrable.

L'étude des tribus et mesures produits (dans le cadre d'un produit fini ou infini) aurait pu être exposée dès le début du cours, puisque le concept d'intégrale n'est pas, strictement parlant, nécessaire à leur introduction. Cependant, j'ai suivi l'usage qui consiste à ne pas séparer cette étude du théorème de Fubini ; en outre la construction est facilitée par le concept de fonction mesurable. Pour compenser ce défaut de plan, j'ai annoncé dès le premier chapitre le résultat d'existence de mesure produit, après le théorème de Carathéodory, et j'en ai donné une démonstration dans un cas particulier. J'ai également annoncé le théorème d'existence de Kolmogorov à cet endroit.

L'introduction de la mesure de Lebesgue est l'occasion d'une discussion assez précise sur la mesurabilité et la non-mesurabilité, en rapport avec les paradoxes de Banach-Tarski. Le lecteur y est encouragé à se méfier de l'axiome du choix, ou même à ne pas l'utiliser du tout. *Toute l'analyse classique peut se construire sans la forme forte de l'axiome du choix.*

Les mesures de Hausdorff sont absentes de la plupart des traités introductifs (à l'exception notable de celui de Billingsley) ; ce sujet est d'habitude réservé à un public plus averti. En raison de l'importance de ce concept dans de nombreuses branches des mathématiques, et de la popularité du concept de mesure fractale ou de dimension fractale, il m'a semblé qu'on ne devrait pas s'en passer, même dans un ouvrage d'introduction.

L'uniforme convexité des espaces de Lebesgue L^p ($p > 1$) est démontrée à partir des inégalités de Hanner plutôt que des inégalités de Clarkson ; on obtient ainsi des estimations essentiellement optimales du module de convexité. On propose au passage une démonstration très simple des inégalités de Hanner, basée sur une application directe de l'inégalité de Jensen.

Dans le chapitre d'analyse fonctionnelle, j'ai inclus des énoncés de compacité dont je sais par expérience qu'ils sont très précieux dans certains problèmes d'analyse, mais dont la démonstration est souvent difficile à trouver dans les ouvrages de

référence : en particulier, le théorème de compacité de Dunford-Pettis, et celui de Schur.

Le Chapitre ?? traite aussi bien du Théorème de Radon–Nikodym (bien représenté dans les ouvrages de référence) que du problème plus général de la désintégration de la mesure, d’ordinaire réservé aux ouvrages probabilistes un peu avancés. J’ai adopté des conventions selon lesquelles la représentation de Radon–Nikodym apparaît vraiment comme un cas particulier de la notion de désintégration ; je fais cependant le lien avec d’autres notions plus générales de désintégration. Ce chapitre continue avec des énoncés de reconstruction de la densité de Radon–Nikodym dans des espaces localement compacts ; j’y présente aussi bien le théorème classique de densité de Lebesgue (reconstruction ponctuelle) que des énoncés légèrement moins précis mais plus simples de reconstruction L^1 .

Le Chapitre XI adopte un style différent des précédents : moins axé sur la rigueur et la généralité, il tente de donner une intuition des applications de l’analyse de Fourier dans des problèmes divers issus de différents champs scientifiques.

Je remercie ceux qui m’ont aidé à la rédaction de ces notes par des commentaires, rectifications et aides ponctuelles, en particulier Luigi Ambrosio, Guillaume Aubrun, Haïm Brézis, Nassif Ghoussoub, Étienne Ghys, François Japiot, Sébastien Martineau, Quentin Mérigot, Jean-Claude Sikorav.

Bibliographie

Les textes suivants sont particulièrement recommandés pour approfondir ce cours.

Cours concis :

Le cours de l'Ecole Polytechnique de **J.-M. Bony**, *Intégration et analyse hilbertienne* (édition 2001), constitue un excellent exposé, concis et clair.

Le cours d'**A. Gramain**, *Intégration* (Hermann, Paris, 1988) est agréable à lire et contient également des rappels sur l'intégrale de Riemann.

Cours plus détaillés sur l'intégration :

Le grand classique est sans conteste l'ouvrage de **W. Rudin**, *Real and Complex Analysis* (McGrawHill, New York, 1987, 3e édition), qui est également une référence précieuse pour l'analyse complexe.

Un exposé clair et précis de la théorie de l'intégration et de méthodes d'"analyse réelle" en général se trouve dans les premiers chapitres du livre de **G. Folland**, *Real analysis*, (Wiley, New York, 1999, 2e édition).

L'ouvrage récent de **E. Lieb et M. Loss**, *Analysis* (American Mathematical Society, Providence, 2001, 2e édition) est clair, pédagogique et original ; ses chapitres 1 à 5 constituent une ouverture vivement recommandée.

Une introduction à la théorie de la mesure peut aussi se trouver dans de nombreux ouvrages de théorie des probabilités, tels celui de **P. Billingsley**, *Probability and Measure* (Wiley, New York, 1979) ; et dans des ouvrages traitant de sujets liés à la théorie de la mesure géométrique. Parmi ces derniers, on recommande la belle synthèse exposée au début de l'ouvrage de recherche de **L. Ambrosio, N. Fusco et D. Pallara**, *Functions of bounded variation and free discontinuity problems* (Oxford University Press, Oxford, 2000).

Compléments sur l'analyse de Fourier :

Deux ouvrages classiques proposent de nombreux développements de l'analyse de Fourier, dans des domaines très divers des mathématiques :

Fourier series and integrals, de **H. Dym et H.P. McKean** (Academic Press, Boston, 1972), particulièrement intéressant pour les liens avec la théorie des groupes et l'analyse complexe ;

Fourier analysis, de **T.W. Körner** (Cambridge University Press, Cambridge, 1993, 5e édition), un livre merveilleux qui se lit comme un roman ou presque.

Perspectives historiques :

L'ouvrage passionnant de **J.-P. Kahane et P.G. Lemarié-Rieusset**, *Séries de Fourier et ondelettes* (Nouvelle Bibliothèque Mathématique, Cassini, Paris, 1998), joue un double rôle en relation avec ce cours. D'une part, la contribution de Kahane est une remarquable synthèse de l'histoire des séries de Fourier, depuis leur création jusqu'à la théorie moderne, en relation avec les progrès de l'intégration. On y assistera en passant à la naissance de la théorie des ensembles, développée par Cantor pour résoudre des problèmes de régularité de séries entières ! D'autre part, la contribution de Lemarié-Rieusset présentera l'histoire des ondelettes et leur théorie moderne.

L'ouvrage de référence en ce qui concerne la naissance et le développement de l'intégrale de Lebesgue, et des théories concurrentes, est le magistral traité de **T. Hawkins**, *Lebesgue's theory of integration* (Chelsea Publishing Company, The Bronx, 1970).

Le Cours de **H. Lebesgue** lui-même à l'Académie des Sciences en 1903-1904, *Leçons sur l'intégration* (deuxième édition Gauthiers-Villars, Paris, 1950) est toujours intéressant à parcourir ! Sa Note aux Comptes Rendus de l'Académie des Sciences, *Sur une généralisation de l'intégrale définie*, datée du 29 avril 1901, peut être considérée comme l'acte de naissance de la théorie moderne de l'intégration.

Autres ouvrages de référence :

Un ouvrage méthodique et concis sur l'intégration dans \mathbb{R}^n et l'étude fine des fonctions (dérivabilité, etc.) est le traité de **L.C. Evans et R. Gariepy**, *Measure theory and fine properties of functions*, CRC Press, Boca Raton, 1992) ; cet ouvrage contient presque tous les outils de théorie de la mesure dont se servent les analystes qui travaillent dans \mathbb{R}^n .

L'ouvrage passionnant de **S. Wagon**, *The Banach-Tarski Paradox* (Cambridge University Press, 1993, 3e édition) fait le point sur les paradoxes et les interrogations axiomatiques liés à la mesurabilité et à la non-mesurabilité, du type de celui de Banach-Tarski.

Les ouvrages de **K. Falconer**, *The Geometry of fractal sets* (Cambridge University Press, 1985) et *Fractal Geometry* (Wiley, 1989) constituent une excellente référence sur les mesures et la dimension de Hausdorff (les résultats principaux en sont repris dans le livre d'Evans et Gariepy déjà mentionné) et leurs applications à l'étude des objets fractals. Le deuxième de ces ouvrages, écrit pour un public non-spécialiste, contient de nombreuses images de courbes fractales et une excellente synthèse sur leur apparition dans divers domaines des mathématiques, de la physique et de la modélisation. Sur ce sujet on pourra également consulter l'ouvrage célèbre de **B. Mandelbrot**, *Les objets fractals* (...).

Un ouvrage très complet, à consulter uniquement en cas de recherche sur un sujet précis : *Real Analysis and Probability*, de **R.M. Dudley** (Cambridge University Press, 2002, 2e édition). On y trouve également de nombreux commentaires historiques et bibliographiques. Un autre livre de référence sur les liens entre théorie de la mesure, analyse fonctionnelle et topologie est la somme de **N. Dunford et J.T. Schwartz**, *Linear Operators* (Interscience Publishers Inc., New York, 1958).

Le livre de **J.C. Oxtoby**, *Measure and Category (A survey of the analogies between topological and measure spaces)* (Springer-Verlag, New York, 1980, 2e édition)

étudie en profondeur le parallèle entre la théorie de la mesure et la théorie de la catégorie de Baire; son propos est illustré par de nombreux exemples et contre-exemples amusants, quelque peu académiques cependant.

La “bible” de la théorie géométrique de la mesure est l’ouvrage de **H. Federer**, *Geometric Measure Theory* (Grundlehren der mathematischen Wissenschaft 153, Springer, Berlin, 1969). Cet ouvrage comprend un exposé de la théorie abstraite de la mesure, et une étude assez systématique des mesures de Hausdorff. Son format imposant et son caractère exhaustif en font une référence peu commode, à ne consulter qu’en cas de besoin. Le livre d’Evans et Gariepy, déjà mentionné, contient un “reader’s digest” de cet ouvrage.

Le traité de référence classique sur les fonctions convexes dans \mathbb{R}^n est l’ouvrage de **R.T. Rockafellar**, *Convex Analysis* (Princeton University, 1997; reproduction de l’édition de 1970).

Pour en savoir plus sur la théorie des probabilités dans des espaces métriques (en particulier Polonais), on pourra consulter, outre l’ouvrage de Dudley déjà mentionné, le livre de **P. Billingsley**, *Weak Convergence of Measures : Applications in Probability* (Society for Industrial and Applied Mathematics, Philadelphia, 1971), ou celui de **K.R. Parthasarathy**, *Probability Measures on Metric Spaces* (Academic Press, New York, 1967). Il est également recommandé de parcourir la synthèse claire et concise que l’on trouve dans l’Appendice D de l’ouvrage de **A. Dembo et O. Zeitouni**, *Large Deviations techniques and applications* (Springer, New York, 1998, 2e édition); l’Appendice A y fait également des rappels de topologie.

Les bases de l’analyse dans les espaces de Banach se trouvent dans de nombreux ouvrages populaires et pédagogiques, par exemple celui de **H. Brézis**, *Analyse fonctionnelle, Théorie et applications* (Troisième Tirage, Masson, Paris, 1992), ou celui de **W. Rudin** déjà cité (un autre ouvrage du même auteur, *Functional Analysis* (McGraw-Hill, New York, 1991, 2e édition) n’est pas recommandé au lecteur non averti, du fait de son caractère général et abstrait). On pourra regretter l’importance accordée par ces auteurs à l’axiome du choix et à ses avatars. Citons également le gros traité de référence de **N. Dunford et J.T. Schwartz**, *Linear Operators* (Interscience Publishers Inc., New York, 1958), et le livre d’**Edwards**, *Functional Analysis, Theory and Applications* (Holt, Rinehart and Winston, New York, 1965), qui présente une mine d’informations sans tomber dans l’abstraction à outrance.

Les propriétés topologiques de l’espace des mesures “de Radon” sont évoquées dans de nombreux ouvrages traitant de la théorie des distributions, ne serait-ce que l’ouvrage fondateur de **L. Schwartz**, *Théorie des distributions* (Hermann, Paris, 1966).

La théorie des mesures de Haar est l’objet du livre de **L. Nachbin**, *The Haar Integral* (University Series in Higher Mathematics, Van Nostrand, Princeton, 1965), autocontenu et agréable. Pour des exposés plus complets sur l’analyse harmonique dans les groupes localement compacts (mesure de Haar, convolution, transformée de Fourier, théorie des représentations, etc.), on pourra consulter **L.H. Loomis**, *An introduction to abstract harmonic analysis* (Van Nostrand, New York, 1953), **E. Hewitt et K.A. Ross**, *Abstract Harmonic Analysis* (Springer-Verlag, Berlin, 1963) et surtout l’ouvrage plus récent de **G.B. Folland**, *A course in abstract harmonic analysis* (CRC Press, Boca Raton, 1995), synthétique et précis. Tous ces ouvrages

utilisent des hypothèses très générales, qui rendent l'exposé quelque peu abstrait et les démonstrations parfois délicates.

Mon traitement de la désintégration de la mesure est inspiré de celui du traité de Dudley, déjà mentionné ; on pourra trouver une approche différente dans **Dellacherie–Meyer** (.....) ou **Stroock–Varadhan** (Section 1.1) (....). Pour le problème de la reconstruction dans \mathbb{R}^n , l'ouvrage d'Evans–Gariepy déjà mentionné va plus loin puisqu'il traite les mesures non doublantes par le lemme de recouvrement de Besicovich (cette approche est propre à l'espace \mathbb{R}^n).

Le livre d'Evans–Gariepy pourra également être consultée pour les théorèmes plus avancés de différentiation presque partout (théorème de Rademacher, théorème d'Alexandrov) ; je recommande cependant le traitement appliqué dans mon propre ouvrage *Optimal transport, old and new* (Théorèmes 10.8 et Premier Appendice du Chapitre 14), où je me suis efforcé de présenter les différents résultats d'une manière aussi cohérente que possible.

Dans le cours du texte, je ferai systématiquement référence à ces textes plutôt qu'aux ouvrages et articles originaux. Je ferai une exception pour l'article suivant, qui contient de nombreuses remarques et références que l'on ne rencontre pas sous forme de livre : **K. Ball, E. Carlen et E. Lieb**, *Sharp uniform convexity and smoothness inequalities for trace norms*, *Invent. Math.* 115 (1994), 463–482 (Sections I et II).

Je désignerai toutes ces références par le nom de leur auteur, suivi le cas échéant d'un numéro : par exemple [Falconer1], [Falconer2] pour désigner les deux ouvrages de Falconer mentionnés plus haut (la numérotation suit l'ordre d'apparition dans le texte ci-dessus).

Notations et conventions

Outre des notations très classiques, j'utiliserai les conventions suivantes :

Logique et axiomatique :

$A \setminus B$: complémentaire de B dans A

$\mathcal{P}(X)$: ensemble des parties de X

1_A : fonction indicatrice de A ; $1_A(x) = 1$ si $x \in A$, 0 si $x \notin A$

Ensembles :

\mathbb{N} : l'ensemble des nombres entiers naturels *non nuls* : $\mathbb{N} = \{1, 2, 3, \dots\}$

\mathbb{Z} : l'ensemble des entiers positifs ou négatifs

\mathbb{R} : l'ensemble des nombres réels

\mathbb{C} : l'ensemble des nombres complexes

ensemble dénombrable = ensemble fini ou en bijection avec \mathbb{N}

Calculs dans \mathbb{R}^n :

$\langle x, y \rangle$: produit scalaire de x et y

$|x|$: norme euclidienne du vecteur x (valeur absolue si $n = 1$)

Fonctions :

f_+ : partie positive de f , i.e. $\max(f, 0)$

f_- : partie négative de f , i.e. $\max(-f, 0)$

Topologie :

\overline{A} : fermeture topologique de A

$\text{Int}(A)$: intérieur topologique de A

$B_r(x) = B(x, r)$: boule ouverte de centre x , de rayon r

$B_{r]}(x) = B(x, r]$: boule fermée de centre x , de rayon r

$d(x, A)$: distance de x à A , i.e. $\inf\{d(x, y); y \in A\}$

Notations de théorie de la mesure :

μ^* : mesure extérieure associée à la fonction d'ensembles μ

$\sigma(\mathcal{F})$: tribu engendrée par \mathcal{F}

$f_{\#}\mathcal{A}$: tribu image de \mathcal{A} par f

$f_{\#}\mu$: mesure image de μ par f

$\mathcal{A} \otimes \mathcal{B}$: tribu produit des tribus \mathcal{A} et \mathcal{B}

$\mu \otimes \nu$: mesure produit (tensoriel) des mesures μ et ν

$C(A_1, \dots, A_N)$: cylindre de base A_1, \dots, A_N

Si μ est une mesure définie sur \mathbb{R} , j'abrègerai souvent $\mu[[a, b]]$ en $\mu[a, b]$, $\mu[[a, b[$ en $\mu[a, b[$, etc.

Convexité :

Φ^* : transformée de Legendre de la fonction Φ

p' : exposant conjugué de p : $p' = p/(p - 1)$

Espaces fonctionnels :

$C(X, \mathbb{R}) = C(X)$: espace des fonctions continues de X dans \mathbb{R}

$C_b(X, \mathbb{R}) = C_b(X)$: espace des fonctions continues bornées

$C_0(X, \mathbb{R}) = C_0(X)$: espace des fonctions continues tendant vers 0 à l'infini

$C_c(X, \mathbb{R}) = C_c(X)$: espace des fonctions continues à support compact

$L^p(X, d\mu)$: espace de Lebesgue des fonctions p -sommables sur (X, μ)

$\|f\|_\infty$: supremum essentiel de $|f|$

Calcul différentiel :

∇T : matrice Jacobienne de $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (gradient de T si $m = 1$)

$\nabla^2 f$: matrice Hessienne de f

Δf : Laplacien de f (trace de $\nabla^2 f$)

“le lecteur” = “le lecteur ou la lectrice”

Introduction : Motivations de l'intégrale de Lebesgue

À la fin du dix-neuvième siècle, les limitations de la théorie d'intégration de Riemann deviennent apparentes et plusieurs mathématiciens célèbres (Jordan, Borel, Young, ...) se mettent en devoir de la généraliser. C'est finalement la théorie de **Lebesgue**, exposée dans une note fondatrice de 1901, puis développée dans le Cours Peccot, qui sera adoptée par l'ensemble de la communauté mathématique. Elle se développe à partir du concept de **mesure**, introduit par **Borel** vers 1895.

La théorie de la mesure et l'intégration de Lebesgue seront ensuite perfectionnées et généralisées par de nombreux mathématiciens au cours du vingtième siècle, en particulier (par ordre chronologique approximatif) Carathéodory, Vitali, Radon, Riesz, Hausdorff, Kolmogorov et Besicovich. L'histoire de la théorie de la mesure est associée au développement de la théorie des probabilités, à celui de l'analyse harmonique moderne et même à celui de la logique axiomatique.

Avant d'entrer dans le vif du sujet, je vais passer en revue certaines des caractéristiques les plus notables de l'intégrale de Lebesgue, par comparaison avec d'autres théories de l'intégration.

Élargissement de la classe des fonctions intégrables

Le cadre classique le plus simple pour définir une intégrale est celui des fonctions en escalier sur un intervalle $[a, b]$, ou sa complétion pour la topologie de la convergence uniforme, l'espace des fonctions réglées (admettant une limite finie à droite et à gauche). La théorie de Riemann permet d'atteindre une plus grande généralité. Cependant, Riemann lui-même a conscience que l'intégrabilité "au sens de Riemann" impose encore des conditions relativement fortes : il démontre qu'une fonction $f : [a, b] \rightarrow \mathbb{R}$ est intégrable si et seulement si, pour tout $\alpha > 0$ donné, on peut choisir une décomposition de $[a, b]$ en sous-intervalles suffisamment fins pour que la somme des longueurs des sous-intervalles sur lesquels l'oscillation de la fonction dépasse α soit arbitrairement petite [Kahane, p. 64].

Plus tard, Lebesgue montrera qu'une fonction $f : [a, b] \rightarrow \mathbb{R}$ est Riemann-intégrable si et seulement si l'ensemble de ses points de discontinuité est *de mesure nulle*, au sens où on peut l'inclure dans une union d'intervalles ouverts dont la somme des longueurs est arbitrairement petite.

Ces conditions peuvent sembler assez faibles, puisqu'elles autorisent par exemple une fonction qui ne serait discontinue qu'en une quantité dénombrable de points. Mais il est facile de construire des fonctions bornées "naturelles" ne remplissant pas ces conditions : le contre-exemple qui vient tout de suite à l'esprit est la fonction indicatrice de \mathbb{Q} , ou sa restriction à un segment. Dans de nombreux problèmes d'analyse, on rencontre des fonctions qui ne sont pas forcément Riemann-intégrables.

Dans la théorie de Lebesgue, la classe des fonctions intégrables est beaucoup plus grande. Par exemple, toute fonction bornée "raisonnable" (en gros, que l'on peut

décrire par un énoncé mathématique) est Lebesgue-intégrable. En outre, sa théorie généralise bien celle de Riemann.

Solution du problème des primitives

C'est par ce problème que Lebesgue motive sa construction dans sa note de 1901. L'intégrale de Riemann permet d'intégrer des fonctions discontinues, mais ne permet pas d'intégrer n'importe quelle fonction dérivée, même bornée ! Si donc f est une fonction continue sur $[a, b]$ et dérivable sur $]a, b[$, il n'est pas garanti que l'identité

$$(1) \quad f(b) - f(a) = \int_a^b f'(t) dt$$

ait un sens. En fait, au début du siècle, divers auteurs (Volterra, Köpcke, Brodén, Schoenflies) construisent des classes de fonctions dérivables, dont la dérivée est bornée mais non Riemann-intégrable [Hawkins, pp. 57 et 108-110].

Au contraire, dans la théorie de Lebesgue, la dérivation et l'intégration deviennent des opérations inverses, sous des hypothèses simples. C'est ainsi que l'identité (1) est automatiquement vérifiée dès que f est continue sur $[a, b]$ et dérivable sur $]a, b[$, de dérivée bornée.

Bon comportement face aux limites

Peut-on échanger les opérations limite et intégrale ? C'est un problème classique, source de milliers d'exercices dans le cadre de l'intégrale de Riemann. On ne peut même pas y formuler le problème de manière suffisamment générale, car une limite de fonctions Riemann-intégrables n'est pas forcément Riemann-intégrable, même si ces fonctions sont uniformément bornées. Pour s'en convaincre, on peut noter que la fonction indicatrice de $[0, 1] \cap \mathbb{Q}$, non intégrable au sens de Riemann, est limite de limites de fonctions continues puisque

$$1_{\mathbb{Q}}(x) = \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} [\cos(2\pi n!x)]^m.$$

Au contraire, Lebesgue parvient à définir un concept de fonctions intégrables qui est stable par passage à la limite dans de nombreuses situations, et pour lequel l'échange intégrale-limite est presque automatique, sous des hypothèses simples et faciles à vérifier. Ce bon comportement par rapport aux limites trouve un intérêt même dans le cadre des fonctions Riemann-intégrables ! Pour s'en convaincre, on pourra méditer sur l'exercice suivant :

Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions continues $[a, b] \rightarrow [0, 1]$, convergeant (ponctuellement) vers 0. Alors $\int_a^b f_n \rightarrow 0$.

Cet énoncé a bien sûr un sens dans le cadre de l'intégrale de Riemann, pourtant sa démonstration au moyen d'outils classiques est délicate (l'hypothèse naturelle dans cette théorie est la convergence uniforme et non la convergence simple) ; alors que la théorie de Lebesgue résout le problème sans douleur !

Mauvais traitement des compensations

Pour puissante qu'elle soit, la théorie de Lebesgue est impuissante à traiter la "semi-convergence" des intégrales, c'est-à-dire les situations où une fonction f se trouve être intégrable du fait de compensations entre valeurs positives et négatives, alors que sa valeur absolue $|f|$ n'est pas intégrable. D'autres théories sont plus habiles à tirer parti des compensations : ainsi les intégrales M (intégrale de Denjoy-Perron) ou M^2 présentées dans [Zygmund, T.2, pp.83-91].

Pourquoi alors n'enseigne-t-on pas ces théories alternatives plutôt que celle de Lebesgue ? D'une part, parce que dans l'immense majorité des applications, le mauvais traitement des intégrales semi-convergentes s'avère sans gravité (il s'agit en fait de situations relativement exceptionnelles) ; d'autre part, parce que la théorie de Lebesgue dispose d'une grande souplesse qui lui permet de se généraliser de manière abstraite, comme il est expliqué ci-après.

Insensibilité à la topologie

Dans l'intégrale de Riemann, un rôle particulier est joué par les propriétés de régularité, en un sens très lâche, des fonctions que l'on veut intégrer (variation importante au voisinage d'un point...) On a déjà mentionné, par exemple, des critères d'intégrabilité faisant intervenir l'ensemble des points de discontinuité. La topologie de l'espace de définition des fonctions (en l'occurrence la droite réelle) intervient donc. Cela se reflète sur les généralisations abstraites : pour adapter la construction de Riemann à des espaces plus généraux, on est tout de suite amené à faire des hypothèses de nature topologique assez fortes. (Ce constat vaut aussi pour les intégrales de Denjoy-Perron ou M^2 .)

Au contraire, comme le comprit Radon vers 1913, l'intégrale construite par Lebesgue peut être adaptée à un cadre extrêmement général, sans qu'aucune hypothèse topologique ne soit faite sur l'espace d'intégration, et l'expérience montre que l'on peut construire ainsi des théories maniabiles. Le cas échéant, cet espace pourra même être un espace fonctionnel de dimension infinie : un exemple classique et très important dans les applications est l'**espace de Wiener**, qui n'est autre que l'espace $C([0, 1], \mathbb{R}^d)$ des fonctions continues sur $[0, 1]$ à valeurs dans \mathbb{R}^d . La **mesure de Wiener** est une mesure "gaussienne" sur cet espace, d'importance capitale en probabilités et en physique (mouvement brownien).

Idée de départ : intégration par tranches

L'idée de départ de Lebesgue semble très simple. Comme dans le cas de l'intégrale de Riemann, il s'agit d'approcher l'aire sous le graphe de la fonction par une union de rectangles. Mais ces rectangles sont définis de manière différente ! Dans le cas de Riemann, on s'intéresse aux variations de la fonction sur son domaine de définition : une base étant donnée, on définit le rectangle comme l'ensemble des points au-dessus de cette base, qui sont situés en-dessous de la courbe. Au contraire, dans le cas de Lebesgue, on définit le rectangle en fonction des *valeurs* de la fonction, sans jamais s'intéresser trop à l'espace de départ. Ce n'est donc pas la base du rectangle que l'on se donnera au départ, mais une variation dans les valeurs atteintes (côté "vertical"). Evidemment, on doit admettre que plusieurs rectangles partagent un même côté vertical. C'est ici que l'intégrale de Lebesgue va gagner toute sa complexité : alors que dans l'intégrale de Riemann, une brique élémentaire est un simple rectangle, dans celle de Lebesgue, il pourra s'agir de plusieurs ou même d'une infinité de rectangles.

L'exemple de la fonction indicatrice de \mathbb{Q} est révélateur : bien que discontinue partout, cette fonction est très facile à décrire en fonction de ses valeurs. Dans la théorie de Riemann, on tenterait vainement de découper le segment $[0, 1]$ en tout petits intervalles où cette fonction ne varie guère ; dans celle de Lebesgue, on partage $[0, 1]$ en seulement deux morceaux qui sont assez complexes (totalement discontinus) mais sur chacun desquels la fonction est effectivement constante.

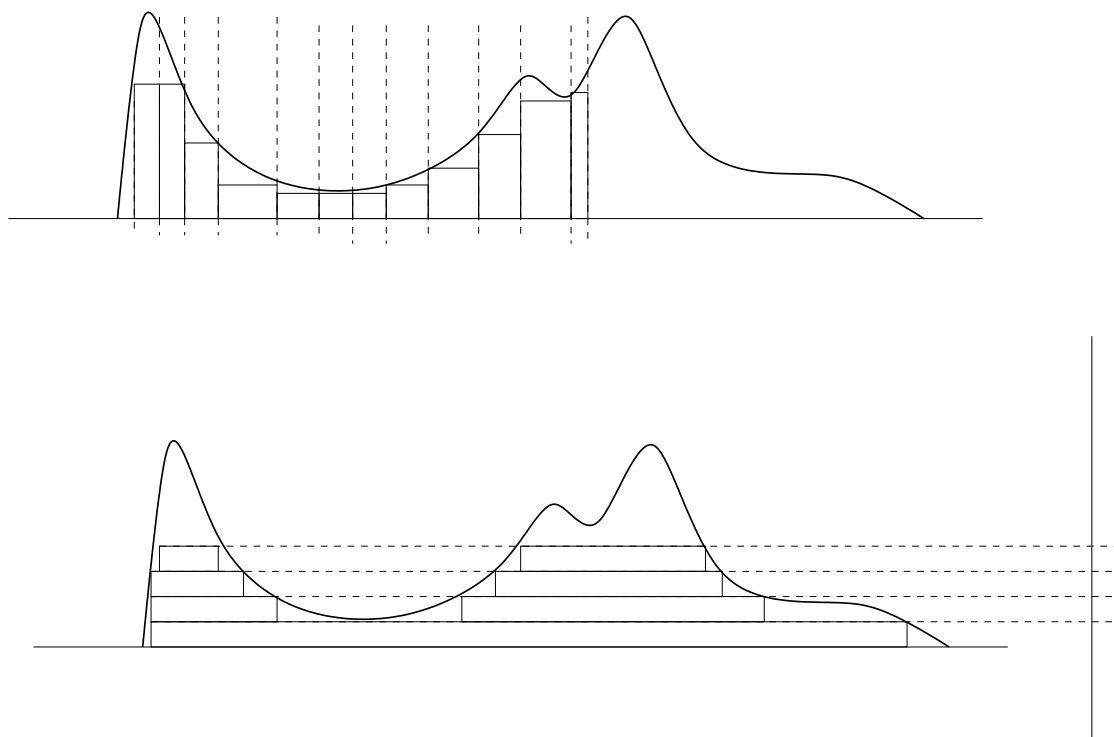


FIG. 1. Procédé de Riemann vs. procédé de Lebesgue

CHAPITRE I

Théorie abstraite de la mesure

Ce chapitre est consacré à une exposition abstraite de la théorie de la mesure, dans une perspective résolument ensembliste où une mesure est vue comme une fonction définie sur les parties mesurables. Dans la section I-1, j'introduirai le concept d'espace mesuré, sur lequel repose la théorie de Lebesgue. La section I-2 est consacrée à des rappels de topologie, avec un accent particulier sur les notions les plus utiles en théorie de la mesure. La section I-3 recense les principales propriétés de "régularité" dont peuvent être dotés les espaces mesurés. Dans la section I-4 se trouvent l'énoncé et la démonstration de l'important théorème de prolongement de Carathéodory, outil-clé de la construction de nombreuses mesures. Cette section est l'occasion de développements plus avancés, tels que l'énoncé du théorème d'existence de Kolmogorov sur des produits infinis (dont la démonstration est remise au Chapitre III). La section I-5 étudie l'opération de complétion d'une mesure. Enfin, la section I-7 est consacrée à l'étude de recouvrements par de petites boules ; son intérêt principal apparaîtra beaucoup plus tard, dans le Chapitre ??.

Seule la section I-1 est indispensable à la compréhension de la suite du cours ; toutes les autres pourront être omises en première lecture, et consultées quand le besoin s'en fera sentir.

I-1. Espaces mesurés

On ne peut guère séparer les notions d'"intégrale" et de "volume". On pense en particulier au volume d'un ensemble comme l'intégrale de la fonction indicatrice de cet ensemble ; et réciproquement, en pratique on calcule une intégrale en découpant l'aire sous la fonction en "petites tranches superposées" dont on estime le volume. En d'autres termes, on doit avoir les relations

$$\left\{ \begin{array}{l} \text{vol}(A) = \int 1_A \\ \int f = \int_0^{+\infty} \text{vol}(\{f \geq t\}) dt \end{array} \right.$$

où $\{f \geq t\}$ est l'ensemble des x tels que $f(x) \geq t$, et f est supposée positive.

La théorie de Lebesgue se développe sur le concept de **mesure**, et c'est ce qui joue le rôle du volume. Les mesures ont été introduites par Borel, quelques années avant les travaux de Lebesgue, afin de quantifier les tailles de certains ensembles, et de construire des fonctions vérifiant certaines propriétés.

Avant de préciser la définition d'une mesure, cherchons à établir un cahier des charges. On souhaite définir une mesure comme une fonction μ qui associe à un ensemble A un "poids" positif (fini ou infini), noté $\mu[A]$ ou $\mu(A)$.

C'est bien le minimum d'imposer qu'une mesure μ soit **additive** : si A et B sont disjoints, alors la mesure de $A \cup B$ doit être la somme des mesures de A et de B .

Cette relation d'additivité est fondamentale et implique toutes les règles de calcul habituel des volumes : par exemple,

- si $A \subset B$, on peut appliquer la relation d'additivité à $B = (B \setminus A) \cup A$ et trouver que

$$\mu[B] = \mu[A] + \mu[B \setminus A] \geq \mu[A];$$

la mesure μ est donc une fonction *croissante* d'ensembles ;

- en utilisant les identités $A \cup B = (A \setminus B) \cup B$ et $A = (A \setminus B) \cup (A \cap B)$, on obtient facilement

$$(2) \quad \mu[A \cap B] < \infty \implies \mu[A \cup B] = \mu[A] + \mu[B] - \mu[A \cap B],$$

formule qui sert souvent dans les calculs de volume.

Enfin, en pratique on connaîtra la valeur de μ sur certains ensembles particuliers, ou bien on imposera certaines propriétés d'invariance. Par exemple, pour définir le volume usuel dans \mathbb{R}^3 il est naturel de demander que le volume d'un pavé soit égal au produit des longueurs de ses côtés (volume euclidien), et d'imposer que le volume soit invariant par rotation et translation.

Ce cahier des charges paraît raisonnable. Le théorème suivant pourra donc apparaître comme un choc : *Il est impossible de démontrer l'existence¹ d'une fonction d'ensembles $\mu : \mathcal{P}(\mathbb{R}^3) \rightarrow [0, +\infty]$ additive, invariante par rotation et translation, telle que $\mu[[0, 1]^3] = 1$.*

Une fois le choc passé, il n'est pas très difficile de trouver un remède. Au lieu de définir une mesure comme une fonction sur $\mathcal{P}(X)$, l'ensemble de *toutes* les parties d'un ensemble X , on va la définir sur *un sous-ensemble* de $\mathcal{P}(X)$, constitué de parties que l'on appelle "mesurables". On aura alors des relations du type de (2), mais seulement quand on reste dans la classe des parties mesurables.

Nous sommes donc menés à nous intéresser aux classes de parties stables par union, intersection, soustraction : on les appellera des **algèbres**. Les algèbres sont le cadre naturel sur lequel définir des fonctions additives d'ensembles.

I-1.1. Algèbres.

DÉFINITION I-1 (Algèbre). *Soit X un ensemble quelconque, et soit $\mathcal{P}(X)$ l'ensemble de toutes les parties de X . Un sous-ensemble \mathcal{A} de $\mathcal{P}(X)$ est appelé une algèbre (ou algèbre de parties de X) si*

- (i) $\emptyset \in \mathcal{A}$;
- (ii) $A \in \mathcal{A} \implies X \setminus A \in \mathcal{A}$;
- (iii) $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$.

REMARQUE I-2. De nombreuses variantes équivalentes de ces trois axiomes sont possibles. Par exemple, à titre d'exercice on pourra vérifier que la réunion des axiomes (i)-(iii) ci-dessus est équivalente à la réunion des quatre axiomes suivants : (i') $\emptyset, X \in \mathcal{A}$, (ii') Si $A, B \in \mathcal{A}$, alors $A \cup B \in \mathcal{A}$, (iii') Si $A, B \in \mathcal{A}$, alors $A \cap B \in \mathcal{A}$, (iv') Si $A, B \in \mathcal{A}$, alors $A \setminus B \in \mathcal{A}$.

REMARQUE I-3. Une algèbre est automatiquement stable par union finie, intersection finie, différence et différence symétrique. (La différence de deux ensembles A et B est $A \setminus B$, tandis que leur différence symétrique est $(A \setminus B) \cup (B \setminus A)$.)

¹Plus précisément : en utilisant l'axiome du choix, on peut montrer qu'une telle fonction n'existe pas. J'en reparlerai dans le Chapitre IV.

EXEMPLES I-4. (i) Algèbres associées à des **partitions**. Si X est un ensemble quelconque, on appelle partition π de X une collection finie de parties non vides, deux à deux disjointes, dont la réunion est X tout entier. Si π est une partition de X , alors l'ensemble \mathcal{A} de toutes les réunions finies d'éléments de π constitue une algèbre de parties de X . Son cardinal est $2^{\#\pi}$, où $\#\pi$ est le cardinal de π . On peut montrer que toute algèbre finie est de cette forme : pour cela, on identifie les éléments de π comme les éléments minimaux de \mathcal{A} .

(ii) Algèbres associées à des familles **stables par intersection**. Une famille \mathcal{F} est dite stable par intersection si l'intersection de deux membres A et B de \mathcal{F} est elle-même un élément de \mathcal{F} (en conséquence de quoi l'intersection d'un nombre fini arbitraire d'éléments de \mathcal{F} est également un élément de \mathcal{F}). Soient X un ensemble quelconque, \mathcal{F} une famille de parties de X , stable par intersection, contenant X et telle que le complémentaire de tout élément de \mathcal{F} est une union finie d'éléments de \mathcal{F} ; alors l'ensemble \mathcal{A} de toutes les *unions finies d'éléments de \mathcal{F}* est une algèbre de parties de X .

(iii) Algèbre engendrée par les **pavés**. C'est un cas particulier de (ii). On se donne $(X_k)_{1 \leq k \leq K}$ une famille finie d'ensembles, et pour chaque X_k on se donne une algèbre \mathcal{A}_k de parties de X_k . On pose $X = \prod X_k$, le problème est de définir une algèbre "naturelle" sur X . Pour cela on considère la famille \mathcal{F} formée des pavés, i.e. les $P = \prod A_k$, où chaque A_k est un élément de \mathcal{A}_k . La famille \mathcal{F} est alors stable par intersection, et le complémentaire d'un pavé peut s'écrire comme une union finie de pavés : par exemple, pour $K = 2$ on a

$$(X_1 \times X_2) \setminus (A_1 \times A_2) = (X_1 \setminus A_1) \times (X_2 \setminus A_2) \cup (X_1 \setminus A_1) \times X_2 \cup X_1 \times (X_2 \setminus A_2).$$

D'après (ii), on sait alors que l'ensemble des unions finies de pavés forme une algèbre de parties de X . Ce point est illustré graphiquement sur la figure 1, dans l'espace produit $X = \mathbb{R}^2$.

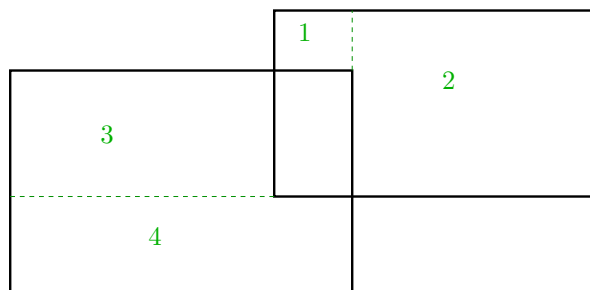


FIG. 1. La différence symétrique de deux pavés est une union finie de pavés

(iv) Algèbre engendrée par les **cylindres**. C'est encore un cas particulier de (ii), d'une importance considérable en théorie des probabilités. Il s'agit de définir une algèbre naturelle sur un produit *infini* d'ensembles dont chacun est muni de sa propre algèbre. Je vais commencer par considérer le cas simple d'un produit dénombrable d'ensembles finis X_k munis de leur algèbre $\mathcal{P}(X_k)$. (Le choix "trivial" où chaque X_k est l'espace $\{0, 1\}$ est déjà non-trivial!) L'espace $X = \prod X_k$ est donc l'espace des suites $(x_k)_{k \in \mathbb{N}}$ telles que $x_k \in X_k$ pour tout k , et on va alors définir un *cylindre élémentaire* comme une partie de X qui "ne

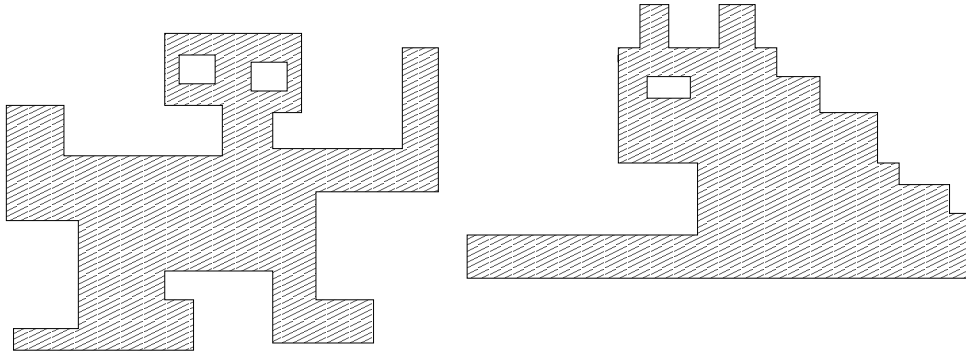


FIG. 2. Deux membres de l'algèbre engendrée par les pavés dans $\mathbb{R} \times \mathbb{R}$

dépend que d'un nombre fini de coordonnées" : pour tout K et tout choix de $(a_1, \dots, a_K) \in X_1 \times \dots \times X_K$, on posera donc

$$C_K(a_1, \dots, a_K) = \{x \in X; 1 \leq k \leq K \Rightarrow x_k = a_k\}.$$

Il est facile de vérifier que la famille de ces cylindres est stable par intersection, et que le complémentaire d'un cylindre élémentaire est une union finie de cylindres élémentaires. Par (ii), on sait donc que l'ensemble des unions finies de cylindres élémentaires forme une algèbre.

Cette construction se généralise comme suit au cas où l'ensemble des indices n'est pas forcément dénombrable. Soit $(X_t)_{t \in T}$ une famille d'ensembles, indexée par un ensemble T quelconque ; pour chaque X_t on se donne une algèbre \mathcal{A}_t de parties de X_t . Pour tout entier K et tout choix de (t_1, \dots, t_K) , dans T^K , pour tout $j \in \{1, \dots, K\}$ on choisit A_{t_j} dans \mathcal{A}_{t_j} et on définit le cylindre élémentaire $C_{(t_1, \dots, t_K)}(A_{t_1}, \dots, A_{t_K})$ par la formule

$$C_{(t_1, \dots, t_K)}(A_{t_1}, \dots, A_{t_K}) = \left\{ x \in \prod_{t \in T} X_t; \quad \forall j \in \{1, \dots, K\}, x_{t_j} \in A_{t_j} \right\}.$$

L'ensemble des unions finies de cylindres élémentaires est alors une algèbre.

I-1.2. Sigma-algèbres. La théorie de Lebesgue n'est pas construite sur la base des algèbres, mais sur celle des σ -algèbres, ou **tribus**. Le préfixe σ indique une propriété de *stabilité vis-à-vis des opérations dénombrables*. La différence entre le concept d'algèbre et celui de σ -algèbre peut s'énoncer comme suit. Rappelons-nous que si $(A_k)_{1 \leq k \leq K}$ est une famille finie d'éléments d'une algèbre \mathcal{A} , alors l'union $\cup A_k$ est toujours un élément de l'algèbre. Dans la définition d'une σ -algèbre, on impose que cette propriété soit également vraie quand on considère une famille **dénombrable** de parties.

DÉFINITION I-5 (σ -algèbre). Soit X un ensemble quelconque, et soit $\mathcal{P}(X)$ l'ensemble de toutes les parties de X . Un sous-ensemble \mathcal{A} de $\mathcal{P}(X)$ est appelé une σ -algèbre (ou σ -algèbre de parties, ou tribu) si

- (i) $\emptyset \in \mathcal{A}$;
- (ii) $A \in \mathcal{A} \Rightarrow X \setminus A \in \mathcal{A}$;
- (iii) $[\forall k \in \mathbb{N}, A_k \in \mathcal{A}] \Rightarrow \bigcup_{k \in \mathbb{N}} A_k \in \mathcal{A}$.

EXEMPLE I-6. L'ensemble des unions finies d'intervalles de \mathbb{R} est une algèbre, mais ce n'est pas une σ -algèbre.

REMARQUE I-7. Une σ -algèbre est automatiquement stable par **intersection dénombrable**, comme on le voit en passant aux complémentaires dans (iii).

REMARQUE I-8. Le lecteur tant soit peu familier avec la topologie aura remarqué une certaine analogie entre la notion de σ -algèbre et celle de topologie. Rappelons que, par définition, une topologie sur un ensemble quelconque X est un sous-ensemble \mathcal{O} de parties de X tel que (i) $\emptyset, X \in \mathcal{O}$, (ii) $O_1, O_2 \in \mathcal{O} \implies O_1 \cap O_2 \in \mathcal{O}$, (iii) $\forall i, O_i \in \mathcal{O} \implies \bigcup O_i \in \mathcal{O}$. Noter que dans (iii), la famille I indexant les O_i est arbitraire (pas nécessairement dénombrable). Les éléments de \mathcal{O} sont appelés des *ouverts*, et leurs complémentaires sont appelés des *fermés*. L'exemple le plus important de topologie est la topologie définie par une distance : on définit un ouvert comme une union de boules ouvertes. Des rappels plus détaillés seront effectués dans la section I-2.

DÉFINITION I-9 (espace mesurable). *On appelle espace mesurable un couple (X, \mathcal{A}) , où $\mathcal{A} \subset \mathcal{P}(X)$ est une σ -algèbre. Les éléments de \mathcal{A} seront alors appelés parties mesurables ou ensembles mesurables.*

Par abus de langage, on dira souvent que X est un espace mesurable. Bien sûr, cette terminologie n'a de sens que si l'on fait référence implicite à une certaine tribu : après tout, n'importe quel espace X est mesurable quand on le munit de la tribu triviale $\mathcal{P}(X)$ (ou de $\{\emptyset, X\}$, non moins triviale).

PROPOSITION I-10 (tribu engendrée). *Soient X un ensemble quelconque, et \mathcal{F} un sous-ensemble quelconque de $\mathcal{P}(X)$. L'intersection de toutes les σ -algèbres contenant \mathcal{F} est une σ -algèbre, et c'est la plus petite qui contienne \mathcal{F} . On l'appelle tribu engendrée par \mathcal{F} et on la notera $\sigma(\mathcal{F})$.*

On laisse en exercice la démonstration de cette proposition. (Notons que l'intersection apparaissant dans l'énoncé n'est pas vide car $\mathcal{P}(X)$ est une σ -algèbre contenant \mathcal{F} .)

Considérer des σ -algèbres plutôt que des algèbres est a priori assez séduisant car on obtient ainsi des familles riches qui se comportent bien vis-à-vis des unions infinies, limites, etc. On sait, par exemple, que tout ouvert de \mathbb{R} peut s'écrire comme union disjointe dénombrable d'intervalles ouverts. La σ -algèbre engendrée par les intervalles ouverts de \mathbb{R} contient donc tous les ensembles ouverts, et c'est par conséquent la σ -algèbre engendrée par les ouverts de \mathbb{R} ; on l'appelle **tribu des boréliens** de \mathbb{R} . Elle contient tous les ouverts, donc tous les fermés, mais aussi les unions dénombrables d'ensembles ouverts ou fermés, les unions dénombrables d'intersections dénombrables d'ensembles ouverts ou fermés, etc. — et plus encore. Cette richesse est à la fois une force et une faiblesse : il est très difficile de “décrire” explicitement un ensemble borélien. Pour autant, la tribu borélienne est strictement plus petite que $\mathcal{P}(\mathbb{R})$ (elle a “seulement” la puissance du continu, son cardinal est 2^{\aleph_1}) : en se limitant à cette tribu, on sélectionne des parties de \mathbb{R} qui ont une certaine “régularité”.

EXEMPLES I-11. (i) **Tribu borélienne abstraite** : soit X un espace topologique abstrait (i.e. quelconque). On définit sa tribu borélienne $\mathcal{B}(X)$ comme

la σ -algèbre engendrée par les ouverts de X . C'est également, bien sûr, la σ -algèbre engendrée par les fermés de X .

(ii) **Tribu borélienne réelle** : Dans le cas où $X = \mathbb{R}^n$, on peut trouver de nombreuses familles génératrices beaucoup plus restreintes que la collection de tous les ouverts ou tous les fermés. Par exemple :

- les pavés fermés $\prod [a_k, b_k]$;
- les pavés ouverts $\prod]a_k, b_k[$;
- les pavés semi-ouverts $\prod [a_k, b_k[$;
- les cubes fermés $\prod [a_k, a_k + c]$ (ou ouverts, ou semi-ouverts) ;
- les cubes fermés dyadiques $\prod [m_k 2^{-\ell_k}, (m_k + 1) 2^{-\ell_k}]$ (ou ouverts, ou semi-ouverts) ;
- les boules ouvertes (ou fermées) dans \mathbb{R}^n .

(iii) **Tribu produit** : Cette construction suit celle de l'Exemple I-4(iii). Soient X et Y deux espaces mesurables, avec leurs tribus respectives \mathcal{A} et \mathcal{B} . On appelle pavé mesurable un ensemble de la forme $A \times B$, où $A \in \mathcal{A}$ et $B \in \mathcal{B}$. La tribu engendrée par les pavés mesurables est appelée tribu produit, et notée $\mathcal{A} \otimes \mathcal{B}$. On ne peut guère la décrire explicitement... Cette construction se généralise facilement au produit d'un nombre *fini* d'espaces mesurables.

(iv) **Tribu cylindrique** : Si maintenant $X = \prod X_t$ est un produit infini (dénombrable ou non) d'espaces X_t dont chacun est muni d'une tribu \mathcal{A}_t , alors on peut munir X de la σ -algèbre engendrée par les cylindres élémentaires, suivant la construction de l'Exemple I-4(iv). Il s'agit en fait de la généralisation naturelle du concept de tribu produit. La tribu cylindrique est la σ -algèbre "classique" que l'on introduit sur un produit infini ; cependant, ce n'est pas la seule possible, et certaines constructions alternatives jouent un rôle important en recherche contemporaine.

EXEMPLE I-12. Considérons $X = \{0, 1\}^{\mathbb{N}}$, où chaque facteur est muni de la tribu triviale. Alors les singletons sont mesurables pour la tribu cylindrique. En effet, si $x \in \{0, 1\}^{\mathbb{N}}$ alors $\{x\} = \cap C_k$ où

$$C_k = \left\{ y \in \{0, 1\}^{\mathbb{N}} ; \quad \forall j \leq k, y_j = x_j \right\}.$$

En revanche, il est impossible² de décrire une partie de $\{0, 1\}^{\mathbb{N}}$ qui ne soit pas mesurable pour la tribu cylindrique.

REMARQUE I-13. La construction de la tribu produit est très similaire à celle de la topologie produit, qui sera rappelée dans la section I-2.4. Ces constructions abstraites sont très simples, mais cachent certaines subtilités, que nous aurons l'occasion de retrouver en étudiant l'intégrale sur les espaces produits, au Chapitre III. Pour se donner une idée des problèmes proprement vertigineux que l'on peut rencontrer, le lecteur peut se poser la question suivante. Étant donnés X et Y deux espaces topologiques, munis de leur tribu borélienne, il existe deux tribus naturelles sur $X \times Y$: la tribu borélienne (pour la topologie produit), et la tribu produit (des tribus boréliennes). Ces deux tribus coïncident-elles ?

²En fait, la non-existence de parties non mesurables de $\{0, 1\}^{\mathbb{N}}$ est cohérente avec les axiomes classiques autres que l'axiome du choix ; j'y reviendrai.

EXERCICE I-14. Démontrer l'assertion faite dans l'Exemple I-11(ii), selon laquelle la tribu borélienne de \mathbb{R}^n est engendrée par les pavés ouverts bornés. Il suffit bien sûr de vérifier que tout ouvert appartient à la tribu \mathcal{T} engendrée par les pavés ouverts. On pourra montrer successivement que

- \mathcal{T} contient les pavés ouverts non bornés ;
- \mathcal{T} contient les pavés fermés ;
- \mathcal{T} contient les pavés semi-ouverts ;
- \mathcal{T} contient les ouverts.

On pourra noter que tout ouvert s'écrit comme une réunion dénombrable disjointe de pavés (obtenus en considérant des maillages de plus en plus fins, par exemple).

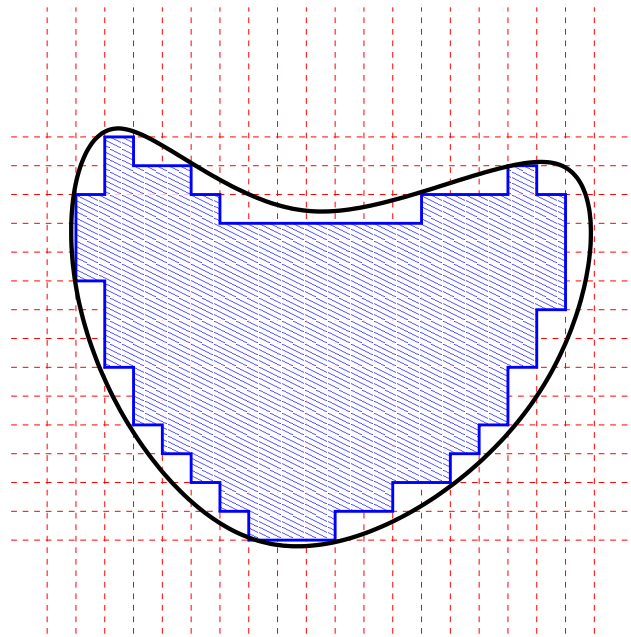


FIG. 3. Approximation d'un ouvert de \mathbb{R}^n par une union de petits pavés

EXERCICE I-15. Soient $m, n \in \mathbb{N}$. En utilisant le résultat de l'exercice I-14, montrer que $\mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n) = \mathcal{B}(\mathbb{R}^{m+n})$.

I-1.3. Mesures. Les σ -algèbres sont le cadre naturel sur lequel sont définies les mesures, au sens de Lebesgue.

DÉFINITION I-16. Soit (X, \mathcal{A}) un espace mesurable. On appelle *mesure* (ou *mesure σ -additive*, ou *mesure positive*) sur X une application μ définie sur \mathcal{A} , à valeurs dans $[0, +\infty]$ et vérifiant l'axiome dit de σ -additivité : pour toute famille dénombrable $(A_k)_{k \in \mathbb{N}}$ d'ensembles mesurables **disjoints**,

$$(3) \quad \mu \left[\bigcup_{k \in \mathbb{N}} A_k \right] = \sum_{k \in \mathbb{N}} \mu[A_k].$$

Le triplet (X, \mathcal{A}, μ) est alors appelé un *espace mesuré*.

Par abus de notation, on dira souvent que (X, μ) est un espace mesuré, la tribu \mathcal{A} étant alors implicite; ou même que X est un espace mesuré, la tribu \mathcal{A} et la mesure μ étant alors implicites.

REMARQUES I-17. (i) La somme apparaissant au membre de droite de (3) converge toujours dans $[0, +\infty]$.

(ii) Comme cas particulier de la définition (choisir $A_k = \emptyset$ pour tout k), on a bien sûr $\mu[\emptyset] = 0$. Les mesures vérifient également les règles habituelles de calcul des volumes : par exemple

$$\begin{aligned} A \subset B &\implies \mu[A] \leq \mu[B], \\ \mu[A \cap B] < \infty &\implies \mu[A \cup B] = \mu[A] + \mu[B] - \mu[A \cap B]. \end{aligned}$$

(iii) On obtient une mesure quand on ajoute deux mesures, ou quand on multiplie une mesure par un nombre réel positif.

(iv) Si (X, μ) est un espace mesuré, et Y une partie mesurable de X , alors μ induit sur Y une structure d'espace mesuré par **restriction** à la tribu des parties mesurables de X qui sont incluses dans Y . La mesure ν , définie comme restriction de μ à Y , est notée

$$\nu = \mu|_Y \quad \text{ou} \quad \nu = \mu|_Y.$$

EXERCICE I-18. Soit μ une mesure. Montrer que pour toute famille dénombrable (A_k) de parties mesurables,

$$\mu \left[\bigcup_{k \in \mathbb{N}} A_k \right] \leq \sum_{k \in \mathbb{N}} \mu[A_k].$$

Grâce aux axiomes des σ -algèbres, les propriétés exprimées en termes d'unions disjointes peuvent être reformulées en termes d'unions croissantes, ou d'intersections décroissantes. Rappelons qu'une famille $(A_k)_{k \in \mathbb{N}}$ d'ensembles est dite *croissante* si on a $A_k \subset A_{k+1}$ pour tout k , et *décroissante* si on a $A_{k+1} \subset A_k$ pour tout k .

PROPOSITION I-19. Soit (X, μ) un espace mesuré.

(i) Soit $(A_k)_{k \in \mathbb{N}}$ une famille croissante de parties mesurables, alors

$$\mu \left[\bigcup_{k \in \mathbb{N}} A_k \right] = \lim_{k \rightarrow \infty} \mu[A_k] = \sup_{k \in \mathbb{N}} \mu[A_k].$$

(ii) Soit $(A_k)_{k \in \mathbb{N}}$ une famille décroissante de parties mesurables, l'un au moins des A_k étant de mesure finie. Alors

$$\mu \left[\bigcap_{k \in \mathbb{N}} A_k \right] = \lim_{k \rightarrow \infty} \mu[A_k] = \inf_{k \in \mathbb{N}} \mu[A_k].$$

DÉMONSTRATION. Démontrons (i). Si $\mu[A_\ell]$ tend vers l'infini quand $\ell \rightarrow \infty$, alors forcément $\mu[\bigcup A_k] = \infty$, et l'assertion est vraie. Dans le cas contraire, la suite croissante $\mu[A_\ell]$ converge vers une limite finie quand $\ell \rightarrow \infty$. On écrit alors $\bigcup A_k$ comme la réunion disjointe de $B_0 = A_0$ et des $B_j = A_j \setminus A_{j-1}$, pour $1 \leq j \leq \infty$. Par σ -additivité,

$$\mu[\bigcup A_k] = \mu[A_0] + \sum_{j \geq 1} \mu[B_j].$$

La série de droite est donc convergente, et on a

$$\mu[A_0] + \sum_{j \geq 1} \mu[B_j] = \lim_{\ell \rightarrow \infty} \left(\mu[A_0] + \sum_{1 \leq j \leq \ell} \mu[B_j] \right) = \lim_{\ell \rightarrow \infty} \mu[A_\ell].$$

Pour démontrer (ii), supposons sans perte de généralité que A_0 est de mesure finie ; alors μ définit (par restriction) une mesure sur A_0 , et on peut appliquer la partie (i) de la proposition à la famille croissante $(A_0 \setminus A_k)$. \square

EXERCICE I-20. Montrer que la propriété (i) est en fait *équivalente* à l'axiome de σ -additivité. Trouver un contre-exemple à la propriété (ii) si l'on ne suppose pas que l'un des A_k est de mesure finie. (On pourra utiliser la fonction “cardinal” sur \mathbb{N} .) Faire l'analogie avec un contre-exemple classique de topologie montrant qu'une intersection décroissante de fermés peut être vide.

REMARQUE I-21. Nous sommes en train de développer une théorie des mesures σ -additives sur des σ -algèbres. Pourquoi ne pas développer une théorie des fonctions *finiment additives* définies sur des algèbres ? Autrement dit, pourquoi utiliser l'axiome de σ -additivité au lieu de l'axiome plus naturel

$$A \cap B = \emptyset \implies \mu[A \cup B] = \mu[A] + \mu[B] \quad ?$$

Une réponse possible est que la relation de σ -additivité sera précieuse et puissante pour résoudre certains problèmes. Mais surtout, il n'existe pas à l'heure actuelle de consensus autour d'une théorie des “mesures finiment additives” (voir les références dans [Dudley, p. 112]). Cela dit, peut-être la théorie de la mesure actuelle sera-t-elle un jour remplacée par une autre théorie plus générale.....

REMARQUE I-22. On peut également poser la question dans le sens contraire : pourquoi se limiter aux mesures σ -additives, et ne pas considérer des mesures qui soient additives pour des familles disjointes d'ensembles, pas nécessairement dénombrables ? La réponse est cette fois assez facile : on tomberait immédiatement sur des paradoxes choquants. Par exemple, la “longueur” d'un singleton de $[0, 1]$ est évidemment 0, l'intervalle $[0, 1]$ est union de singletons, mais sa longueur est 1 et non pas 0...

EXEMPLES I-23. Si la construction des σ -algèbres est souvent un exercice facile, la construction des mesures peut poser des problèmes considérables. Voici quelques mesures célèbres.

- (i) L'exemple non trivial le plus simple d'une mesure est ce que l'on appelle une **masse de Dirac** : soit X un espace mesurable, et x_0 un point de X , on note δ_{x_0} la mesure définie par

$$\delta_{x_0}[A] = \begin{cases} 1 & \text{si } x_0 \in A, \\ 0 & \text{si } x_0 \notin A. \end{cases}$$

On peut se représenter cette mesure comme une “masse ponctuelle” située au point x_0 . Malgré sa simplicité, la masse de Dirac peut être considérée comme la “brique élémentaire des mesures : de larges classes de mesures peuvent en effet être “approchées” par des combinaisons de masses de Dirac.

- (ii) La **mesure de comptage** n'est autre que la fonction “cardinal”, à valeurs dans $\mathbb{N} \cup \{+\infty\}$.

- (iii) La **mesure de Lebesgue** dans \mathbb{R}^n est la mesure de référence naturelle dans un cadre euclidien. Elle correspond aux notions habituelles de longueur ($n = 1$), surface ($n = 2$) ou volume ($n = 3$). On peut la définir en spécifiant le volume des pavés, ou celui des boules, selon les formules dont nous avons l'habitude. Son existence n'a rien d'évident et nous parlerons en détail de sa construction dans le chapitre IV.

- (iv) Les **mesures de Hausdorff** permettent de définir des notions de “volume d -dimensionnel” dans un espace de dimension n : longueur d'une courbe tracée dans \mathbb{R}^3 , etc. A chaque dimension $d \geq 0$ (entière ou non) est associée une mesure de Hausdorff ; quand d n'est pas entier on parle souvent de mesure fractale. La mesure de Hausdorff n -dimensionnelle dans \mathbb{R}^n coïncide avec la mesure de Lebesgue, ce qui n'est pas évident a priori ; quant à la mesure de Hausdorff 0-dimensionnelle, ce n'est autre que la mesure de comptage. Les mesures de Hausdorff se définissent naturellement dans des espaces métriques arbitraires et pas seulement dans \mathbb{R}^n .

- (v) La **mesure de Haar** est une autre généralisation de la mesure de Lebesgue ; on la construit sur un *groupe topologique localement compact* ; on reviendra sur ces concepts. La mesure de Haar est caractérisée par certaines propriétés, dont la principale est l'invariance vis-à-vis de l'action du groupe (disons action à gauche), c'est-à-dire l'action des translations $\tau_a : x \mapsto (a.x)$. Par exemple, la mesure de Haar sur le groupe localement compact \mathbb{R}^n n'est autre que la mesure de Lebesgue. Un autre exemple est le tore $\mathbb{R}^n/\mathbb{Z}^n$, que l'on peut identifier à $[0, 1]^n$; il s'agit d'un groupe compact, et sa mesure de Haar est encore la (restriction de la) mesure de Lebesgue.

- (vi) La **mesure de volume**, sur une variété riemannienne (M, g) de classe C^2 (se reporter à un cours de géométrie différentielle pour des explications !), est encore une autre généralisation de la mesure de Lebesgue. On peut la définir comme la mesure de Hausdorff n -dimensionnelle, où n est la dimension de la variété, et on a fait de M un espace métrique en le munissant de la distance géodésique ; mais on dispose aussi d'une formule explicite, que l'on peut écrire dans des cartes, en fonction du déterminant de la métrique g et de la mesure de Lebesgue.

- (vii) La **mesure de Wiener**, définie sur $C([0, 1], \mathbb{R}^d)$, est liée au mouvement brownien. On peut la comprendre intuitivement ainsi : soit une particule partant d'un point donné (l'origine) et décrivant une trajectoire brownienne, observée sur l'intervalle de temps $[0, T]$. La probabilité pour que la trajectoire de cette particule possède une certaine propriété (P) est la mesure (de Wiener) de l'ensemble de tous les chemins dans $C([0, T], \mathbb{R}^d)$ qui possèdent la propriété (P). Depuis les travaux d'Einstein et de Smoluchowski au début du siècle, on connaissait les valeurs de cette mesure sur tous les cylindres, un cylindre étant l'ensemble de toutes les fonctions f telles que $(f(t_1), \dots, f(t_k))$ appartient à un certain ensemble mesurable de \mathbb{R}^k , pour toute famille finie arbitraire de temps (t_1, \dots, t_k) . Il a cependant fallu attendre 1921 pour que Wiener réussisse à démontrer l'*existence* de cette mesure sur la tribu engendrée par les cylindres.

I-2. Rappels de topologie

Dans la section précédente, les concepts de nature topologique ne jouaient aucun rôle ; mais dans toute la suite du cours, ils interviendront de manière importante. Je vais brièvement passer en revue les notions qui nous seront utiles ; le lecteur pourra dans la plupart des cas compléter lui-même les démonstrations, ou bien les trouver dans la plupart des ouvrages d'introduction à la topologie.

Préliminaires

Soit X un ensemble quelconque ; on le munit d'une structure topologique en lui associant une famille \mathcal{O} de parties de X , appelées **ouverts**, telles que

- (i) l'ensemble vide et X sont des ouverts,
- (ii) l'intersection de deux ouverts est un ouvert,
- (iii) la réunion d'une famille quelconque d'ouverts est un ouvert.

L'intersection d'un nombre fini d'ouverts, une union quelconque d'ouverts sont donc des ouverts.

Si une partie V contient un ouvert contenant un élément x , on dit que V est **voisinage** de x .

Le complémentaire d'un ouvert est appelé **fermé**. L'ensemble vide, l'espace X tout entier, l'union d'un nombre fini d'ouverts, une intersection quelconque de fermés sont fermés.

On dit qu'un espace topologique X est **séparé** (terminologie anglo-saxonne : espace de Hausdorff) si, étant donnés deux éléments distincts x et y de X , on peut toujours leur trouver des voisinages disjoints.

Soit \mathcal{F} une famille de parties de X . L'intersection de toutes les topologies sur X contenant \mathcal{F} est une topologie, et c'est la plus petite qui fasse de tous les éléments de \mathcal{F} des ouverts. On l'appelle topologie engendrée par \mathcal{F} .

Un ensemble $A \subset X$ étant donné, on définit l'**adhérence** \overline{A} de A comme le plus petit fermé contenant A , c'est-à-dire l'intersection de tous les fermés contenant A ; et l'**intérieur** de A comme le plus grand ouvert contenu dans A , c'est-à-dire l'union de tous les ouverts contenus dans A . Une partie A est dite dense dans X si $\overline{A} = X$.

Un sous-ensemble arbitraire A de X peut être vu comme un espace topologique si on le munit de la topologie définie par les intersections de A avec les ouverts de X . Cette topologie est dite **topologie induite**.

On dit qu'une suite $(x_n)_{n \in \mathbb{N}}$ converge vers $x \in X$ si, pour tout voisinage V de x on peut trouver $N \in \mathbb{N}$ tel que $n \geq N \implies x_n \in V$. En d'autres termes, pour n assez grand, x_n est confiné dans n'importe quel voisinage de x fixé a priori. On dit alors que x est la limite de la suite (x_n) et on note $x_n \longrightarrow x$. On appelle valeur d'adhérence de (x_n) tout $x \in X$ qui peut être obtenu comme limite d'une suite extraite $(x_{k(n)})_{n \in \mathbb{N}}$.

Etant donnés deux espaces topologiques X et Y , on dit qu'une fonction $f : X \rightarrow Y$ est **continue** si $f^{-1}(O)$ est un ouvert de X , pour tout ouvert O de Y . Si X n'est pas a priori muni d'une topologie, on appelle topologie engendrée par f la plus petite topologie qui rende f continue : c'est l'ensemble de toutes les images réciproques d'ouverts de Y par f .

Un espace topologique X est dit **connexe** si on ne peut le séparer en deux ouverts disjoints non vides. Un espace topologique arbitraire étant donné, on peut toujours le décomposer en composantes connexes, qui sont les plus grands (au sens de l'inclusion, vue comme ordre partiel) ensembles connexes contenus dans X . Un

point $x \in X$ est dit **isolé** si le singleton $\{x\}$ est un ouvert, ou de manière équivalente si sa composante connexe est réduite à lui-même.

La notion de topologie peut être vue comme une abstraction extrême de la notion de “proximité” : quand on se donne un voisinage V d’un point x , on définit en quelque sorte les points qui sont proches de x (qui appartiennent à ce voisinage). Les composantes connexes sont en quelque sorte des parcelles disjointes de l’espace, telles que l’on puisse se déplacer “continûment” à l’intérieur d’une même parcelle, sans pouvoir passer de parcelle en parcelle.

Les topologies que l’on rencontre le plus souvent sont celles qui sont engendrées par une **métrique** ; elles font l’objet de la section suivante.

De manière générale, de nombreuses métriques peuvent être associées à une même topologie (elles sont alors dites équivalentes du point de vue topologique) ; ce qui laisse penser que le concept de topologie est plus satisfaisant que celui de métrique. Cependant,

- (a) en général, une topologie abstraite peut présenter des propriétés “pathologiques” qui sont interdites si la topologie provient d’une métrique ;
- (b) une métrique permet de quantifier la notion de “proximité”, ce qui peut être précieux dans des problèmes pratiques ;
- (c) en analyse³, on peut (presque) toujours en pratique se ramener à des topologies métriques.

I-2.1. Espaces métriques. Soit X un ensemble quelconque ; on dit qu’une application $d : X \times X \rightarrow \mathbb{R}_+$ est une métrique, ou distance, si elle satisfait aux deux axiomes

- (i) $\forall x, y, z \in X, \quad d(x, z) \leq d(x, y) + d(y, z) ;$
- (ii) $[d(x, y) = 0] \iff x = y.$

On appelle **espace métrique** un couple (X, d) , où d est une distance sur l’ensemble X . Par abus de langage, on dira souvent que X est un espace métrique, la distance d étant alors implicite.

Soient (X, d) un espace métrique, $x \in X$ et $r \geq 0$. On définit la **boule ouverte** $B_r(x) = B(x, r)$, centrée en x et de rayon r , et la **boule fermée** $B_{r]}(x) = B(x, r]$, par les formules

$$B_r(x) := \{y \in X; d(x, y) < r\}; \quad B_{r]}(x) := \{y \in X; d(x, y) \leq r\}.$$

Si (X, d) est un espace métrique, on introduit une topologie séparée sur X en définissant les ouverts comme les unions de boules ouvertes. On a alors les propriétés suivantes :

- les boules ouvertes sont ouvertes, les boules fermées sont fermées ;
- un ensemble V est voisinage d’un point x si et seulement si il existe $r > 0$ tel que $B_r(x) \subset V$;
- un ensemble O est ouvert si et seulement si il est voisinage de tous ses points ;
- une fonction $f : (X, d) \rightarrow (X', d')$ définie entre deux espaces métriques est continue si et seulement si

$$\forall x \in X \quad \forall \varepsilon > 0 \quad \exists \delta > 0; \quad d(x, y) \leq \delta \implies d'(f'(x), f'(y)) \leq \varepsilon.$$

³Ce n’est pas vrai en géométrie algébrique, où la topologie de Zariski, d’usage fréquent, est résolument non métrique.

Les suites sont souvent d'une aide précieuse dans les espaces métriques ; plusieurs des propriétés que nous avons mentionnées précédemment admettent des caractérisations simples en termes de suites convergentes. Ainsi, notons que

- une suite $(x_n)_{n \in \mathbb{N}}$ à valeurs dans X converge vers $x \in X$ si et seulement si $d(x_n, x) \rightarrow 0$ quand $n \rightarrow \infty$;

- l'adhérence d'un ensemble $A \subset X$ est l'ensemble de toutes les limites de suites à valeurs dans A ; en particulier, un ensemble F est fermé si et seulement si il est stable par passage à la limite : les assertions $x_n \in F$, $x_n \rightarrow x \in X$ impliquent $x \in F$;

- une fonction f définie entre espaces métriques est continue si et seulement si elle préserve les limites : $f(x_n) \rightarrow f(x)$ dès que $x_n \rightarrow x$.

Si A est une partie d'un espace métrique (X, d) et $x \in X$, on pose $d(x, A) = \inf\{d(x, y) ; y \in A\}$: c'est la distance de x à A . L'adhérence de A est l'ensemble des points qui sont à distance nulle de A ; en particulier, si F est un fermé, les assertions $d(x, F) = 0$ et $x \in F$ sont équivalentes.

I-2.2. Régularité des espaces topologiques. La zoologie des espaces topologiques est très riche ; les propriétés d'usage le plus courant sont la *séparabilité*, la *compacité* et la *complétude*.

Un espace topologique X est dit **séparable** s'il admet une famille dénombrable dense. Dire qu'un espace métrique X est séparable revient à dire qu'il existe une famille $(x_k)_{k \in \mathbb{N}}$, telle que pour tout $x \in X$ il existe une application $n \rightarrow k(n)$, définie de \mathbb{N} dans \mathbb{N} , telle que $x_{k(n)} \rightarrow x$.

Une suite $(x_n)_{n \in \mathbb{N}}$ dans un espace métrique (X, d) est dite **de Cauchy** si $d(x_m, x_n) \rightarrow 0$ quand $(m, n) \rightarrow \infty$; en d'autres termes,

$$\forall \varepsilon > 0 \quad \exists N \in \mathbb{N}; \quad n, m \geq N \implies d(x_m, x_n) \leq \varepsilon.$$

Un espace métrique (X, d) est dit **complet** si toute suite de Cauchy dans X converge.

Un espace topologique X séparé est dit **compact** si, de tout recouvrement de X par une famille d'ouverts $(O_i)_{i \in I}$, on peut extraire un sous-recouvrement fini, i.e. il existe une famille finie $J \subset I$ telle que $X \subset \bigcup_{j \in J} O_j$. Si X est un espace topologique séparé et K une partie de X , on dit que K est un compact de X si la topologie induite par X sur K en fait un espace topologique compact.

Si X est un espace topologique séparé, et A un sous-ensemble de X , on dit que A est **précompact**⁴ si son adhérence \overline{A} est un compact.

Un espace métrique (X, d) étant donné, on appelle diamètre de X le supremum de d sur $X \times X$; l'espace X est dit **borné** si son diamètre est fini.

Les propriétés suivantes se prouvent sans difficulté :

- un espace compact est fermé, et une partie fermée d'un compact est compacte ;

⁴J'adopte ici la terminologie anglo-saxonne (precompact) au lieu de la terminologie française courante "relativement compact". En français on réserve d'habitude le terme "précompact" pour ce qui sera appelé plus loin "totalement borné". La différence n'est pas cruciale : Dans un espace métrique séparable complet, il est équivalent de dire qu'une partie est relativement compacte, ou qu'elle est totalement bornée ; Cf. le Théorème I-32.

- l'union de deux compacts est un compact, et le produit de deux compacts aussi ; plus généralement, une union finie de compacts et un produit fini de compacts sont des compacts (on reparlera plus tard de produits infinis) ;

- l'image par une fonction continue d'un compact est un compact ; en particulier, une fonction continue sur un compact, à valeurs dans \mathbb{R} , est bornée et atteint ses bornes ;

- si K_1 et K_2 sont deux compacts dans un espace métrique (X, d) , alors il existe $x_1 \in K_1$ et $x_2 \in K_2$ tels que

$$d(x_1, x_2) = \inf \left\{ d(y_1, y_2), \quad y_1 \in K_1, y_2 \in K_2 \right\}$$

(conséquence de ce que la fonction distance atteint son infimum sur le compact $K_1 \times K_2$) ;

- un espace métrique compact est complet et borné ;

- un espace métrique X est compact si et seulement si toute suite $(x_n)_{n \in \mathbb{N}}$ à valeurs dans X admet une valeur d'adhérence (théorème de **Bolzano-Weierstrass**) ;

- une fonction f continue entre un espace métrique compact (X, d) et un espace métrique (Y, d') est automatiquement **uniformément continue** : pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que

$$\forall x, y \in X \quad d(x, y) \leq \delta \implies d'(f(x), f(y)) \leq \varepsilon.$$

On peut alors définir le module de continuité de f par

$$m_f(\delta) := \sup \left\{ d'(f(x), f(y)) ; d(x, y) \leq \delta \right\} :$$

c'est une fonction croissante, continue, vérifiant $m_f(0) = 0$.

EXEMPLES I-24. Une partie finie est compacte. Une partie bornée et fermée de \mathbb{R}^n , ou de n'importe quelle variété de dimension finie, est compacte. L'espace \mathbb{R}^n tout entier est complet mais pas compact.

Intuitivement, une partie compacte est une partie “petite” au sens imagé où, quand on cherche à l'explorer (au moyen des valeurs d'une suite, par exemple), on revient toujours sur ses pas (il existe un x près duquel on revient une infinité de fois). La topologie moderne s'est développée sur la base des concepts d'ouvert et de compact.

Les définitions suivantes introduisent les hypothèses de régularité les plus utilisées en théorie de la mesure.

DÉFINITION I-25 (σ -compacité). *Un espace topologique X est dit σ -compact s'il est union dénombrable de compacts.*

DÉFINITION I-26 (compacité locale). *Un espace topologique X est dit **localement compact** si tout $x \in X$ admet un voisinage compact.*

DÉFINITION I-27 (espace polonais). *Un espace topologique X est dit **polonais** si c'est un espace métrique séparable et complet.*

EXEMPLE I-28. L'espace \mathbb{R}^n , et plus généralement n'importe quelle variété riemannienne lisse complète de dimension finie (munie de sa distance géodésique), sont tous localement compacts, σ -compacts et polonais. En revanche, l'espace de Wiener $C([0, 1], \mathbb{R}^n)$, muni de la norme du supremum, est polonais mais n'est ni localement compact, ni σ -compact (tous les compacts y sont d'intérieur vide).

REMARQUE I-29. Les spécialistes de théorie de la mesure moderne utilisent des notions plus fines : espaces de Lusin, espaces de Suslin, etc., dont les espaces polonais sont des cas particuliers.

I-2.3. Théorèmes d'extension et de séparation. Voici maintenant quelques théorèmes d'extension qui nous serviront dans la suite du cours :

- Si (X, d) et (Y, d') sont des espaces métriques *complets*, et f est une application *uniformément continue* sur une partie A de X , à valeurs dans Y , alors elle admet un unique prolongement continu de \overline{A} dans Y . Ce prolongement peut se définir par l'identité $f(x) = \lim f(x_n)$, où x_n est une suite arbitraire convergeant vers x .

- Si (X, d) est un espace métrique et F_0, F_1 sont deux fermés disjoints de X , alors la fonction f valant 0 sur F_0 et 1 sur F_1 admet un prolongement continu à X tout entier, à valeurs dans $[0, 1]$. Pour le voir, il suffit de poser

$$f(x) := \frac{d(x, F_0)}{d(x, F_0) + d(x, F_1)}$$

(Si F_0 (resp. F_1) est vide, on pose $f \equiv 1$ (resp. $f \equiv 0$).) Bien sûr, on en déduit qu'une fonction qui vaut a_0 sur F_0 et a_1 sur F_1 admet un prolongement continu à valeurs dans $[a_0, a_1]$.

- Si (X, d) est un espace métrique et A un ensemble quelconque de X , alors toute fonction Lipschitzienne $f : A \rightarrow \mathbb{R}$ admet un prolongement Lipschitzien à X tout entier (a priori non unique). Pour le voir, il suffit de poser

$$f(x) = \inf_{y \in A} [d(x, y) + f(y)].$$

- Si (X, d) est un espace métrique et F est un fermé de X , alors toute fonction continue $f : F \rightarrow \mathbb{R}$ admet un prolongement continu (a priori non unique) à X tout entier, vérifiant $\sup_X f = \sup_F f$, $\inf_X f = \inf_F f$. C'est le **Théorème d'extension de Tietze-Urysohn**, dont je vais esquisser la démonstration.

PREUVE DU THÉORÈME DE TIETZE-URYSOHN. En décomposant f en f_+ et f_- , on se ramène au cas où f est à valeurs positives.

Si f est bornée, on peut sans perte de généralité supposer que f est à valeurs dans $[0, 1]$. On construit alors une série d'approximations continues à f , comme suit. On introduit d'abord une fonction g_1 continue, à valeurs dans $[0, 1/2]$, qui vaut identiquement 0 sur le fermé $\{f = 0\}$ et identiquement $1/2$ sur le fermé $\{f \geq 1/2\}$. On a alors $0 \leq f - g_1 \leq 1/2$ sur F . On introduit ensuite g_2 , à valeurs dans $[0, 1/4]$, identiquement égale à 0 sur le fermé $\{f - g_1 = 0\}$ et identiquement égale à $1/4$ sur le fermé $\{f - g_1 \geq 1/4\}$; on a alors $0 \leq f - g_1 - g_2 \leq 1/4$ sur F .

Par récurrence, on construit ainsi une suite de fonctions $(g_n)_{n \in \mathbb{N}}$ telle que $|g_n| \leq 2^{-n}$ et $|f - (g_0 + \dots + g_n)| \leq 2^{-n}$ sur F . Cette série converge uniformément dans X tout entier vers une fonction continue g , qui coïncide avec f sur F .

Si maintenant f est non bornée, on applique le résultat précédent à $\tilde{f} := f/(1 + f)$, construisant ainsi une fonction continue \tilde{g} sur X , à valeurs dans $[0, 1]$. L'ensemble

$F' := \{\tilde{g} = 1\}$ est un fermé disjoint de F , on peut donc trouver une fonction h continue, à valeurs dans $[0, 1]$, valant 0 sur F' et 1 sur F ; la fonction $h\tilde{g}$ est alors à valeurs dans $[0, 1]$, et $g := (h\tilde{g})/(h\tilde{g} - 1)$ est une extension continue de f . \square

Enfin, voici pour conclure un résultat utile de séparation : *soient F_0 et F_1 des fermés disjoints dans un espace métrique ; alors on peut trouver des ouverts disjoints O_0 et O_1 tels que $F_0 \subset O_0$, $F_1 \subset O_1$.* En effet, on sait construire une fonction continue f à valeurs dans $[0, 1]$, valant 0 sur F_0 et 1 sur F_1 ; il suffit de poser $O_0 = \{x; f(x) < 1/3\}$, $O_1 = \{x; f(x) > 2/3\}$.

I-2.4. Espaces produits. Si $(X_t)_{t \in T}$ est une famille d'espaces topologiques, indexée par un ensemble quelconque, considérons leur produit cartésien $X = \prod X_t$: c'est l'ensemble des fonctions x définies sur T , telles que $x(t) \in X_t$ pour tout t . Dans cet ensemble on peut définir les **cylindres** : si A_{t_k} est une partie de X_{t_k} , pour $1 \leq k \leq K$, alors

$$C_{(t_1, \dots, t_K)}(A_{t_1}, \dots, A_{t_K}) := \{x \in X; x_{t_k} \in A_{t_k}\}.$$

On peut alors définir la **topologie produit** comme la topologie engendrée par les ouverts cylindriques, i.e. les ouverts de la forme $C(O_{t_1}, \dots, O_{t_K})$, chaque O_{t_k} étant un ouvert de X_{t_k} .

Si l'ensemble T est quelconque, la topologie produit est en général non métrique ; un exemple typique est $[0, 1]^{[0, 1]}$. En revanche, un produit *dénombrable* d'espaces métriques reste un espace métrique. Précisons un peu les choses. Soit $(X_k, d_k)_{k \in \mathbb{N}}$ une famille dénombrable d'espaces métriques, et $X = \prod X_k$. La topologie produit sur X est engendrée par les cylindres de la forme $C(A_1, \dots, A_n)$, où chaque A_i est un ouvert de X_i . Dire qu'une suite $(x^k)_{k \in \mathbb{N}}$ d'éléments de X converge vers $x \in X$ revient donc à dire que pour tout indice i , la suite $(x_i^k)_{k \in \mathbb{N}}$ converge vers x_i quand $k \rightarrow \infty$. Il est alors facile de vérifier que X est un espace métrique quand on le munit de la distance

$$d_X(x, y) = \sup_{i \in \mathbb{N}} 2^{-i} \frac{d_i(x_i, y_i)}{1 + d_i(x_i, y_i)}.$$

(Noter que la distance $d_i/(1 + d_i)$ est topologiquement équivalente à d_i , tout en étant automatiquement bornée.) En pratique, la convergence dans l'espace X est la "convergence composante par composante".

THÉORÈME I-30 (produits dénombrables d'espaces métriques). *Soit $(X_k, d_k)_{k \in \mathbb{N}}$ une famille dénombrable d'espaces métriques, et soit $X = \prod X_k$, muni de la topologie produit. Alors*

- (i) *Si chaque X_k est compact, alors X est compact ;*
- (ii) *Si chaque X_k est polonais, alors X est polonais.*

REMARQUE I-31. Si l'on admet l'axiome du choix, alors l'énoncé (i) se généralise à des familles arbitraires de compacts, éventuellement non dénombrables ; c'est un célèbre résultat de topologie générale appelé **théorème de Tychonov**. La démonstration de ce théorème est un remarquable exercice d'"abstract nonsense", que l'on pourra trouver dans de nombreux ouvrages de référence, par exemple [Dunford-Schwartz]. La preuve de l'énoncé (i) est beaucoup plus simple.

DÉMONSTRATION. (i) C'est l'occasion d'introduire le concept d'**extraction diagonale** (ou argument diagonal de Cantor), qui servira souvent par la suite. Soit $(x^n)_{n \in \mathbb{N}}$ une suite d'éléments de X ; chaque x^n est une suite $(x_k^n)_{k \in \mathbb{N}}$, avec $x_k^n \in X_k$.

De la famille des $(x_1^n)_{n \in \mathbb{N}}$, à valeurs dans le compact K_1 , on peut extraire une sous-suite convergente, notée $x_1^{\varphi_1(n)}$. De la famille $x_2^{\varphi_1(n)}$, à valeurs dans le compact K_2 , on peut extraire une sous-suite convergente notée $x_2^{\varphi_1 \circ \varphi_2(n)}$. Par récurrence, on construit des applications strictement croissantes $\varphi_k : \mathbb{N} \rightarrow \mathbb{N}$, telles que la suite $x_k^{\varphi_1 \circ \dots \circ \varphi_k(n)}$ est convergente dans X_k . On pose alors

$$\varphi(n) = \varphi_1 \circ \varphi_2 \circ \dots \circ \varphi_n(n).$$

Pour tout $k \leq n$, on peut écrire

$$\varphi(n) = (\varphi_1 \circ \dots \circ \varphi_k) \circ \psi_k(n),$$

où ψ_k est une fonction croissante. Il s'ensuit que, pour $n \geq k$, la suite $(x_k^{\varphi(n)})$ est extraite de $x_k^{\varphi_1 \circ \dots \circ \varphi_k(n)}$, et converge donc dans X_k vers une limite $x_k \in X_k$. Par définition de la topologie cylindrique, la suite (de suites) $x^{\varphi(n)}$ converge vers la suite $x = (x_k)_{k \in \mathbb{N}}$. La suite $(x^n)_{n \in \mathbb{N}}$ admet donc x pour valeur d'adhérence, ce qui prouve la compacité de X .

(ii) Soit d_n une métrique rendant X_n complet ; on vérifie alors que la métrique

$$d(x, y) := \sup_{n \in \mathbb{N}} \frac{2^{-n} d_n(x_n, y_n)}{1 + d_n(x_n, y_n)}$$

métrise la topologie produit de X et le rend complet.

Par ailleurs, si $(x_k^n)_{k \in \mathbb{N}}$ est une suite dense dans X_n , on vérifie que la famille des suites

$$(x_1^{k_1}, x_2^{k_2}, \dots, x_N^{k_N}, x_{N+1}^1, x_{N+2}^1, x_{N+3}^1, \dots)$$

où $(k_1, \dots, k_N) \in \mathbb{N}^N$ et $N \in \mathbb{N}$, est dénombrable, et dense dans X (c'est une suite de suites !). L'espace X est donc séparable. \square

I-2.5. Quel cadre topologique pour la théorie de la mesure ? Même si leurs fondements axiomatiques présentent des similitudes, théorie de la mesure et topologie font souvent mauvais ménage, et peuvent mener à des descriptions qualitatives très différentes. Ainsi, un ensemble peut être “gros” au sens de la topologie (par exemple un ensemble gras au sens de Baire, i.e. une intersection dense d'ouverts) et très petit au sens de la mesure (i.e. de mesure très petite) ; et inversement.

Dans de nombreux domaines, les notions de “généricité” au sens topologique et au sens de la théorie de la mesure sont différentes ; un exemple célèbre est le théorème KAM en mécanique Hamiltonienne, pour lequel l'instabilité générique (au sens topologique) est compatible avec une stabilité très probable.

Pour autant, théorie de la mesure et topologie ne sont pas des concepts étrangers. Nous verrons dans la section suivante que les propriétés topologiques d'un espace métrique déterminent en partie ses propriétés en tant qu'espace mesuré, quand on le munit de la tribu borélienne. Il est donc légitime de se demander s'il existe un cadre topologique naturel pour développer la théorie de la mesure.

Dans les années 1950 et 1960, il a pu sembler qu'un tel cadre était celui des **espaces localement compacts**, non nécessairement métriques. Plusieurs résultats fondamentaux ont été établis pour ces espaces ; parmi les plus remarquables se trouvent le Théorème de Représentation de Riesz, dont je parlerai au Chapitre II ; et le Théorème de Haar, que nous rencontrerons vers la fin du cours. La théorie de la mesure dans les espaces localement compacts occupait alors une place importante

dans nombre de traités de référence tels que ceux de Bourbaki et Halmos, et même dans l'ouvrage plus concis de Rudin, dont la popularité reste intacte.

En revanche, les probabilistes n'ont jamais pu admettre ce cadre, qui exclut les espaces fonctionnels naturels tels que l'espace de Wiener (connu depuis les années 1920). Depuis Kolmogorov, l'essentiel de la théorie des probabilités a été développé dans le cadre des espaces métriques séparables, le plus souvent complets ; autrement dit, des **espaces polonais**.

Deux théories concurrentes se sont donc développées parallèlement au cours du vingtième siècle : les espaces localement compacts d'un côté, les espaces polonais de l'autre. Pour apprécier un peu mieux cette distinction, voici quelques exemples :

- un espace métrique compact est bien sûr localement compact, et automatiquement polonais ;
- l'espace euclidien \mathbb{R}^n , ou plus généralement n'importe quelle variété riemannienne complète, est à la fois un espace localement compact et un espace polonais ;
- l'espace $[0, 1]^{[0, 1]}$, muni de la topologie de la convergence simple, est compact mais n'est pas polonais (car non métrique) ;
- l'espace $C([0, 1]; \mathbb{R})$, muni de la topologie de la convergence uniforme, est polonais mais non localement compact.

Il semble clair aujourd'hui que le cadre des espaces polonais est plus naturel et concerne une communauté scientifique considérablement plus importante. En outre, la théorie de la mesure dans des espaces non métriques (même compacts) mène à diverses pathologies [Dudley, Appendice E]. Pour toutes ces raisons, il semble maintenant que le "bon" contexte mathématique de la théorie de la mesure soit celui des espaces polonais. Cela n'empêche pas que des hypothèses *additionnelles* de compacité locale aient des conséquences fort utiles.

Pour autant, je n'ai pas banni complètement de ces notes les énoncés faisant intervenir des espaces localement compacts abstraits ; la principale motivation en est la volonté de préserver toute la splendeur (!) des théorèmes de Riesz et de Haar. De manière générale, je proposerai donc des preuves complètes dans le cas des espaces polonais, et des preuves presque complètes dans le cas localement compact ; le lecteur curieux pourra en reconstituer les détails.

I-2.6. Pourquoi les espaces polonais sont-ils agréables ? Les axiomes des espaces polonais en font des espaces particulièrement bien adaptés à la propriété de σ -additivité des mesures : dans de nombreux problèmes, on peut se ramener à une question portant sur une famille **dénombrable** d'ensembles simples tels que des boules. Voici un bon exemple.

THÉORÈME I-32 (ouverts et compacts d'un polonais). *Soit (X, d) un espace métrique séparable, soit $(x_n)_{n \in \mathbb{N}}$ une suite dense dans X , et soit $(\varepsilon_k)_{k \in \mathbb{N}}$ une suite de nombres positifs décroissant vers 0. Soit \mathcal{B}_k l'ensemble des boules ouvertes de rayon x_n , de rayon ε_k , et \mathcal{B} la réunion des \mathcal{B}_k . Alors*

(i) *une partie O de X est ouverte si et seulement si elle est union dénombrable d'éléments de \mathcal{B} ;*

(ii) *si X est en outre complet (donc polonais), alors une partie K de X est compacte si et seulement si elle s'écrit comme une intersection d'unions finies d'adhérences*

d'éléments de \mathcal{B}_k ; en d'autres termes,

$$(4) \quad K = \bigcap_{k \in \mathbb{N}} \bigcup_{1 \leq j \leq N_k} \overline{B_{\varepsilon_k}(x_{n(k,j)})}.$$

PREUVE DU THÉORÈME I-32. (i) Il est clair qu'une union d'éléments de \mathcal{B} est ouverte. Pour vérifier la réciproque, il suffit de prouver que tout $x \in O$ appartient à un élément de \mathcal{B} . Puisque O est ouvert, il existe $r > 0$ tel que $B_r(x) \subset O$. Soit k tel que $\varepsilon_k < r/2$, et soit n tel que $d(x_n, x) < \varepsilon_k$. Nous allons constater que $B_{\varepsilon_k}(x_n) \subset O$, ce qui conclura l'argument. Soit donc y tel que $d(x_n, y) < \varepsilon_k$; alors $d(x, y) \leq d(x, x_n) + d(x_n, y) < \varepsilon_k + \varepsilon_k < r$, donc $y \in B_r(x)$ et donc $y \in O$.

(ii) Montrons qu'une partie K de la forme (4) est compacte. Une union finie de parties fermées étant fermée, K s'écrit comme une intersection (dénombrable) de parties fermées ; c'est donc un fermé. Si l'on parvient à montrer que toute suite à valeurs dans K converge dans X , on saura que la limite est aussi dans K , et le critère de Bolzano-Weierstrass assurera que K est compact.

Soit donc $(y_\ell)_{\ell \in \mathbb{N}}$ une suite à valeurs dans K . Par hypothèse, elle prend ses valeurs dans un nombre fini de boules de rayon ε_1 ; l'une de ces boules au moins contient donc une infinité de termes de la suite. On peut donc extraire de (y_ℓ) une sous-suite dont tous les éléments sont à distance au plus ε_1 les uns des autres. Mais la suite (y_ℓ) ainsi extraite est également à valeurs dans une union finie de boules de rayon ε_2 ; l'une de ces boules au moins contient donc une infinité de termes de la suite, et on peut extraire à nouveau une sous-suite dont tous les éléments sont à distance au plus ε_2 les uns des autres. On continue ainsi le processus : par un procédé d'*extraction diagonale*, il est possible de construire une suite extraite, toujours notée (y_ℓ) , telle que tous les y_ℓ pour $\ell \geq k$ sont à distance au plus ε_k les uns des autres. C'est donc une suite de Cauchy, et grâce à l'hypothèse de complétude elle converge dans X , ce qui conclut l'argument.

Réciproquement, soit K un compact, montrons qu'il peut s'écrire sous la forme (4). Pour tout k , et pour tout $x \in K$, on peut trouver n tel que $d(x, x_n) < \varepsilon_k$; on peut donc inclure K dans l'union des $B_{\varepsilon_k}(x_n)$, avec $d(x_n, K) < \varepsilon_k$. Par compacité on peut extraire un sous-recouvrement ouvert. Il existe donc des éléments $x_{n(k,j)}$ ($1 \leq j \leq N_k$) tels que K soit inclus dans l'union des $B_{\varepsilon_k}(x_{n(k,j)})$, a fortiori dans l'union des $\overline{B_{\varepsilon_k}(x_{n(k,j)})}$, avec $d(x_{n(k,j)}, K) < \varepsilon_k$. Cela étant valable pour tout k , on a

$$K \subset K' := \bigcap_{k \in \mathbb{N}} \bigcup_{1 \leq j \leq N_k} \overline{B_{\varepsilon_k}(x_{n(k,j)})}.$$

Soit maintenant $y \in K'$, et soit $k \in \mathbb{N}$. Par hypothèse, il existe x_n tel que $y \in \overline{B_{\varepsilon_k}(x_n)}$, avec $d(x_n, K) < \varepsilon_k$. On a alors $d(x_n, y) \leq \varepsilon_k$, et donc $d(y, K) \leq \varepsilon_k + \varepsilon_k = 2\varepsilon_k$. Puisque k était arbitraire, $d(y, K) = 0$, ce qui entraîne $y \in K$. On conclut que $K' = K$, ce qui était notre but. \square

I-2.7. Pourquoi les espaces localement compacts sont-ils agréables ? La popularité des espaces localement compacts tient pour beaucoup à ce que dans de tels espaces on peut, dans de nombreuses situations, remplacer les ensembles ouverts (identifiés à leurs fonctions indicatrices) par des "fonctions plateaux", à valeurs dans $[0, 1]$, **continues** et à support compact. C'est ce qu'expriment les deux théorèmes suivants [Rudin, Théorèmes 2.12 et 2.13].

THÉORÈME I-33 (lemme d'Urysohn). *Soit X un espace séparé localement compact, O un ouvert et K un compact de X , $K \subset O$. Alors il existe une fonction f , continue, à valeurs dans $[0, 1]$, qui vaut identiquement 1 au voisinage de K , et dont le support est compact et inclus dans O . En particulier,*

$$1_K \leq f \leq 1_O.$$

THÉORÈME I-34 (partition de l'unité). *Soient X un espace séparé localement compact, K un compact de X , et O_1, \dots, O_n une collection finie d'ouverts, tels que $K \subset \cup O_k$. Alors il existe des fonctions continues f_1, \dots, f_n , à valeurs dans $[0, 1]$, telles que chaque f_k ait son support compact et inclus dans O_k , et*

$$x \in K \implies \sum_k f_k(x) = 1.$$

Si l'on se donne un recouvrement quelconque \mathcal{O} de K par des ouverts, alors on peut en extraire un sous-recouvrement fini $\{O_1, \dots, O_n\}$, et la conclusion précédente reste vraie.

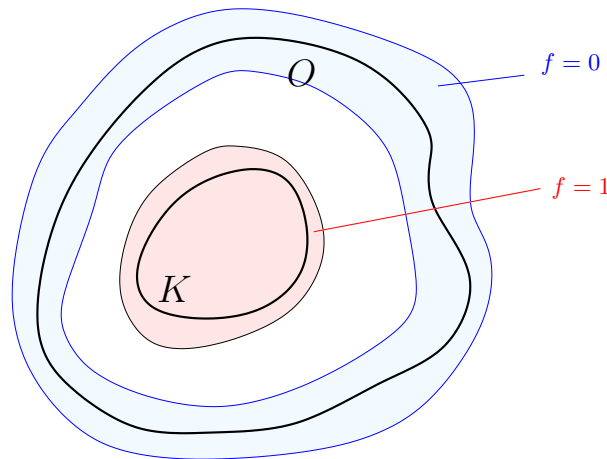


FIG. 4. Le lemme d'Urysohn

Le lemme suivant est également utile. On peut le voir comme une conséquence directe du lemme d'Urysohn (pourquoi ?), mais on peut aussi considérer qu'il précède logiquement cet énoncé.

LEMME I-35 (voisinages compacts). *Soient X un espace localement compact, K un compact de X , et O un voisinage ouvert de K . Alors il existe un compact K' et un ouvert O' de X tels que*

$$K \subset O' \subset K' \subset O.$$

PREUVE DU LEMME I-35 DANS LE CAS MÉTRIQUE. Soit $F := X \setminus O$. Pour tout $x \in K$, on a $d(x, F) > 0$; on peut donc trouver $r_x > 0$ tel que la boule $B_{r_x}(x)$ n'intersecte pas $X \setminus O$, et soit par conséquent incluse dans O . Par hypothèse, il existe également un voisinage compact K_x de x . Considérons $C_x := \overline{B_{r_x/2}}(x) \cap K_x$: c'est un voisinage compact de x , inclus dans O . Soit V_x un sous-ensemble ouvert de C_x contenant x : les V_x recouvrent K ; par compacité on peut donc en extraire

un sous-recouvrement fini V_{x_1}, \dots, V_{x_K} , et la famille des C_{x_k} recouvre également K . L'union des C_{x_k} est alors un voisinage compact de K , inclus dans O . \square

PREUVE DU THÉORÈME I-33 DANS LE CAS MÉTRIQUE. Soient K' et O' comme dans le Lemme I-35. On pose alors

$$f(x) := \frac{d(x, X \setminus K')}{d(x, X \setminus K') + d(x, O')}$$

(noter que le dénominateur ne s'annule jamais). \square

PREUVE DU THÉORÈME I-34 DANS LE CAS MÉTRIQUE. On ne démontrera que la première partie de l'énoncé, la deuxième en découlant facilement. Pour tout j , on note F_j le complémentaire dans X de l'union des ensembles O_i , pour $i \neq j$: c'est alors un fermé, et

$$K \cap F_j \subset \left(\bigcap_i O_i \right) \setminus \left(\bigcap_{i \neq j} O_i \right) \subset O_j.$$

Grâce au Théorème I-33, on peut trouver une fonction f_j à support compact dans O_j , à valeurs dans $[0, 1]$, qui vaille identiquement 1 sur un voisinage de $K \cap F_j$. On définit alors

$$f_j(x) := \begin{cases} f_j(x) & \text{si } x \in K \cap F_j \\ \frac{f_j(x)}{\sum_{i=1}^n f_i(x)} & \text{si } x \in K \cap \left(\bigcup_{i \neq j} O_i \right) \end{cases}$$

et on vérifie que cette famille remplit toutes les conditions requises. \square

Voici pour finir un théorème énonçant que l'on peut "épuiser" un espace localement compact et σ -compact par une famille de compacts "gentiment emboîtés" (souvent appelée *suite exhaustive* de compacts) :

THÉORÈME I-36. *Soit X un espace localement compact et σ -compact. Alors X peut s'écrire comme l'union croissante d'une famille de compacts $(K_n)_{n \in \mathbb{N}}$, telle que K_{n+1} soit voisinage de K_n . Pour tout compact K de X , il existe alors $n_0 \in \mathbb{N}$ tel que K soit inclus dans l'intérieur de K_n pour $n \geq n_0$.*

DÉMONSTRATION. Par hypothèse, on peut trouver des compacts $(C_n)_{n \in \mathbb{N}}$ dont l'union est X entier. Posons $K_1 = C_1$. On va montrer qu'il existe un compact K_2 qui contienne un voisinage de K_1 , et C_2 .

Pour chaque n on peut construire un ouvert O_n et un compact C'_n , de sorte que $C_n \subset O_n \subset C'_n$. Il est clair que les O_n recouvrent X . En particulier, le compact C_1 est recouvert par un nombre fini des O_n : il existe $N_1 \in \mathbb{N}$ tel que $C_1 \subset O_1 \cup O_2 \cup \dots \cup O_{N_1}$. En particulier, le compact $C'_1 \cup \dots \cup C'_{N_1}$ est voisinage de C_1 . Il suffit alors de poser $K_2 := C'_1 \cup \dots \cup C'_{N_1}$.

En répétant ce raisonnement, on construit par récurrence une suite $(K_n)_{n \in \mathbb{N}}$ telle que K_{n+1} soit voisinage de K_n et contienne $C_1 \cup \dots \cup C_{n+1}$. La réunion des K_n est donc X tout entier.

Soit V_n l'intérieur de K_n . Puisque V_n contient K_{n-1} , l'union croissante des V_n est X tout entier. Si K est un compact de X , il est donc inclus dans V_n pour n assez grand, ce qui conclut la preuve. \square

I-3. Régularité des espaces mesurés

Comme on l'a dit, les mesures peuvent être étudiées dans un cadre très général. Cependant, dans la pratique on utilise souvent certaines propriétés bien utiles de "régularité", en un sens vague, qui mêlent des hypothèses topologiques et des hypothèses de théorie de la mesure.

I-3.1. Vocabulaire de base.

DÉFINITION I-37 (finitude et σ -finitude). *Une mesure μ sur un espace mesuré X est dite finie si X est de mesure finie ; elle est dite σ -finie si X peut s'écrire comme une union dénombrable d'ensembles A_k de mesure finie.*

REMARQUE I-38. Certains théorèmes importants d'intégration sur les espaces produits utiliseront crucialement des hypothèses de σ -finitude.

DÉFINITION I-39 (probabilité). *Un espace mesuré (X, μ) est appelé espace de probabilité si $\mu[X] = 1$.*

REMARQUE I-40. Cette définition semble particulièrement triviale, mais il ne faut pas oublier qu'il a fallu attendre longtemps avant que l'on comprenne que la théorie de la mesure était un cadre conceptuel naturel pour développer la théorie des probabilités. Cette intuition est due à Kolmogorov et (indépendamment) Ulam.

DÉFINITION I-41 (atome). *On dit que $x \in X$ est un atome pour la mesure μ si $\mu[\{x\}] > 0$.*

Le choix de cette terminologie est transparent : un atome est une partie que l'on ne peut découper, au sens de la théorie de la mesure, en morceaux plus petits.

DÉFINITION I-42 (concentration, négligeabilité). *Soit (X, \mathcal{A}, μ) un espace mesuré.*

(i) *Soit C un sous-ensemble quelconque de X ; on dit que μ est concentrée sur C si, pour toute partie mesurable A contenant C , on a $\mu[X \setminus A] = 0$.*

(ii) *Soit N un sous-ensemble quelconque de X ; on dit que N est μ -négligeable si N est contenu dans un ensemble mesurable A tel que $\mu[A] = 0$.*

(iii) *Soit C un sous-ensemble quelconque de X ; on dit que μ charge C si, pour toute partie mesurable A contenant C , on a $\mu[C] > 0$.*

REMARQUE I-43. Ces définitions se simplifient quand on se restreint à des parties mesurables : si A est mesurable,

- μ est concentrée sur A si et seulement si $\mu[X \setminus A] = 0$;
- A est μ -négligeable si et seulement si $\mu[A] = 0$;
- μ charge C si et seulement si $\mu[C] > 0$.

EXEMPLE I-44. La mesure δ_x est concentrée sur $\{x\}$, qui n'est pas forcément mesurable.

Intuitivement, les ensembles négligeables sont ceux qui ne devraient jouer aucun rôle en intégration. Cependant, il est parfois délicat de traduire cette intuition quand ces ensembles ne sont pas mesurables. Ceci motive la notion suivante :

DÉFINITION I-45 (complétude). *Un espace mesuré est dit complet s'il possède la propriété suivante : si A est négligeable et B est inclus dans A , alors B est mesurable (et donc automatiquement négligeable).*

- REMARQUES I-46. (i) Le sens du mot “complet” est différent de celui qu’il a dans “espace métrique complet”.
- (ii) La complétude est une propriété subtile, parfois bien commode, parfois source de complications infinies. On verra plus tard que l’on peut toujours “compléter” une mesure.

I-3.2. Mesures de Borel et régularité.

DÉFINITION I-47 (mesure de Borel). *Soit X un espace topologique. La tribu $\mathcal{B}(X)$ engendrée par les ouverts de X est appelée tribu borélienne de X , et les mesures définies sur cette tribu sont dites mesures de Borel.*

La définition suivante est particulièrement importante. Elle exprime le fait que les ensembles boréliens, pour compliqués qu’ils soient, peuvent être approchés au sens de la mesure, de l’intérieur par des ensembles “topologiquement petits” (des compacts), et de l’extérieur par des ensembles “topologiquement gros” (des ouverts).

DÉFINITION I-48 (régularité). *Soit (X, μ) un espace topologique et μ une mesure définie sur une tribu \mathcal{A} de X contenant la tribu borélienne. On dit que μ est régulière si elle vérifie la propriété caractéristique suivante : pour tout ensemble mesurable $A \in \mathcal{A}$ on a*

$$\begin{aligned}\mu[A] &= \inf \{ \mu[O]; O \text{ ouvert}, A \subset O \} \\ &= \sup \{ \mu[K]; K \text{ compact}, K \subset A \}.\end{aligned}$$

La régularité d’une mesure implique automatiquement que les ensembles mesurables se décomposent en une partie “régulière” et une partie de mesure nulle, au sens de la proposition suivante.

PROPOSITION I-49 (mesurabilité, F_σ et G_δ). *Soit X un espace topologique et soit μ une mesure régulière sur X , définie sur une σ -algèbre \mathcal{A} contenant la tribu borélienne. Alors, toute partie $A \in \mathcal{A}$ (et en particulier tout Borélien), de mesure finie, peut s’écrire sous la forme $F \cup N$, où F est une union dénombrable de fermés (un F_σ) et N un ensemble négligeable; elle peut également s’écrire sous la forme $G \setminus N'$, où G est une intersection dénombrable d’ouverts (un G_δ) et N' un ensemble négligeable.*

DÉMONSTRATION. Par régularité, on peut trouver une suite de compacts K_j et d’ouverts O_j tels que $K_j \subset A \subset O_j$ et $\mu[K_j] \rightarrow \mu[A]$, $\mu[O_j] \rightarrow \mu[A]$. Quitte à poser $K'_1 = K_1$, $K'_j = K'_{j-1} \cup K_j$, $O'_1 = O_1$, $O'_j = O'_{j-1} \cap O_j$, on peut supposer que les K_j sont croissants et les O_j décroissants. On pose alors $F = \bigcup K_j$ et $G = \bigcap O_j$. La conclusion découle de la σ -additivité de μ . \square

La régularité va souvent de pair avec la propriété suivante :

DÉFINITION I-50 (finitude sur les compacts). *Une mesure de Borel μ sur un espace topologique X est dite finie sur les compacts⁵ si pour tout compact K de X on a $\mu[K] < +\infty$.*

⁵Une mesure de Borel finie sur les compacts est parfois appelée mesure de Radon; parfois on impose aussi des propriétés de régularité, au sens indiqué plus loin [Evans-Gariepy].

EXEMPLE I-51. La mesure de comptage sur \mathbb{R} n'est ni finie sur les compacts, ni régulière, puisque la mesure de tout segment non trivial, et de tout ouvert non trivial, est $+\infty$.

Enfin la régularité est liée de manière quelque peu subtile à la propriété de σ -additivité, comme le montre l'énoncé suivant (que l'on peut considérer à ce stade comme une curiosité, mais qui s'avèrera utile dans le Chapitre ??) :

PROPOSITION I-52. Soient \mathcal{A} et \mathcal{B} deux algèbres, avec $\mathcal{A} \subset \mathcal{B}$, et $\mu : \mathcal{B} \rightarrow \mathbb{R}_+$ une fonction additive d'ensembles, vérifiant la propriété de régularité intérieure partielle

$$(5) \quad \forall B \in \mathcal{B} \quad \mu[B] = \sup \left\{ \mu[K]; K \text{ compact}, K \subset B, K \in \mathcal{A} \right\}.$$

Alors μ est σ -additive sur \mathcal{A} .

REMARQUE I-53. La Remarque VI-62 montrera que l'hypothèse 5 est cruciale.

PREUVE DE LA PROPOSITION I-52. Soit μ vérifiant les hypothèses du théorème. Par additivité on a, pour toute famille $(A_k)_{k \in \mathbb{N}}$ d'ensembles mesurables disjoints, $\sum_{1 \leq k \leq N} \mu[A_k] \leq \mu[\cup_{k \in \mathbb{N}} A_k]$, d'où

$$\sum_{k \in \mathbb{N}} \mu[A_k] \leq \mu[\bigcup_{k \in \mathbb{N}} A_k].$$

Si μ n'est pas σ -additive, il existe des $(A_k)_{k \in \mathbb{N}}$ disjoints tels que $\sum_{k \in \mathbb{N}} \mu[A_k] < \mu[\cup_{k \in \mathbb{N}} A_k]$, en d'autres termes on peut trouver $\delta > 0$ tel que pour tout $N \in \mathbb{N}$,

$$\sum_{k=1}^N \mu[A_k] \leq \mu[\cup_{k \in \mathbb{N}} A_k] - \delta = \sum_{k=1}^N \mu[A_k] + \mu[\bigcup_{k \geq N+1} A_k];$$

d'où $\mu[\cup_{k \geq N+1} A_k] \geq \delta$ (ici on utilise le fait que μ est à valeurs finies). Posons $C_\ell = \cup_{k \geq \ell+1} A_k$: alors $(C_\ell)_{\ell \in \mathbb{N}}$ est une suite décroissante de compacts d'intersection vide, et $\mu[C_\ell] \geq \delta$ pour tout ℓ .

Pour chaque $\ell \in \mathbb{N}$, soit K_ℓ un compact tel que $K_\ell \subset C_\ell$, $K_\ell \in \mathcal{A}$, et $\mu[C_\ell \setminus K_\ell] \leq 2^{-(\ell+1)}\delta$. Alors $K_1 \cap \dots \cap K_\ell \in \mathcal{A}$ et

$$\mu[K_1 \cap \dots \cap K_\ell] \geq \mu[C_\ell] - \sum 2^{-(j+1)}\delta \geq \delta - \delta/2 = \delta/2;$$

en particulier $K_1 \cap \dots \cap K_\ell$ est non vide. Les compacts $(K_1 \cap \dots \cap K_\ell)_{\ell \in \mathbb{N}}$ forment une suite décroissante de compacts non vides, leur intersection est donc non vide, en contradiction avec le fait que les C_ℓ eux-mêmes sont d'intersection vide. \square

I-3.3. Théorèmes de régularité automatique. Les espaces polonais d'une part, les espaces localement compacts d'autre part, jouissent de propriétés bien commodes. En particulier, on a le résultat suivant, qui peut paraître surprenant au premier abord :

THÉORÈME I-54 (Régularité des mesures sur les espaces polonais). Soit X un espace polonais muni d'une mesure borélienne μ , σ -finie. Alors μ est automatiquement régulière, et concentrée sur un ensemble σ -compact.

Dans le cas où $\mu[X] < +\infty$, la dernière assertion de ce théorème est connue sous le nom de **lemme d'Ulam**. Voici un corollaire immédiat du Théorème I-54.

COROLLAIRE I-55 (Régularité des mesures sur \mathbb{R}^n). Soit μ une mesure de Borel sur \mathbb{R}^n , finie sur les compacts ; alors μ est régulière.

Avant de démontrer le Théorème I-54, mentionnons sans démonstration une variante qui s'applique à des espaces topologiques localement compacts [Rudin, Théorème 2.18], et implique également le Corollaire I-55.

THÉORÈME I-56 (Régularité des mesures dans un espace localement compact). *Soit X un espace séparé localement compact, dans lequel tout ouvert est σ -compact, muni d'une mesure de Borel μ , finie sur les compacts. Alors μ est automatiquement régulière.*

REMARQUE I-57. Il est facile de vérifier que les ouverts de \mathbb{R}^n , ou d'une variété de dimension finie, sont σ -compacts. Par ailleurs, on trouvera dans [Rudin, Chapitre 2, exercice 18] un contre-exemple montrant que la conclusion du Théorème I-56 n'est pas forcément vraie sans l'hypothèse quelque peu étrange de σ -compacité des ouverts.

Voici maintenant une preuve du Théorème I-54, inspirée de celle que l'on trouve par exemple dans [Dudley, p. 225]. La démonstration utilisera un résultat dont la preuve se trouve plus loin (Théorème I-68), mais bien sûr il n'y a pas de cercle vicieux !

PREUVE DU THÉORÈME I-54. 1. Dans le début de cette preuve, on suppose que μ est finie. Soit $\varepsilon > 0$, montrons qu'il existe un compact K_ε tel que $\mu[X \setminus K_\varepsilon] \leq \varepsilon$. De ce résultat il sera facile de déduire (exercice) que μ est concentrée sur l'ensemble σ -compact $S := \bigcup_{k \in \mathbb{N}} K_{1/k}$.

Par hypothèse, il existe une suite $(x_n)_{n \in \mathbb{N}}$ dense dans X . En particulier,

$$X = \bigcup_{n \geq 1} B(x_n, 1),$$

et donc $\mu[X] = \lim_{n \rightarrow \infty} \mu[\bigcup_{k \leq n} B(x_k, 1)]$. Comme $\mu[X] < +\infty$, on peut donc trouver n_1 tel que

$$\mu[X \setminus (\bigcup_{k \leq n_1} B(x_k, 1))] \leq \frac{\varepsilon}{2}.$$

De même, pour tout j on peut trouver n_j tel que

$$\mu[X \setminus (\bigcup_{k \leq n_j} B(x_k, 1/j))] \leq \frac{\varepsilon}{2^j}.$$

Posons

$$K_\varepsilon := \bigcap_{j \geq 1} \left(\bigcup_{k \leq n_j} \overline{B(x_k, 1/j)} \right).$$

D'une part, K_ε est totalement borné : en effet, si $\delta > 0$ est donné on peut choisir $j \geq 1/\delta$ et on a

$$K_\varepsilon \subset \bigcup_{k \leq n_j} \overline{B(x_k, 1/j)} \subset \bigcup_{k \leq n_j} \overline{B(x_k, \delta)}.$$

D'autre part, K_ε est fermé car intersection de fermés ; comme l'espace ambiant X est complet, il s'ensuit que K_ε est également complet. Etant complet et totalement borné, il est compact. Enfin,

$$\begin{aligned} \mu[X \setminus K_\varepsilon] &= \mu[\bigcup_j (X \setminus \bigcup_{k \leq n_j} \overline{B(x_k, 1/j)})] \\ &\leq \mu[\bigcup_j (X \setminus \bigcup_{k \leq n_j} B(x_k, 1/j))] \\ &\leq \sum_{j \geq 1} \mu[X \setminus (\bigcup_{k \leq n_j} B(x_k, 1/j))] \\ &\leq \varepsilon \sum_{j \geq 1} 2^{-j} = \varepsilon. \end{aligned}$$

2. Montrons maintenant que μ est régulière. Soit A un ensemble mesurable, et $\varepsilon > 0$, nous voulons montrer qu'il existe un ouvert O contenant A et un compact K inclus dans A tels que

$$\mu[O] - \varepsilon \leq \mu[A] \leq \mu[K] + \varepsilon.$$

On va d'abord supposer que X est compact : il y a alors identité entre compacts et fermés. Définissons \mathcal{F} comme la famille de toutes les parties boréliennes A de X telles que, pour tout $\varepsilon > 0$ on peut trouver un ouvert O contenant A , et un fermé F contenu dans A , satisfaisant à

$$(6) \quad \mu[O] - \varepsilon \leq \mu[A] \leq \mu[F] + \varepsilon.$$

Clairement, notre but est de montrer que \mathcal{F} coïncide avec l'ensemble de la tribu des boréliens. Il est très facile de montrer que \mathcal{F} est stable par passage au complémentaire. Il est également stable par union croissante : en effet, si A_k est une famille croissante d'éléments de \mathcal{F} , et $A = \cup A_k$, on peut choisir k_0 tel que $\mu[A_{k_0}] \geq \mu[A] - \varepsilon/2$, et un compact K contenu dans A_{k_0} tel que $\mu[A_{k_0}] \leq \mu[K] + \varepsilon/2$, ce qui impliquera

$$\mu[A] \leq \mu[A_{k_0}] + \varepsilon/2 \leq \mu[K] + \varepsilon.$$

On peut également, pour tout $k \in \mathbb{N}$, introduire un ouvert O_k contenant A_k , tel que $\mu[O_k] \leq \mu[A_k] + \varepsilon 2^{-k}$; alors $O = \cup O_k$ vérifie $A \subset O$ et $O \setminus A \subset \cup (O_k \setminus A)$, donc

$$\mu[O \setminus A] \leq \sum_{k \in \mathbb{N}} \varepsilon 2^{-k} = \varepsilon.$$

3. Si l'on montre que \mathcal{F} contient tous les ensembles ouverts, le Lemme de Classe monotone (Théorème I-68, démontré plus loin) impliquera que \mathcal{F} est la tribu borélienne tout entière. Si A est ouvert, l'inégalité de gauche dans (6) est trivialement vérifiée par $O = A$; pour montrer l'inégalité de droite, il suffit de prouver l'existence d'une famille croissante d'ensembles fermés F_k inclus dans A tels que $\mu[F_k] \rightarrow \mu[A]$. Introduisons, pour $k \in \mathbb{N}$,

$$F_k := \left\{ x \in A; d(x, X \setminus A) \geq 1/k \right\}.$$

Soit $x \in A$; comme A est ouvert, on peut inclure dans A une boule ouverte centrée en x , et donc x est à une distance positive de $X \setminus A$. Le point x appartient donc à F_k pour k assez grand, on conclut que l'union des F_k est A tout entier. Les F_k formant une famille croissante, on a donc $\mu[F_k] \rightarrow \mu[A]$. Or chaque F_k est fermé, puisque image réciproque de l'intervalle fermé $[1/k, +\infty[$ par l'application continue (et même 1-Lipschitzienne) $x \mapsto d(x, A)$.

4. Eliminons maintenant l'hypothèse de compacité de X . Soient A mesurable et $\varepsilon > 0$. Comme μ est concentrée sur un ensemble σ -compact, on peut trouver un compact $X' \subset X$ tel que $\mu[X \setminus X'] \leq \varepsilon/2$. L'espace X' est métrique, séparable et complet, il est en outre compact, on sait donc que la restriction de μ à X' est régulière. Il existe donc un ouvert O' de X' , contenant $A' = X' \cap A$, et un compact K' de X' , contenu dans A' , tels que

$$\mu[O'] - \varepsilon/2 \leq \mu[A'] \leq \mu[K'] + \varepsilon/2.$$

Comme intersection de compacts, K' est automatiquement compact. Par ailleurs, on peut écrire $O' = X' \cap O$ pour un certain ouvert O contenant A . Mais alors,

$$\mu[O] \leq \mu[O'] + \mu[X \setminus X'] \leq \mu[O'] + \varepsilon/2 \leq \mu[A'] + \varepsilon \leq \mu[A] + \varepsilon,$$

et

$$\mu[A] \leq \mu[A'] + \mu[X \setminus X'] \leq \mu[A'] + \varepsilon/2 \leq \mu[K'] + \varepsilon.$$

On en déduit que μ est bien régulière.

5. Eliminons finalement l'hypothèse de finitude de μ . Par hypothèse, il existe une famille $(X_n)_{n \geq 1}$ de parties de X de mesure finie, dont l'union est X entier. Sans perte de généralité, on peut supposer que X_n est une famille croissante. Définissons une famille de mesures μ_n par

$$\mu_n[A] := \mu[A \cap X_n].$$

Il est clair que μ_n est finie ; d'après le morceau de démonstration déjà effectué, elle est donc concentrée sur un ensemble σ -compact S_n . Par conséquent, μ est concentrée sur l'union des S_n , qui est une union dénombrable d'unions dénombrables de compacts ; et donc une union dénombrable de compacts. On vérifie enfin la propriété de régularité. Soit A un ensemble mesurable de X , et soit $\varepsilon > 0$. Pour tout n , on peut trouver un compact K_n et un ouvert O_n tels que

$$K_n \subset A \cap X_n \subset O_n,$$

avec $\mu[O \setminus (A \cap X_n)] < \varepsilon 2^{-n}$; et $\mu[(A \cap X_n) \setminus K_n] < \varepsilon/2$. On pose $O = \bigcup O_n$: alors O est un ouvert contenant $\bigcup (A \cap X_n) = A$, et

$$\mu[O \setminus A] = \mu[(\bigcup O_n) \setminus (\bigcup (A \cap X_n))] \leq \sum_n \mu[O_n \setminus (A \cap X_n)] < \varepsilon (\sum 2^{-n}) = \varepsilon.$$

Pour l'autre sens, distinguons deux cas. Si $\mu[A] = +\infty$, alors $\mu[A \cap X_n] \rightarrow +\infty$, et l'inégalité $\mu[K_n] \geq \mu[A \cap X_n] - \varepsilon$ implique $\mu[K_n] \rightarrow \infty$. Si en revanche $\mu[A] < +\infty$, alors on peut trouver $N \in \mathbb{N}$ tel que

$$\mu[A \cap X_N] \geq \mu[A] - \frac{\varepsilon}{2}.$$

Il s'ensuivra

$$\mu[K_N] \geq \mu[A \cap X_N] - \frac{\varepsilon}{2} \geq \mu[A] - \varepsilon.$$

Dans tous les cas, on a bien $\mu[A] = \sup\{\mu[K]\}$, où K décrit l'ensemble des compacts inclus dans A . Ceci achève de prouver la régularité. \square

I-3.4. Concentration et caractère diffus. Une mesure borélienne μ sur un espace mesuré X peut être concentrée sur un petit sous-ensemble de X , ou au contraire “voir tout X ”. La notion de **support** permet de préciser cette intuition : le support est le plus petit fermé sur lequel μ est concentrée.

THÉORÈME I-58 (support). *Soit X un espace topologique séparé et μ une mesure borélienne régulière sur X . On peut alors définir le support de μ comme le complémentaire du plus grand ouvert sur lequel μ est identiquement nulle.*

COROLLAIRE I-59. *Si (X, μ) est un espace polonais σ -fini, alors on peut définir le support de μ .*

REMARQUE I-60. On pourra comparer la notion de support d'une mesure à celle de support d'une fonction continue à valeurs réelles, que l'on définit comme le plus petit fermé en-dehors duquel f est identiquement nulle.

DÉMONSTRATION DU THÉORÈME I-58. Soit Ω la réunion de tous les ouverts $\omega \subset X$ tels que $\mu[\omega] = 0$. Par construction Ω contient tout ouvert où μ s'annule; le but est de montrer que $\mu[\Omega] = 0$, et bien sûr cela n'est pas évident car c'est une union a priori non dénombrable.

Supposons que $\mu[\Omega]$ soit strictement positif; par régularité il existe alors un compact $K \subset \Omega$ tel que $\mu[K] > 0$. Pour tout $x \in K$ il existe un ouvert $\omega = \omega_x$ contenant x , tel que $\mu[\omega_x] = 0$. Par compacité on peut trouver $J \in \mathbb{N}$ et $x_1, \dots, x_J \in K$ tels que $K \subset \cup\{\omega_{x_j}; 1 \leq j \leq J\}$. Alors $\mu[K] \leq \sum \mu[\omega_{x_j}] = 0$, ce qui est en contradiction avec l'hypothèse. On conclut effectivement que $\mu[\Omega] = 0$. \square

REMARQUE I-61. Le corollaire I-59 découle directement du Théorème I-58 et du théorème de régularité automatique dans les espaces polonais (Théorème I-54). Cependant on peut aussi le démontrer plus simplement, en notant qu'un espace polonais admet une base \mathcal{B} dénombrable d'ouverts (comme dans le Théorème I-32). Le support de μ peut alors être construit comme le complémentaire de l'union (forcément dénombrable) de tous les ouverts ω de \mathcal{B} tels que $\mu[\omega] = 0$.

Si une mesure est "diffuse", on peut le quantifier par la propriété d'avoir un support plein ($\text{Spt } \mu = X$). On peut également raffiner en introduisant la propriété de **doublément**.

DÉFINITION I-62 (doublément). Soient (X, d) un espace métrique muni de sa tribu borélienne, et μ une mesure de Borel sur X . Soit $C \geq 0$ une constante; on dit que μ est C -doublante si, pour tout $x \in X$ et $r > 0$ on a

$$\mu[B_{2r}(x)] \leq C \mu[B_r(x)].$$

On dit que μ est doublante si elle est C -doublante pour un $C \geq 0$.

REMARQUE I-63. Il est équivalent de définir ce concept en termes de boules ouvertes ou de boules fermées. En effet, par σ -additivité,

$$\mu[B_r(x)] = \lim_{n \rightarrow \infty} \mu[B_{r-n^{-1}}(x)], \quad \mu[B_r](x) = \lim_{n \rightarrow \infty} \mu[B_{r+n^{-1}}(x)].$$

EXEMPLES I-64. Nous verrons plus tard que la mesure de Lebesgue en dimension n est 2^n -doublante. La mesure de volume sur une variété riemannienne compacte (ou plus généralement, de courbure minorée, en un sens adéquat) est également doublante. En revanche la mesure $\sum_{n \in \mathbb{N}} \delta_n$ sur \mathbb{N} (ou \mathbb{R}) n'est pas doublante : par exemple, la boule $B_{1/2}(1/2)$ a pour masse 0, alors que la boule $B_1(1/2)$ a pour masse 2.

Intuitivement, une mesure doublante est "bien répartie" : quand on fait varier le rayon d'une boule, on n'observe pas de variations trop brusques de la mesure... Le théorème suivant précise cette idée.

THÉORÈME I-65. Soit μ une mesure régulière doublante sur un espace métrique (X, d) , non identiquement égale à 0 ou $+\infty$. Alors

- pour tout $x \in X$ et $r > 0$ on a $0 < \mu[B_r(x)] < +\infty$;
- la mesure μ ne peut avoir d'atome en dehors des points isolés de X .

DÉMONSTRATION. Soit C telle que μ soit C -doublante. Supposons d'abord qu'il existe $x \in X$ et $r > 0$ avec $\mu[B_r(x)] = 0$. Alors $\mu[B_{2r}(x)] \leq C\mu[B_r(x)] = 0$. Par récurrence, $\mu[B_{2^k r}(x)] = 0$ pour tout k . Puisque X est l'union croissante des boules $B_{2^k r}(x)$, $\mu[X] = 0$ par σ -additivité, ce qui est contraire à l'hypothèse. On conclut que $\mu[B_r(x)] > 0$.

Puisque μ n'est pas identiquement $+\infty$, il existe au moins un $x_0 \in X$ tel que $\mu[\{x_0\}] < +\infty$. Par régularité, il existe un ouvert contenant x_0 dont la mesure soit finie, et donc une boule $B_{r_0}(x_0)$ dont la mesure soit finie. Par le même raisonnement que ci-dessus, toutes les boules $B_{2^k r_0}(x_0)$ sont de mesure finie. Si $x \in X$ et $r > 0$ sont données, on peut toujours trouver k tel que $B_r(x) \subset B_{2^k r_0}(x_0)$, ce qui implique que $B_r(x)$ est aussi de mesure finie.

Soit maintenant $x \in X$, qui ne soit pas un point isolé; montrons que $\mu[\{x\}] = 0$. Supposons par l'absurde que $\mu[\{x\}] = \alpha > 0$, et soit $\delta > 0$, à choisir plus tard. Par régularité on peut trouver $\varepsilon > 0$ tel que $\mu[B_\varepsilon(x)] \leq \alpha + \delta$. Comme x n'est pas isolé, la boule $B_{\varepsilon/2}(x)$ ne se réduit pas à x ; soit donc $y \neq x$ tel que $d(x, y) < \varepsilon/2$. On pose $r := d(x, y)$. La boule $B_r(y)$ est tout entière contenue dans $B_\varepsilon(x)$ (pour tout $z \in B_r(y)$ on a $d(x, z) \leq d(x, y) + d(y, z) < r + r < \varepsilon$). Comme par ailleurs $B_r(y)$ ne contient pas x , on a

$$\alpha + \mu[B_r(y)] = \mu[\{x\}] + \mu[B_r(y)] = \mu[\{x\} \cup B_r(y)] \leq \mu[B_\varepsilon(x)] \leq \alpha + \delta.$$

On en déduit que $\mu[B_r(y)] \leq \delta$. D'autre part, $x \in B_{2r}(y)$, d'où $\mu[B_{2r}(y)] \geq \alpha$. La mesure μ étant C -doublante, on a

$$\alpha \leq \mu[B_{2r}(y)] \leq C\mu[B_r(y)] \leq C\delta.$$

On obtient une contradiction en choisissant $\delta = \alpha/(2C)$. □

I-4. Prolongement de mesures

Comme on l'a déjà dit, on veut souvent définir a priori la valeur d'une mesure sur une certaine classe d'ensembles : par exemple, les pavés dans \mathbb{R}^2 ou \mathbb{R}^d . En général, il n'est pas évident que l'on puisse le faire, c'est-à-dire qu'il existe une mesure qui attribue des valeurs spécifiées a priori sur certains ensembles. Un tel résultat est appelé **théorème de prolongement** (ou d'extension).

Le théorème de prolongement le plus célèbre et le plus utile a été démontré vers 1914 par Carathéodory, qui à cette occasion a développé le concept important de **mesure extérieure**⁶, ou mesure de Carathéodory.

Avant de démontrer ce théorème, nous allons examiner un résultat beaucoup plus simple, qui est l'outil le plus utilisé pour montrer que deux mesures coïncident sur une tribu entière : le **lemme de classe monotone** (déjà utilisé dans la preuve du Théorème I-54).

I-4.1. Lemme de classe monotone.

DÉFINITION I-66 (classe monotone). *On appelle classe monotone une famille \mathcal{C} de parties d'un ensemble X , stable par limite croissante et par différence :*

$$A_k \in \mathcal{C}, A_k \subset A_{k+1} \implies \bigcup_{k \in \mathbb{N}} A_k \in \mathcal{C},$$

$$[A, B \in \mathcal{C}, A \subset B] \implies B \setminus A \in \mathcal{C}.$$

⁶que certains auteurs appellent tout simplement "mesure" [Evans-Gariepy]

REMARQUE I-67. Noter que dans cette définition on a imposé $B \setminus A \in \mathcal{C}$ seulement dans le cas où $A \subset B$.

Bien sûr, une σ -algèbre est une classe monotone ; et réciproquement, il ne manque pas grand chose à une classe monotone pour être une σ -algèbre : seulement la stabilité par intersection, et la condition de contenir X . Le lemme suivant donne une condition suffisante pour qu'il y ait identité.

THÉORÈME I-68 (Lemme de classe monotone). *Soit \mathcal{F} une famille de parties d'un ensemble X , stable par intersection finie. Soit \mathcal{C} la plus petite classe monotone contenant \mathcal{F} ; on suppose que $X \in \mathcal{C}$. Alors \mathcal{C} coïncide avec la tribu $\sigma(\mathcal{F})$ engendrée par \mathcal{F} .*

DÉMONSTRATION. Il suffit bien sûr de vérifier que \mathcal{C} est une tribu : en effet, toute tribu contenant \mathcal{F} doit forcément contenir \mathcal{C} .

Notre but est donc de vérifier que pour tout A et B dans \mathcal{C} , on a $A \cap B \in \mathcal{C}$; cette propriété, combinée aux axiomes de classe monotone, garantira que \mathcal{C} est une tribu. Cependant ce résultat de stabilité par intersection semble a priori délicat car nous n'avons aucun moyen de décrire \mathcal{C} ; nous allons contourner cette difficulté en utilisant un raisonnement classique.

Soit d'abord $A \in \mathcal{F}$, et

$$\mathcal{C}_A = \{B \in \mathcal{C}; A \cap B \in \mathcal{C}\}.$$

Par hypothèse, \mathcal{C}_A contient \mathcal{F} . D'autre part, on vérifie aisément que \mathcal{C}_A est stable par différence et limite croissante ; il s'ensuit que c'est une classe monotone contenant \mathcal{F} , et c'est donc \mathcal{C} tout entière. On a donc démontré que pour tout $A \in \mathcal{F}$, et $B \in \mathcal{C}$, on a $A \cap B \in \mathcal{C}$.

Maintenant, pour $A \in \mathcal{C}$, on définit à nouveau

$$\mathcal{C}_A = \{B \in \mathcal{C}; A \cap B \in \mathcal{C}\}.$$

La première étape nous montre que \mathcal{C}_A contient \mathcal{F} . On conclut comme précédemment que $\mathcal{C}_A = \mathcal{C}$. \square

Le lemme de classe monotone est un outil d'usage universel en théorie de la mesure, car ses hypothèses s'accordent bien avec la propriété de σ -additivité des mesures (finies). En effet, si A et B sont deux ensembles mesurables et μ une mesure, en général on ne sait pas calculer $\mu[A \cap B]$ en fonction de $\mu[A]$ et $\mu[B]$; mais si $A \subset B$ on sait que $\mu[B \setminus A] = \mu[B] - \mu[A]$. De même, si l'on a une suite croissante d'ensembles mesurables A_k , on sait que $\mu[\cup A_k] = \lim \mu[A_k]$.

I-4.2. Théorème de prolongement de Carathéodory. Soient \mathcal{A} une algèbre de parties de X , et μ une fonction additive sur \mathcal{A} , i.e. une fonction positive vérifiant l'axiome d'additivité $\mu[A \cup B] = \mu[A] + \mu[B]$ pour A et B disjoints. Peut-on étendre μ en une mesure σ -additive sur la σ -algèbre engendrée par \mathcal{A} ? Pour cela il faut bien sûr que μ soit σ -additive sur \mathcal{A} lui-même : supposant qu'un élément A de \mathcal{A} s'écrive comme union disjointe d'éléments A_k de \mathcal{A} (par exemple, un rectangle dans \mathbb{R}^2 peut s'écrire comme une union dénombrable de rectangles disjoints, d'une infinité de manières différentes), on doit avoir

$$\mu[A] = \sum_{k \in \mathbb{N}} \mu[A_k].$$

THÉORÈME I-69 (théorème de prolongement de Carathéodory). *Soient X un ensemble, \mathcal{A} une algèbre de parties de X , et μ une fonction positive σ -additive sur \mathcal{A} , telle que X peut s'écrire comme union dénombrable de $A_k \in \mathcal{A}$ avec $\mu[A_k] < +\infty$. Alors il existe un unique prolongement de μ en une mesure (σ -additive) définie sur la σ -algèbre $\sigma(\mathcal{A})$.*

REMARQUE I-70. La propriété de σ -additivité sur \mathcal{A} n'est pas forcément évidente à vérifier. Si $\mu[X] < +\infty$, elle est équivalente à la condition suivante, parfois plus commode : pour toute suite décroissante $(A_k)_{k \in \mathbb{N}}$ d'éléments de \mathcal{A} ,

$$(\cap A_k = \emptyset) \implies \lim_{k \rightarrow \infty} \mu[A_k] = 0.$$

Si $\mu[X] = +\infty$, il existe aussi une reformulation analogue, mais elle est un tout petit peu plus délicate. Par hypothèse, X est union croissante de parties X_k appartenant à \mathcal{A} , de mesure finie. Soit B un élément de \mathcal{A} de mesure infinie, on peut écrire B comme l'union croissante des $\mu[B \cap X_k]$, et une condition nécessaire à la σ -additivité sur \mathcal{A} est $\lim_{k \rightarrow \infty} \mu[B \cap X_k] = \mu[B] = +\infty$. Il est en fait assez facile de montrer que l'hypothèse de σ -additivité sur \mathcal{A} est équivalente à la conjonction des deux hypothèses suivantes :

(i) pour toute suite décroissante $(A_k)_{k \in \mathbb{N}}$ d'éléments de \mathcal{A} ,

$$(\mu[A_0] < +\infty \text{ et } \cap A_k = \emptyset) \implies \lim_{k \rightarrow \infty} \mu[A_k] = 0;$$

(ii) pour tout $B \in \mathcal{A}$ tel que $\mu[B] = +\infty$, on a

$$\lim_{k \rightarrow \infty} \mu[B \cap X_k] = +\infty.$$

On peut trouver des preuves du Théorème I-69 dans diverses sources, par exemple [Bony, section 1.6], [Gramain, section VI.1] ou [Dudley, Théorème 3.1.4]. Je vais maintenant adapter ces preuves pour démontrer un énoncé plus général, qui contient le Théorème de Carathéodory comme cas particulier. L'intérêt propre de l'énoncé généralisé apparaîtra par la suite.

THÉORÈME I-71 (théorème de Carathéodory généralisé). *Soient X un ensemble, et \mathcal{F} une famille de parties de X , stable par intersection finie. Soit μ une fonction définie sur \mathcal{F} , à valeurs dans $[0, +\infty]$. Alors*

(i) *Si X est union dénombrable d'une famille croissante d'éléments X_k de \mathcal{F} tels que $\mu[X_k] < +\infty$, alors il existe au plus un prolongement de μ en une mesure sur $\sigma(\mathcal{F})$;*

(ii) *Soit, pour toute partie A de X ,*

$$\mu^*[A] := \inf \left\{ \sum_{k=1}^{\infty} \mu[A_k]; A_k \in \mathcal{F}; A \subset \cup A_k \right\}.$$

On suppose que

$$(7) \quad \forall A, B \in \mathcal{F}, \quad \mu[A \cap B] + \mu^*[A \setminus B] = \mu[A].$$

Alors μ^ définit sur $\sigma(\mathcal{F})$ une mesure qui prolonge μ . En outre, cette mesure est σ -additive sur la σ -algèbre \mathcal{M} , contenant $\sigma(\mathcal{F})$, définie par*

$$\mathcal{M} := \{A \subset X; \forall B \subset X, \mu^*[B \cap A] + \mu^*[B \setminus A] = \mu^*[B]\}.$$

Les éléments de \mathcal{M} sont dits μ -mesurables; la tribu \mathcal{M} , munie de μ^* , est automatiquement complète.

(iii) On suppose maintenant que non seulement \mathcal{F} est stable par intersection finie, mais qu'en outre **le complémentaire de tout élément de \mathcal{F} peut s'écrire comme une union finie disjointe d'éléments de \mathcal{F}** . Alors la condition (7) est satisfaite si et seulement si μ est σ -additive sur \mathcal{F} ; i.e. pour toute famille dénombrable $(A_n)_{n \in \mathbb{N}}$ d'éléments disjoints de \mathcal{F} tels que $\cup A_n$ appartient à \mathcal{F} , on a

$$\mu\left[\bigcup_{n \in \mathbb{N}} A_n\right] = \sum_{n \in \mathbb{N}} \mu[A_n].$$

En particulier, μ admet un prolongement σ -additif à $\sigma(\mathcal{F})$ si et seulement si μ est σ -additive sur \mathcal{F} .

REMARQUES I-72. (i) La fonction μ^* est appelée **mesure extérieure** associée à μ (et à la famille \mathcal{F}). Bien noter qu'elle est définie pour **toute** partie A de X . Ce n'est pas a priori une mesure; en revanche elle est croissante, et vérifie l'axiome de **sous-additivité dénombrable**: pour toute famille dénombrable $(A_k)_{k \in \mathbb{N}}$ de parties de X ,

$$\mu^*[\cup A_k] \leq \sum_k \mu^*[A_k].$$

Noter que dans cette définition il est inutile de supposer les A_k disjoints. Par extension, on appelle mesure extérieure n'importe quelle application définie sur l'ensemble des parties d'un ensemble, à valeurs dans $[0, +\infty]$, qui soit croissante, attribue la valeur 0 à l'ensemble vide, et vérifie l'axiome de sous-additivité dénombrable.

(ii) Il est crucial, dans la définition de la mesure extérieure, d'autoriser une union dénombrable et pas seulement une union finie de A_k . Pour s'en convaincre, on peut penser au cas de la mesure de Lebesgue sur l'intervalle $[0, 1]$, que l'on peut construire à partir de la mesure extérieure associée à l'ensemble \mathcal{F} des sous-intervalles de $[0, 1]$, et à la fonction μ = "longueur". En effet, si l'on cherche à recouvrir $\mathbb{Q} \cap [0, 1]$ par une famille finie d'intervalles, la somme des longueurs de ces intervalles est forcément supérieure ou égale à 1. En revanche, pour tout ε on peut recouvrir $\mathbb{Q} \cap [0, 1]$ par une famille **dénombrable** d'intervalles dont la somme des longueurs est plus petite que ε .

(iii) Sans l'hypothèse de " σ -finitude" faite en (i), il n'y a pas forcément unicité du prolongement. Quand elle prolonge effectivement μ , la mesure extérieure est alors le plus grand prolongement possible [Gramain, p. 116].

(iv) Une partie \mathcal{F} qui vérifie les hypothèses de la partie (iii) du théorème, à savoir : \mathcal{F} est stable par intersection binaire; et le complémentaire de tout élément de \mathcal{F} est réunion d'un nombre fini d'éléments disjoints de \mathcal{F} ; est parfois appelé une **semi-algèbre**. Une algèbre étant un cas particulier de semi-algèbre, la conclusion de ce théorème implique bien sûr celle du Théorème I-69.

La partie (ii) du Théorème I-71 est assez subtile. En revanche, la partie (i) est une application simple du Lemme de classe monotone.

DÉMONSTRATION DU THÉORÈME I-71. Commençons par démontrer (i) sous l'hypothèse $\mu[X] < +\infty$. Le Lemme de classe monotone (Théorème I-68) s'applique puisque X est par hypothèse limite croissante d'éléments de \mathcal{F} . Donc la σ -algèbre $\sigma(\mathcal{F})$ n'est autre que la classe monotone \mathcal{C} engendrée par \mathcal{F} . Soient μ et $\tilde{\mu}$ deux prolongements possibles. On pose

$$\mathcal{B} = \{C \in \mathcal{C}; \mu[A] = \tilde{\mu}[A]\}.$$

Par hypothèse, \mathcal{B} contient \mathcal{F} . Comme μ et $\tilde{\mu}$ sont compatibles avec les opérations de limite croissante et de soustraction, au sens où

$$(8) \quad A \subset B \implies \mu[B \setminus A] = \mu[B] - \mu[A], \quad \text{etc.,}$$

on voit que \mathcal{B} est une classe monotone. Il s'ensuit que $\mathcal{B} = \mathcal{C}$, ce qui conclut la preuve de (i). Notons que l'hypothèse de finitude a été utilisée implicitement quand nous avons écrit (8), qui n'aurait guère de sens si $\mu[B] = \mu[A] = +\infty$.

Passons maintenant à la démonstration de (i) dans le cas général. Soient μ et $\tilde{\mu}$ deux prolongements possibles; par le raisonnement précédent on sait que $\mu[X_k \cap A] = \tilde{\mu}[X_k \cap A]$ pour tout A mesurable. La famille (X_k) étant croissante, la famille $(X_k \cap A)$ l'est aussi, et son union est $X \cap A = A$. Par σ -additivité, on peut passer à la limite quand $k \rightarrow \infty$ et obtenir $\mu[A] = \tilde{\mu}[A]$, ce qui conclut la preuve de (i).

On s'attelle ensuite à la démonstration de (ii). Comme dans l'énoncé, on définit

$$\mathcal{M} := \{A \subset X; \forall B \subset X, \mu^*[B \cap A] + \mu^*[B \setminus A] = \mu^*[B]\}.$$

Nous allons montrer que \mathcal{M} est une σ -algèbre, et μ^* une mesure sur \mathcal{M} ; cette σ -algèbre est en outre complète. Cet énoncé est indépendant de l'hypothèse (7). On divise la preuve en sept étapes.

1. μ^* est croissante. En effet, si $A' \subset A$, l'infimum qui définit $\mu^*[A']$ est pris sur une classe de familles de parties plus vaste que celui qui définit $\mu^*[A]$.

2. μ^* est dénombrablement sous-additive. En d'autres termes, si $(A_k)_{k \in \mathbb{N}}$ est une famille de parties de X , on a

$$\mu^*[\cup A_k] \leq \sum \mu^*[A_k].$$

Si $\mu^*[A_k] = +\infty$ pour un certain A_k , alors bien sûr il n'y a rien à démontrer. Dans le cas contraire, par définition de la borne inférieure, pour tout k on peut trouver une famille $(F_{jk})_{j \in \mathbb{N}}$ d'éléments de \mathcal{F} tels que

$$A_k \subset \cup_j F_{jk}, \quad \sum_j \mu[F_{jk}] \leq \mu^*[A_k] + \varepsilon 2^{-k},$$

où $\varepsilon > 0$ est arbitrairement petit. En particulier,

$$(\cup A_k) \subset \bigcup_{jk} F_{jk}, \quad \sum_{jk} \mu[F_{jk}] \leq \sum_k \mu^*[A_k] + \varepsilon.$$

Il s'ensuit que $\mu^*[\cup A_k] \leq \sum \mu^*[A_k] + \varepsilon$, et on obtient la conclusion souhaitée en faisant tendre ε vers 0.

Notons en particulier que pour toutes parties A et B de X ,

$$\mu^*[B] \leq \mu^*[B \cap A] + \mu^*[B \setminus A].$$

Pour prouver l'appartenance d'une partie A à \mathcal{M} , il suffit donc d'établir l'inégalité inverse.

3. \mathcal{M} est une algèbre. D'une part, il est clair que \emptyset appartient à \mathcal{M} ; et il est évident que \mathcal{M} est stable par passage au complémentaire. Il suffit donc de vérifier que \mathcal{M} est stable par intersection. Soient A_1 et A_2 deux éléments de \mathcal{M} , et soit B une partie quelconque de X . Notre but est de montrer que

$$\mu^*[B \cap (A_1 \cap A_2)] + \mu^*[B \setminus (A_1 \cap A_2)] \leq \mu^*[B].$$

Pour cela, on note que $B \setminus (A_1 \cap A_2) = ((B \cap A_1) \setminus A_2) \cup (B \setminus A_1)$, d'où, par sous-additivité,

$$\mu^*[B \setminus (A_1 \cap A_2)] \leq \mu^*[(B \cap A_1) \setminus A_2] + \mu^*[B \setminus A_1].$$

En combinant cette inégalité avec l'appartenance de A_1 et de A_2 à \mathcal{M} , on obtient

$$\begin{aligned} \mu^*[B \cap A_1 \cap A_2] + \mu^*[B \setminus (A_1 \cap A_2)] &\leq \mu^*[(B \cap A_1) \cap A_2] + \mu^*[(B \cap A_1) \setminus A_2] + \mu^*[B \setminus A_1] \\ &= \mu^*[B \cap A_1] + \mu^*[B \setminus A_1] = \mu^*[B], \end{aligned}$$

ce qui conclut l'argument.

4. μ^* est additive sur \mathcal{M} . En effet, considérons A et B deux éléments disjoints de \mathcal{M} ; puisque $B \in \mathcal{M}$, on a

$$\mu^*[A \cup B] = \mu^*[(A \cup B) \cap A] + \mu^*[(A \cup B) \setminus A] = \mu^*[A] + \mu^*[B],$$

ce qui prouve l'additivité de μ^* .

Par ailleurs, si A_1 et A_2 sont deux éléments disjoints de \mathcal{M} , alors pour toute partie B de X ,

$$\mu^*[B \cap (A_1 \cup A_2)] = \mu^*[(B \cap (A_1 \cup A_2)) \cap A_1] + \mu^*[(B \cap (A_1 \cup A_2)) \setminus A_1] = \mu^*[B \cap A_1] + \mu^*[B \cap A_2];$$

par récurrence on obtient que si on se donne des éléments A_k de \mathcal{M} , disjoints et en nombre fini, alors, pour toute partie B de X on a

$$\mu^*[B \cap (\cup A_k)] = \sum \mu^*[B \cap A_k].$$

5. \mathcal{M} est une σ -algèbre. Pour montrer cela, il nous suffit de vérifier que pour toute famille dénombrable de parties A_k **disjointes**, éléments de \mathcal{M} , et pour toute partie B de X ,

$$\mu^*[B \cap (\cup A_k)] + \mu^*[B \setminus (\cup A_k)] \leq \mu^*[B].$$

Posons $A^n = \cup_{k=1}^n A_k$, $A^\infty = \cup_{k=1}^\infty A_k$. D'après l'identité établie à l'étape 4, et la croissance de μ^* , on sait que pour tout n ,

$$\mu^*[B] = \mu^*[B \cap A^n] + \mu^*[B \setminus A^n] \geq \sum_{k=1}^n \mu^*[B \cap A_k] + \mu^*[B \setminus A^\infty].$$

En passant à la limite $n \rightarrow \infty$, et en utilisant la sous-additivité, on trouve

$$\begin{aligned} \mu^*[B] &\geq \sum_{k=1}^\infty \mu^*[B \cap A_k] + \mu^*[B \setminus A^\infty] \\ &\geq \mu^*[B \cap A^\infty] + \mu^*[B \setminus A^\infty] \geq \mu^*[B]. \end{aligned}$$

Les trois membres de l'inégalité sont donc égaux, ce qui prouve que $A^\infty \in \mathcal{M}$. En particulier,

$$\mu^*[B \cap A^\infty] + \mu^*[B \setminus A^\infty] = \mu^*[B].$$

6. μ^* est σ -additive sur \mathcal{M} . Pour s'en convaincre, il suffit de poser $B = X$ dans l'égalité précédente.

7. (X, \mathcal{M}, μ^*) est un espace complet. Soit $A \in \mathcal{M}$ avec $\mu^*[A] = 0$; en particulier, pour tout $B \subset X$ on a $\mu^*[B \setminus A] = \mu^*[B]$. Soient $A' \subset A$ et $B \subset X$, alors $\mu^*[A' \cap B] \leq \mu^*[A] = 0$, et $\mu^*[B] \geq \mu^*[B \setminus A'] \geq \mu^*[B \setminus A] = \mu^*[B]$. On conclut que $\mu^*[B \setminus A'] = \mu^*[B]$, et $\mu^*[A'] = 0$, donc A' est μ -mesurable.

Récapitulons : nous avons défini une σ -algèbre \mathcal{M} et une mesure μ^* sur \mathcal{M} . Pour conclure la preuve de (ii), il nous suffit de prouver que \mathcal{M} contient \mathcal{F} (ce qui impliquera que $\sigma(\mathcal{F}) \subset \mathcal{M}$), et que μ^* coïncide avec μ sur \mathcal{F} . C'est ici que l'hypothèse (7) va intervenir.

Posons $B = \emptyset$ dans (7), on trouve $\mu^*[A] = \mu[A]$ pour tout $A \in \mathcal{F}$, ce qui montre que la restriction de μ^* à \mathcal{F} est bien μ .

Soient maintenant $A \in \mathcal{F}$, et $B \subset X$. Soit (A_k) une famille d'éléments de \mathcal{F} recouvrant B ; la famille $(A_k \cap A)$ recouvre alors $B \cap A$, et tous ses éléments appartiennent à \mathcal{F} grâce à la propriété de stabilité par intersection finie. En appliquant successivement la définition de μ^* , sa sous-additivité et l'hypothèse (7), on trouve

$$\begin{aligned} \mu^*[B \cap A] + \mu^*[B \setminus A] &\leq \sum_k \mu[A_k \cap A] + \mu^*[B \setminus A] \\ &\leq \sum_k (\mu[A_k \cap A] + \mu^*[A_k \setminus A]) = \sum_k \mu[A_k]. \end{aligned}$$

En prenant la borne inférieure sur tous les recouvrements (A_k) admissibles, on parvient à

$$\mu^*[B \cap A] + \mu^*[B \setminus A] \leq \mu^*[B];$$

l'inégalité réciproque est toujours vérifiée, il y a donc égalité, ce qui signifie que $A \in \mathcal{M}$. La preuve est complète.

Passons enfin à la partie (iii) du Théorème I-71. Si le critère (7) est vérifié, alors μ se prolonge en une mesure σ -additive μ^* , et en particulier elle est σ -additive sur \mathcal{F} ; c'est bien sûr la réciproque qui est délicate. On va donc supposer que μ est σ -additive sur \mathcal{F} et établir (7). On procède en trois étapes.

1. μ est croissante sur \mathcal{F} . Soient A et B deux éléments de \mathcal{F} avec $A \subset B$. Par hypothèse on peut écrire $X \setminus A = \cup C_j$, où les C_j sont des éléments de \mathcal{F} , disjoints; alors

$$B = A \cup (B \setminus A) = A \cup \bigcup_j (A \cap C_j),$$

où le membre de droite est une union d'éléments disjoints de \mathcal{F} ; par additivité de μ on a

$$\mu[B] = \mu[A] + \sum_j \mu[A \cap C_j] \geq \mu[A],$$

ce qui prouve que μ est bien croissante.

2. μ coïncide avec μ^* sur \mathcal{F} . D'après la définition de μ^* , on a toujours $\mu^*[A] \leq \mu[A]$ pour tout $A \in \mathcal{F}$. D'autre part, soit $A \in \mathcal{F}$ et soit $(A_n)_{n \in \mathbb{N}}$ un recouvrement arbitraire de A par des éléments de \mathcal{F} , c'est-à-dire $A \subset \cup A_n$. On pose

$$A'_1 = A_1, \quad A'_2 = A_2 \setminus A_1, \quad A'_3 = A_3 \setminus (A_1 \cup A_2), \quad A'_4 = A_4 \setminus (A_1 \cup A_2 \cup A_3), \quad \text{etc.}$$

Vérifions que **chacun des A'_j peut s'écrire comme une union finie de parties disjointes appartenant à \mathcal{F}** . En effet, par hypothèse, pour chaque j on peut trouver des parties disjointes B_{j,i_j} , $1 \leq i_j \leq N_j$, appartenant à \mathcal{F} , telles que

$$\begin{aligned} A'_j &= A_j \cap (X \setminus A_1) \cap (X \setminus A_2) \cap \dots \cap (X \setminus A_{j-1}) \\ &= A_j \cap \left(\bigcup_{i_1=1}^{N_1} B_{1,i_1} \right) \cap \left(\bigcup_{i_2=1}^{N_2} B_{2,i_2} \right) \cap \dots \cap \left(\bigcup_{i_{j-1}=1}^{N_{j-1}} B_{j-1,i_{j-1}} \right) \\ &= \bigcup_{i_1, \dots, i_{j-1}} A_j \cap B_{1,i_1} \cap \dots \cap B_{j-1,i_{j-1}}. \end{aligned}$$

Pour chaque j , les B_{j,i_j} sont disjoints; cela entraîne que pour deux choix différents du multi-indice (i_1, \dots, i_j) , les parties $A \cap A_j \cap B_{1,i_1} \cap \dots \cap B_{j-1,i_{j-1}}$ correspondantes sont disjointes. Pour chaque j donné, la réunion de toutes ces parties constitue A'_j , et par construction les A'_j sont deux à deux disjoints. On conclut que toutes les parties

$$A_j \cap B_{1,i_1} \cap \dots \cap B_{j-1,i_{j-1}},$$

que l'on renumérote C_{jk} ($1 \leq k \leq M_j$) sont disjointes. Par construction, leur union est égale à $\cup A_n$; et grâce à la stabilité de \mathcal{F} par intersection finie, toutes ces parties sont des éléments de \mathcal{F} . En outre, pour tout j on a

$$A_j = \bigcup_{1 \leq \ell \leq j, 1 \leq k \leq M_\ell} A_j \cap C_{\ell k};$$

d'où, par σ -additivité de μ sur \mathcal{F} ,

$$\mu[A_j] = \sum_{1 \leq \ell \leq j, 1 \leq k \leq M_\ell} \mu[A_j \cap C_{\ell k}],$$

et en particulier (en ne conservant que les termes en $\ell = j$) on a

$$\sum_{1 \leq k \leq M_j} \mu[C_{jk}] \leq \mu[A_j].$$

En sommant sur tous les indices j , on obtient

$$\sum_{jk} \mu[C_{jk}] \leq \sum_j \mu[A_j],$$

puis, comme μ est croissante,

$$\sum_{jk} \mu[A \cap C_{jk}] \leq \sum_j \mu[A_j].$$

Les parties $A \cap C_{jk}$ sont deux à deux disjointes, appartiennent à \mathcal{F} , et leur union est $A \cap (\cup A_n) = A$. Par σ -additivité, le membre de gauche de l'égalité précédente est donc $\mu[A]$. En conclusion, on a montré que pour tout recouvrement arbitraire de A par une famille (A_j) d'éléments de \mathcal{F} , on avait

$$\mu[A] \leq \sum_{j \in \mathbb{N}} \mu[A_j].$$

Par définition de μ^* , on a donc $\mu[A] \leq \mu^*[A]$, ce qui achève de prouver que μ et μ^* coïncident sur \mathcal{F} .

3. L'égalité (7) est satisfaite. Puisque μ^* est sous-additive, un corollaire de l'étape précédente est

$$\mu[A] \leq \mu[A \cap B] + \mu^*[A \setminus B],$$

pour toutes parties A et B dans \mathcal{F} . Pour prouver la validité de (7), il suffit d'établir l'inégalité inverse. Pour cela, on remarque que $A \setminus B = A \cap (X \setminus B)$ est l'intersection de A avec une union finie disjointe d'éléments de \mathcal{F} ; et peut donc s'écrire comme une union finie disjointe d'éléments D_k de \mathcal{F} . Alors

$$\mu[A \cap B] + \mu^*[\cup D_k] \leq \mu[A \cap B] + \sum_k \mu^*[D_k] = \mu[A \cap B] + \sum_k \mu[D_k].$$

Mais $A \cap B$ et les D_k sont des parties disjointes dont la réunion est A ; toujours par additivité de μ , la dernière somme est donc égale à $\mu[A]$. On conclut que

$$\mu[A \cap B] + \mu^*[A \setminus B] \leq \mu[A],$$

ce qui conclut la preuve de (7) et du Théorème I-71. \square

EXERCICE I-73. Réécrire la preuve précédente en cherchant à la simplifier pour démontrer directement le Théorème I-69 et non sa version généralisée (Théorème I-71).

REMARQUE I-74. Dans la définition de \mathcal{M} , dans le cas où X est de masse totale finie ($\mu[X] < +\infty$), on pourrait se contenter de définir les parties mesurables A par l'égalité

$$\mu^*[A] + \mu^*[X \setminus A] = \mu[X].$$

Ainsi, intuitivement, une partie A de X est μ -mesurable si l'on parvient à l'approcher extérieurement, au sens de la mesure μ , par des unions d'éléments de \mathcal{F} , l'approximation étant suffisamment précise pour que la mesure extérieure ne comptabilise aucune masse appartenant à $X \setminus A$. Notons que Lebesgue utilisait déjà cette construction.

Le théorème de Carathéodory est capital en théorie de la mesure; pour l'illustrer on va mentionner dès à présent deux problèmes importants qu'il permet de résoudre. Dans l'immédiat, on n'en donnera que des preuves partielles; les preuves complètes viendront plus tard dans le cours.

I-4.3. Produits infinis. Le théorème suivant est une conséquence importante du théorème de prolongement de Carathéodory.

THÉORÈME I-75 (produit infini de probabilités). *Soit $(X_k, \mu_k)_{k \in \mathbb{N}}$ une famille dénombrable d'espaces de probabilités. Pour tout m et toute famille A_1, \dots, A_m de parties mesurables de X_1, \dots, X_m respectivement, on pose*

$$C(A_1, \dots, A_m) = A_1 \times \dots \times A_m \times \prod_{k=m+1}^{\infty} X_k.$$

On définit alors

$$\mu^\infty[C(A_1, \dots, A_m)] = \prod_{k=1}^m \mu_k[A_k].$$

Cette fonction μ^∞ se prolonge en une unique mesure de probabilité sur le produit infini $\prod A_k$, muni de la tribu engendrée par les cylindres $C(A_1, \dots, A_m)$, $m \in \mathbb{N}$.

Dans le cas où les X_k sont des ensembles de cardinal fini, la démonstration de ce théorème est très simple et nous allons la présenter tout de suite. Le cas général [Dudley, p. 257-259] est plus subtil, et nous attendrons d'être plus aguerris pour la présenter : Cf. Théorème III-97 au Chapitre III. Ce théorème peut se généraliser de diverses manières, mais la conclusion est en général fausse si l'on n'impose pas de restriction sur les quantités $\mu_k[X_k]$.

DÉMONSTRATION DU THÉORÈME I-75 POUR DES ESPACES DE CARDINAL FINI. Sans perte de généralité, on suppose que

$$X_k = \{0, \dots, N_k\}; \quad \mu_k = \sum_{\ell=1}^{N_k} \alpha_\ell^k \delta_\ell; \quad \sum_{\ell=1}^{N_k} \alpha_\ell^k = 1,$$

et chaque X_k est muni de la tribu triviale $\mathcal{P}(X_k)$. Les cylindres sont de la forme $C = A_1 \times A_2 \times \dots \times A_k \times X_{k+1} \times X_{k+2} \times \dots$; on définit alors $\mu^\infty[C] = \prod_{j=1}^k \mu_j[A_j]$. Il est facile de vérifier que le complémentaire d'un cylindre produit est une union finie de cylindres produits. En outre, $\mu^\infty[X] = 1$. Pour appliquer le Théorème I-71 et conclure à l'unicité d'un prolongement σ -additif de μ^∞ , il suffit de vérifier la σ -additivité de μ^∞ sur la famille des cylindres produits. Soit donc $(C_n)_{n \in \mathbb{N}}$ une famille de cylindres produits disjoints, dont l'union est un cylindre produit $C = \prod A_k$. Par le Lemme I-76 ci-dessous, il n'y a qu'un nombre fini de C_n non vides; l'identité $\mu^\infty[\cup C_n] = \sum \mu^\infty[C_n]$ est donc conséquence de l'additivité de μ^∞ . \square

LEMME I-76 (absence de recouvrement dénombrable par des cylindres). *Soient $(X_k)_{k \in \mathbb{N}}$ une famille d'ensembles finis, et $(C_n)_{n \in \mathbb{N}}$ une famille de cylindres disjoints de $\prod X_k$, telle que $\cup C_n = \prod X_k$. Alors il n'y a qu'un nombre fini de C_n non vides.*

Nous allons donner deux démonstrations de ce lemme. La première est simple et élémentaire, le principe rappelle un peu celui que nous utiliserons dans la suite pour démontrer le Théorème I-75 dans un cadre général. La deuxième, plus compacte, illustrera l'intérêt d'un raisonnement topologique, et préparera la voie à la démonstration du Théorème I-77 ci-dessous; elle présupposera quelques connaissances en topologie.

PREMIÈRE DÉMONSTRATION. Chacun des cylindres C_n est union disjointe de "cylindres élémentaires", de la forme $(a_1, a_2, \dots, a_K) \times \prod_{k \geq K+1} X_k$; on dit qu'un tel cylindre est d'ordre K et a pour base $\{a_1, \dots, a_K\}$. Il suffit de démontrer le lemme dans le cas où tous les C_k sont des cylindres élémentaires. On retire les cylindres non vides; en particulier les C_k seront supposés tous distincts. On va supposer qu'il y a une infinité de C_k , et arriver à une contradiction.

Considérons l'ensemble des cylindres d'ordre 1. Si chaque élément de X_1 est le premier élément d'un cylindre d'ordre 1, alors il y a exactement $|X_1|$ cylindres d'ordre 1, et leur réunion finie couvre X , il n'y a donc qu'un nombre fini de C_k , ce qui est impossible. Il existe donc un sous-ensemble non vide $Y_1 = \{u_1, \dots, u_\ell\}$ de X_1 , tel que le cylindre de base Y_1 n'intersecte aucun des C_k d'ordre 1, et doit donc être recouvert par les cylindres d'ordre 2 ou plus. Le premier élément de chacun des cylindres d'ordre 2 ou plus est l'un des u_j ; comme les u_j sont en nombre fini et qu'il y a une infinité de cylindres, l'un au moins des u_j apparaît une infinité de fois en premier élément de l'un des cylindres d'ordre 2 ou plus. Appelons-le y_1 : le cylindre de base y_1 est donc recouvert par une infinité de cylindres, dont la première composante est toujours y_1 .

On montre alors, par un raisonnement similaire, qu'il existe un élément y_2 de X_2 , qui n'est le deuxième élément d'aucun cylindre d'ordre 2, tel que le cylindre de base $\{y_1, y_2\}$ est recouvert par une infinité de cylindres d'ordre 3 ou plus, dont les deux premières composantes sont $\{y_1, y_2\}$.

Par récurrence, on construit ainsi une suite $(y_k)_{k \in \mathbb{N}}$ dans $\prod X_k$, telle que pour tout $j \in \mathbb{N}$, $\{y_1, y_2, \dots, y_j\}$ n'est le premier élément d'aucun cylindre d'ordre j parmi les C_n . En particulier, cette suite n'appartient à aucun des cylindres C_n , ce qui fournit une contradiction. \square

DEUXIÈME DÉMONSTRATION. On munit chaque X_k de la topologie discrète, toute partie de X_k est alors ouverte; par définition, la topologie produit est alors engendrée par les cylindres, qui sont en particulier des ouverts. D'autre part, C est un produit (infini) de compacts, donc compact par le théorème de Tychonov (dans la version simple où on considère un produit dénombrable d'espaces produits). De la famille (C_k) on peut donc extraire un sous-recouvrement fini; comme ils sont disjoints, seul un nombre fini d'entre eux est non vide. \square

I-4.4. Théorème de prolongement de Kolmogorov. Ce théorème est fondamental en théorie des probabilités, et tout particulièrement des processus stochastiques. Il s'agit essentiellement d'une généralisation du précédent.

THÉORÈME I-77 (théorème de prolongement de Kolmogorov). *Soit T un ensemble arbitraire, et $(X_t)_{t \in T}$ une famille d'espaces polonais; on définit*

$$X := \prod_{t \in T} X_t.$$

Pour toute partie finie $F = \{t_1, \dots, t_K\} \subset T$, on définit $X_F := X_{t_1} \times \dots \times X_{t_K}$; et pour tout Borélien A_F de X_F , on définit le cylindre

$$C(A_F) := \{x \in X; (x_{t_1}, \dots, x_{t_K}) \in A_F\}.$$

On munit X de la tribu engendrée par tous les cylindres $C(A_F)$, pour toutes les parties finies F de T . On se donne une fonction μ , qui pour toute partie finie F de T définit une mesure de probabilité sur la tribu des cylindres $C(A_F)$. Alors μ se prolonge en une unique mesure de probabilité sur $\prod X_t$.

REMARQUES I-78. (i) Bien noter la condition de **compatibilité** implicite dans ce théorème : si $F \subset F'$, tout cylindre $C(A_F)$ peut aussi être vu comme un $C(A_{F'})$: il suffit de prendre $A_{t'} = X_{t'}$ pour tous les $t' \in T' \setminus T$. Les nombres $\mu[C(A_F)]$ et $\mu[C(A_{F'})]$ doivent alors bien sûr coïncider ! On parle de "système de marginales" compatible.

(ii) En théorie des processus stochastiques, l'espace T est d'habitude un morceau de \mathbb{R}_+ , interprété comme l'espace des **temps**. Le cas particulier où la famille T n'est autre que \mathbb{N} (penser à des temps discrets) relève également du théorème dit de Ionescu Tulcea, souvent utilisé en théorie des probabilités. Cependant, le théorème de Kolmogorov ne nécessite aucune hypothèse de régularité sur T . Pour certaines généralisations, on pourra consulter par exemple [Dudley, p. 441].

(iii) Après avoir étudié les propriétés principales de l'intégration produit, nous démontrerons ce théorème au Chapitre III (voir le Théorème III-111).

(iv) Dans la pratique, les modalités de la construction de la probabilité μ sont rarement utiles ; c'est seulement le résultat d'existence que l'on utilise.

I-4.5. Critère de Carathéodory. Comme nous l'avons vu, le théorème de prolongement de Carathéodory construit une tribu \mathcal{M} sur laquelle la mesure extérieure μ^* est automatiquement σ -additive. Le critère de Carathéodory est une condition d'apparence relativement simple qui entraîne que \mathcal{M} contient la tribu borélienne. C'est le critère que l'on utilise traditionnellement pour construire les mesures de Hausdorff dans \mathbb{R}^n , que nous étudierons plus tard.

THÉORÈME I-79 (Critère de Carathéodory). *Soit (X, d) un espace métrique, et soit μ^* une mesure extérieure sur X , au sens de la Remarque I-72 (i) ; on définit la tribu \mathcal{M} comme dans l'énoncé du Théorème I-71. Si, pour toutes parties A et B de X telles que*

$$d(A, B) := \inf \{d(x, y) ; x \in A, y \in B\} > 0,$$

on a

$$\mu^*[A \cup B] = \mu^*[A] + \mu^*[B],$$

alors la tribu \mathcal{M} contient la tribu borélienne de X .

DÉMONSTRATION. Il suffit de démontrer que la tribu \mathcal{M} contient tous les fermés. Soient donc A un ensemble fermé, et B un ensemble arbitraire, on veut prouver que

$$\mu^*[B] = \mu^*[B \cap A] + \mu^*[B \setminus A].$$

Grâce à la sous-additivité de μ^* , il suffit d'établir

$$\mu^*[B] \geq \mu^*[B \cap A] + \mu^*[B \setminus A].$$

Sans perte de généralité, on suppose que $\mu^*[B] < +\infty$. Pour tout $n \geq 1$, on définit

$$A_n := \{x \in B ; d(x, A) \leq 1/n\},$$

et on note que $\cup A_n = A$ puisque A est fermé. Alors $d(B \setminus A_n, B \cap A) \geq 1/n > 0$, d'où

$$\mu^*[B \setminus A_n] + \mu^*[B \cap A] = \mu^*[(B \setminus A_n) \cup (B \cap A)] \leq \mu^*[B].$$

Il suffit donc de prouver que

$$\mu^*[B \setminus A_n] \xrightarrow{n \rightarrow \infty} \mu^*[B \setminus A].$$

Soit alors $A_{n,n+1} := A_n \setminus A_{n+1}$; en particulier $B = B \setminus A_n \cup (\cup_{k \geq n} A_{k,k+1})$. Par sous-additivité de μ^* ,

$$\mu^*[B \setminus A_n] \leq \mu^*[B \setminus A] \leq \mu^*[B \setminus A_n] + \sum_{k=n}^{\infty} \mu^*[A_{k,k+1}],$$

et la conclusion s'ensuivra si l'on peut démontrer

$$\sum_{k \geq 1} \mu^*[A_{k,k+1}] < +\infty.$$

Comme $d(A_{i,i+1}, A_{j,j+1}) > 0$ dès que $j \geq i + 2$, on peut établir par récurrence

$$\sum_{k=1}^N \mu^*[A_{2k,2k+1}] = \mu^*\left[\bigcup_{k=1}^N A_{2k,2k+1}\right] \leq \mu^*[A] < +\infty;$$

et, de même,

$$\sum_{k=1}^N \mu^*[A_{2k+1,2k+2}] \leq \mu^*[A] < +\infty.$$

On conclut facilement en faisant tendre N vers l'infini. \square

I-5. Complétion de mesures

Si le prolongement d'une mesure est une opération délicate, sa complétion en revanche est très facile.

THÉORÈME I-80. *Soit (X, \mathcal{A}, μ) un espace mesuré, et soit $\overline{\mathcal{A}}$ la famille des parties E de X qui s'écrivent $A \cup B'$, où $A \in \mathcal{A}$ et B' est inclus dans une partie négligeable $B \in \mathcal{A}$, telle que $\mu[B] = 0$. Alors $\overline{\mathcal{A}}$ est une tribu, et μ admet un prolongement unique $\overline{\mu}$ à $\overline{\mathcal{A}}$, tel que $(X, \overline{\mathcal{A}}, \overline{\mu})$ est un espace mesuré complet.*

DÉMONSTRATION. Ce théorème pourra être démontré en guise d'exercice : il suffit de poser $\overline{\mu}[E] = \mu[A]$, et de vérifier les axiomes de σ -additivité. La seule subtilité consiste à montrer que cette définition est licite. Pour cela, on pourra remarquer que si $E = A_1 \cup B'_1 = A_2 \cup B'_2$ (avec des notations évidentes), alors $A_1 \setminus A_2 \subset E \setminus A_2 \subset B'_2 \subset B_2$ est de mesure nulle, donc $\mu[A_1] = \mu[(A_1 \cap A_2)] + \mu[A_1 \setminus A_2] = \mu[A_1 \cap A_2] \leq \mu[A_2]$. Par symétrie, $\mu[A_1] = \mu[A_2]$. \square

REMARQUE I-81. La simplicité de l'énoncé et de sa preuve masque le fait que les ensembles ainsi complétés peuvent être extrêmement compliqués. Comme nous le verrons en parlant d'intégration sur des espaces produits, il ne faut pas croire que l'opération de complétude est inoffensive.

Le théorème I-71 fournissait déjà un prolongement complet de la mesure μ ; il n'est pas clair a priori que ce soit le même que celui qui est fourni par le Théorème I-80, car une mesure admet en général plusieurs prolongements complets. Il y a cependant unicité quand on impose certaines conditions de régularité.

THÉORÈME I-82 (unicité de la complétion régulière). *Soit X un espace topologique et μ une mesure de Borel sur X , définie sur la tribu borélienne \mathcal{A} . On suppose que X est σ -fini et que μ est régulière. Alors l'espace $(X, \overline{\mathcal{A}}, \overline{\mu})$ défini dans le Théorème I-80 est l'unique prolongement complet de (X, \mathcal{A}, μ) en un espace mesuré complet et régulier.*

DÉMONSTRATION. Soit $(X, \tilde{\mathcal{A}}, \tilde{\mu})$ un prolongement complet régulier de (X, \mathcal{A}, μ) . Comme X est σ -fini, tout élément A de $\tilde{\mathcal{A}}$ peut s'écrire comme union dénombrable d'éléments A_k de $\tilde{\mathcal{A}}$ avec $\tilde{\mu}[A_k] < +\infty$. Par la Proposition I-49, pour chaque A_k il existe des Boréliens B_k et C_k tels que $\tilde{\mu}[B_k] = \tilde{\mu}[A_k] = \tilde{\mu}[C_k]$, et $B_k \subset A_k \subset C_k$. En particulier, A_k s'écrit comme l'union d'un Borélien et d'un ensemble inclus dans un Borélien $\tilde{\mu}$ -négligeable, ce qui montre que $\tilde{\mathcal{A}}$ contient $\overline{\mathcal{A}}$. On conclut que les deux tribus sont égales, et la fin de la démonstration en découle aisément. \square

En conclusion : tant que l'on travaille avec des mesures régulières, on n'a pas à se poser de questions sur le procédé de complétion.

I-6. Application : Construction de la mesure de Lebesgue

Le théorème suivant sera notre première application du théorème de prolongement de Carathéodory. On munira bien sûr \mathbb{R} de sa topologie habituelle.

THÉORÈME I-83 (mesure de Lebesgue en dimension 1). *Il existe une unique mesure borélienne λ sur \mathbb{R} telle que la mesure d'un intervalle $[a, b]$ ($a < b$) soit égale à sa longueur $b - a$. On l'appelle mesure de Lebesgue.*

*La complétion $\bar{\lambda}$ de λ , également appelée mesure de Lebesgue, est définie sur la tribu des **ensembles Lebesgue-mesurables**, qui est constituée de toutes les parties E de \mathbb{R} telles qu'il existe des ensembles Boréliens A et B tels que*

$$A \subset E \subset B; \quad \lambda[B \setminus A] = 0.$$

La construction présentée ci-dessous était déjà celle qu'utilisait Lebesgue.

DÉMONSTRATION DU THÉORÈME I-83. La famille des intervalles est stable par intersection finie (l'intersection de deux intervalles est un intervalle), et \mathbb{R} est l'union des intervalles $[-k, k]$ pour $k \in \mathbb{N}$; l'unicité de la mesure de Lebesgue est donc une conséquence du Théorème I-71(i).

L'existence demandera plus de travail. Considérons la famille \mathcal{F} de tous les intervalles de \mathbb{R} . Un intervalle $I \subset \mathbb{R}$ étant donné, on définit $\lambda[I]$ comme étant la longueur $|I|$ de I . L'intersection de deux intervalles est un intervalle, et le complémentaire d'un intervalle est la réunion de deux intervalles; nous sommes donc dans les conditions d'application du Théorème de prolongement I-71 (iii). Pour prouver que λ se prolonge en une mesure sur la tribu engendrée par \mathcal{F} , qui n'est autre que $\mathcal{B}(\mathbb{R})$, il suffit de vérifier la σ -additivité de λ . C'est un exercice qui s'énonce ainsi : *Étant donnée une famille d'intervalles $(I_k)_{k \in \mathbb{N}}$ disjoints, dont la réunion est un intervalle I de \mathbb{R} , prouver que*

$$(9) \quad \sum_{k \in \mathbb{N}} |I_k| = |I|.$$

Admettons provisoirement ce résultat; on peut alors appliquer le Théorème I-71 (iii) pour construire la mesure de Lebesgue via le concept de mesure extérieure de Carathéodory. Les propriétés de la complétion résultent alors des Théorèmes I-80 et I-82. \square

DÉMONSTRATION DE (9). Si l'on sait prouver (9) dans le cas où I est un intervalle borné, le cas général s'ensuivra : en effet, pour tout entier ℓ on peut poser $I^\ell = I \cap [-\ell, \ell]$, $I_k^\ell = I_k \cap [-\ell, \ell]$ et il est très facile de vérifier que

$$|I| = \sum_{\ell} |I^\ell|, \quad |I_k| = \sum_{\ell} |I_k^\ell|.$$

On suppose donc que I est borné. Si I n'est pas fermé, on peut toujours adjoindre à I un ou deux points (qui sont des intervalles particuliers, de longueur nulle!) : cela ne change ni $|I|$, ni $\sum |I_k|$. Il nous suffit donc de prouver (9) dans le cas particulier où $I = [a, b]$. Sans perte de généralité (le problème étant invariant par translation et dilatation), on pourra même supposer $I = [0, 1]$. Il est facile de vérifier que pour tout k ,

$$|I_1| + |I_2| + \dots + |I_k| \leq 1,$$

en particulier

$$\sum_{k \in \mathbb{N}} |I_k| \leq 1.$$

C'est bien sûr l'inégalité inverse qui est (légèrement) plus subtile.

Si $A \subset [0, 1]$ est réunion d'un nombre *fini* d'intervalles disjoints, on définira $|A|$ comme la somme des longueurs de ces intervalles; il est intuitivement évident (mais un tout petit peu fastidieux à vérifier) que cette définition est indépendante du choix de la décomposition de A en intervalles disjoints (par exemple, si on écrit $[a, c] = [a, b] \cup [b, c]$ on a $c - a = (b - a) + (c - b)$). On vérifie en outre que si $A = I_1 \cup \dots \cup I_n$, alors $|A| \leq |I_1| + \dots + |I_n|$, que les intervalles I_k soient disjoints ou non.

Soit $\varepsilon > 0$, arbitrairement petit. Pour tout k , on définit un intervalle J_k , ouvert dans $[0, 1]$, contenant I_k , tel que $|J_k \setminus I_k| \leq 2^{-k}\varepsilon$ (par exemple, si $I_k = [a, b]$ on pourra choisir $J_k =]a - 2^{k+1}\varepsilon, b + 2^{k+1}\varepsilon[\cap [0, 1]$). Les intervalles ouverts J_k recouvrent $[0, 1]$ tout entier, puisque les I_k forment déjà un recouvrement de $[0, 1]$. Par compacité, on peut en extraire un sous-recouvrement fini : il existe $K \in \mathbb{N}$ tel que $[0, 1] \subset J_1 \cup J_2 \cup \dots \cup J_K$. En particulier,

$$1 \leq \sum_{k=1}^K |J_k| \leq \sum_{k=1}^K (|I_k| + 2^{-k}\varepsilon) \leq \left(\sum_k |I_k| \right) + \varepsilon.$$

En faisant tendre ε vers 0, on obtient bien $1 \leq \sum_k |I_k|$, comme on le souhaitait. \square

I-7. Complément : Recouvrement et remplissage

Cette section pourra être omise en première lecture; outre qu'elle répond à certaines questions naturelles sur les liens entre ensembles mesurables et boules, elle s'avèrera d'un grand intérêt dans certains chapitres ultérieurs.

Pour étudier une mesure "localement" au voisinage d'un point, on considère la mesure de petites boules centrées en ce point; on développera ce point de vue dans le Chapitre ?? . Il est donc assez naturel de s'intéresser à des recouvrements d'ensembles par des petites boules. Dans le cadre de la théorie de la mesure, on ne sait gérer les mesures de familles d'ensembles que lorsqu'ils sont disjoints; le problème d'identifier des "sous-recouvrements disjoints" est donc assez naturel. Cependant, si un ensemble A est recouvert par des boules, on ne peut en général en tirer un sous-recouvrement disjoint; au mieux on peut espérer extraire une sous-famille disjointe, qui recouvre "presque" l'ensemble A , au sens où elle continue à en recouvrir une proportion non négligeable. Ce problème est l'objet de divers **lemmes de recouvrement**. On va ici considérer le plus simple d'entre eux, le **Lemme de Vitali**. Je vais l'énoncer avec des boules fermées, mais on pourrait aussi bien le faire avec des boules ouvertes.

Dans l'énoncé suivant, si $B = B[x, r]$ est une boule fermée de centre x et de rayon r et λ est un nombre positif, on note λB la boule de centre x et de rayon λr . (En général cette convention n'a de sens que si l'on considère B comme un couple (x, r) , de sorte que la valeur de r est uniquement déterminée par la boule choisie; dans un espace métrique général il est très possible que $B[x, r] = B[x, r']$ sans pour autant que r soit égal à r' !)

THÉORÈME I-84 (Lemme de recouvrement de Vitali). *(i) Soient X un espace métrique séparable, et \mathcal{B} une famille de boules fermées dans X , de rayon non nul et*

majoré. Alors de \mathcal{B} on peut extraire une famille dénombrable $\tilde{\mathcal{B}}$ de boules disjointes telles que

$$\bigcup_{B \in \mathcal{B}} B \subset \bigcup_{B \in \tilde{\mathcal{B}}} 4B.$$

(ii) En outre si μ est une mesure borélienne C -doublante sur X , on a

$$\mu\left[\bigcup_{B \in \tilde{\mathcal{B}}} B\right] \geq C^{-2} \mu\left[\bigcup_{B \in \mathcal{B}} B\right].$$

Avant d'aborder la preuve du Théorème I-84 proprement dite, commençons par deux lemmes très simples :

LEMME I-85. Soient B et B' deux boules fermées d'un espace métrique, de rayons respectifs r et r' , telles que

$$B' \cap B \neq \emptyset; \quad r' \geq \frac{2}{3}r.$$

Alors, avec les notations du Théorème I-84, on a $B \subset 4B'$.

LEMME I-86. Si X est un espace métrique séparable et \mathcal{B} est une famille quelconque de boules de rayon non nul, alors on peut extraire de \mathcal{B} une sous-famille disjointe maximale \mathcal{M} , ce qui veut dire que toute sous-famille de \mathcal{B} strictement plus grande que \mathcal{M} ne peut être disjointe.

PREUVE DU LEMME I-85. On écrit $B = B_{r\downarrow}(x)$, $B' = B_{r'\downarrow}(x')$. Par hypothèse il existe $z \in B \cap B'$. Alors, pour tout $y \in B$ on a $d(x', y) \leq d(x', z) + d(z, y) \leq r' + 2r \leq r' + 3r' = 4r'$. (Faire un dessin!) \square

PREUVE DU LEMME I-86. Cet énoncé relève a priori de la théorie des ensembles, et peut d'ailleurs se déduire facilement de l'axiome du choix (ou de manière équivalente du principe de maximalité de Hausdorff). Cependant l'hypothèse de séparabilité permet d'éviter l'usage de l'axiome du choix, comme on va le voir.

Soit $(z_n)_{n \in \mathbb{N}}$ une suite dense dans X . On va construire par récurrence la famille \mathcal{M} , comme suit.

Si z_1 appartient à l'un des éléments de \mathcal{B} , on choisit dans \mathcal{B} une boule B_1 contenant z_1 et on pose $\tilde{\mathcal{B}}_1 = \{B_1\}$. Dans le cas contraire, on pose $\tilde{\mathcal{B}}_1 = \emptyset$.

Si z_2 appartient à l'un des éléments de \mathcal{B} qui n'intersectent aucun élément de $\tilde{\mathcal{B}}_1$, on choisit dans \mathcal{B} une boule B_2 contenant z_2 et n'intersectant aucun élément de $\tilde{\mathcal{B}}_1$; on pose alors $\tilde{\mathcal{B}}_2 = \tilde{\mathcal{B}}_1 \cup \{B_2\}$. Dans le cas contraire, on pose $\tilde{\mathcal{B}}_2 = \tilde{\mathcal{B}}_1$.

Et ainsi de suite : si z_k appartient à l'un des éléments de \mathcal{B} qui n'intersectent aucun élément de $\tilde{\mathcal{B}}_{k-1}$, on pose $\tilde{\mathcal{B}}_k = \tilde{\mathcal{B}}_{k-1} \cup \{B_k\}$, où B_k est une boule de \mathcal{B} contenant z_k et n'intersectant aucun élément de $\tilde{\mathcal{B}}_{k-1}$; dans le cas contraire, on pose $\tilde{\mathcal{B}}_k = \tilde{\mathcal{B}}_{k-1}$.

Soit $\mathcal{M} = \{B_{i_1}, B_{i_2}, \dots, B_{i_j}, \dots\}$ l'union de toutes les familles \mathcal{B}_k ainsi construites. Il est clair que \mathcal{M} est dénombrable. Si B_k et B_ℓ appartiennent à \mathcal{M} , supposons par exemple $\ell > k$, alors B_ℓ a été choisie parmi les boules n'intersectant pas les éléments de $\tilde{\mathcal{B}}_{\ell-1}$, en particulier n'intersectant pas B_k ; donc la famille \mathcal{M} est disjointe.

Il reste à montrer que cette famille est maximale. Soit donc B un élément de \mathcal{B} n'appartenant pas à \mathcal{M} , montrons que $\mathcal{M} \cup \{B\}$ n'est pas disjointe. Puisque B est de rayon strictement positif, il existe un $z_m \in B$. À l'étape m de la construction,

- soit une boule B_m contenant z_m a été choisie et incluse dans la famille \mathcal{M} ; mais alors $B_m \cap B \neq \emptyset$;

- soit on n'a pas fait de tel choix, ce qui veut dire que toutes les boules de \mathcal{B} contenant z_m (en particulier la boule B) intersectaient déjà un élément de l'ensemble $\tilde{\mathcal{B}}_{m-1}$.

Dans les deux cas, B rencontre un élément de \mathcal{M} , ce qui achève la démonstration de la maximalité. \square

DÉMONSTRATION DU THÉORÈME I-84. Commençons par traiter le cas simple où il n'y a qu'un nombre fini de boules. On peut alors classer les boules par ordre décroissant du rayon : $r(B_1) \geq r(B_2) \geq r(B_3) \geq \dots$. On construit alors l'ensemble $\tilde{\mathcal{B}}$ selon la même procédure que précédemment : au début on pose $\tilde{\mathcal{B}}_1 = \{B_1\}$, puis si B_2 n'intersecte pas B_1 on pose $\tilde{\mathcal{B}}_2 = \{B_1, B_2\}$, sinon on pose $\tilde{\mathcal{B}}_2 = \{B_1\}$; et ainsi de suite. À l'étape k , si B_k n'intersecte aucun élément de $\tilde{\mathcal{B}}_{k-1}$ on pose $\tilde{\mathcal{B}}_k = \tilde{\mathcal{B}}_{k-1} \cup \{B_k\}$, et sinon on pose $\tilde{\mathcal{B}}_k = \tilde{\mathcal{B}}_{k-1}$.

On définit alors $\tilde{\mathcal{B}} = \cup \tilde{\mathcal{B}}_k$. Soit $B = B_k$ un élément quelconque de \mathcal{B} . S'il n'est pas dans $\tilde{\mathcal{B}}$, c'est qu'il intersecte une boule B_j de $\tilde{\mathcal{B}}_{k-1}$ avec bien sûr $j < k$, donc $r(B_j) \geq r(B_k)$. Posons $B' = B_j$, on a alors $B \cap B' \neq \emptyset$ et (avec les mêmes notations que dans le Lemme I-85) $r' \geq r$; d'où $B \subset 3B'$ par un argument similaire à celui du Lemme I-85.

Dans le cas général cependant, il est impossible d'ordonner les boules par ordre de rayon décroissant (l'énoncé autorise même une infinité non dénombrable de boules...). Il convient donc de modifier légèrement la stratégie. Si $B = B[x, r]$ est une boule de rayon r , on note $r = r(B)$. Par hypothèse il existe $R > 0$ tel que toutes les quantités $r(B)$ soient majorées par R . Pour tout $j \in \mathbb{N}$ on définit

$$\mathcal{B}_j := \left\{ B \in \mathcal{B}; \quad \left(\frac{2}{3}\right)^j R \leq 2r(B) \leq \left(\frac{2}{3}\right)^{j-1} R \right\}.$$

On choisit grâce au Lemme I-86 une famille dénombrable disjointe maximale $\tilde{\mathcal{M}}_1$ dans \mathcal{B}_1 .

On choisit ensuite une famille dénombrable disjointe maximale $\tilde{\mathcal{M}}_2$ dans

$$\mathcal{Z}_2 = \left\{ B \in \mathcal{B}_2; \quad \forall B' \in \mathcal{M}_1, \quad B' \cap B = \emptyset \right\}.$$

On continue de même : au rang k , on choisit une famille dénombrable disjointe maximale $\tilde{\mathcal{M}}_k$ dans

$$\mathcal{Z}_k = \left\{ B \in \mathcal{B}_k; \quad \forall B' \in \bigcup_{j \leq k-1} \tilde{\mathcal{B}}_j, \quad B' \cap B = \emptyset \right\}.$$

On pose enfin $\tilde{\mathcal{B}} = \cup \mathcal{M}_k$. Il est facile de montrer que cette famille est disjointe; il reste à vérifier que toute boule de \mathcal{B} est incluse dans $4B'$ pour une certaine boule $B' \in \tilde{\mathcal{B}}$.

Soit donc $B \in \mathcal{B}$. Si $B \in \tilde{\mathcal{B}}$, le résultat est évident. Sinon, introduisons k tel que $B \in \mathcal{B}_k$. Puisque $\tilde{\mathcal{B}}_k$ est une famille disjointe maximale dans l'ensemble

$$\left\{ B \in \mathcal{B}_k; \quad \forall B' \in \bigcup_{j \leq k-1} \tilde{\mathcal{B}}_j, \quad B' \cap B = \emptyset \right\},$$

il n'y a que deux possibilités :

- soit B n'appartient pas à \mathcal{Z}_k , ce qui veut dire que B intersecte un élément de l'un des $\widetilde{\mathcal{M}}_j$ pour $j \leq k-1$;
- soit B appartient à \mathcal{Z}_k , et alors la famille obtenue en adjoignant B à \mathcal{M}_k n'est pas disjointe, ce qui veut dire que B intersecte un élément de \mathcal{M}_k .

Dans tous les cas, B intersecte un élément B' de $\bigcup_{j \leq k} \widetilde{\mathcal{M}}_j$; en particulier $r(B') \geq (2/3)r(B)$. On applique alors le Lemme I-85 pour conclure que $B \subset 4B'$. Ceci conclut la preuve de (i).

Passons maintenant à (ii) : Pour cela on écrit

$$\mu\left[\bigcup_{B \in \mathcal{B}} B\right] \leq \mu\left[\bigcup_{B \in \widetilde{\mathcal{B}}} 4B\right] \leq \sum_{B \in \widetilde{\mathcal{B}}} \mu[4B] \leq C^2 \sum_{B \in \widetilde{\mathcal{B}}} \mu[B] = C^2 \mu\left[\bigcup_{B \in \widetilde{\mathcal{B}}} B\right],$$

où l'avant-dernière inégalité découle de la propriété de C -doublement, et la dernière provient de ce que la famille \mathcal{B} est *disjointe*. \square

Voici maintenant un corollaire frappant et utile du Lemme de Vitali; il énonce que l'on peut remplir, au sens de la théorie de la mesure, un ouvert par de petites boules (fermées) *disjointes* :

COROLLAIRE I-87. *Soient X un espace métrique séparable, μ une mesure borélienne sur X , doublante et finie sur les boules fermées de X . Alors, pour tout ouvert O de X et pour tout $\delta > 0$ on peut trouver une famille dénombrable \mathcal{G} de boules fermées disjointes $B[x, r] \subset O$, de rayon $r \leq \delta$, telles que*

$$\mu\left[O \setminus \bigcup_{B \in \mathcal{G}} B\right] = 0.$$

EXEMPLE I-88. Nous verrons plus tard que la mesure naturelle dans \mathbb{R}^n , la mesure de Lebesgue, est 2^n -doublante. Il s'ensuivra que tout ouvert de \mathbb{R}^n est, à un ensemble de mesure de Lebesgue nulle près, union dénombrable de boules euclidiennes fermées disjointes.

PREUVE DU COROLLAIRE I-87. 1. Traitons d'abord le cas où O est inclus dans une boule fermée, en particulier $\mu[O]$ est fini et μ est C -doublante sur O . Soit \mathcal{B} l'ensemble de toutes les boules fermées de rayon au plus δ , incluses dans O . Puisque O est ouvert, la réunion de tous les éléments de \mathcal{B} est exactement O . Par le Théorème I-84, il existe une famille dénombrable disjointe $\widetilde{\mathcal{B}} \subset \mathcal{B}$ telle que

$$\mu\left[\bigcup_{B \in \widetilde{\mathcal{B}}} B\right] \geq C^{-2} \mu[O];$$

d'où

$$\mu\left[O \setminus \bigcup_{B \in \widetilde{\mathcal{B}}} B\right] \leq (1 - C^{-2}) \mu[O].$$

Par σ -additivité, il existe une sous-famille finie $\mathcal{B}' \subset \widetilde{\mathcal{B}}$, telle que

$$\mu\left[O \setminus \bigcup_{B \in \mathcal{B}'} B\right] \leq \left(1 - \frac{C^{-2}}{2}\right) \mu[O].$$

On pose $O_1 := O \setminus \bigcup_{B \in \mathcal{B}'} B$: comme intersection finie d'ouverts, c'est un ouvert, il est inclus dans O et de mesure au plus $\lambda \mu[O]$ avec $\lambda = (1 - C^{-2}/2) < 1$.

On itère alors la construction : par récurrence on construit une suite décroissante d'ouverts O_k , tel que $O_{k-1} \setminus O_k$ est une union finie de boules fermées, et $\mu[O_k] \leq$

$\lambda \mu[O_{k-1}]$. Par σ -additivité, l'intersection des O_k est de mesure nulle, et son complémentaire dans O est une union dénombrable de boules fermées.

2. Considérons maintenant le cas général où O n'est pas inclus dans une boule. Soit x_0 un élément quelconque de X , pour tout $r > 0$ on pose $S_r = \{x \in X; d(x_0, x) = r\}$. (C'est la sphère de centre x_0 et de rayon r .) Puisque X est l'union des boules $B(x_0, k)$, $k \in \mathbb{N}$, elle est σ -finie; en particulier il y a au plus une infinité dénombrable de $r > 0$ tels que $\mu[S_r] > 0$. Fixons une fois pour toute une suite $r_k \rightarrow \infty$ ($k \in \mathbb{N}$) telle que $\mu[S_{r_k}] = 0$. On pose alors $C_0 = B(x_0, r_1)$, et pour $k \in \mathbb{N}$, $C_k = B(x_0, r_{k+1}) \setminus B(x_0, r_k)$. (Les C_k sont donc des coronnes ouvertes disjointes.) Le complémentaire des C_k dans X est de mesure nulle, en particulier O est, à un ensemble de mesure nulle près, l'union disjointe des ouverts $O_k = O \cap C_k$. Par la première partie de la preuve, chacun des O_k peut s'écrire, à un ensemble de mesure nulle près, comme une union disjointe de boules fermées $B_{k,j}$ ($j \in \mathbb{N}$). Ceci conclut la preuve du théorème. \square

REMARQUE I-89. Le Lemme de Vitali est le lemme de recouvrement le plus simple et le plus connu; il en existe cependant bien d'autres, utilisés dans des situations variées. Parmi les plus intéressants, mentionnons

- le **Lemme de recouvrement de Besicovich** [Evans-Gariepy pp. 30–35] : Soit $\mathcal{B} = \{B(x_\alpha, r_\alpha)_{\alpha \in A}\}$ une collection de boules de rayon borné dans l'espace Euclidien \mathbb{R}^n , et soit C l'ensemble de leurs centres. Alors il existe une constante K , ne dépendant que de n , et des sous-familles disjointes dénombrables $\mathcal{B}_1, \dots, \mathcal{B}_K$ de \mathcal{B} , qui recouvrent l'ensemble C . Ce lemme, qui exploite la structure particulière de l'espace \mathbb{R}^n , permet d'étudier des mesures non nécessairement doublantes : par exemple, on peut l'utiliser pour montrer que le Corollaire I-87 reste vrai si O est un ouvert de \mathbb{R}^n et μ une mesure *arbitraire*.

- le **Lemme de recouvrement de Whitney**, très utile en analyse harmonique, qui permet de remplir un ouvert O de \mathbb{R}^n par une famille dénombrable de cubes C_k , dont les côtés sont parallèles aux axes, dont les intérieurs sont disjoints (au sens où leurs intérieurs sont disjoints), et dont les diamètres sont à peu près proportionnels à leur distance au bord de O :

$$\text{diam}(C_k) \leq d(C_k, \mathbb{R}^n \setminus O) \leq 4 \text{diam}(C_k).$$

Pour en savoir plus, on pourra consulter le passionnant ouvrage d'**E.M. Stein**, *Singular Integrals and Differentiability Properties of Functions* (Princeton University Press, New Jersey, 1970), pp. 16–18 et Chapitre VI.

CHAPITRE II

Intégration selon Lebesgue et selon Riesz

Dans ce chapitre, nous allons définir l'intégrale de Lebesgue pour une large classe de fonctions, que nous appellerons **sommables** au sens de Lebesgue. Nous nous placerons dans un cadre abstrait, qui inclura comme cas particulier l'intégration des fonctions continues. Le point de départ sera la notion de fonction mesurable, introduite dans la section II-1 ; on définira ensuite l'intégrale dans la section II-2, et on vérifiera dans la section II-3 qu'elle constitue une forme linéaire.

Il existe un autre point de vue pour l'intégration, qui consiste à prendre les fonctions continues comme point de départ. Le **théorème de Riesz**, qui est l'objet de la section II-4, assure que ces deux points de vue sont équivalents (modulo quelques subtilités), sous certaines hypothèses topologiques sur l'espace ambiant. La plus restrictive de ces hypothèses est la condition de **compacité locale**, qui est satisfaite par \mathbb{R}^n ou par n'importe quelle variété Riemannienne de dimension finie, mais mise en défaut par de nombreux espaces intéressants, au premier rang desquels se trouve l'espace de Wiener.

Ce chapitre se conclut par quelques mots sur l'intégration à valeurs vectorielles, qui sera abordée plus en détail dans un chapitre ultérieur.

II-1. Fonctions mesurables

On cherche à définir une large classe de fonctions susceptibles d'être intégrées, que nous appellerons **fonctions mesurables**. Il est souhaitable que cette classe de fonctions contienne au moins les fonctions indicatrices d'ensembles mesurables, et (dans un cadre topologique) les fonctions continues.

En topologie, où les parties ouvertes jouent un rôle privilégié, on dit qu'une fonction entre deux espaces topologiques est continue si l'image réciproque de tout ouvert est un ouvert. En théorie de la mesure, on adopte une démarche similaire pour définir les fonctions mesurables.

II-1.1. Définition.

DÉFINITION II-1 (fonctions mesurables). *Soient X et Y deux espaces mesurables, et $f : X \rightarrow Y$. On dit que f est mesurable si l'image réciproque $f^{-1}(B)$ de n'importe quelle partie mesurable $B \subset Y$ est une partie mesurable de X .*

EXEMPLE II-2. L'exemple le plus simple est la fonction indicatrice $f = 1_A$ d'un ensemble mesurable A : c'est la fonction qui vaut 1 sur A et 0 sur le complémentaire. L'image réciproque d'un ensemble *quelconque* par f est l'un des quatre ensembles $\emptyset, X, A, X \setminus A$, qui sont tous bien sûr mesurables.

Nous verrons dans la suite des critères pratiques de mesurabilité, permettant de construire de très nombreuses fonctions mesurables. En fait, comme on le commentera dans la Remarque II-19, la mesurabilité est la règle plutôt que l'exception.

REMARQUES II-3. (i) Soient X un espace mesurable, Y un ensemble quelconque, et $f : X \rightarrow Y$ une fonction quelconque. Il est toujours possible de munir Y d'une σ -algèbre pour laquelle f soit une fonction mesurable. Il suffit pour cela de **définir** une partie mesurable comme une partie dont l'image réciproque est mesurable. On pourra vérifier à titre d'exercice que cela définit bien une tribu. Cette tribu est **la plus grande** qui rende f mesurable : toute tribu plus petite a la même propriété. On l'appelle tribu image de \mathcal{A} par f et on la note souvent $f_{\#}\mathcal{A}$.

(ii) Soient maintenant X un ensemble quelconque, Y un espace mesurable, et $f : X \rightarrow Y$ une fonction quelconque. Il est encore possible de munir X d'une σ -algèbre pour laquelle f soit une fonction mesurable. Il suffit pour cela de considérer les images réciproques des ensembles mesurables. Cette tribu est **la plus petite** qui rende f mesurable : toute tribu plus grande a la même propriété. On l'appelle tribu engendrée par f ; nous aurons l'occasion de revenir sur cette notion.

(iii) De même que l'image d'un ouvert par une application continue n'est en général pas ouverte, l'image d'un ensemble mesurable par une application mesurable f n'est en général pas mesurable. C'est cependant le cas si $f : X \rightarrow Y$ est une application **bijective** entre deux espaces Polonais X et Y munis de leur tribu borélienne : sous ces hypothèses, l'image de tout ensemble mesurable par f est mesurable (voir [Parthasarathy]). En particulier, la réciproque d'une bijection mesurable entre espaces Polonais est automatiquement mesurable.

Le critère pratique qui suit est d'usage constant.

PROPOSITION II-4 (critère pratique de mesurabilité). *Soient (X, \mathcal{A}) et (Y, \mathcal{B}) deux espaces mesurables, et $f : X \rightarrow Y$. On suppose que la tribu \mathcal{B} est engendrée par une famille \mathcal{F} de parties de Y : $\mathcal{B} = \sigma(\mathcal{F})$. Alors f est mesurable si et seulement si pour tout $F \in \mathcal{F}$, $f^{-1}(F)$ est mesurable.*

DÉMONSTRATION. Définissons

$$\mathcal{C} := \{B \in \mathcal{B}; f^{-1}(B) \in \mathcal{A}\} = f_{\#}\mathcal{A}.$$

En utilisant les formules

$$f^{-1}(Y \setminus B) = f^{-1}(Y) \setminus f^{-1}(B), \quad f^{-1}\left(\bigcup B_k\right) = \bigcup f^{-1}(B_k),$$

on montre que \mathcal{C} est une tribu. Comme elle contient \mathcal{F} , elle contient également $\sigma(\mathcal{F}) = \mathcal{B}$. \square

EXEMPLES II-5. (i) L'exemple le plus courant est le suivant : si X et Y sont deux espaces topologiques, munis de leur tribu borélienne, et $f : X \rightarrow Y$ est une fonction quelconque, alors f est mesurable si et seulement si l'image réciproque de tout ouvert de Y est un Borélien de X . On dit alors que f est **borélienne**. En particulier, toute fonction continue est borélienne.

(ii) Dans le cas où $Y = \mathbb{R}^n$, pour montrer qu'une fonction $f : X \rightarrow Y$ est mesurable, il suffit de vérifier que l'image réciproque de tout pavé est un Borélien de X . Par exemple, si on réussit à montrer que c'est une union dénombrable d'intersections dénombrables d'unions dénombrables de fermés...

- (iii) Dans le cas où $Y = \mathbb{R}$, il suffit de vérifier que l'image réciproque de tout intervalle semi-ouvert, de la forme $I = [y, +\infty[$, est un Borélien. En particulier, toute fonction semi-continue inférieurement (ou supérieurement) est borélienne. Mais en général la classe des fonctions boréliennes est beaucoup plus large.
- (iv) Dans le cas où Y est un espace produit, muni de la topologie produit, il suffit de vérifier que l'image réciproque de tout pavé est mesurable. Pour un produit infini, il suffit de vérifier que l'image réciproque de tout cylindre est mesurable.

On va voir au paragraphe suivant comment on peut, via des opérations simples, construire de très nombreuses fonctions boréliennes qui ne sont pas du tout continues, ni continues par morceaux. Toutefois, sous certaines hypothèses topologiques, les fonctions boréliennes peuvent être bien **approchées** par des fonctions continues, et encore mieux par des fonctions semi-continues; les théorèmes de Lusin et Vitali-Carathéodory, que nous évoquerons dans la section II-4.4, en sont une bonne illustration.

II-1.2. Stabilité des fonctions mesurables. Il est très facile de construire des fonctions discontinues en manipulant des fonctions continues et un peu de théorie des ensembles : il suffit par exemple de définir une fonction séparément sur $[a, b]$ et $]b, c]$. En revanche, en manipulant des fonctions mesurables et des ensembles mesurables, on ne peut guère construire que des fonctions mesurables !

PROPOSITION II-6 (restriction). *Soient (X, \mathcal{A}) et (Y, \mathcal{B}) deux espaces mesurables, et A une partie mesurable de X . Soit $f : X \rightarrow Y$ une application mesurable. On munit A de la tribu induite par \mathcal{A} , i.e. l'ensemble de tous les éléments de \mathcal{A} qui sont inclus dans A . Alors la restriction de f à A est une application mesurable de A dans Y .*

DÉMONSTRATION. C'est un simple jeu de maniement des axiomes. \square

PROPOSITION II-7 (recollement). *Dans un espace mesurable X , soit $(A_k)_{k \in \mathbb{N}}$ une famille dénombrable de parties mesurables disjointes, telle que $X = \cup A_k$. Soit également Y un espace mesurable. Sur chaque A_k (considéré comme espace mesurable), on se donne une fonction mesurable $f_k : A_k \rightarrow Y$. Soit f la fonction qui pour tout k coïncide avec f_k sur A_k . Alors f est mesurable.*

DÉMONSTRATION. L'image réciproque d'un ensemble mesurable B par f est l'union des ensembles mesurables $f_k^{-1}(B)$, c'est donc un ensemble mesurable. \square

PROPOSITION II-8 (produit infini de fonctions mesurables). *Soient $(X_t)_{t \in T}$ et $(Y_t)_{t \in T}$ des espaces mesurés, dépendant d'un paramètre $t \in T$, et soient $(f_t)_{t \in T}$ des fonctions mesurables de X_t dans Y_t respectivement. On munit $X = \prod X_t$ et $Y = \prod Y_t$ de la tribu produit, i.e. la plus petite tribu qui rende mesurable tous les cylindres. Alors l'application $f = \prod f_t$ est mesurable de X dans Y .*

En particulier, si f_1, \dots, f_k sont des applications mesurables définies sur X_1, \dots, X_k respectivement, à valeurs dans Y_1, \dots, Y_k respectivement, alors l'application $f = (f_1, \dots, f_k)$ est mesurable de $X_1 \times \dots \times X_k$ dans $Y_1 \times \dots \times Y_k$.

DÉMONSTRATION. Par le critère II-4, il suffit de montrer que l'image réciproque de tout cylindre est mesurable; on peut même se limiter à le faire pour les cylindres

engendrés par des pavés :

$$C(B_{t_1}, \dots, B_{t_K}) = \left\{ y \in Y; \forall k \in \{1, \dots, K\}, y_{t_k} \in B_{t_k} \right\},$$

où K est arbitraire et chaque B_{t_k} est une partie mesurable de Y_{t_k} (pour alléger l'écriture, on n'a pas incorporé dans les notations le fait que le cylindre dépend explicitement du choix de t_1, \dots, t_K , et pas seulement des ensembles B_{t_1}, \dots, B_{t_K}). On vérifie aisément que

$$f^{-1}(C(B_{t_1}, \dots, B_{t_K})) = C(A_{t_1}, \dots, A_{t_K}), \quad A_{t_k} = f_k^{-1}(B_{t_k}).$$

C'est en particulier un ensemble mesurable. \square

PROPOSITION II-9 (composition). *Soient $f : X \rightarrow Y$ et $g : Y \rightarrow Z$ deux applications mesurables entre espaces mesurables, alors leur composition $g \circ f$ est mesurable.*

DÉMONSTRATION. C'est une conséquence immédiate de la définition. \square

COROLLAIRE II-10. *Soient X, Y et Z des espaces mesurables, tels que Y et Z sont des espaces topologiques munis de leur tribu borélienne, soient $f : X \rightarrow Y$ une application mesurable, et $\varphi : Y \rightarrow Z$ une application continue. Alors $\varphi \circ f : X \rightarrow Z$ est une application mesurable.*

DÉMONSTRATION. C'est une conséquence de la proposition précédente, combinée avec l'exemple II-5(i). \square

A l'aide de ces critères simples, il est facile de trouver beaucoup d'opérations élémentaires qui préservent la notion de mesurabilité. La proposition suivante rassemble les plus courantes.

PROPOSITION II-11 (opérations élémentaires). *Soient X un espace mesuré, et f, g deux fonctions mesurables de X dans \mathbb{R} (muni de la tribu borélienne). Alors les fonctions $f + g$, $f - g$, fg , $\min(f, g)$, $\max(f, g)$ et, si g ne s'annule pas, f/g sont mesurables.*

DÉMONSTRATION. On applique le corollaire II-10 avec les applications continues addition, soustraction, etc. Noter que $\min(f, g) = (f + g)/2 - |f - g|/2$. \square

Outre ces opérations élémentaires, une opération fréquemment utilisée pour définir des fonctions est la **limite** (ou ses avatars tels que série, etc.). Pour parler de limites, il sera bien commode de remplacer \mathbb{R} par $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, dont on fait un espace topologique en décidant que les intervalles $]a, +\infty[$ et $]a, +\infty]$ ($a \in \mathbb{R}$) engendrent les ouverts. La restriction de cette topologie à \mathbb{R} est la topologie usuelle ; en effet, les intervalles $]a, +\infty[$ engendrent la topologie usuelle de \mathbb{R} . On note que $\overline{\mathbb{R}}$ est un espace métrique compact. On munira toujours $\overline{\mathbb{R}}$ de sa tribu borélienne. La raison pour considérer $\overline{\mathbb{R}}$ plutôt que \mathbb{R} est simple : bien souvent, une fonction admettra en certains points une limite dans $\overline{\mathbb{R}}$ et non dans \mathbb{R} .

PROPOSITION II-12 (Boréliens de $\overline{\mathbb{R}}$). *Les Boréliens de $\overline{\mathbb{R}}$ sont engendrés par les intervalles de la forme $[a, +\infty]$ ($a \in \mathbb{R}$), ou par les intervalles de la forme $]a, +\infty]$.*

DÉMONSTRATION. Il suffit par exemple de vérifier que la tribu engendrée par les intervalles $[a, +\infty]$ contient tous les intervalles de la forme $]a, +\infty[$. En prenant le complémentaire de $[a, +\infty]$ on obtient un intervalle $[-\infty, a[$; en prenant l'intersection de $[-\infty, b[$ et de $[a, +\infty]$ on obtient un intervalle $[a, b[$; en prenant l'union

des $[a + 1/k, b[$ ($k \in \mathbb{N}$) on obtient l'intervalle ouvert $]a, b[$; enfin $]a, +\infty[$ est l'union dénombrable des intervalles $]a, k[$. \square

REMARQUE II-13. Attention aux opérations dans $\overline{\mathbb{R}}$: l'opération $(x, y) \rightarrow x + y$ n'est **pas** bien définie dans $\overline{\mathbb{R}}$ tout entier à cause de l'indétermination $(+\infty) + (-\infty)$. Ainsi, si f et g sont deux fonctions à valeurs dans $\overline{\mathbb{R}}$, on ne peut pas affirmer que $f + g$ soit mesurable. Si on sait que f et g prennent leurs valeurs dans $\mathbb{R} \cup \{+\infty\}$, alors $f + g$ est bien mesurable; mais $f - g$ n'est pas forcément défini...

THÉORÈME II-14 (stabilité par limite). (i) Soit $(f_n)_{n \in \mathbb{N}}$ une famille de fonctions mesurables sur un espace mesurable X , à valeurs dans $\overline{\mathbb{R}}$. Alors les fonctions $\liminf f_n$ et $\limsup f_n$ sont mesurables.

(ii) En particulier, si la famille (f_n) converge simplement, i.e. pour chaque x la suite $f_n(x)$ converge dans $\overline{\mathbb{R}}$, alors la fonction $\lim f_n$ est mesurable.

(iii) Si la famille (f_n) converge dans $\overline{\mathbb{R}}$ sur une partie C de X , et g est une fonction mesurable quelconque de X (par exemple la fonction nulle), alors la fonction f définie par

$$\begin{cases} f(x) = \lim f_n(x) & \text{si } x \in C \\ f(x) = g(x) & \text{sinon} \end{cases}$$

est mesurable de X dans $\overline{\mathbb{R}}$.

Les énoncés (ii) et (iii) restent vrais si l'on remplace $\overline{\mathbb{R}}$ par \mathbb{R} .

REMARQUE II-15. On rappelle que la limite supérieure (respectivement inférieure) d'une suite d'éléments de $\overline{\mathbb{R}}$ est sa plus grande (respectivement plus petite) valeur d'adhérence dans $\overline{\mathbb{R}}$. La limite supérieure et la limite inférieure existent toujours dans $\overline{\mathbb{R}}$, pas forcément dans \mathbb{R} .

DÉMONSTRATION. Soit (f_n) une famille de fonctions mesurables, on va montrer par exemple que $\limsup f_n$ est mesurable. Il suffit de montrer que pour tout $\alpha \in \mathbb{R}$, l'ensemble $C_\alpha := \{x \in X; \limsup f_n(x) \geq \alpha\}$, avec $\alpha \in \mathbb{R}$, est mesurable. Dire que x appartient à C_α , c'est dire que pour tout k , $f_n(x)$ prend une infinité de fois une valeur supérieure ou égale à $\alpha - 1/k$, ou encore : pour tout k , pour tout n il existe $m \geq n$ tel que $f_m(x) \geq \alpha - 1/k$. Autrement dit,

$$C_\alpha = \bigcap_{k \in \mathbb{N}} \bigcap_{n \in \mathbb{N}} \bigcup_{m \in \mathbb{N}} f_m^{-1}([\alpha - 1/k, +\infty]),$$

et C_α est bien mesurable, ce qui conclut la démonstration du point (i).

On en déduit en particulier que l'ensemble C des points de convergence est mesurable : en effet, C est l'union de trois ensembles qui d'après (i) sont tous mesurables :

- l'ensemble où $f_n \rightarrow +\infty$, i.e. $\liminf f_n = +\infty$;
- l'ensemble où $f_n \rightarrow -\infty$, i.e. $\limsup f_n = -\infty$;
- l'ensemble où f_n converge dans \mathbb{R} , i.e. $\liminf f_n < +\infty$ et $\limsup f_n > -\infty$ et $\limsup f_n - \liminf f_n = 0$.

On peut donc considérer C comme espace mesuré, et on en déduit facilement (ii) grâce à la Proposition II-6. On prouve (iii) de même, en appliquant la proposition II-7. \square

EXERCICE II-16. D'après le théorème précédent, si une famille de fonctions (f_n) converge en tout point dans \mathbb{R} , la limite est une fonction mesurable. Démontrer ce résultat sans faire référence à des \liminf et \limsup , via le critère : *une suite de nombres réels converge si et seulement si elle est de Cauchy*. En déduire une généralisation du point (ii) dans laquelle l'espace d'arrivée \mathbb{R} est remplacé par un espace métrique complet arbitraire.

EXEMPLES II-17. (i) Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction dérivable, alors sa dérivée est mesurable. En effet, c'est la limite simple de la suite de fonctions continues

$$g_k(x) = k[f(x + 1/k) - f(x)].$$

- (ii) Une série de Fourier convergente définit une fonction mesurable. Rappelons que de telles fonctions peuvent ne pas être continues.
- (iii) L'intégrale de Riemann étant définie par un procédé de limite, les fonctions définies comme des intégrales de Riemann à paramètre sont mesurables. Nous verrons au Chapitre III que cet énoncé se généralise à des intégrales à paramètre définies dans le cadre de la théorie de Lebesgue.

Citons pour conclure un dernier résultat de stabilité de la mesurabilité, que nous admettrons (voir la démonstration dans [Parthasarathy]).

THÉORÈME II-18 (mesurabilité de l'inverse). *Soient X et Y deux espaces Polonais, munis de leurs tribus boréliennes respectives, et f une bijection mesurable de X dans Y . Alors f^{-1} est mesurable.*

REMARQUES II-19. (i) En conséquence des théorèmes de stabilité que nous avons énoncés, la plupart (sinon la totalité) des fonctions que l'on est amené à construire pour résoudre des problèmes concrets dans \mathbb{R} ou \mathbb{R}^d sont mesurables. Cela ne veut pas dire qu'il faille forcément se dispenser de montrer la mesurabilité, quand on applique un théorème du cours ! Cependant, dans la quasi-totalité des problèmes que l'on rencontre en analyse réelle, toutes les fonctions sont mesurables et, le plus souvent, on ne se donne même pas la peine de le mentionner. Cette règle admet quelques exceptions ; nous en reparlerons dans le chapitre IV.

- (ii) La situation est un peu différente en théorie des probabilités, surtout dans l'étude des processus stochastiques. D'une part, dans ce contexte on considère souvent des produits infinis, parfois indexés par un ensemble non dénombrable, pour lesquels la mesurabilité peut être une propriété non triviale. D'autre part, dans ce domaine on est souvent amené à définir des tribus plus ou moins grandes, emboîtées les unes dans les autres, et la mesurabilité par rapport aux plus ces tribus pourra être facilement violée. Voici un exemple classique : soit $X = \{0, 1\}^{\mathbb{N}}$, que l'on interprète comme l'espace des suites de résultats obtenus quand on fait une infinité dénombrable de tirages pile ou face (0 pour face, 1 pour pile). On décide naturellement que toutes les parties de $\{0, 1\}$ sont mesurables. Il est naturel d'introduire une fois pour toutes la tribu produit (infini) sur X , qui rend mesurables toutes les applications coordonnées $\sigma_k : x \mapsto x_k$. Cependant, dans une perspective stochastique, on préfère souvent munir X de la famille de tribus emboîtées $(\mathcal{A}_n)_{n \in \mathbb{N}}$, où \mathcal{A}_n est la tribu engendrée par les applications $\sigma_1, \dots, \sigma_n$ (constituée des parties cylindriques dont la base est un sous-ensemble arbitraire des n premiers facteurs). Un indice n étant fixé,

il existe bien sûr de très nombreuses fonctions qui ne sont pas \mathcal{A}_n -mesurables (à commencer par σ_{n+1}).

- (iii) Si X est un espace topologique, muni de sa tribu borélienne, on a vu que toute limite simple d'une suite de fonctions continues est mesurable. Il est naturel de se demander si la réciproque est vraie ; dans un tel cas on pourrait définir les fonctions mesurables comme les limites de fonctions continues. La réponse est négative : on ne peut pas en général approcher une fonction mesurable par des fonctions continues. Cependant, sous certaines hypothèses topologiques (compacité locale, par exemple), la réponse devient positive si on s'autorise à oublier un ensemble de mesure nulle : c'est un corollaire du **théorème de Lusin**, que nous démontrerons à la fin de ce chapitre. Nous mentionnerons également le **théorème de Vitali-Carathéodory**, qui montre que l'on peut approcher une fonction mesurable par des fonctions semi-continues plus grandes, ou plus petites, au prix d'une erreur arbitrairement petite sur l'intégrale.

II-1.3. Tribu engendrée par une fonction mesurable.

THÉORÈME II-20 (tribu engendrée par une fonction). *(i) Soient X un espace quelconque et Y un espace mesurable, et soit f une application quelconque de X dans Y . Il existe alors une plus petite tribu sur X qui rende f mesurable ; on la note $\sigma(f)$. Elle est faite de tous les ensembles $f^{-1}(B)$, où B est une partie mesurable quelconque de Y .*

Si au départ X est un espace mesurable, muni d'une tribu \mathcal{A} , et f est mesurable, alors $\sigma(f) \subset \mathcal{A}$.

(ii) Plus généralement, soient X un espace quelconque, $(Y_t)_{t \in T}$ une famille d'espaces mesurables indexés par un ensemble T arbitraire ; pour tout $t \in T$ on se donne une fonction $f_t : X \rightarrow Y_t$. Alors il existe une plus petite tribu sur X qui rende mesurables toutes les applications f_t ; on la note $\sigma((f_t)_{t \in T})$. Si X est au départ un espace mesurable, muni d'une tribu \mathcal{A} , et chacune des f_t est mesurable, alors $\sigma((f_t)_{t \in T}) \subset \mathcal{A}$.

DÉMONSTRATION. L'énoncé (i) est une conséquence immédiate des définitions, et des formules

$$f^{-1}\left(\bigcap B_k\right) = \bigcap f^{-1}(B_k); \quad f^{-1}\left(\bigcup B_k\right) = \bigcup f^{-1}(B_k).$$

Pour l'énoncé (ii), on construit la tribu $\sigma((f_t)_{t \in T})$ comme l'intersection de toutes les tribus contenant toutes les tribus $\sigma(f_t)$. \square

Intuitivement, la tribu engendrée par une fonction mesurable f est faite des parties dont la définition “ne fait intervenir que les valeurs de f ” ; une fonction mesurable pour la tribu $\sigma(f)$ est donc une fonction qui “ne dépend que de f ” – une propriété qu'il peut être utile de formaliser dans des contextes très variés. Essayons de caractériser ces fonctions. Pour se convaincre que le problème est assez subtil, expliquer pourquoi la “démonstration” ci-dessous est incomplète.

PRÉTENDU THÉORÈME II-21 (fonctions mesurables pour $\sigma(f)$). *Soient X un espace quelconque, et Y et Z deux espaces mesurables. On se donne f une fonction quelconque de X dans Y , et on munit X de la tribu $\sigma(f)$. On suppose en outre que les singletons de Z sont mesurables. Alors les fonctions mesurables de X dans Z*

sont exactement les fonctions de la forme $\Phi \circ f$, où Φ est une fonction mesurable de Y dans Z .

PRÉTENDUE DÉMONSTRATION. Soit g une fonction mesurable de X dans Z . On pose $B = f(X)$. On va construire une fonction mesurable Φ sur B telle que $g = \Phi \circ f$. On pourra ensuite attribuer une valeur quelconque à Φ sur $Y \setminus B$.

Soit maintenant $z \in Z$, par hypothèse $\{z\}$ est mesurable, et $A_z := g^{-1}(\{z\})$ est un élément de $\sigma(f)$, il s'écrit donc $f^{-1}(B_z)$ avec B_z mesurable, que l'on peut choisir inclus dans B . Les B_z sont deux à deux disjoints : si $y \in B_z \cap B_{z'}$ avec $z \neq z'$, alors on écrit $y = f(x)$, d'où $x \in f^{-1}(B_z) \cap f^{-1}(B_{z'}) = g^{-1}(\{z\}) \cap g^{-1}(\{z'\}) = \emptyset$. Ils recouvrent par ailleurs B , puisque tout élément y de B s'écrit sous la forme $f(x)$, on peut alors poser $z = g(x)$ et on a $x \in A_z$, d'où $x \in f^{-1}(B_z)$, d'où $f(x) \in B_z$. Tout $y \in B$ appartient donc à un unique B_z , et on peut alors poser $\Phi(y) = z$. \square

Nous verrons plus loin comment démontrer rigoureusement un énoncé un peu moins ambitieux.

II-1.4. Fonctions mesurables et complétion. Le théorème suivant fait le lien entre fonctions mesurables pour une tribu, et fonctions mesurables pour la tribu complétée.

THÉORÈME II-22 (mesurabilité pour la tribu complétée). *Soit (X, \mathcal{A}, μ) une tribu, et $\overline{\mathcal{A}}$ la complétion de \mathcal{A} pour μ ; soit f une fonction mesurable pour la tribu $\overline{\mathcal{A}}$. Alors il existe une fonction g , mesurable pour la tribu \mathcal{A} , telle que $f = g$ μ -presque partout.*

Nous démontrerons ce théorème dans la section II-2.2, grâce à une très efficace méthode d'approximation des fonctions mesurables.

II-2. Construction de l'intégrale

De la même façon que les ensembles mesurables sont ceux dont on définit la mesure, les fonctions mesurables sont celles dont on espère définir l'intégrale.

Cependant, nous ne parviendrons pas à définir une intégrale pour **toutes** les fonctions mesurables. L'intégrale peut bien sûr diverger (valoir formellement $+\infty$), mais ce n'est pas très grave : on peut attribuer la valeur $+\infty$. Plus grave est la possibilité de compensation entre valeurs positives et valeurs négatives. Dans le cas de l'intégrale de Riemann, on connaît bien les difficultés liées au maniement des intégrales **semi-convergentes**, pour la définition desquelles divers procédés de limite ad hoc (qui s'étendent mal à des espaces plus généraux que \mathbb{R}) sont nécessaires. Dans la théorie de Lebesgue, nous éluderons cette difficulté en ne considérant que des intégrales absolument convergentes.

On commencera par définir l'intégrale d'une fonction positive : cette quantité sera **toujours** bien définie, pourvu que l'on admette la valeur $+\infty$. Comme dans l'énoncé du théorème II-14, il est commode de considérer des fonctions à valeurs dans $[0, +\infty]$.

II-2.1. Fonctions simples. Comme dans la théorie de Riemann, nous allons définir la valeur de l'intégrale d'une fonction en l'approchant par des fonctions particulièrement simples, pour lesquelles la valeur de l'intégrale est indiscutable.

DÉFINITION II-23 (fonction simple). Soit X un espace mesurable. On appelle fonction étagée, ou fonction simple, une fonction de la forme

$$f = \sum_{k=1}^N \alpha_k 1_{A_k},$$

où les α_k sont des nombres positifs et les A_k sont des parties mesurables formant une partition de X .

REMARQUE II-24. La condition de positivité est imposée uniquement parce que nous avons pour but de définir d'abord l'intégrale des fonctions positives.

Les critères suivants sont presque évidents et laissés en exercice.

PROPOSITION II-25 (reformulation de la simplicité). (i) Soit $(A_k)_{1 \leq k \leq N}$ une famille finie de parties mesurables d'un espace mesurable X , et $(\alpha_k)_{1 \leq k \leq N}$ une famille de nombres réels positifs; alors $f = \sum \alpha_k 1_{A_k}$ est simple.

(ii) Une fonction f est simple si et seulement si elle est mesurable, positive et prend un nombre fini de valeurs.

La définition de l'intégrale d'une fonction simple tombe sous le sens :

DÉFINITION II-26 (intégrale d'une fonction simple). Une fonction simple f étant donnée sur l'espace mesuré (X, μ) , avec les notations de la Définition II-23, on pose

$$\int f d\mu = \sum_{k=1}^N \alpha_k \mu[A_k],$$

avec la convention $0 \times (+\infty) = 0$.

REMARQUE II-27. Cette définition est indépendante du choix des α_k et des A_k , comme on pourra le vérifier en exercice (fastidieux et pas si facile). Si l'on impose que les α_k soient tous distincts et les A_k disjoints, on a alors unicité de la représentation de f sous la forme $\sum \alpha_k 1_{A_k}$.

PROPOSITION II-28 (additivité de l'intégrale des fonctions simples). Soient f et g deux fonctions simples, alors pour tous $\alpha, \beta \in \mathbb{R}_+$, la fonction $\alpha f + \beta g$ est simple, et

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

En particulier, si $f = \sum_{1 \leq k \leq N} \alpha_k 1_{A_k}$, alors

$$\int f d\mu = \sum \alpha_k \mu[A_k],$$

que les A_k forment une partition ou non.

DÉMONSTRATION. Écrivons $f = \sum a_j 1_{A_j}$, $g = \sum b_k 1_{B_k}$, où les (A_j) et les B_k forment deux partitions de X . Alors $\alpha f + \beta g = \sum_{jk} (\alpha a_j + \beta b_k) 1_{A_j \cap B_k}$ est bien une

fonction simple, et la valeur de son intégrale est

$$\begin{aligned}
 \sum_{j,k} (\alpha a_j + \beta b_k) \mu[A_j \cap B_k] &= \alpha \sum_j a_j \left(\sum_k \mu[A_j \cap B_k] \right) + \beta \sum_k b_k \left(\sum_j \mu[A_j \cap B_k] \right) \\
 &= \alpha \sum_j a_j \mu[A_j \cap (\cup_k B_k)] + \beta \sum_k b_k \mu[(\cup_j A_j) \cap B_k] \\
 &= \alpha \sum_j a_j \mu[A_j] + \beta \sum_k b_k \mu[B_k].
 \end{aligned}$$

□

II-2.2. Approximation des fonctions mesurables. Il est bon de savoir que toute fonction mesurable positive peut être approchée par des fonctions simples. Plus précisément, nous allons démontrer l'énoncé suivant.

THÉORÈME II-29 (approximation par des fonctions simples). *Soit f une fonction mesurable sur un espace X , à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$. Alors il existe une suite croissante $(\varphi_n)_{n \in \mathbb{N}}$ de fonctions simples, qui converge simplement vers f .*

Si f est bornée par M , on peut en outre imposer $\varphi_n - \varphi_{n-1} \leq M/2^n$, $\varphi_0 = 0$; et si f est non bornée, on peut imposer $\varphi_n - \varphi_{n-1} \leq 1_{|f| \geq n} + 2^{-n}$.

Si X est muni d'une mesure σ -finie μ , on peut en outre imposer à chaque φ_n d'être nulle en-dehors d'un ensemble de mesure finie.

DÉMONSTRATION. Soit $\delta_n = 2^{-n}$, on pose $\varphi_n(x) = k\delta_n$ si $f(x) \in [k\delta_n, (k+1)\delta_n[$ et $f(x) < n$; $\varphi_n(x) = n$ si $f(x) \geq n$. Il est facile de vérifier que $\varphi_n(x)$ converge vers $f(x)$ pour tout x . D'autre part, si $f(x) \in [k\delta_n, (k+1)\delta_n[$, alors $f(x) \in [2k\delta_{n+1}, (2k+2)\delta_{n+1}[$, donc $\varphi_{n+1}(x)$ vaudra soit $2k\delta_{n+1}$, soit $(2k+1)\delta_{n+1}$, soit $(2k+2)\delta_{n+1}$, et dans tous les cas sera supérieur ou égal à $\varphi_n(x)$.

Dans le cas où on se donne une mesure σ -finie μ , on a $X = \cup X_n$, avec $\mu[X_n] < +\infty$, et on peut poser $\tilde{\varphi}_n = \varphi_n 1_{X_n}$ pour prouver la dernière partie de l'énoncé. □

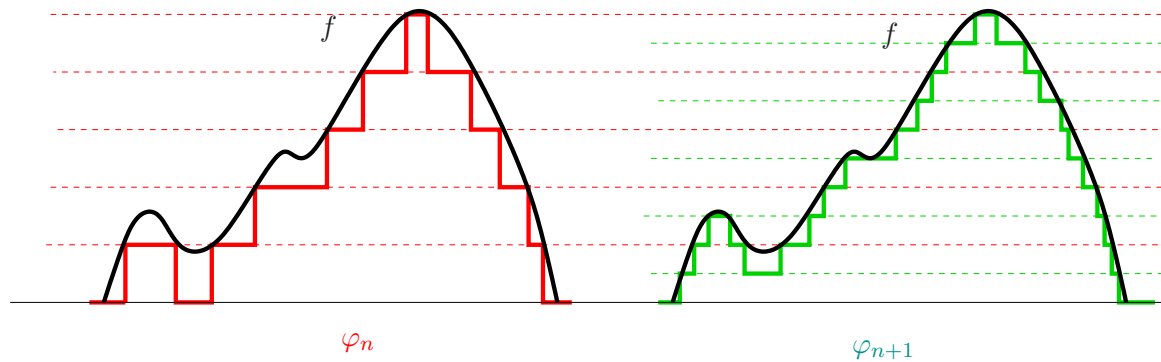


FIG. 1. Approximation d'une fonction mesurable par des fonctions simples

COROLLAIRE II-30 (une fonction mesurable est combinaison infinie de fonctions indicatrices). *Soit f une fonction mesurable sur un espace X , à valeurs dans*

$\mathbb{R}_+ \cup \{+\infty\}$. Alors il existe des nombres positifs $(c_k)_{k \geq 1}$ et des ensembles mesurables $(A_k)_{k \in \mathbb{N}}$ tels que

$$f = \sum_{k=1}^{\infty} c_k 1_{A_k}.$$

EXERCICE II-31. Utiliser une variante de la construction précédente pour montrer que l'on peut choisir la famille (c_k) a priori parmi l'ensemble des suites qui convergent vers 0, et dont la série diverge (par exemple, $c_k = 1/k$ fait l'affaire).

Pour illustrer l'intérêt de ce théorème d'approximation, nous allons l'utiliser pour résoudre le problème abordé dans le paragraphe II-1.3, dans un cadre légèrement restreint ; et pour démontrer le Théorème II-22.

THÉORÈME II-32 (fonctions mesurables pour $\sigma(f)$). Soient X un espace quelconque, Y un espace mesurable, et $Z = \mathbb{R}$ (muni de sa tribu borélienne). On se donne f une fonction quelconque de X dans Y , et on munit X de la tribu $\sigma(f)$. Alors les fonctions mesurables de X dans Z sont exactement les fonctions de la forme $\Phi \circ f$, où Φ est une fonction mesurable de Y dans Z .

DÉMONSTRATION. 1. Soit g une fonction simple de X dans \mathbb{R}_+ . Comme g prend un nombre fini de valeurs, la tentative de démonstration présentée au paragraphe II-1.3 aboutit (pourquoi ?) et permet de construire une fonction Φ mesurable, telle que $g = \Phi \circ f$.

2. Si maintenant g est une fonction positive arbitraire, on peut construire une famille g_n de fonctions simples, $\sigma(f)$ -mesurables, convergeant simplement vers g . En particulier, il existe Φ_n mesurable tel que $g_n = \Phi_n \circ f$. La fonction $\Phi := \limsup \Phi_n$ est mesurable, et pour tout $x \in X$ on a

$$g(x) = \lim g_n(x) = \lim \Phi_n(f(x)) = \limsup \Phi_n(f(x)) = (\limsup \Phi_n)(f(x)) = \Phi(f(x)).$$

3. Enfin, si g est une fonction arbitraire, on peut écrire $g_{\pm} = \Phi_{\pm} \circ f$, d'où $g = (\Phi_+ - \Phi_-) \circ f$. \square

REMARQUE II-33. Cette preuve se généralise au cas où Z est un espace Polonais arbitraire. En effet, on a vu dans l'Exercice II-16 que toute limite de fonctions mesurables, à valeurs dans un espace métrique **complet**, est mesurable. D'autre part, une fonction mesurable $f : X \rightarrow Z$ étant donnée, où Z est un espace métrique **séparable**, on peut modifier la preuve du théorème d'approximation II-29 pour montrer que $f = \lim f_n$, où chaque f_n ne prend qu'au plus une infinité **dénombrable** de valeurs. Ces énoncés sont suffisants pour conclure.

DÉMONSTRATION DU THÉORÈME II-22. En décomposant f en parties positives et négatives, on se ramène au cas où f est positive. Soit f une fonction mesurable pour la tribu $\overline{\mathcal{A}}$; d'après le Corollaire II-30, on peut écrire

$$f = \sum_{k=1}^{\infty} c_k 1_{A_k},$$

où les c_k sont des nombres positifs, et les A_k sont des éléments de $\overline{\mathcal{A}}$. Par définition de la tribu complétée, pour tout k on peut écrire $A_k = B_k \cup E_k$, $E_k \subset N_k$, avec $B_k, E_k \in \mathcal{A}$ et $\mu[N_k] = 0$. On pose alors $N := \bigcup N_k$, et $g = \sum c_k 1_{B_k}$. \square

Le théorème d'approximation des fonctions mesurables par des fonctions simples rend plausible l'idée que l'on peut construire l'intégrale des fonctions mesurables positives à partir de l'intégrale des fonctions simples; c'est effectivement ce que nous allons faire.

II-2.3. Intégrale des fonctions positives.

DÉFINITION II-34 (intégrale d'une fonction positive). *Soit (X, μ) un espace mesuré, et f une fonction mesurable sur X , à valeurs dans $\mathbb{R}_+ \cup \{+\infty\}$. On appelle intégrale de f pour la mesure μ , et on note*

$$\int f d\mu = \int_X f d\mu = \int_X f(x) d\mu(x) = \int_X f(x) \mu(dx)$$

le nombre

$$\sup \left\{ \int g d\mu; \quad g \text{ simple}; \quad 0 \leq g \leq f \right\} \in [0, +\infty].$$

REMARQUE II-35. (i) Le supremum est pris sur une classe de fonctions non vide, puisque la fonction nulle est admissible.

(ii) Si f est une fonction simple, cette définition coïncide avec celle que nous avons donnée au paragraphe précédent.

(iii) Nous verrons dans le chapitre suivant que

$$\int f d\mu = \lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} \frac{1}{2^n} \mu \left[\left\{ x; f(x) \geq \frac{k}{2^n} \right\} \right],$$

ce qui justifie l'intuition suggérée par la figure 1. En ce sens, l'intégrale de Lebesgue est bien un procédé de **sommation par tranches**.

(iv) Si A est une partie mesurable de X , on peut considérer A comme un espace mesuré et définir $\int_A f$ comme l'intégrale de la restriction de f à A .

DÉFINITION II-36 (fonctions intégrables). *Si f est une fonction mesurable positive d'intégrale finie, on dit qu'elle est intégrable, ou sommable.*

La proposition suivante rassemble quelques propriétés élémentaires de l'intégrale. Comme les ensembles négligeables (de mesure nulle) ne jouent aucun rôle dans la valeur de l'intégrale, il est commode de l'exprimer en utilisant la terminologie ci-après.

DÉFINITION II-37 (presque partout). *Soit (X, μ) un espace mesuré. On dit qu'une propriété est vraie presque partout, ou μ -presque partout, ou $d\mu$ -presque partout, ou p.p., ou μ -p.p., ou $d\mu$ -p.p., si l'ensemble des éléments de X qui ne vérifient pas cette propriété est de mesure nulle.*

PROPOSITION II-38 (Propriétés élémentaires de l'intégrale des fonctions positives). *Soient (X, μ) un espace mesuré, et f, g deux fonctions positives mesurables sur X .*

(i) *Si $\int f d\mu < +\infty$, alors f est finie presque partout;*

(ii) *Si $\int f d\mu = 0$, alors f est nulle presque partout;*

(iii) *Si $f \leq g$ presque partout, alors $\int f \leq \int g$. En particulier, si g est sommable, alors f l'est aussi. Par ailleurs, si $f = g$ presque partout, alors $\int f = \int g$.*

(iv) Si A et B sont deux parties mesurables disjointes, et f est une fonction mesurable positive, alors

$$\int_{A \cup B} f = \int_A f + \int_B f.$$

DÉMONSTRATION. (i) Si $A = f^{-1}(+\infty)$ est de mesure strictement positive, alors la famille de fonctions simples $(k1_A)_{k \in \mathbb{N}}$ montre que le supremum dans la définition de l'intégrale de f est infini, d'où $\int f = +\infty$.

(ii) Supposons que $\int f = 0$; soit $\varepsilon > 0$. Si la mesure de $F_\varepsilon := \{x; f(x) \geq \varepsilon\}$ était strictement positive, on pourrait construire une fonction simple valant ε sur F_ε , positive et d'intégrale strictement positive, minorant f , donc l'intégrale de f serait strictement positive. C'est faux par hypothèse, donc F_ε est de mesure nulle. En conséquence, l'ensemble des points où f n'est pas nulle est de mesure nulle, car c'est la réunion dénombrable des $F_{1/k}$ ($k \in \mathbb{N}$), qui sont tous de mesure nulle.

Enfin, (iii) est évident par construction : si $f \leq g$ presque partout, soit φ une fonction simple minorant f , on redéfinit φ sur l'ensemble négligeable où $f > g$, en lui attribuant la valeur 0 sur cet ensemble. La fonction ainsi obtenue est simple, minore g et a même intégrale que φ . On passe ensuite au supremum sur toutes les fonctions simples φ minorant f . \square

II-2.4. Intégrale des fonctions sommables.

DÉFINITION II-39 (fonctions sommables). On appelle fonction sommable une fonction mesurable à valeurs dans \mathbb{R} telle que $|f|$ est sommable. Alors la partie positive $f_+ = \max(f, 0)$ de f , et sa partie négative $f_- = \max(-f, 0)$, étant majorées par $|f|$, sont toutes deux sommables, et on pose

$$\int f d\mu = \left(\int f_+ d\mu \right) - \left(\int f_- d\mu \right).$$

REMARQUE II-40. Plus généralement, on peut définir $\int f d\mu$ dans $\overline{\mathbb{R}}$ dès que l'une au moins des fonctions f_+ et f_- est sommable.

La proposition suivante est presque une trivialité.

PROPOSITION II-41 (propriétés élémentaires de l'intégrale des fonctions sommables). Soit (X, μ) un espace mesuré, et f une fonction mesurable de X dans \mathbb{R} . Alors,

(i) Si $|f|$ est majoré par une fonction sommable, alors f est sommable;

(ii) Si f est sommable, alors

$$\left| \int f \right| \leq \int |f|;$$

(iii) Si f et g sont deux fonctions sommables, et $f \leq g$ presque partout, alors $\int f \leq \int g$.

(iv) Si A et B sont deux parties mesurables disjointes, et f est une fonction sommable, alors

$$\int_{A \cup B} f = \int_A f + \int_B f.$$

EXEMPLE II-42. Si X est de mesure finie, toute fonction bornée est sommable. En effet, $|f|$ est alors majoré par une fonction constante c , dont l'intégrale vaut $c\mu[X]$.

II-3. L'intégrale est une forme linéaire positive

On rappelle qu'une forme linéaire L sur un espace vectoriel E est une application linéaire de E dans \mathbb{R} compatible avec les opérations d'addition et de multiplication par un scalaire. Quand l'espace E est un espace de fonctions à valeurs réelles, on dit que L est **positive** si elle prend des valeurs positives sur toutes les fonctions positives.

Dans le cas présent, il est trivial que l'intégrale d'une fonction sommable positive est positive. Il est à peine moins trivial que $\int(\lambda f) = \lambda \int f$ pour toute fonction f sommable et pour tout scalaire λ . La relation capitale $\int(f + g) = (\int f) + (\int g)$ est en revanche beaucoup plus subtile !

II-3.1. Addition des fonctions positives.

THÉORÈME II-43 (addition des intégrales des fonctions positives). *Soient f et g deux fonctions positives mesurables sur un espace mesuré (X, μ) . Alors*

$$\int (f + g) d\mu = \int f d\mu + \int g d\mu.$$

En particulier, $f + g$ est sommable si (et seulement si) f et g le sont.

DÉMONSTRATION. 1. Si $\int f = +\infty$ ou $\int g = +\infty$, alors par comparaison on sait que $\int(f + g) = +\infty$ et il n'y a rien à démontrer. Supposons donc que ces deux intégrales sont finies. Soient φ et ψ des fonctions simples telles que $0 \leq \varphi \leq f$, $0 \leq \psi \leq g$, et $\int f \leq \int \varphi + \varepsilon$, $\int g \leq \int \psi + \varepsilon$. D'après la Proposition II-28, $\varphi + \psi$ est simple, et

$$\int f + \int g \leq \int \varphi + \int \psi + 2\varepsilon = \int (\varphi + \psi) + 2\varepsilon \leq \int (f + g) + 2\varepsilon.$$

En faisant tendre ε vers 0, on obtient

$$\int f + \int g \leq \int (f + g).$$

2. Par la Proposition II-29, on peut trouver des suites croissantes (φ_k) et (ψ_k) de fonctions simples telles que $0 \leq \varphi_k \leq f$, $0 \leq \psi_k \leq g$, convergeant simplement vers f et g respectivement. Soit $\delta \in (0, 1)$ arbitraire, on pose

$$A_k := \left\{ x; \varphi_k(x) \geq (1 - \delta)f(x), \psi_k(x) \geq (1 - \delta)g(x) \right\}.$$

Les A_k forment une famille croissante; si $f(x) + g(x) < +\infty$, alors $x \in A_k$ pour k assez grand (c'est évident si $f(x) = 0$, tandis que si $f(x) \neq 0$ la limite de $\varphi_k(x)$ est strictement plus grande que $(1 - \delta)f(x)$). On a donc $X = \cup A_k \cup Z$, où Z est l'ensemble des x pour lesquels $f(x) = +\infty$ ou $g(x) = +\infty$. Puisque f et g sont sommables, Z est un ensemble négligeable.

Soit χ une fonction simple minorant $f + g$. Par additivité de l'intégrale des fonctions simples,

$$\int f + \int g \geq \int \varphi_k + \int \psi_k \geq \int_{A_k} \varphi_k + \int_{A_k} \psi_k = \int_{A_k} (\varphi_k + \psi_k).$$

Par la définition de A_k et la positivité de l'intégrale, on a

$$\int_{A_k} (\varphi_k + \psi_k) \geq \int_{A_k} (1 - \delta)\chi = (1 - \delta) \int_{A_k} \chi.$$

Ecrivons χ sous la forme $\sum_{1 \leq j \leq J} \alpha_j 1_{B_j}$. Alors on a, puisque les A_k forment une famille croissante et que $\mu[X \setminus (\cup A_k)] = 0$,

$$\int_{A_k} \chi = \sum_j \alpha_j \mu[A_k \cap B_j] \xrightarrow{k \rightarrow \infty} \sum_j \alpha_j \mu[(\cup A_k) \cap B_j] = \sum_j \alpha_j \mu[(\cup A_k) \cap B_j] = \int \chi.$$

En passant à la limite quand $k \rightarrow \infty$ dans l'inégalité

$$\int f + \int g \geq (1 - \delta) \int_{A_k} \chi,$$

on trouve donc

$$\int f + \int g \geq (1 - \delta) \int_{X \setminus Z} \chi = (1 - \delta) \int_X \chi.$$

En faisant tendre δ vers 0, on obtient

$$\int f + \int g \geq \int \chi.$$

Puisque χ est une fonction simple arbitraire minorant $f + g$, on a finalement

$$\int f + \int g \geq \int (f + g).$$

□

REMARQUE II-44. Dans le chapitre suivant, nous utiliserons un raisonnement exactement similaire pour démontrer le théorème dit de convergence dominée de Lebesgue. En fait, dans la plupart des ouvrages de référence on démontre d'abord le théorème de convergence dominée, et on en déduit ensuite l'additivité de l'intégrale.

II-3.2. Généralisation : fonctions sommables. Comme conséquence facile du Théorème II-43, nous allons établir la linéarité de l'intégrale des fonctions sommables.

THÉORÈME II-45 (linéarité de l'intégrale). *Soient (X, μ) un espace mesuré ; f, g deux fonctions sommables sur X , à valeurs dans \mathbb{R} ; et α, β deux scalaires. Alors $\alpha f + \beta g$ est sommable, et*

$$\int (\alpha f + \beta g) d\mu = \alpha \left(\int f d\mu \right) + \beta \left(\int g d\mu \right).$$

En particulier, l'intégrale est une forme linéaire positive sur l'espace vectoriel des fonctions sommables.

DÉMONSTRATION. On note d'abord que $\int (\alpha f) d\mu = \alpha \left(\int f d\mu \right)$, et $\int (\beta g) d\mu = \beta \left(\int g d\mu \right)$. Il suffit donc de montrer que si f et g sont sommables de signe quelconque, alors $\int (f + g) d\mu = \left(\int f d\mu \right) + \left(\int g d\mu \right)$. Pour cela on écrit

$$(f + g)_+ - (f + g)_- = f + g = (f_+ - f_-) + (g_+ - g_-),$$

d'où (quand f et g sont finies, ce qui est vrai en-dehors d'un ensemble de mesure nulle)

$$(f + g)_+ + f_- + g_- = f_+ + g_+ + (f + g)_-;$$

on intègre alors les deux membres en utilisant le Théorème II-43 :

$$\int (f + g)_+ d\mu + \int f_- d\mu + \int g_- d\mu = \int f_+ d\mu + \int g_+ d\mu + \int (f + g)_- d\mu.$$

Toutes ces quantités sont finies puisque $|f|$ et $|g|$ sont intégrables, on en déduit donc

$$\int (f+g)_+ d\mu - \int (f+g)_- d\mu = \int f_+ d\mu - \int f_- d\mu + \int g_+ d\mu - \int g_- d\mu,$$

soit $\int (f+g) d\mu = \int f d\mu + \int g d\mu$.

Le raisonnement précédent montre bien que l'intégrale par rapport à μ est une forme linéaire. La positivité quant à elle est évidente : si f est mesurable positive, alors $\int f d\mu \geq 0$. \square

DÉFINITION II-46 (espace de Lebesgue). *L'espace vectoriel des fonctions sommables est noté L^1 , ou $L^1(X)$, ou $L^1(\mu)$, ou $L^1(d\mu)$, ou $L^1(X, \mu)$, ou $L^1(X, d\mu)$ et appelé espace de Lebesgue d'exposant 1.*

REMARQUE II-47. Soient f et g deux fonctions sommables, et $\lambda \in \mathbb{R}$. Alors

$$\begin{aligned} \int |\lambda f| &= \int |\lambda| |f| = |\lambda| \int |f|; \\ \int |f+g| &\leq \int |f| + |g| = \int |f| + \int |g|. \end{aligned}$$

L'application

$$f \longmapsto \int |f|,$$

définie sur $L^1(d\mu)$, est donc proche de satisfaire les axiomes requis par une norme : il lui manque seulement la propriété $\int |f| = 0 \implies f = 0$. Mais cette dernière identité est clairement fautive : on sait que $\int |f| = 0$ si et seulement si la fonction f est nulle hors d'un ensemble de mesure nulle (pour μ), ce qui n'impose pas à f d'être identiquement nulle.

Si l'on veut transformer L^1 en espace vectoriel, muni de la norme $\int |f|$, il faut donc *quotienter par la relation d'équivalence "coïncider presque partout"*. Deux fonctions qui ne diffèrent que par un ensemble de mesure nulle seront alors considérées "identiques". Insistons sur le fait que cette opération de quotient n'est utile que si l'on veut mettre à profit la structure d'espace vectoriel normé de l'espace ainsi obtenu.

II-3.3. Action sur les fonctions continues. On a vu que si X est un espace mesuré de mesure finie, alors les fonctions bornées sont intégrables. Définissons la **norme de la convergence uniforme** sur l'espace $C_b(X)$ des fonctions continues bornées de X dans \mathbb{R} par la formule

$$\|f\|_\infty := \sup_{x \in X} |f(x)|.$$

On a alors l'énoncé suivant.

PROPOSITION II-48 (l'intégrale appartient à $(C_b)^*$). *Soit μ une mesure de Borel sur un espace X de mesure finie. Alors μ définit une forme linéaire positive continue sur l'espace vectoriel $C_b(X)$ des fonctions continues bornées sur X , normé par la norme de la convergence uniforme.*

REMARQUE II-49. On rappelle que la norme de la convergence uniforme est définie par

$$\|f\|_\infty := \sup_{x \in X} |f(x)|.$$

DÉMONSTRATION. Bien sûr, toute fonction continue est borélienne, donc mesurable, et toute fonction mesurable bornée est intégrable. En outre, on a

$$\left| \int f \right| \leq \int |f| \leq \int \|f\|_\infty = \mu[X] \|f\|_\infty.$$

Il s'ensuit que l'intégrale est bien une forme linéaire **continue** sur $C_b(X)$, dont la norme est majorée par le nombre positif $\mu[X]$. \square

On peut construire ainsi de nombreuses formes linéaires continues sur $C_b(X)$, mais il n'est pas clair que ce soient les seules, même dans des cas simples comme $X = \mathbb{R}^d$. Pour s'en convaincre, le lecteur averti pourra utiliser le théorème d'extension de Hahn-Banach pour prolonger l'application "limite à l'infini", définie sur le sous-espace fermé des fonctions qui convergent à l'infini, en une application "limite à l'infini" définie pour **toutes** les fonctions continues bornées ! Une telle application n'est pas une mesure. C'est en revanche une "mesure finiment additive" !

Cependant, si X est **compact**, le théorème de représentation de Riesz assure que toutes les formes linéaires continues sur $C(X)$ correspondent à des mesures. Dans ce cas, bien sûr, toutes les fonctions continues sont bornées. Nous allons donner dans la prochaine section un énoncé un peu plus général.

REMARQUE II-50. Un lecteur aux fortes convictions pourra objecter aux arguments ci-dessus, car l'existence de l'application "limite à l'infini" repose sur l'utilisation du théorème de Hahn-Banach, qui lui-même utilise l'axiome du choix. Nous reparlerons bientôt de ces subtilités logiques. En tous les cas, la construction précédente montre au moins qu'il est impossible d'exclure l'existence de formes linéaires continues qui ne soient pas des mesures, sauf à changer l'axiomatique traditionnelle.

II-4. Le théorème de Riesz

Ce théorème est fondamental car il présente une autre approche, alternative à celle de Lebesgue, de la construction d'une intégrale abstraite.

II-4.1. Énoncé du théorème. Commençons par quelques définitions.

DÉFINITION II-51 (espaces de fonctions continues). *Soit X un espace topologique arbitraire. Si $f : X \rightarrow \mathbb{R}$ est une fonction continue, on appelle support de f le plus petit fermé en-dehors duquel f est identiquement nulle. On note $C(X)$ l'espace des fonctions **continues** de X dans \mathbb{R} , $C_b(X)$ l'espace des fonctions **continues bornées** sur X , et $C_c(X)$ l'espace des fonctions **continues à support compact** sur X . Enfin, on note $C_0(X)$ l'espace des fonctions f continues sur X qui **tendent vers 0 à l'infini**, au sens où pour tout $\varepsilon > 0$ on peut trouver un compact $K \subset X$ en-dehors duquel $|f| \leq \varepsilon$. Clairement,*

$$C_c(X) \subset C_0(X) \subset C_b(X) \subset C(X).$$

Les espaces $C_c(X)$, $C_0(X)$ et $C_b(X)$, munis de la norme de la convergence uniforme, sont des espaces vectoriels normés.

REMARQUE II-52. L'espace $C_c(X)$ n'est pas a priori *complet* (au sens usuel) : par exemple, on peut facilement construire une fonction sur \mathbb{R} , à support non compact, qui soit limite uniforme de fonctions continues à support compact. Dans un espace localement compact, la complétion de $C_c(X)$ est l'espace $C_0(X)$ des fonctions continues qui tendent vers 0 à l'infini. L'espace $C_b(X)$, en revanche, est complet.

La définition suivante, interne à ce cours, servira uniquement dans cette section, pour abréger quelques formulations.

DÉFINITION II-53 (presque régularité). *Soit μ une mesure de Borel sur un espace topologique X . On conviendra de dire que μ est presque régulière si pour tout Borélien A de X ,*

$$\mu[A] = \inf \{ \mu[O]; \ O \text{ ouvert}; \ A \subset O \}$$

(régularité extérieure) et pour tout **ouvert** B de X ,

$$\mu[B] = \sup \{ \mu[K]; \ K \text{ compact}; \ K \subset B \}.$$

THÉORÈME II-54 (théorème de représentation de Riesz). *Soit X un espace topologique séparé, localement compact. Alors on peut identifier (mettre en correspondance bijective)*

- d'une part, les formes linéaires Λ sur $C_c(X)$, positives;
 - d'autre part, les mesures de Borel μ sur X , presque régulières et finies sur les compacts;
- via la formule

$$\Lambda f = \int f d\mu.$$

Avant de continuer, voici une liste de commentaires sur cet énoncé, qui admet quelques variantes plus ou moins subtiles.

- REMARQUES II-55 (Commentaires sur le Théorème de Riesz). (i) L'hypothèse de compacité locale est fondamentale. L'espace de Wiener $W = \{ \gamma \in C([0, 1]; \mathbb{R}^n); \ \gamma(0) = 0 \}$ ne la remplit pas : on vérifie aisément que ses compacts sont tous d'intérieur vide. L'espace $C_c(W)$ est donc réduit à $\{0\}$! Pourtant il existe des mesures non triviales sur W , telles que les mesures de Dirac, ou la célèbre mesure de Wiener.
- (ii) Dans l'énoncé, on ne peut pas remplacer $C_c(X)$ par l'espace plus gros $C_b(X)$. On peut en revanche le remplacer sans dommage par l'espace $C_0(X)$, complétion de $C_c(X)$.
- (iii) Dans la définition de "presque régularité" on a imposé que l'identité $\mu[A] = \sup_{K \subset A} \mu[K]$ soit vérifiée pour tout ouvert. En fait cette identité sera alors vérifiée automatiquement pour tout ensemble mesurable de mesure finie. En particulier, si une mesure produite par le Théorème de Riesz est de masse totale finie, alors elle est régulière. Cette remarque s'avèrera utile plus tard dans la démonstration du Théorème VI-61 ; il ne faut cependant pas y attacher une grande importance, car en pratique, dans la grande majorité des cas la régularité est automatique, par exemple grâce aux Théorèmes II-58 et II-59 présentés dans la sous-section suivante.
- (iv) On peut, si on le souhaite, compléter la mesure μ grâce au Théorème I-80, et obtenir donc une mesure complète.
- (v) La preuve ne nécessite pas vraiment la linéarité de l'application Λ : il suffit de savoir que Λ est une fonctionnelle positive, croissante ($f \leq g \implies \Lambda f \leq \Lambda g$) et sur-additive ($\Lambda(f+g) \geq \Lambda f + \Lambda g$) sur l'espace des fonctions continues positives à support compact. Cette remarque aussi sera utile pour la démonstration du Théorème VI-61.

- (vi) Si l'on réfléchit un peu à l'énoncé, on a l'impression que l'hypothèse de presque régularité peut être évitée : en effet, toute mesure de Borel finie μ sur les compacts définit bien une forme linéaire positive Λ sur $C_c(X)$. Cependant, si l'on n'impose pas la presque régularité, rien ne garantit a priori l'*unicité* de la mesure μ correspondant à Λ .

Voici maintenant deux remarques d'ordre plus général :

REMARQUE II-56. Si $\mu[X] = +\infty$, la forme linéaire Λ définie par μ n'est pas **continue** sur $C_c(X)$ considéré comme espace vectoriel normé (norme de la convergence uniforme). En revanche, on peut munir $C_c(X)$ d'une topologie alternative bien choisie, de sorte que Λ soit une forme linéaire continue en un sens bien précis. Je n'en dirai pas plus sur ce problème, dont la solution peut être considérée comme le point de départ de la théorie des distributions [Schwartz].

REMARQUE II-57. Le nom de "théorème de représentation de Riesz" est également donné à un autre théorème, très différent (description du dual d'un espace de Hilbert, voir Chapitre VI). Cette coïncidence n'a rien de surprenant, Riesz étant, avec Banach, l'un des principaux fondateurs de l'analyse fonctionnelle moderne.

Avant de passer à la preuve du Théorème II-54, je vais maintenant donner deux énoncés simplifiés.

II-4.2. Énoncés simplifiés. L'hypothèse de régularité est souvent vérifiée automatiquement, sous des hypothèses peu contraignantes sur X . On pourra donc retenir les variantes explicitées ci-après, qui n'utilisent pas explicitement ce concept.

THÉORÈME II-58 (théorème de représentation de Riesz, version simplifiée). *Soit X un espace topologique séparé, localement compact, dans lequel tout ouvert est union dénombrable de compacts. Alors on peut identifier*

- *d'une part, les formes linéaires Λ positives sur $C_c(X)$;*
- *d'autre part, les mesures boréliennes μ sur X , finies sur les compacts ;*

via la formule

$$\Lambda f = \int f d\mu.$$

Ces mesures sont automatiquement régulières.

THÉORÈME II-59 (théorème de représentation de Riesz, cas métrique compact). *Soit X un espace topologique métrique compact. Alors on peut identifier*

- *d'une part, les formes linéaires Λ positives sur $C(X)$;*
- *d'autre part, les mesures boréliennes μ finies sur X ;*

via la formule

$$\Lambda f = \int f d\mu.$$

Ces mesures sont automatiquement régulières, et ces formes linéaires sont automatiquement continues.

DÉMONSTRATION. Les Théorèmes II-58 et II-59 s'obtiennent en combinant le théorème de représentation de Riesz avec les théorèmes de régularité I-56 et I-54, respectivement. \square

On trouvera une démonstration compacte (!), en apparence assez délicate, du Théorème de Riesz, dans [Rudin, pp. 40-47]. Pour les besoins de ce cours, nous allons établir ce théorème comme une conséquence relativement simple du Théorème de Carathéodory généralisé établi au Chapitre I, Théorème I-71 ; on remarquera d'ailleurs que la démonstration de [Rudin] reprend plusieurs des arguments utilisés dans la preuve du Théorème I-71. Notons au passage l'intérêt du Théorème I-71 : démontrer le théorème de Riesz via le Théorème de Carathéodory lui-même, sous la forme du Théorème I-69, est un casse-tête formidable ! Une variante de cette dernière démarche est menée à bien dans [Dudley], via un intermédiaire délicat appelé théorème de Daniell-Stone, qui traite de prolongement des fonctionnelles linéaires positives (voir aussi [Rudin, p. 398] ; *in fine*, la démonstration du Théorème de Riesz y fait intervenir le théorème de convergence uniforme de Dini.

II-4.3. Preuve du théorème de Riesz. Soit X un espace séparé, localement compact, et soit μ une mesure finie sur les compacts. Si f est une fonction continue à support compact K , elle est bornée par la fonction sommable $\|f\|_\infty 1_K$, donc sommable. La forme linéaire Λ définie par $\Lambda f := \int f d\mu$ est donc bien définie sur $C_c(X)$, et elle est évidemment positive.

C'est bien sûr la réciproque qui est délicate. Soit Λ une forme linéaire positive sur $C_c(X)$, montrons qu'il existe au plus une mesure μ , satisfaisant aux hypothèses du Théorème de Riesz, qui puisse la représenter. Soient μ_1 et μ_2 deux mesures admissibles, et soit K un compact. Comme μ_1 est finie sur les compacts, et presque régulière, au sens de la Définition II-53, on sait qu'il existe un ouvert O contenant K tel que $\mu_1[K] \geq \mu_1[O] - \varepsilon$, où $\varepsilon > 0$ est arbitrairement petit. Par le lemme d'Urysohn, on peut construire une fonction continue φ encadrée par les fonctions indicatrices 1_K et 1_O . On a donc

$$\mu_2[K] = \int 1_K d\mu_2 \leq \int \varphi d\mu_2 = \int \varphi d\mu_1 \leq \int 1_O d\mu_1 = \mu_1[O] \leq \mu_1[K] + \varepsilon.$$

On conclut en faisant tendre ε vers 0 que $\mu_2[K] \leq \mu_1[K]$, et par symétrie $\mu_1[K] = \mu_2[K]$. Il s'ensuit que μ_1 et μ_2 coïncident sur les compacts ; comme elles sont presque régulières, elles coïncident également sur les ouverts, et par suite (toujours par presque régularité) sur tous les ensembles mesurables. Cela prouve l'unicité de μ .

Passons maintenant à la construction de μ . L'idée est encore une fois d'approcher les fonctions indicatrices des ouverts par des fonctions continues. Pour tout ensemble ouvert O , on pose donc

$$\mu[O] = \sup \{ \Lambda f ; f \in C_c(X) ; 0 \leq f \leq 1_O \}.$$

Le problème est maintenant de prolonger μ à la tribu borélienne tout entière. Comme l'ensemble \mathcal{F} de tous les ouverts de X est stable par intersection finie, le Théorème I-71(ii) assure l'existence d'un tel prolongement si la condition (7) est satisfaite pour tous A, B ouverts.

Dans un premier temps, vérifions que μ est dénombrablement sous-additive sur l'ensemble des ouverts : si $(O_k)_{k \in \mathbb{N}}$ est une famille d'ouverts, et $O := \bigcup O_k$, alors $\mu[O] \leq \sum_k \mu[O_k]$. En effet, soit f une fonction à support compact, $0 \leq f \leq 1_O$, et soit K son support. K étant inclus dans l'union des O_k , on peut appliquer le théorème I-34 de partition de l'unité pour trouver des fonctions continues $\chi_{i_1}, \dots, \chi_{i_n}$, telles que

$0 \leq \chi_{i_j} \leq 1$, χ_{i_j} a son support inclus dans O_{i_j} et $\sum \chi_{i_j} = 1$ sur K . En particulier,

$$f = (\sum \chi_{i_j})f = \sum_j g_j,$$

où chaque fonction g_j est à support compact dans O_{i_j} , et prend ses valeurs dans $[0, 1]$. On en déduit que

$$\Lambda f = \sum_j \Lambda g_j \leq \sum_j \mu[O_{i_j}] \leq \sum_k \mu[O_k].$$

En passant au supremum sur f , on conclut que

$$\mu[O] \leq \sum_k \mu[O_k].$$

Comme la famille \mathcal{F} est également stable par union dénombrable, la définition de la mesure extérieure se simplifie : dans le contexte présent,

$$\mu^*[A] = \inf \{ \mu[O]; O \text{ ouvert, } A \subset O \}.$$

En particulier, il est clair que μ^* coïncide avec μ sur \mathcal{F} . Donc, si A et B sont deux ouverts de X , l'inégalité

$$\mu[A] \leq \mu[A \cap B] + \mu^*[A \setminus B]$$

est conséquence de la sous-additivité de μ^* . Il nous reste uniquement à vérifier l'inégalité inverse, à savoir : *pour tous ouverts A et B de X ,*

$$(10) \quad \mu[A \cap B] + \mu^*[A \setminus B] \leq \mu[A].$$

Nous allons démontrer cette inégalité en deux étapes.

Etape 1 : μ est sur-additive (et donc additive) sur \mathcal{F} . Soient U et V deux ouverts **disjoints**, nous allons voir que

$$\mu[U] + \mu[V] \leq \mu[U \cup V],$$

ce qui est un cas particulier de (10). Soient f et g deux fonctions continues à supports compacts inclus dans U et V respectivement, à valeurs dans $[0, 1]$. Les supports de f et g étant disjoints, la fonction continue $f + g$ est toujours à valeurs dans $[0, 1]$; et bien sûr, son support est inclus dans $U \cup V$. Il s'ensuit

$$\Lambda f + \Lambda g = \Lambda(f + g) \leq \mu[U \cup V].$$

On conclut en passant au supremum sur toutes les fonctions f et g admissibles.

Etape 2 : *cas général.* C'est seulement à ce niveau de la construction que l'hypothèse de compacité locale va intervenir. Soient deux ouverts A et B de X , et soit f une fonction à support compact $K \subset A \cap B$, à valeurs dans $[0, 1]$. D'après le Lemme I-35, il existe un compact K' et un ouvert O' tels que

$$K \subset O' \subset K' \subset A \cap B.$$

En particulier,

$$\Lambda f \leq \mu[O'].$$

Par ailleurs,

$$A \setminus B = A \setminus (A \cap B) \subset A \setminus K',$$

et $A \setminus K'$ est un ouvert, d'où

$$\mu^*[A \setminus B] \leq \mu^*[A \setminus K'] = \mu[A \setminus K'].$$

On a finalement

$$\Lambda f + \mu^*[A \setminus B] \leq \mu[O'] + \mu[A \setminus K'].$$

Les ouverts O' et $A \setminus K'$ sont disjoints et leur union est incluse dans A ; grâce au résultat de l'Etape 1, on peut compléter l'inégalité précédente comme suit :

$$\Lambda f + \mu^*[A \setminus B] \leq \mu[O'] + \mu[A \setminus K'] = \mu[O' \cup (A \setminus K')] \leq \mu[A],$$

ce qui achève la démonstration du théorème.

Les Remarques II-55 ne nécessitent pas de justification, sauf le point (iii) que je vais maintenant considérer.

DÉMONSTRATION DE LA REMARQUE II-55(III). Soit \mathcal{A} la famille de toutes les parties mesurables, et \mathcal{B} l'ensemble de toutes les parties $A \in \mathcal{A}$ tels que (a) $\mu[A] < \infty$; (b) $\mu[A] = \sup \{\mu[K]; K \text{ compact} \subset A\}$. Le but est de montrer que \mathcal{B} est exactement l'ensemble de toutes les parties de mesure finie. Nous allons procéder en deux temps.

1. On vérifie que $A \cap C \in \mathcal{B}$, pour tout compact C et pour tout $A \in \mathcal{A}$. Pour cela, on introduit

$$\mathcal{C} := \{A \in \mathcal{A}; A \cap C \in \mathcal{B} \text{ pour tout compact } C\}.$$

Il est clair que \mathcal{C} contient X ; et plus généralement tous les fermés (car l'intersection d'un fermé et d'un compact est compacte). Si l'on montre que \mathcal{C} est une classe monotone, alors le Théorème I-68 (Lemme de classe monotone) impliquera que \mathcal{C} coïncide avec la tribu engendrée par les fermés, qui est \mathcal{A} tout entière.

Montrons donc que \mathcal{C} est une classe monotone. Si $(A_\ell)_{\ell \in \mathbb{N}}$ est une famille croissante d'éléments de \mathcal{C} , et C est un compact, pour tout $\ell \in \mathbb{N}$ et pour tout $\varepsilon > 0$ on peut trouver un compact K_ℓ tel que $\mu[(A_\ell \cap C) \setminus K_\ell] \leq 2^{-\ell}\varepsilon$. Quitte à remplacer K_ℓ par $K_1 \cup \dots \cup K_\ell$, on peut supposer que la suite (K_ℓ) est croissante; et on a toujours $\mu[(A_\ell \cap C) \setminus K_\ell] \leq \varepsilon$. La suite $(\mu[K_\ell])_{\ell \in \mathbb{N}}$ est croissante et majorée par $\mu[A \cap C]$, elle converge donc, et il existe $\ell_0 \in \mathbb{N}$ tel que pour tout $\ell \geq \ell_0$ on ait $\mu[K_\ell \setminus K_{\ell_0}] \leq \varepsilon$. On conclut facilement que $\mu[(A_\ell \cap C) \setminus K_{\ell_0}] \leq 2\varepsilon$, et la même estimation vaut pour $A \cap C$, où A est l'union des A_ℓ . La conclusion est que \mathcal{C} est stable par limite croissante.

Soient maintenant A et B dans \mathcal{C} , soit C un compact et soit $\varepsilon > 0$. Soit $K \subset A \cap C$ un compact tel que $\mu[(A \cap C) \setminus K] \leq \varepsilon$. Par ailleurs il existe un ouvert O contenant $B \cap C$ tel que $\mu[O \setminus (B \cap C)] \leq \varepsilon$. On en déduit que $\mu[(A \setminus B) \cap O \setminus (K \setminus O)] \leq \mu[(A \cap C) \setminus K] + \mu[O \setminus (B \cap C)] \leq 2\varepsilon$. On en déduit que $A \setminus B \in \mathcal{C}$, et \mathcal{C} est donc stable par différence. Ceci achève de prouver que \mathcal{C} est une classe monotone, et conclut l'argument.

2. On vérifie que tout $A \in \mathcal{A}$ de mesure finie est en fait un élément de \mathcal{B} . Soit en effet A une telle partie, et soit $\varepsilon > 0$. Il existe un ouvert O contenant A tel que $\mu[O] \leq \mu[A] + \varepsilon$. Il existe un compact K contenu dans O tel que $\mu[O] \geq \mu[K] - \varepsilon$; en particulier $\mu[O \setminus K] \leq 2\varepsilon$. Puisque $A \cap K \in \mathcal{B}$ (par l'étape 1), il existe un compact K' contenu dans $A \cap K$ tel que $\mu[(A \cap K) \setminus K'] \leq \varepsilon$. On en déduit que $\mu[A \setminus K'] \leq \mu[(A \cap K) \setminus K'] + \mu[A \setminus K] \leq \mu[(A \cap K) \setminus K'] + \mu[O \setminus K] \leq \varepsilon + 2\varepsilon = 3\varepsilon$, ce qui achève l'argument. \square

II-4.4. Complément : approximation des fonctions mesurables par des fonctions continues ou semi-continues. Le théorème de Riesz montre que sous certaines hypothèses topologiques, on peut choisir les fonctions continues comme point de départ de la théorie de l'intégration, au lieu des fonctions simples. On peut se demander si cette idée peut être poursuivie plus loin, et s'il est possible d'établir un analogue du théorème d'approximation par des fonctions simples, exprimé en termes de fonctions continues. Les théorèmes de Lusin et de Vitali-Carathéodory donnent une réponse positive à cette question. Tous deux s'autorisent une erreur arbitrairement petite, au sens de la mesure.

THÉORÈME II-60 (Théorème de Lusin). *Soit X un espace topologique séparé localement compact, et soit μ une mesure de Borel régulière sur X . Soit $f : X \rightarrow \mathbb{R}$ une fonction mesurable, nulle en-dehors d'un ensemble de mesure finie. Alors,*

(i) *pour tout $\varepsilon > 0$ il existe une fonction continue f_ε , à support compact, telle que*

$$\inf f \leq \inf f_\varepsilon \leq \sup f_\varepsilon \leq \sup f$$

et f_ε coïncide avec f en-dehors d'un ensemble de mesure inférieure ou égale à ε .

(ii) *En-dehors d'un ensemble de mesure nulle, f est limite d'une suite (f_n) de fonctions continues à support compact, prenant toutes leurs valeurs dans $[\inf f, \sup f]$.*

En utilisant de manière anticipée le théorème de convergence dominée, qui sera démontré dans le chapitre suivant, on peut déduire du théorème précédent le corollaire suivant :

COROLLAIRE II-61 (Densité des fonctions continues). *Soit X un espace topologique séparé localement compact, et soit μ une mesure de Borel régulière sur X , σ -finie. Alors, pour toute fonction intégrable f sur X on peut trouver une suite $(f_n)_{n \in \mathbb{N}}$ de fonctions continues à support compact, telle que*

$$\int |f_n - f| d\mu \xrightarrow{n \rightarrow \infty} 0.$$

DÉMONSTRATION DU THÉORÈME II-60. Nous allons démontrer les deux énoncés en même temps, en construisant une famille $(f_n)_{n \in \mathbb{N}}$ de fonctions continues à support compact, toutes comprises entre $\inf f$ et $\sup f$, telles que $\mu[\{x; f(x) \neq f_n(x)\}] \leq 1/n$, et pour presque tout $x \in X$, $f(x) = f_n(x)$ pour n assez grand.

Supposons d'abord que f est la fonction indicatrice d'un ensemble mesurable A de mesure finie. Comme μ est régulière, il existe une suite $(K_n)_{n \in \mathbb{N}}$ de compacts inclus dans A , et une suite $(O_n)_{n \in \mathbb{N}}$ d'ouverts contenant A , tels que $\mu[O_n \setminus K_n] \leq 1/n$. Sans perte de généralité, on peut supposer la suite (K_n) croissante et la suite (O_n) décroissante. Par le Lemme d'Urysohn I-33, pour chaque n il existe une fonction continue φ_n , à support compact dans O_n , à valeurs dans $[0, 1]$, identiquement égale à 1 sur K_n . La fonction φ_n coïncide avec f en-dehors de $O_n \setminus K_n$, qui est de mesure inférieure ou égale à $1/n$. En outre, si l'on pose $G = \cap O_n$ et $F = \cup K_n$, $N = G \setminus F$, alors $\mu[N] = 0$; tout $x \in X \setminus N$ appartient à K_n pour n assez grand, ou à $X \setminus O_n$ pour n assez grand, et dans tous les cas on a alors $\varphi_n(x) = f(x)$. En conclusion, $(\varphi_n)_{n \in \mathbb{N}}$ remplit le cahier des charges.

Par combinaison linéaire, le résultat s'étend instantanément au cas où f est une fonction simple, nulle en-dehors d'un ensemble de mesure finie. Soit maintenant f une fonction mesurable positive bornée, nulle en-dehors d'un ensemble S de mesure

finie; sans perte de généralité on suppose $f \leq 1$; on sait alors que f est limite d'une suite croissante $(g_k)_{k \in \mathbb{N}}$ de fonctions simples, nulles en-dehors de S , telles que $g_k - g_{k-1} \leq 2^{-k}$. Chacune de ces fonctions est également limite d'une famille $(\varphi_{k,n})_{n \geq 1}$ de fonctions continues à support compact, telles qu'il existe une famille décroissante de parties mesurables $A_{k,n}$ vérifiant

$$\{x; \varphi_{k,n}(x) \neq g_k(x) - g_{k-1}(x)\} \subset A_{k,n}; \quad \mu[A_{k,n}] \leq 2^{-k}/n.$$

On définit alors

$$f_n(x) = \sum_{k=0}^{\infty} \varphi_{k,n}(x).$$

Par convergence uniforme, f_n est continue; elle est bornée par $\sup f$, et coïncide avec $\sum (g_k - g_{k-1}) = f$ en-dehors de l'ensemble $A_n := \cup_k A_{k,n}$, dont la mesure est au plus $\sum_{k=0}^{\infty} \varepsilon 2^{-k}/n = 2/n$. En outre, la famille des A_n est décroissante, son intersection est donc de mesure nulle, et tout $x \in X \setminus (\cap A_n)$ vérifie $f_n(x) = f(x)$ pour $n = n(x)$ assez grand. La famille (f_n) remplit donc toutes les conditions souhaitées.

Si f est positive mais non bornée, on définit $E_m := \{x \in X; f(x) \geq m\}$. Comme f est mesurable et à valeurs dans \mathbb{R} , l'intersection décroissante des E_m est vide, et par σ -additivité $\mu[E_m] \rightarrow 0$ quand $m \rightarrow \infty$. On peut alors effectuer un raisonnement similaire au raisonnement ci-dessus. Enfin, si f n'est pas positive, on sépare f en partie positive et négative f_+ et f_- , et on conclut en appliquant le théorème à f_+ et f_- séparément. \square

DÉMONSTRATION DU COROLLAIRE II-61. Par hypothèse, on peut écrire X comme la réunion croissante des X_k ($k \in \mathbb{N}$), avec X_k mesurable et $\mu[X_k] < +\infty$. Soit $g_n = f 1_{X_n} 1_{|f| \leq n}$. Puisque f est intégrable, $|f|$ est fini μ -presque partout, et donc g_n converge presque partout vers f . Le théorème III-11 de convergence dominée de Lebesgue implique alors

$$\int |f - g_n| d\mu \longrightarrow 0.$$

Pour chaque n , la fonction g_n est nulle en-dehors de l'ensemble de mesure finie X_n , et bornée par n . Par le théorème de Lusin, on peut trouver une fonction f_n continue à support compact, bornée par n , qui coïncide avec g_n en-dehors d'un ensemble A_n de mesure inférieure à $1/n^2$). En particulier,

$$\int |g_n - f_n| \leq \sup(|f_n| + |g_n|) \mu[A_n] \leq \frac{2n}{n^2} \longrightarrow 0.$$

Il s'ensuit que $\int |f - f_n| \rightarrow 0$. \square

REMARQUE II-62. Ce théorème de densité est très général, mais pas très explicite. Nous verrons plus tard que dans le cas où $X = \mathbb{R}^n$, on peut construire, grâce à l'opération de **convolution**, des approximations beaucoup plus explicites d'une fonction intégrable.

Dans le théorème de Lusin, il est en général impossible d'imposer $f_\varepsilon \leq f$, alors que l'on peut le faire quand on approche f par une famille de fonctions simples. Le théorème suivant remédie partiellement à ce problème.

THÉORÈME II-63 (Théorème de Vitali-Carathéodory). *Soit X un espace topologique localement compact, et soit μ une mesure de Borel régulière sur X . Soit $f : X \rightarrow \mathbb{R}$ une fonction mesurable. Alors, pour tout $\varepsilon > 0$ il existe des fonctions f^+*

et f^- , telles que $f^- \leq f \leq f^+$, f^- est semi-continue supérieurement et majorée, f^- est semi-continue inférieurement et minorée, et

$$\int f^+ d\mu - \varepsilon \leq \int f \leq \int f^- d\mu + \varepsilon.$$

DÉMONSTRATION. Quitte à séparer f en parties positive et négative, on peut supposer $f \geq 0$. En approchant f par une suite de fonctions simples, on constate que l'on peut écrire

$$f = \sum_{i=1}^{\infty} c_i 1_{E_i}, \quad \sum_{i=1}^{\infty} c_i \mu[E_i] < +\infty.$$

Pour chaque i on choisit un ouvert O_i contenant E_i , et un compact K_i contenu dans E_i , tels que

$$\mu[O_i \setminus K_i] \leq \varepsilon/2^{i+1}.$$

On pose alors

$$f^- = \sum_{i=1}^{\infty} c_i 1_{K_i}, \quad f^+ = \sum_{i=1}^N c_i 1_{O_i},$$

où N est choisi de telle sorte que

$$\sum_{i=N+1}^{\infty} c_i \mu[E_i] \leq \varepsilon/2.$$

On vérifie facilement que f^+ et f^- vérifient toutes les conditions requises. \square

II-5. Intégration à valeurs vectorielles

Jusqu'ici, nous avons seulement cherché à intégrer des fonctions à valeurs dans \mathbb{R} ou $\overline{\mathbb{R}}$. Il est facile d'en déduire une théorie de l'intégration des fonctions à valeurs dans un espace vectoriel de dimension finie, par exemple \mathbb{R}^n ou \mathbb{C} : il suffit d'"intégrer composante par composante". On démontre facilement la proposition suivante.

PROPOSITION II-64 (intégration à valeurs dans un espace vectoriel de dimension finie). Soient (X, μ) un espace mesuré, et $E = \mathbb{R}^n$ (resp. $E = \mathbb{C}$), muni d'une norme (resp. du module complexe) $|\cdot|$. On dit qu'une fonction mesurable $f : X \rightarrow E$ est intégrable, ou sommable, si la fonction $|f|$, définie sur X et à valeurs dans \mathbb{R}_+ , est sommable. Une base $(e_k)_{1 \leq k \leq n}$ de E en tant que \mathbb{R} -espace vectoriel étant donnée, on peut décomposer la fonction f sous la forme

$$f = \sum f_k e_k,$$

où les fonctions f_k sont mesurables de X dans \mathbb{R} . Si f est sommable, toutes les fonctions f_k le sont, et on définit

$$\int_X f d\mu = \sum \left(\int_X f_k d\mu \right) e_k.$$

Le vecteur ainsi défini ne dépend pas du choix de la base (e_k) , et l'opération d'intégration ainsi construite satisfait aux règles de calcul suivantes : pour toutes fonctions sommables f et g , et pour tout $\lambda \in \mathbb{R}$ (resp. $\lambda \in \mathbb{C}$),

$$\int (\lambda f) d\mu = \lambda \int f d\mu,$$

$$\int (f + g) d\mu = \int f d\mu + \int g d\mu,$$

$$\left| \int f d\mu \right| \leq \int |f| d\mu.$$

En outre, si $E = \mathbb{R}^n$ et $|\cdot|$ est la norme euclidienne (resp. $E = \mathbb{C}$ et $|\cdot|$ est le module), il ne peut y avoir égalité dans la dernière inégalité que si l'image de f est, hors d'un ensemble négligeable, contenue dans une droite de \mathbb{R}^n (resp. de \mathbb{C} , vu comme \mathbb{R} -espace vectoriel).

On peut maintenant se poser la question de l'intégration de fonctions à valeurs dans des espaces vectoriels plus généraux, éventuellement de dimension infinie. C'est ce que l'on appelle la **théorie de l'intégration à valeurs vectorielles**, ou théorie de l'**intégrale de Bochner**. Cette question est assez naturelle quand on considère des fonctions à plusieurs variables comme des fonctions d'une variable à valeurs vectorielles, démarche classique en théorie des équations aux dérivées partielles par exemple, ou en théorie de l'interpolation.

La définition de la sommabilité tombe sous le sens : une fonction mesurable f de X dans un espace vectoriel abstrait E muni d'une norme $\|\cdot\|$ est dite intégrable si la fonction $\|f\|$ est intégrable sur X . Cependant, il est beaucoup plus délicat de définir l'intégrale de f : on ne peut pas, en général, se contenter de la définir composante par composante.

On peut développer la théorie de l'intégration à valeurs vectorielles d'une manière assez similaire à la théorie de l'intégration des fonctions à valeurs réelles, avec quelques subtilités cependant. Par exemple, pour passer de l'intégrale des fonctions à valeurs positives à l'intégrale des fonctions à valeurs réelles, nous avons tout simplement utilisé le fait que tout nombre réel peut s'écrire comme différence de deux nombres positifs. Mais si l'espace des valeurs est un espace fonctionnel E , il n'est pas forcément vrai que toute fonction de E puisse s'écrire comme différence de deux fonctions de E à valeurs positives. On peut parfois élargir E pour s'y ramener ; on peut aussi développer la théorie directement, sans passer par les fonctions positives. Ce faisant, on peut être amené à modifier légèrement la définition des fonctions simples pour que leur intégration soit toujours possible, en considérant uniquement des combinaisons linéaires de fonctions indicatrices d'ensembles **de mesure finie**, etc.

Il arrive aussi que les deux démarches (passage par les fonctions à valeurs positives, ou approche directe) soient possibles ; au Chapitre ?? on étudiera un exemple, le Théorème de Désintégration de la mesure, qu'il est plus facile de traiter de manière directe.

Les résultats principaux de la théorie de l'intégrale de Bochner sont très similaires aux résultats classiques que nous avons étudiés jusqu'à présent, ne réservent guère de surprise, et peuvent presque toujours être formulés *in fine* dans le langage de l'intégrale classique : ainsi, l'assertion $u \in L^1(X; C(Y))$ peut se transcrire en

$$\int_X \left(\sup_Y |u(x, y)| \right) d\mu(x) < +\infty;$$

quant à l'intégrale vectorielle $\int_X u(x, \cdot) d\mu(x)$, on peut toujours se la représenter comme la fonction qui à y associe $\int_X u(x, y) d\mu(x)$. Avec de tels réflexes, le lecteur ne devrait pas rencontrer de difficultés dans l'interprétation des résultats de l'intégration à valeurs vectorielles.

Sans développer la théorie complète, nous montrerons au chapitre VII comment définir une théorie simple et satisfaisante d'intégration à valeurs vectorielles qui couvre la plupart des espaces habituels. Pour cela, nous attendrons d'être un peu plus aguerris dans le domaine de l'analyse fonctionnelle.

CHAPITRE III

Théorèmes fondamentaux d'intégration

Passons maintenant à l'étude des propriétés fondamentales de l'intégrale, telle que nous l'avons définie au chapitre précédent. Par "fondamentales" il faut juste comprendre les propriétés qui servent constamment.

Nous étudierons dans un cadre très général les outils-clés de la théorie de la mesure : (i) des théorèmes de passage à la limite sous l'intégrale, (ii) des théorèmes de changement de variable vus sous l'angle des mesures, (iii) des théorèmes sur l'intégrale produit, et (iv) quelques inégalités élémentaires portant sur des expressions intégrales. Ce matériel est développé dans les sections III-1 à III-4, qui est probablement la partie la plus importante du cours. Les sections III-5 et III-6 traitent de sujets plus avancés : l'équi-intégrabilité, la tension et la construction de mesures produits avec un nombre infini de facteurs ; elles pourront être omises en première lecture, et consultées en cas de besoin.

Dans tout ce chapitre nous travaillerons avec des mesures fixées une fois pour toutes : typiquement, un théorème fera intervenir une mesure fixée. Au contraire, dans le Chapitre VI, nous développerons un autre point de vue en considérant des *espaces* de mesures.

Pour illustrer certains théorèmes ou énoncer certains contre-exemples, nous admettrons l'existence de la mesure de Lebesgue sur \mathbb{R} , qui sera étudiée plus en détail au Chapitre IV ; pour l'instant, il suffira de savoir que la mesure de Lebesgue d'un intervalle de \mathbb{R} est simplement sa longueur, et que l'intégrale associée prolonge l'intégrale de Riemann des fonctions continues par morceaux.

III-1. Comportement face aux limites

Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables, définies sur un espace mesuré (X, μ) , à valeurs réelles. Nous allons considérer quatre problèmes différents :

★ On suppose d'abord que la suite converge en un sens très fort : de manière monotone, par exemple en croissant. Peut-on passer à la limite sous le signe de sommation ? Le **Théorème de convergence monotone de Beppo Levi** assure que c'est toujours possible.

★ On suppose maintenant que la suite converge, sans que la convergence soit monotone ; on sait alors que sa limite est mesurable. Peut-on passer à la limite sous le signe \int ? Dans de nombreuses situations, le **Théorème de convergence dominée de Lebesgue** permet de le faire.

★ Enfin, on considère le cas où la suite (f_n) ne converge pas nécessairement ; on sait alors que sa \liminf et sa \limsup sont mesurables. Peut-on relier les intégrales de ces fonctions à l'intégrale des f_n ? C'est à ce problème que répond le **Lemme de Fatou**.

★ Quand on s'intéresse aux fonctions continues, une hypothèse très forte que l'on étudie souvent est la convergence **uniforme**, qui permet en particulier de passer

à la limite dans l'intégrale de Riemann. Nous n'aurons pas de théorème de limite d'intégrale sous hypothèse de convergence uniforme, car toutes ces situations seraient en pratique couvertes par le théorème de convergence dominée, qui suppose seulement la convergence simple. On peut se demander à quel point la notion de convergence uniforme est plus forte que la notion de convergence simple. Le **Théorème d'Egorov** implique que, du point de vue de la mesure, la différence n'est pas si grande.

III-1.1. Convergence monotone. De même que toutes les propriétés cruciales des mesures découlent de la propriété de σ -additivité, toutes les propriétés importantes de passage à la limite dans l'intégrale découlent du théorème suivant, appelé théorème de convergence monotone de Beppo Levi, ou tout simplement théorème de convergence monotone. On rappelle qu'une suite (f_n) de fonctions à valeurs dans $\overline{\mathbb{R}}$ est dite croissante si, pour tout x , la suite $(f_n(x))$ est croissante.

THÉORÈME III-1 (théorème de convergence monotone de Beppo Levi). *(i) Soit $(f_n)_{n \in \mathbb{N}}$ une suite croissante de fonctions mesurables sur un espace mesuré (X, μ) , à valeurs dans $[0, +\infty]$. Alors*

$$(11) \quad \int_X \lim_{n \rightarrow \infty} f_n d\mu = \lim_{n \rightarrow \infty} \int_X f_n d\mu.$$

En particulier, la fonction $(\lim f_n)$, définie sur X et à valeurs dans $[0, +\infty]$, est sommable si et seulement si la limite des $\int f_n$ est finie.

(ii) La même conclusion est vraie si (f_n) est une suite croissante (resp. décroissante) de fonctions mesurables à valeurs dans $\overline{\mathbb{R}}$, pourvu que l'une des fonctions f_n soit minorée (resp. majorée) par une fonction sommable.

Le corollaire qui suit s'obtient en remarquant que les sommes partielles d'une série à termes positifs forment une famille croissante.

COROLLAIRE III-2 (interversion de série et sommation pour des fonctions positives). *Soit (f_n) une famille de fonctions mesurables, définies sur un espace mesuré (X, μ) , à valeurs dans $[0, +\infty]$. Alors*

$$(12) \quad \int \left(\sum_{n \in \mathbb{N}} f_n \right) d\mu = \sum_{n \in \mathbb{N}} \int f_n d\mu.$$

EXEMPLE III-3. Soient $(\mu^m)_{m \in \mathbb{N}}$ une suite croissante de mesures. On se donne des parties disjointes $(A_j)_{j \in \mathbb{N}}$, et on note $A = \cup A_j$. Pour tout m on a

$$\mu^m[A] = \sum_{j \in \mathbb{N}} \mu^m[A_j],$$

et donc

$$\lim_{m \rightarrow \infty} \mu^m[A] = \sum_{j \in \mathbb{N}} \left(\lim_{m \rightarrow \infty} \mu^m[A_j] \right).$$

Il s'ensuit que *une limite croissante de mesures est une mesure.*

REMARQUES III-4. (i) Soit $(A_n)_{n \in \mathbb{N}}$ une famille croissante d'ensembles mesurables, $A = \cup A_n$, et soit $f_n = 1_{A_n}$; il est facile de vérifier que la fonction indicatrice 1_A est la limite croissante des f_n , et la formule (11) devient donc

$$\mu[\cup A_k] = \lim_{n \rightarrow \infty} \mu[A_n].$$

Si en revanche les A_n sont supposés disjoints, on vérifie sans peine que 1_A est la somme de la série des f_n , et la formule (12) se transforme en

$$\mu[\cup A_n] = \sum_n \mu[A_n].$$

On retrouve donc les deux formulations habituelles de la σ -additivité de μ . En conclusion, le théorème de convergence monotone n'est autre que **la relation de σ -additivité exprimée en termes de fonctions plutôt que d'ensembles mesurables**.

- (ii) Il est clair que les énoncés précédents sont également valables si les conditions de croissance ou de décroissance ne sont vérifiées que presque partout.

DÉMONSTRATION DU THÉORÈME III-1. Il est facile de voir que l'énoncé (ii) est une conséquence de l'énoncé (i) : si (f_n) est une suite croissante de fonctions, avec $f_{k_0} \geq g$ sommable pour un certain k_0 , alors la famille $(f_{k_0} - g)$ vérifie les hypothèses de (i), et comme g est sommable on a

$$\int \lim_{n \rightarrow \infty} f_n = \int \lim_{n \rightarrow \infty} (f_n - g) + \int g = \lim_{n \rightarrow \infty} \int (f_n - g) + \int g = \lim_{n \rightarrow \infty} \int f_n.$$

On traite l'autre cas en changeant f_n en $-f_n$. Il nous suffit donc d'établir (i).

Soit $f := \lim f_n$; d'après nos hypothèses on a $f_n \leq f$, et il s'ensuit donc $\int f_n \leq \int f$. La suite $(\int f_n)$ étant croissante, elle converge dans $\overline{\mathbb{R}}$, et on a

$$\lim \int f_n \leq \int f.$$

Il nous reste à établir l'inégalité inverse, qui est le coeur du problème. Nous allons pour cela reprendre l'argument déjà utilisé dans la preuve de l'additivité de l'intégrale.

Soit χ une fonction simple qui minore f , et soit $\delta \in]0, 1[$, on pose

$$A_n := \{x \in X; f_n(x) \geq (1 - \delta)\chi(x)\}.$$

Par croissance de f_n , les parties A_n forment une famille croissante; en traitant à part les x tels que $\chi(x) = 0$, on vérifie sans peine la réunion des A_n est l'espace X tout entier. Si l'on écrit $\chi = \sum_{1 \leq j \leq J} \alpha_j 1_{B_j}$,

$$\int \chi 1_{A_n} = \int \sum_{j=1}^J \alpha_j 1_{A_n \cap B_j} = \sum_{j=1}^J \alpha_j \mu[A_n \cap B_j] \xrightarrow{n \rightarrow \infty} \sum_{j=1}^J \alpha_j \mu[X \cap B_j] = \int \chi.$$

D'autre part, par positivité de l'intégrale,

$$\int f_n \geq \int f_n 1_{A_n} \geq (1 - \delta) \int \chi 1_{A_n}.$$

En passant à la limite dans les deux membres, on trouve

$$\lim_{n \rightarrow \infty} \int f_n \geq (1 - \delta) \int \chi.$$

En prenant le supremum sur toutes les fonctions simples χ minorant f , et en faisant tendre δ vers 0, on aboutit bien à

$$\lim_{n \rightarrow \infty} \int f_n \geq \int f.$$

□

EXERCICE III-5. Retrouver l'additivité de l'intégrale en combinant le théorème de convergence monotone et la Proposition II-29.

Voici maintenant une conséquence simple et importante du théorème de convergence monotone.

PROPOSITION III-6 (l'intégrale restreinte définit une mesure). *Soit f une fonction mesurable définie sur un espace mesuré (X, μ) , à valeurs dans $[0, +\infty]$. Alors l'application*

$$(13) \quad f\mu : A \longmapsto f\mu[A] := \int_A f \, d\mu = \int_X f 1_A \, d\mu$$

définit une mesure sur la σ -algèbre des parties mesurables. En outre, elle vérifie

$$\mu[A] = 0 \implies f\mu[A] = 0.$$

DÉMONSTRATION. Soit $(A_k)_{k \in \mathbb{N}}$ une famille de parties disjointes, et A leur union. Comme on l'a déjà mentionné, on vérifie sans peine que $1_A = \sum 1_{A_k}$, et il en découle que

$$f 1_A = \sum (f 1_{A_k}).$$

Le Corollaire III-2 implique donc

$$\sum_k \int (f 1_{A_k}) \, d\mu = \int \left(\sum_k f 1_{A_k} \right) \, d\mu = \int (f 1_A) \, d\mu,$$

soit

$$\sum_k \int_{A_k} f \, d\mu = \int_{\cup A_k} f \, d\mu.$$

Cette propriété de σ -additivité implique que $f\mu$ est bien une mesure. \square

La dernière propriété du Corollaire III-6 est importante et mérite un nom :

DÉFINITION III-7 (absolue continuité). *Soient μ et ν deux mesures définies sur une σ -algèbre commune. On dit que ν est absolument continue par rapport à μ , et on note parfois $\nu \ll \mu$, si pour tout A mesurable,*

$$\mu[A] = 0 \implies \nu[A] = 0.$$

Nous verrons au Chapitre ?? que, sous certaines conditions, les mesures absolument continues par rapport à une mesure μ sont exactement les mesures $f\mu$. Notons une dernière propriété importante de ces mesures :

PROPOSITION III-8 (changement de densité de référence). *Sur (X, μ) un espace mesuré,*

(i) *Soient f et g deux fonctions mesurables à valeurs dans $[0, +\infty]$. Alors*

$$\int f g \, d\mu = \int f \, d(g\mu);$$

(ii) *Soient h et g deux fonctions mesurables à valeurs dans $[0, +\infty]$, telles que*

$$g(x) \in \{0, +\infty\} \implies h(x) = 0.$$

Alors, avec les conventions $1/0 = +\infty$, $0 \times +\infty = 0/0 = +\infty/+\infty = 0$, on a

$$\int h \, d\mu = \int \frac{h}{g} \, d(g\mu).$$

DÉMONSTRATION. (i) Soit (f_n) une suite de fonctions simples convergeant en croissant vers f . Par convergence monotone, $\int f_n g d\mu$ converge vers $\int f g d\mu$ et $\int f_n d(g\mu)$ vers $\int f d(g\mu)$. Il suffit donc de prouver le résultat quand f est une fonction simple, et par linéarité il suffit de le prouver quand f est de la forme 1_A . On reconnaît alors la définition de la mesure $g\mu$.

(ii) Sous les hypothèses que nous avons faites sur les valeurs de g et h , on a l'égalité

$$h = \left(\frac{h}{g}\right) g,$$

ce qui permet d'appliquer (i) avec $f = h/g$. \square

III-1.2. Lemme de Fatou.

THÉORÈME III-9 (lemme de Fatou). (i) Soient (f_n) une suite de fonctions définies sur un espace mesuré (X, μ) , à valeurs dans $[0, +\infty]$. Alors

$$\int (\liminf_{n \rightarrow \infty} f_n) d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu.$$

(ii) Cette conclusion est toujours valable si les f_n sont toutes minorées par une fonction sommable.

DÉMONSTRATION. Là encore, l'énoncé (ii) est une conséquence immédiate de l'énoncé (i), que nous allons démontrer. Soit $g_n := \inf_{k \geq n} f_k$. On vérifie facilement que g_n est mesurable, et définit une suite **croissante** qui converge partout vers $f := \liminf f_k$. Bien sûr, $g_n \leq f_n$. En appliquant le théorème de convergence monotone et en passant à la \liminf , on trouve

$$\int f = \lim_{n \rightarrow \infty} \int g_n \leq \liminf_{n \rightarrow \infty} \int f_n.$$

\square

REMARQUE III-10. Il est très facile de construire des exemples où

$$\int \liminf f_n < \liminf \int f_n,$$

même si la convergence a lieu presque partout. Voici trois situations typiques, sur l'espace \mathbb{R} muni de la mesure de Lebesgue. Soit φ une fonction continue, positive, nulle en-dehors de l'intervalle $[0, 1]$, non identiquement nulle. Pour $n \geq 1$ on définit

$$\begin{cases} f_n(x) = n\varphi(nx); \\ g_n(x) = n^{-1}\varphi(n^{-1}x); \\ h_n(x) = \varphi(x - n). \end{cases}$$

Alors les suites de fonctions (f_n) , (g_n) et (h_n) convergent vers 0 partout sur \mathbb{R} , pourtant il est facile de montrer, par des changements de variables évidents, que $\int f_n = \int g_n = \int h_n = 1$. On dit que la suite (f_n) illustre un phénomène de **concentration** (toute la masse de la suite se concentre près de 0), la suite (g_n) un phénomène d'**évanescence** (toute la masse part à l'infini de manière diffuse), tandis que la suite (h_n) présente un comportement de **bosse glissante** (la masse "glisse" à l'infini, sans s'étaler).

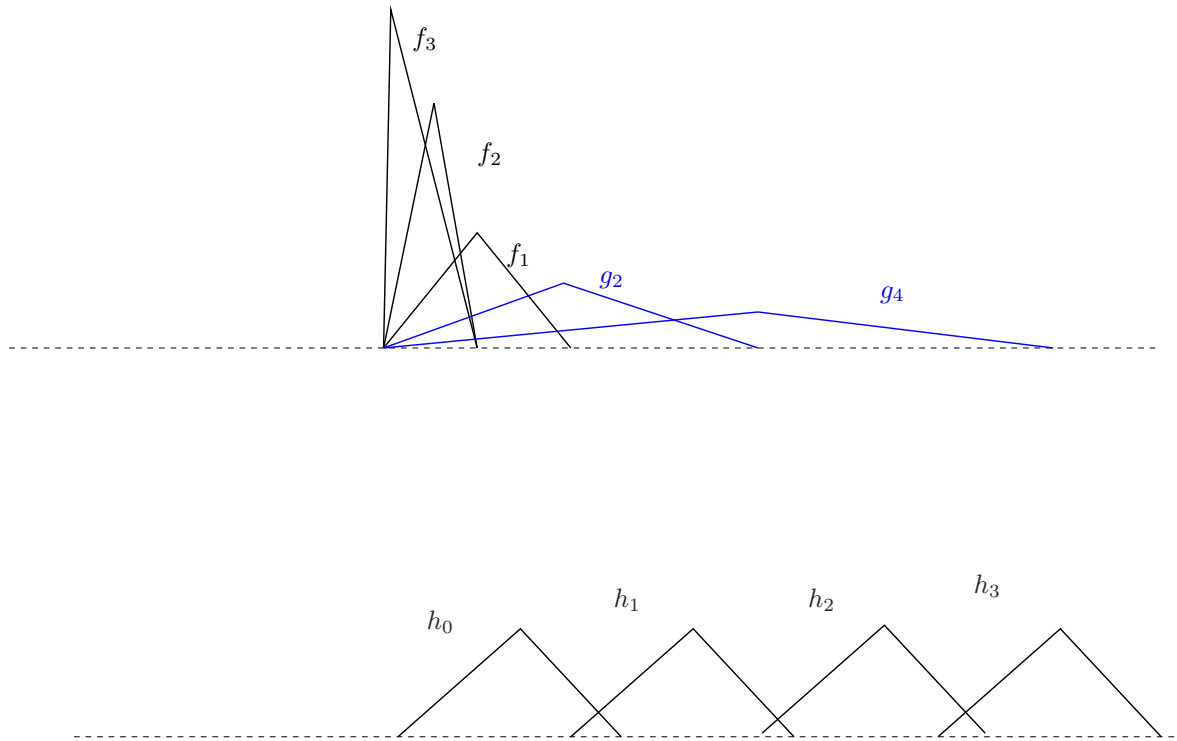


FIG. 1. concentration des f_n , évanescence des g_n , bosse glissante h_n

III-1.3. Convergence dominée.

THÉORÈME III-11 (théorème de convergence dominée de Lebesgue). *Soit $(f_n)_{n \in \mathbb{N}}$ une famille de fonctions définies sur un espace mesuré (X, μ) , à valeurs dans $\overline{\mathbb{R}}$. On suppose que (f_n) est dominée, i.e.*

(i) *Il existe g sommable tel que $|f_n| \leq g$ μ -presque partout pour tout n ;*

alors

$$(14) \quad \int_X \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int_X f_n d\mu \leq \limsup_{n \rightarrow \infty} \int_X f_n d\mu \leq \int_X \limsup_{n \rightarrow \infty} f_n d\mu.$$

En particulier, si (f_n) est dominée et vérifie

(ii) *f_n converge presque partout vers f ,*

alors f est intégrable et on a

$$(15) \quad \lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0; \quad \text{et} \quad \int_X (\lim f_n) d\mu = \lim_{n \rightarrow \infty} \int_X f_n d\mu.$$

En outre, on peut remplacer dans cet énoncé l'hypothèse de domination par la condition plus faible

(i') *Pour tout n il existe g_n sommable tel que $|f_n| \leq g_n$ μ -presque partout, et $\int (\lim g_n) = \lim \int g_n < +\infty$.*

En introduisant les sommes partielles de séries de fonctions, on déduit facilement de ce théorème le corollaire suivant :

COROLLAIRE III-12 (interversion de série et sommation sous hypothèse de domination). Soit $(f_n)_{n \in \mathbb{N}}$ une famille de fonctions mesurables définies sur un espace mesuré (X, μ) , à valeurs dans $\overline{\mathbb{R}}$. Si

$$\int \left(\sum_{n \in \mathbb{N}} |f_n| \right) d\mu < +\infty,$$

alors chaque f_n est sommable, la série de terme général $\int f_n d\mu$ converge, et

$$\int \left(\sum_{n \in \mathbb{N}} f_n \right) d\mu = \sum_{n \in \mathbb{N}} \int f_n d\mu.$$

REMARQUES III-13. (i) La fonction $(\lim f_n)$ n'est définie qu'en dehors d'un ensemble de mesure nulle; pour que l'énoncé précédent ait un sens, il convient donc de l'étendre en une fonction mesurable sur X tout entier, grâce au Théorème II-14(iii).

(ii) La condition (ii) est appelée condition de **domination** : toutes les f_n sont dominées par une fonction intégrable g . Elle est équivalente à l'hypothèse

$$\int \sup_{n \in \mathbb{N}} |f_n| d\mu < +\infty.$$

Bien sûr, cette condition n'était pas satisfaite par les exemples présentés dans la Remarque III-10.

(iii) L'hypothèse de domination de la suite (f_n) peut être affaiblie comme suit : de toute suite extraite de (f_n) on peut extraire une sous-suite dominée. En effet, si une suite (x_n) de nombres réels est telle que toute suite extraite admet une sous-suite convergeant vers $\ell \in \mathbb{R}$, alors la suite (x_n) tout entière converge vers ℓ (en l'occurrence, $\ell = \int f$).

EXEMPLE III-14. Sachant que l'intégrale de Lebesgue généralise l'intégrale de Riemann, on déduit facilement du Théorème III-11 l'énoncé suivant : soit (f_n) une suite de fonctions continues par morceaux sur $[a, b] \subset \mathbb{R}$, bornée uniformément, et convergeant simplement vers une fonction f . Alors $\lim \int f_n(x) dx = \int f(x) dx$. En effet, l'hypothèse de borne uniforme revient à une hypothèse de domination par une fonction constante, qui est clairement intégrable sur un intervalle borné.

DÉMONSTRATION DU THÉORÈME III-11. Nous allons partir de l'hypothèse (i') ; on note $g = \lim g_n$. Chaque f_n est bien sûr intégrable puisque sa valeur absolue est majorée par une fonction intégrable ; en outre $\liminf \int f_n$ et $\limsup \int f_n$ sont majorées en valeur absolue par $\limsup \int g_n = \int g < +\infty$. Enfin $\liminf f_n$ et $\limsup f_n$ sont majorées en valeur absolue par g ; ce sont donc également des fonctions intégrables.

L'énoncé que nous voulons démontrer est une conséquence simple du Lemme de Fatou. En effet, la fonction $g_n + f_n$ est positive, donc

$$\int \liminf (g_n + f_n) \leq \liminf \int (g_n + f_n).$$

En combinant cela avec l'hypothèse (i'), on obtient

$$\begin{aligned} \int g + \int (\liminf f_n) &= \int (g + \liminf f_n) = \int \liminf (g_n + f_n) \\ &\leq \liminf \int (g_n + f_n) = \int g + \liminf \int f_n. \end{aligned}$$

On conclut que

$$\int (\liminf f_n) \leq \liminf \int f_n.$$

En changeant f_n en $-f_n$, on obtient de même

$$\limsup \int f_n \leq \int (\limsup f_n).$$

Les inégalités (14) sont donc bien satisfaites.

Supposons maintenant que f_n converge presque partout vers f , et montrons que $\int |f_n - f| \rightarrow 0$; puisque f est intégrable, il en résultera par inégalité triangulaire que $\int f_n$ converge bien vers $\int f$. On applique l'énoncé précédent en remplaçant f_n et g_n par $\tilde{f}_n = |f_n - f|$ et $\tilde{g}_n = g_n + |f|$. L'hypothèse (i') est bien vérifiée, et on a donc

$$0 \leq \limsup \int |f_n - f| \leq \int \limsup |f_n - f| = 0,$$

d'où la conclusion. \square

EXERCICE III-15. Retrouver en cas particulier de ce théorème le critère connu : une série (x_n) absolument convergente de nombres réels est commutativement convergente, i.e. $(x_{\sigma(n)})$ converge pour toute bijection $\sigma : \mathbb{N} \rightarrow \mathbb{N}$, et la valeur de la somme ne dépend pas de σ .

En guise d'application du théorème de convergence dominée, voici un théorème simple de continuité des intégrales à paramètre.

THÉORÈME III-16 (continuité des intégrales à paramètre). Soient (X, \mathcal{A}, μ) un espace mesuré, et Z un espace métrique. On se donne $f : X \times Z \rightarrow [-\infty, +\infty]$ une fonction telle que

- (i) Pour tout $z \in Z$, $x \mapsto f(x, z)$ est mesurable ;
- (ii) Pour tout $x \in X$, $z \mapsto f(x, z)$ est continue ;
- (iii) Pour tout $(x, z) \in X \times Z$ on a $|f(x, z)| \leq g(x)$, où g est une fonction mesurable telle que $\int g(x) d\mu(x) < +\infty$.

Alors l'application $\varphi(z) = \int f(x, z) d\mu(x)$ est bien définie et continue sur Z .

DÉMONSTRATION. L'hypothèse (iii) implique la sommabilité de $|f(\cdot, z)|$ pour tout z ; l'intégrale $\int f(x, z) d\mu(x)$ est donc bien définie. Soit z_n une suite convergeant vers z ; le problème est de montrer que $\varphi(z_n) \rightarrow \varphi(z)$. Posons $f_n(x) = f(x, z_n)$, et $f(x) = f(x, z)$. Par (ii), on a convergence (partout) de f vers f_n ; et par (iii) la famille (f_n) est dominée par g . La conclusion s'ensuit du théorème de convergence dominée, appliqué à la famille (f_n) . \square

REMARQUE III-17. La question de la mesurabilité et de l'intégrabilité des intégrales à paramètre sera abordée plus loin dans ce chapitre; voir le Théorème III-50.

Voici un corollaire pratique du Théorème III-16 :

COROLLAIRE III-18 (Dérivation des intégrales à paramètre). Soient (X, \mathcal{A}, μ) un espace mesuré, I un intervalle de \mathbb{R} , et $f : X \times I \rightarrow \mathbb{R}$. On suppose que

- (i) Pour tout $t \in I$, $x \mapsto f(x, t)$ est mesurable et $\int |f(x, t)| d\mu(x) < +\infty$;
- (ii) Pour tout $x \in X$, $t \mapsto f(x, t)$ est continûment différentiable ;
- (iii) Pour tout $(x, t) \in X \times I$,

$$\left| \frac{\partial f}{\partial t}(x, t) \right| \leq g(x),$$

où g est une fonction mesurable telle que $\int g(x) d\mu(x) < +\infty$.

Alors $F : t \mapsto \int f(x, t) d\mu(x)$ est dérivable sur I , et pour tout $t \in I$ on a

$$F'(t) = \int_X \frac{\partial f}{\partial t}(x, t) d\mu(x).$$

DÉMONSTRATION. Soit $t \in I$ fixé, et $\varepsilon > 0$ tel que $[t - \varepsilon, t + \varepsilon] \subset I$. Pour $x \in X$ et $s \in [-\varepsilon, \varepsilon]$ on définit

$$h(x, s) = \begin{cases} \frac{f(x, t - s) - f(x, t)}{s} & \text{si } 0 < |s| < \varepsilon \\ \frac{\partial f}{\partial t}(x, t) & \text{si } s = 0. \end{cases}$$

La fonction $h(x, s)$ est alors mesurable en x , continue en s , et majorée uniformément par $g(x)$ en vertu du théorème des accroissements finis. D'après le Théorème III-16, $\lim_{s \rightarrow 0} \int h(x, s) d\mu(x) = \int h(x, 0) d\mu(x)$, ce qui est équivalent au résultat recherché. \square

EXERCICE III-19. Énoncer et prouver une variante du Corollaire III-18 qui s'applique à des fonctions convexes plutôt que lipschitziennes, et qui soit basée sur le Théorème de Convergence Monotone plutôt que sur le Théorème de Convergence Dominée.

III-1.4. Que penser de l'hypothèse de domination ? Dans le théorème de convergence dominée, la condition de domination peut sembler un peu forte, mais les Exemples III-10 montrent que l'on ne peut l'éliminer purement et simplement de l'énoncé du Théorème III-11. On peut se demander toutefois si on peut la remplacer par une hypothèse moins contraignante. En particulier, existe-t-il des situations où la convergence des intégrales est vraie sans qu'il y ait domination ? Voici deux exemples faisant intervenir la mesure de Lebesgue sur \mathbb{R} :

EXEMPLES III-20 (la convergence peut avoir lieu sans domination). (i) Soit (a_n) une famille de nombres positifs, tendant vers 0, dont la somme diverge, et soit, sur \mathbb{R} , la fonction $f_n = a_n 1_{[n, n+1[}$. Alors f_n converge vers 0, et l'intégrale de f_n également ; cependant la fonction $\sup f_n$ est la fonction constante par morceaux valant a_n sur l'intervalle $[n, n+1[$, qui n'est pas intégrable. La suite (f_n) n'est donc pas dominée.

En revanche, on peut extraire de (f_n) une sous-suite qui vérifie l'hypothèse de domination. Et même, de toute sous-suite de (f_n) on peut extraire une sous-suite qui soit dominée.

(ii) Soit f_n définie sur \mathbb{R} par :

$$f_n(x) = \begin{cases} -n & \text{si } -\frac{1}{n} \leq x < 0; \\ +n & \text{si } 0 < x \leq \frac{1}{n}; \\ 0 & \text{sinon.} \end{cases}$$

Alors chaque f_n est sommable, d'intégrale nulle, et f_n converge simplement vers la fonction nulle, mais la suite (f_n) n'est pas dominée, ni aucune de ses sous-suites extraites. En effet, si une sous-suite extraite, toujours dénotée (f_n) , était dominée, alors il en serait de même de $(f_n 1_{x \geq 0})$, et l'intégrale de f_n sur \mathbb{R}_+ convergerait vers 0 ; or elle est toujours égale à 1.

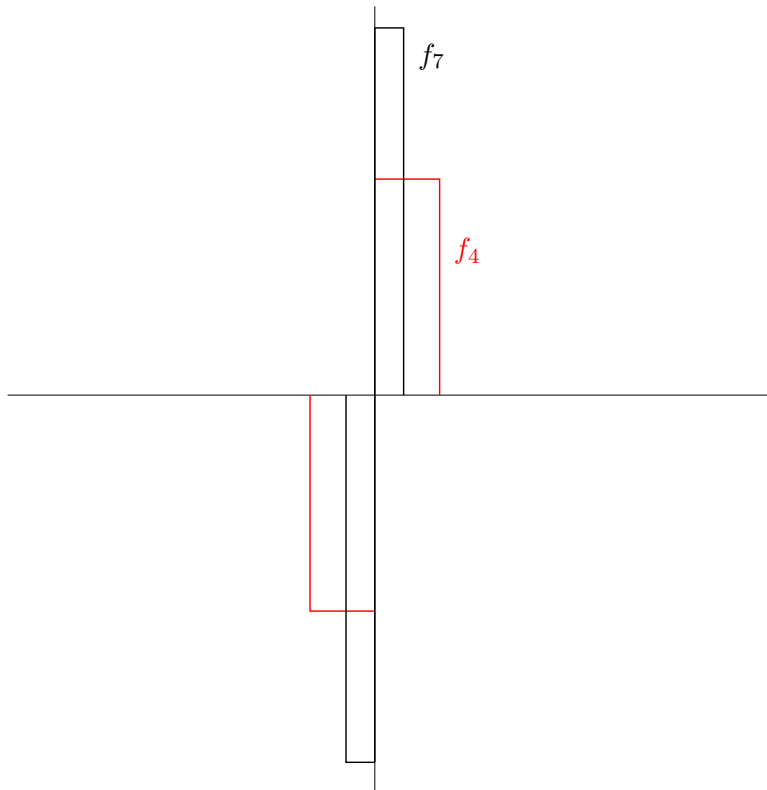


FIG. 2. compensation entre grandes valeurs positives et négatives

Dans le deuxième exemple, on peut attribuer le phénomène de non-domination au fait que de grandes valeurs positives et de grandes valeurs négatives se compensent. La théorie de Lebesgue est impuissante à exploiter de tels phénomènes. En revanche, dès que l'on exclut cette possibilité, par exemple en minorant f par une fonction intégrable, la domination devient la règle, pourvu que l'on autorise l'extraction de sous-suites comme dans le premier exemple ci-dessus.

THÉORÈME III-21 (en l'absence de fortes compensations, la domination est nécessaire). *Soit (f_n) une famille de fonctions mesurables, définies sur un espace mesuré (X, μ) , à valeurs dans $\overline{\mathbb{R}}$, convergeant presque partout vers une fonction sommable.*

(i) *On suppose que la famille (f_n) est uniformément minorée par une fonction intégrable, et que $\lim(\int f_n) = \int(\lim f_n)$. Alors, il existe une sous-suite extraite de (f_n) , notée $(f_{n'})$, et une fonction g intégrable, telle que $|f_{n'}| \leq g$ presque partout.*

(ii) *Si*

$$\lim_{n \rightarrow \infty} \int |f_n| = \int \lim_{n \rightarrow \infty} |f_n|,$$

alors il existe une sous-suite extraite de (f_n) qui vérifie l'hypothèse de domination.

Il est clair que l'énoncé (ii) est une conséquence de (i). Nous démontrerons ce théorème dans le paragraphe suivant, comme conséquence du Théorème d'Egorov.

REMARQUE III-22. L'énoncé (ii) du Théorème III-21 peut se démontrer, dans le cas particulier où les f_n tendent vers 0, comme une conséquence de la *complétude* de l'espace L^1 , dont nous reparlerons au Chapitre VI.

En combinant les Théorèmes III-11 et III-21, on obtient facilement le corollaire suivant.

COROLLAIRE III-23 (en l'absence de fortes compensations, l'échange limite-somme est presque équivalent à la domination). *Soient μ un espace mesuré, et $(f_n)_{n \in \mathbb{N}}$ une famille de fonctions mesurables, définies de X dans $\overline{\mathbb{R}}$, uniformément minorée par une fonction sommable. On suppose que f_n converge presque partout vers une fonction sommable. Alors les deux énoncés*

$$\text{“ } \lim \int f_n d\mu = \int (\lim f_n) d\mu \text{ ”}$$

et

“De toute suite extraite $(f_{n'})$ on peut extraire une sous-suite dominée”

sont équivalents.

III-1.5. Théorème d'Egorov. Comment faire le lien entre la notion naturelle de convergence dans la théorie de Lebesgue, c'est-à-dire la convergence presque partout, et la notion naturelle de convergence des fonctions continues, c'est-à-dire la convergence uniforme? Par définition, la convergence uniforme implique la convergence simple, en particulier presque partout. Du point de vue des fonctions continues, la différence entre les deux notions est considérable : par exemple, une limite simple de fonctions continues n'est en général pas continue. Cependant, du point de vue de la théorie de la mesure, la différence n'est pas si grande, au sens de l'énoncé suivant.

THÉORÈME III-24 (Théorème d'Egorov). *Soit (X, μ) un espace mesuré, $\mu[X] < +\infty$, et soit $(f_n)_{n \in \mathbb{N}}$ une famille de fonctions mesurables, définies sur X , à valeurs dans \mathbb{R} . On suppose que (f_n) converge presque partout dans \mathbb{R} . Alors, (f_n) converge uniformément en-dehors d'un ensemble de mesure arbitrairement petite. En d'autres termes, pour tout $\varepsilon > 0$ il existe un ensemble mesurable $A_\varepsilon \subset X$ tel que $\mu[A_\varepsilon] < \varepsilon$, et (f_n) converge uniformément vers sa limite sur $X \setminus A_\varepsilon$.*

EXEMPLE III-25. Un exemple traditionnel de suite qui converge simplement mais non uniformément est la famille des fonctions $f_n : x \mapsto x^n$ sur $[0, 1]$. Cette suite converge simplement vers la fonction valant 1 en 1, et 0 ailleurs ; pour tout n on a $\sup_{[0,1]} |f_n - f| = 1$, la convergence n'est donc pas uniforme. Cependant, elle est uniforme sur tout intervalle $[0, 1 - \varepsilon]$, $\varepsilon > 0$.

REMARQUE III-26. Puisque la convergence uniforme laisse stable la classe des fonctions continues, le théorème d'Egorov admet le corollaire suivant :

COROLLAIRE III-27 (une limite de fonctions continues est presque continue). *Soit X un espace topologique, muni de sa tribu borélienne, et soit μ une mesure de Borel finie sur X . Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions continues à valeurs dans \mathbb{R} , convergeant simplement vers une fonction $f : X \rightarrow \mathbb{R}$. Alors f est continue en-dehors d'un ensemble de mesure arbitrairement petite.*

On retrouve ainsi un énoncé très proche du théorème de Lusin II-60. Cet énoncé n'implique pas le théorème de Lusin, car il ne s'applique qu'aux limites de fonctions continues, et pas à des fonctions mesurables arbitraires ; en revanche il est valable sans hypothèse topologique sur X .

REMARQUE III-28. Il est impossible de supprimer l'hypothèse de finitude de μ dans le Théorème III-24 ; pour s'en convaincre, on peut considérer le cas où μ est la mesure de comptage sur \mathbb{N} , et la suite de fonctions f_n est définie par $f_n(k) = 1_{k \geq n}$.

PREUVE DU THÉORÈME D'EGOROV. Quitte à poser $f_n(x) = 0$ sur le complémentaire de l'ensemble où (f_n) ne converge pas, on peut supposer que (f_n) converge partout vers une fonction mesurable f à valeurs dans \mathbb{R} . Soit, pour tout $k \in \mathbb{N}$, $n \in \mathbb{N}$, l'ensemble mesurable

$$S_{n,k} := \bigcap_{i,j \geq n} \left\{ x \in X ; |f_j(x) - f_i(x)| \leq 1/k \right\}.$$

Pour tout k , la famille $(S_{n,k})$ est croissante en n , et par hypothèse,

$$\forall k, \quad \bigcup_n S_{n,k} = X.$$

Pour tout k , on peut donc trouver $n = n_k$ tel que

$$\mu[X \setminus S_{n_k,k}] < \varepsilon 2^{-k}.$$

Posons

$$S := \bigcap_{k \geq 1} S_{n_k,k}.$$

Si $x \in S$, alors $x \in S_{n_k,k}$ pour tout k , ce qui veut dire que pour tout k il existe n_k , **dépendant seulement de k et pas de x** , tel que pour tous $i, j \geq n_k$, $|f_j(x) - f_i(x)| \leq 1/k$. En faisant tendre i vers l'infini dans cet énoncé, on voit que pour tout k il existe n_k tel que pour tous $i, j \geq n_k$, $|f_j(x) - f(x)| \leq 1/k$. En d'autres termes, (f_n) converge uniformément sur S . D'autre part,

$$\mu[X \setminus S] \leq \sum_{k=1}^{\infty} \mu[X \setminus S_{n_k,k}] < \varepsilon.$$

L'ensemble $A_\varepsilon = X \setminus S$ vérifie donc la conclusion du théorème. \square

Pour illustrer l'efficacité du théorème d'Egorov, montrons comment on peut en déduire le théorème de convergence dominée de Lebesgue, et comment on peut l'utiliser pour démontrer le Théorème III-21. En fait on aurait pu présenter toute cette section en prenant comme point de départ le théorème d'Egorov plutôt que le théorème de convergence monotone.

NOUVELLE DÉMONSTRATION DU THÉORÈME III-11. Soit (f_n) une suite de fonctions convergeant presque partout, dominée par la fonction intégrable g . Soit Z l'ensemble négligeable où g vaut $+\infty$, on redéfinit $f_n(x) = 0$ et $f(x) = 0$ pour tout $x \in Z$, sans changer les valeurs des intégrales des f_n ou de f , ni l'hypothèse de convergence presque partout. D'autre part, de la domination il s'ensuit que $f_n(x) = 0$ dès que $g(x) = 0$. On peut donc appliquer la Proposition III-8 :

$$\int f_n d\mu = \int h_n d\nu, \quad \int f d\mu = \int h d\nu,$$

$$h_n = \frac{f_n}{g}, \quad h = \frac{f}{g}, \quad \nu = g\mu.$$

L'ensemble des points où g s'annule est de mesure nulle pour ν ; en-dehors de cet ensemble, h_n converge vers $h := f/g$. Par ailleurs, ν est une mesure finie. On peut donc appliquer le théorème d'Egorov à la famille (h_n) et à la mesure ν , et on trouve que pour tout $\varepsilon > 0$ il existe A_ε tel que $\nu[A_\varepsilon] < \varepsilon$, et h_n converge uniformément vers h sur $X \setminus A_\varepsilon$. Par hypothèse de domination, h_n est borné par 1, donc h également. D'où

$$\left| \int_{X \setminus A_\varepsilon} h_n d\nu \right| \leq \varepsilon, \quad \left| \int_{X \setminus A_\varepsilon} h d\nu \right| \leq \varepsilon.$$

On conclut que

$$\left| \int h_n d\nu - \int h d\nu \right| \leq \left| \int_{X \setminus A_\varepsilon} (h_n - h) d\nu \right| + 2\varepsilon.$$

Pour tout ε fixé, grâce à la convergence uniforme on a

$$\left| \int_{X \setminus A_\varepsilon} (h_n - h) d\nu \right| \leq \left(\sup_{x \in X \setminus A_\varepsilon} |h_n - h| \right) \nu[X] \xrightarrow{n \rightarrow \infty} 0.$$

Il s'ensuit que

$$\limsup_{n \rightarrow \infty} \left| \int h_n d\nu - \int h d\nu \right| \leq 2\varepsilon.$$

On conclut en faisant tendre ε vers 0. □

DÉMONSTRATION DU THÉORÈME III-21. Comme nous l'avons dit, il suffit de démontrer la partie (i) de ce théorème; on se donne donc une famille $(f_n)_{n \in \mathbb{N}}$ de fonctions positives, convergeant presque partout vers une limite f , intégrable, telle que $\int f_n \rightarrow \int f$.

Soit A un ensemble mesurable arbitraire. En appliquant le Lemme de Fatou et l'inégalité $\liminf a_n + \liminf b_n \leq \liminf (a_n + b_n)$, laissée en exercice, on a

$$\int_X f = \int_A f + \int_{X \setminus A} f \leq \liminf_{n \rightarrow \infty} \int_A f_n + \liminf_{n \rightarrow \infty} \int_{X \setminus A} f_n \leq \liminf_{n \rightarrow \infty} \int_X f_n = \int_X f.$$

Les deux membres étant égaux, il y a égalité à chaque étape, d'où

$$(16) \quad \int_A f \, d\mu = \liminf_{n \rightarrow \infty} \int_A f_n \, d\mu.$$

Soit $\nu := f\mu$; comme f est sommable, la mesure ν est finie. Pour tout $k \in \mathbb{N}$, on pose

$$B_k := \{x; f(x) \leq 1/k\}.$$

Les B_k forment une famille décroissante, dont l'intersection est l'ensemble où f s'annule, de mesure nulle pour ν . Pour $\varepsilon > 0$ arbitrairement petit, on peut donc choisir k assez grand pour que $\nu[B_k] \leq \varepsilon$. Par le Théorème d'Egorov, on sait également qu'il existe E tel que $\nu[E] < \varepsilon$ et f_n converge uniformément vers f en-dehors de E . Si l'on pose $C_\varepsilon = B_k \cup E$, on a construit un ensemble de ν -mesure plus petite que 2ε , tel que pour tout $x \in X \setminus C$ on ait $f(x) > 1/k$ et f_n converge uniformément vers f sur $X \setminus C$. En particulier, pour tout $n \geq m$ assez grand, on aura

$$x \in X \setminus C \implies f_n(x) \leq 2f(x).$$

D'après (16), appliqué à $A = C$, on sait que $\liminf \int_A f_n \leq 2\varepsilon$. En particulier, on peut trouver $N \geq m$ tel que

$$\int_C f_N \leq 4\varepsilon.$$

Récapitulons : pour tout $\varepsilon > 0$, pour tout $p \in \mathbb{N}$, nous pouvons construire un ensemble C et un entier $N \geq p$ tels que $\int_{C_\varepsilon} f_N \leq 4\varepsilon$, et $f_N \leq 2f$ en-dehors de C . On répète cette construction avec $\varepsilon = 2^{-k}$: n_k étant donné, on construit $C = C_k$ et $N = n_{k+1} \geq n_k$ tels que

$$\int_{C_k} f_{n_k} \leq 4 \cdot 2^{-k}; \quad x \in X \setminus C_k \implies f_{n_k}(x) \leq 2f(x).$$

On définit alors

$$g := 2f + \sum_{k \in \mathbb{N}} f_{n_k} 1_{C_k}.$$

Par construction, g majore tous les f_{n_k} ; d'autre part, g est sommable car f elle-même est sommable, et

$$\int \sum_{k \in \mathbb{N}} f_{n_k} 1_{C_k} = \sum_{k \in \mathbb{N}} \int_{C_k} f_{n_k} \leq 4 \sum_{k \in \mathbb{N}} 2^{-k} < +\infty.$$

□

III-1.6. Formule de sommation par tranches. Nous allons maintenant présenter une application importante des théorèmes vus précédemment. Soit λ la mesure de Lebesgue sur \mathbb{R} , telle que nous l'avons construite dans la section I-6. Si f est mesurable, on notera $\{f > t\} = \{x; f(x) > t\}$.

THÉORÈME III-29 (Formule de sommation par tranches). *Soient (X, μ) un espace mesuré, et f une fonction mesurable positive; alors*

$$\begin{aligned} \int_X f(x) \, d\mu(x) &= \int_{\mathbb{R}_+} \mu[\{f > t\}] \, d\lambda(t) \\ &= \lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} \frac{1}{2^n} \mu \left[\left\{ x; f(x) \geq \frac{k}{2^n} \right\} \right]. \end{aligned}$$

REMARQUE III-30. Cet énoncé justifie en quelque sorte le dessin de la figure 1 présenté dans l'introduction.

DÉMONSTRATION DU THÉORÈME III-29. Dans le cas où $f = 1_A$, A étant un ensemble mesurable quelconque, les trois quantités ci-dessus valent $\mu[A]$ et sont donc égales.

Considérons ensuite le cas où f est une fonction simple, prenant donc un nombre fini de valeurs non nulles, toutes de la forme $k/2^{n_0}$. Pour n fixé, on pose $A_{k,n} = \{f \geq k/2^n\}$. Dès que $n \geq n_0$, on peut écrire $f = \sum 2^{-n} 1_{A_{k,n}}$, et pour tout $t \in [(k-1)2^{-n}, k2^{-n}[$ on a $\{f > t\} = \mu[A_{k,n}]$. Alors il est facile de se convaincre que les trois quantités apparaissant dans l'énoncé du Théorème III-29 sont encore égales.

Soit enfin f une fonction mesurable positive. Par le Théorème II-29, on peut construire une suite (f_n) de fonctions simples telles que $0 \leq f_n \leq f$, f_n converge en croissant vers f , et f_n prend ses valeurs dans $\mathbb{N}/2^n$. D'après le résultat précédent, on sait que

$$\int_X f_n d\mu = \int_{\mathbb{R}} \mu[\{f_n > t\}] d\lambda(t).$$

Par le Théorème de Convergence Monotone, $\int f_n d\mu$ converge vers $\int f d\mu$. D'autre part, il est équivalent de dire que $f(x) > t$ ou que $f_n(x) > t$ pour n assez grand ; en particulier, $\{f > t\}$ est l'union croissante des $\{f_n > t\}$. Par σ -additivité,

$$\mu[\{f > t\}] = \lim_{n \rightarrow \infty} \mu[\{f_n > t\}].$$

On peut alors appliquer le Théorème de Convergence Monotone une seconde fois, à la suite de fonctions (dans la variable t !) $\mu[\{f_n > t\}]$, pour découvrir que

$$\int_{\mathbb{R}} \mu[\{f_n > t\}] d\lambda(t) \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}} \mu[\{f > t\}] d\lambda(t).$$

On conclut que

$$\int_X f d\mu = \int_{\mathbb{R}} \mu[\{f > t\}] d\lambda(t).$$

Enfin, posons $\phi(t) = \mu[\{f > t\}]$, et soit $\phi_n(t)$ la fonction (constante par morceaux) égale à $\phi(k2^{-n})$ sur chaque intervalle $](k-1)2^{-n}, k2^{-n}[$ (on pose $\phi_n(0) = \phi(0)$). La fonction ϕ étant décroissante, on a $0 \leq \phi_n \leq \phi$, et on vérifie facilement que ϕ_n converge en croissant vers ϕ . On peut donc encore appliquer le Théorème de Convergence Monotone pour obtenir

$$\int_{\mathbb{R}} \phi(t) d\lambda(t) = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} \phi_n(t) d\lambda(t),$$

ce qui revient à

$$\int_{\mathbb{R}} \mu[\{f > t\}] d\lambda(t) = \lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} \frac{1}{2^n} \mu \left[\left\{ x; f(x) \geq \frac{k}{2^n} \right\} \right].$$

□

La formule de sommation par tranches admet une généralisation importante :

THÉORÈME III-31 (Somme par tranches, encore). *Soient (X, μ) un espace mesuré, f une fonction mesurable positive, et ν une mesure de Borel sur \mathbb{R}_+ . Pour tout $r \geq 0$, on définit $\Phi(r) = \nu[0, r[$. Alors,*

$$\begin{aligned} \int_X \Phi(f(x)) d\mu(x) &= \int_{\mathbb{R}} \mu[\{f > t\}] d\nu(t) \\ &= \lim_{n \rightarrow \infty} \sum_{k \in \mathbb{N}} \nu[(k-1)2^{-n}, k2^{-n}] \mu \left[\left\{ x; f(x) \geq \frac{k}{2^n} \right\} \right] \end{aligned}$$

REMARQUE III-32. On retrouve le Théorème III-29 via le cas particulier $\nu = \lambda$.

EXEMPLE III-33. Soit ϕ une fonction positive continue par morceaux sur \mathbb{R}_+ , et Φ sa primitive (avec $\Phi(0) = 0$). Alors

$$\int_X \Phi(f(x)) d\mu(x) = \int_{\mathbb{R}} \mu[\{f > t\}] \phi(t) d\lambda(t).$$

Par exemple,

$$(17) \quad \int_X |f|^p d\mu = \int_{\mathbb{R}} \mu[\{f > t\}] p t^{p-1} dt.$$

Nous démontrerons le Théorème III-31 plus tard. Il est clair qu'il suffit d'établir l'égalité $\int_X \Phi(f(x)) d\mu(x) = \int_{\mathbb{R}} \mu[\{f > t\}] d\nu(t)$; la suite de la conclusion en découle facilement. On donnera d'abord une démonstration dans le cas particulier où X est σ -fini, comme conséquence du Théorème de Fubini; il s'agit de la démonstration la plus simple. Le cas général sera ensuite traité comme conséquence d'un théorème de changement de variables.

III-2. Intégration sur les espaces produits

La théorie abstraite de l'intégrale de Lebesgue permet d'aborder efficacement les intégrales multiples, pourvu que l'on prenne garde à quelques subtilités.

III-2.1. Rappels et compléments sur la tribu produit.

DÉFINITION III-34 (tribu produit). *Soient (X, \mathcal{A}) et (Y, \mathcal{B}) deux espaces mesurés. On appelle tribu produit de \mathcal{A} et \mathcal{B} , et on note $\mathcal{A} \otimes \mathcal{B}$, la σ -algèbre engendrée par les pavés $A \times B$, où A et B sont des parties mesurables de X et Y respectivement.*

PROPOSITION III-35 (génération de la tribu produit). *Soient X et Y deux ensembles. On se donne \mathcal{F} une famille de parties de X , et \mathcal{G} une famille de parties de Y . On suppose que X est union dénombrable d'éléments de \mathcal{F} , et Y union dénombrable d'éléments de \mathcal{G} . Alors la famille $\mathcal{F} \otimes \mathcal{G}$ des pavés $A \times B$, où $A \in \mathcal{F}$ et $B \in \mathcal{G}$, génère la tribu produit $\sigma(\mathcal{F}) \otimes \sigma(\mathcal{G})$. En d'autres termes,*

$$\sigma(\mathcal{F} \otimes \mathcal{G}) = \sigma(\mathcal{F}) \otimes \sigma(\mathcal{G}).$$

DÉMONSTRATION. Une inclusion est immédiate : $\sigma(\mathcal{F} \otimes \mathcal{G})$ est la tribu engendrée par la famille des pavés de la forme $A \times B$, où $A \in \mathcal{F}$ et $B \in \mathcal{G}$; alors que $\sigma(\mathcal{F}) \otimes \sigma(\mathcal{G})$ est engendrée par la famille des pavés de la forme $A \times B$, où $A \in \sigma(\mathcal{F})$ et $B \in \sigma(\mathcal{G})$. Il s'ensuit que

$$\sigma(\mathcal{F} \otimes \mathcal{G}) \subset \sigma(\mathcal{F}) \otimes \sigma(\mathcal{G}).$$

C'est l'inclusion réciproque qu'il faut établir. Pour cela, il suffit de montrer que

$$\forall A \in \sigma(\mathcal{F}), \quad \forall B \in \sigma(\mathcal{G}), \quad A \times B \in \sigma(\mathcal{F} \otimes \mathcal{G}).$$

Pour cela, on remarque tout d'abord que pour tous $A \in \mathcal{F}$, $B \in \mathcal{G}$, les ensembles $A \times Y$ et $X \times B$ appartiennent à $\sigma(\mathcal{F} \otimes \mathcal{G})$: en effet, on peut les écrire comme unions dénombrables d'éléments de $\mathcal{F} \otimes \mathcal{G}$. A partir de là, la démonstration suit encore une fois un schéma classique, que nous avons déjà utilisé dans la preuve du Théorème I-68. On montre dans un premier temps que $A \times B \in \sigma(\mathcal{F} \otimes \mathcal{G})$ pour tout $A \in \sigma(\mathcal{F})$, $B \in \mathcal{G}$; pour cela on vérifie que, B étant fixé dans \mathcal{G} , l'ensemble des A tels que $A \times B \in \sigma(\mathcal{F} \otimes \mathcal{G})$ est une σ -algèbre contenant \mathcal{F} . Enfin on montre que $A \times B \in \sigma(\mathcal{F} \otimes \mathcal{G})$ pour tout $A \in \sigma(\mathcal{F})$, $B \in \sigma(\mathcal{G})$, par un argument similaire. \square

Dans un cadre abstrait, la tribu produit peut être très difficile à décrire. Mais dans le cadre des tribus boréliennes, le problème se simplifie grâce à la proposition suivante.

PROPOSITION III-36 (produits de tribus boréliennes). *Soient X et Y deux espaces topologiques, munis de leurs tribus boréliennes respectives $\mathcal{B}(X)$ et $\mathcal{B}(Y)$. Si X et Y sont des espaces métriques séparables, alors*

$$\mathcal{B}(X \times Y) = \mathcal{B}(X) \otimes \mathcal{B}(Y).$$

DÉMONSTRATION. 1. Appliquons la Proposition III-35 avec \mathcal{F} la famille des ouverts de X , et \mathcal{G} la famille des ouverts de Y : on obtient que $\mathcal{B}(X) \otimes \mathcal{B}(Y)$ est la tribu engendrée par les ouverts de la forme $A \times B$ où A est un ouvert de X et B un ouvert de Y . En particulier,

$$\mathcal{B}(X) \otimes \mathcal{B}(Y) \subset \mathcal{B}(X \times Y).$$

Cette conclusion ne fait pas appel à l'hypothèse de séparabilité, qui sera utilisée seulement pour établir l'inclusion inverse.

2. Comme X est métrique séparable, il contient une **base dénombrable d'ouverts** : les boules ouvertes $B(x_k, 1/n)$, où (x_k) est une suite dense. Cela veut dire que tout ouvert U est réunion dénombrable de telles boules : comme dans la preuve du Théorème I-32, il suffit d'écrire

$$U = \bigcup_{B(x_k, 1/n) \subset O} B(x_k, 1/n).$$

En effet, si $x \in U$, alors il existe une boule $B(x, r)$ incluse dans U , et une sous-suite extraite de (x_k) qui converge vers x , et en particulier $x \in B(x_k, 1/n) \subset O$ dès que $d(x_k, x) \leq r/4$ et $1/n \leq r/2$.

Soient maintenant O un ouvert de $X \times Y$, et $(x, y) \in O$. Par définition de la topologie produit, il existe un ouvert $O' = U \times V$ inclus dans O et contenant x , où U est un ouvert de X et V un ouvert de Y . En particulier, $(x, y) \in B(x_k, 1/n) \times B(y_\ell, 1/m)$ pour k, ℓ, m, n bien choisis. La conclusion est que O s'écrit comme une union dénombrable de $B(x_k, 1/n) \times B(y_\ell, 1/m)$; en particulier O appartient à la tribu produit $\mathcal{B}(X) \otimes \mathcal{B}(Y)$. Par définition de la tribu borélienne, $\mathcal{B}(X \times Y) \subset \mathcal{B}(X) \otimes \mathcal{B}(Y)$, ce qui conclut la preuve. \square

EXEMPLE III-37. $\mathcal{B}(\mathbb{R}^{m+n}) = \mathcal{B}(\mathbb{R}^m) \otimes \mathcal{B}(\mathbb{R}^n)$.

REMARQUE III-38. La complétion en revanche passe mal au produit tensoriel. Soient \mathcal{A} et \mathcal{B} deux tribus sur X et Y respectivement, et $\overline{\mathcal{A}}, \overline{\mathcal{B}}$ leurs tribus complétées, construites à l'aide du Théorème I-80. Soit d'autre part $\overline{\mathcal{A} \otimes \mathcal{B}}$ la complétion de la tribu produit $\mathcal{A} \otimes \mathcal{B}$. En général,

$$\overline{\mathcal{A} \otimes \mathcal{B}} \neq \overline{(\mathcal{A} \otimes \mathcal{B})}.$$

Nous verrons au Chapitre IV que ces deux tribus sont distinctes même dans le cas où $\mathcal{A} = \mathcal{B}$ est la tribu borélienne sur $[0, 1]$.

III-2.2. Applications partielles. On parle ici d'application partielle dans le même sens que “dérivée partielle”, i.e. quand on considère une fonction de deux variables comme fonction d'une seule de ces variables, l'autre étant fixée.

La terminologie suivante n'est pas courante, et nous servira seulement à fixer les idées.

DÉFINITION III-39 (tranche). Soit C un ensemble mesurable dans un espace produit $X \times Y$, muni de la tribu produit. On appelle tranche de C le long de Y toute partie de la forme

$$C_x := \{y \in Y; (x, y) \in C\}.$$

PROPOSITION III-40 (les tranches sont mesurables). Soient (X, \mathcal{A}) et (Y, \mathcal{B}) deux espaces mesurables, on munit $X \times Y$ de la tribu produit $\mathcal{A} \otimes \mathcal{B}$. Alors, pour toute partie C mesurable de $X \times Y$, et pour tout $x \in X$, la tranche C_x est une partie mesurable de Y .

DÉMONSTRATION. Soit $x \in X$, on définit

$$\mathcal{C} := \{C \subset X \times Y; C_x \in \mathcal{B}\}.$$

Il est clair que \mathcal{C} est une tribu; en fait c'est la tribu image de \mathcal{A} par l'application $\varphi_x : y \mapsto (x, y)$. Si $P = A \times B$ est un pavé, alors P_x vaut soit B (si $x \in A$), soit \emptyset (si $x \notin A$), et dans les deux cas c'est une partie mesurable de Y . Il s'ensuit que \mathcal{C} contient tous les pavés, et donc toute la tribu produit. \square

REMARQUE III-41. La conclusion de la proposition précédente est mise en défaut par des tribus d'usage courant qui sont plus grandes que la tribu produit — ne serait-ce que la tribu des ensembles Lebesgue-mesurables dans $\mathbb{R} \times \mathbb{R}$, comme nous le verrons au chapitre suivant. C'est la complétude qui est en cause.

Le théorème simple ci-dessous est le premier pas vers la construction des intégrales multiples : étant donnée une fonction de plusieurs variables, il permettra d'intégrer d'abord par rapport à une variable.

THÉORÈME III-42 (mesurabilité par rapport à une composante). Soient X et Y deux espaces mesurables, et $f : X \times Y \rightarrow \overline{\mathbb{R}}$ une fonction mesurable pour la tribu produit sur $X \times Y$. Alors, pour tout $x \in X$, la fonction $y \mapsto f(x, y)$ est mesurable de Y dans $\overline{\mathbb{R}}$.

DÉMONSTRATION. On sait que f est limite d'une suite de fonctions simples f_n . Chaque f_n s'écrit sous la forme $\sum \lambda_k 1_{C_k}$, l'application partielle $y \mapsto f_n(x, y)$ n'est autre que $\sum \lambda_k 1_{(C_k)_x}$. Par la Proposition III-40, cette application est mesurable sur Y ; comme elle ne prend qu'un nombre fini de valeurs elle est simple. En conséquence, $f(x, \cdot) = \lim f_n(x, \cdot)$ est également limite de fonctions simples, donc mesurable. \square

III-2.3. Définition de la mesure produit. Dans \mathbb{R}^2 , il est naturel de définir l'aire d'un rectangle comme le produit des longueurs des côtés. On généralise cette démarche à un cadre abstrait.

THÉORÈME III-43 (mesure produit). (i) Soient (X, \mathcal{A}, μ) et (Y, \mathcal{B}, ν) deux espaces mesurés, σ -finis. On munit $X \times Y$ de la tribu produit. Alors il existe une unique mesure θ sur $X \times Y$ telle que

$$\forall (A, B) \in \mathcal{A} \times \mathcal{B}, \quad \theta[A \times B] = \mu[A] \times \nu[B].$$

Cette mesure est appelée mesure produit de μ par ν et notée $\mu \otimes \nu$.

(ii) En outre, si \mathcal{F} (resp. \mathcal{G}) est une famille de parties de X (resp. Y), stable par intersection finie, telle que $\mathcal{A} = \sigma(\mathcal{F})$ (resp. $\mathcal{B} = \sigma(\mathcal{G})$), X est union dénombrable croissante d'éléments de \mathcal{F} (resp. Y est union dénombrable d'éléments de \mathcal{G}), alors la mesure produit sur $X \times Y$ est caractérisée par la propriété

$$\forall (A, B) \in \mathcal{F} \times \mathcal{G}, \quad \theta[A \times B] = \mu[A] \times \nu[B].$$

DÉMONSTRATION. Nous allons prouver ce théorème par application du Théorème de prolongement I-71. On définit \mathcal{F} comme l'ensemble de tous les pavés $P = A \times B$, où $A \in \mathcal{A}$ et $B \in \mathcal{B}$, et $\theta[P] = \mu[A] \times \nu[B]$. Il est clair que \mathcal{F} est stable par intersection. D'autre part, si (X_k) (resp. Y_k) est une suite croissante d'ensembles mesurables dont l'union est X (resp. Y), alors $X \times Y$ est réunion croissante des ensembles $X_k \times Y_k$, qui vérifient $\theta[X_k \times Y_k] < +\infty$. L'unicité du prolongement éventuel de θ est donc assurée par la partie (i) du Théorème I-71.

Soient $A_1 \times B_1$ et $A_2 \times B_2$ deux pavés; leur intersection $(A_1 \cap A_2) \times (B_1 \cap B_2)$ est un pavé; et leur différence est l'union de deux pavés disjoints, $(A_1 \setminus A_2) \times B_1$, et $(A_2 \setminus A_1) \times (B_1 \setminus B_2)$ (faire un dessin ou se rappeler la figure 1!). Les hypothèses de la partie (iii) du Théorème I-71 sont donc vérifiées, et nous avons seulement à vérifier la σ -additivité de θ sur l'ensemble des pavés.

Soit $P = A \times B$ un pavé, et $(P_k)_{k \in \mathbb{N}}$ un recouvrement de P par des pavés disjoints de la forme $A_k \times B_k$ (comme suggéré sur la figure III-2.3); notre but est de prouver que $\theta[P] = \sum \theta[P_k]$.

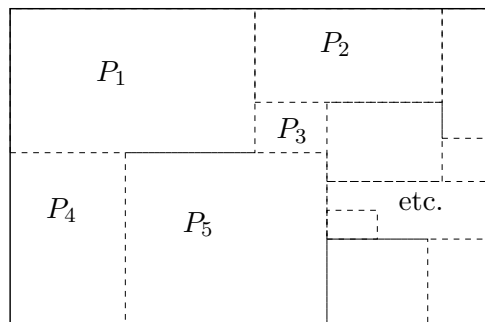


FIG. 3. Recouvrement (infini) de P par des pavés P_k

Quitte à intersecter P et les P_k avec les pavés $X_n \times Y_n$, on peut toujours supposer que $\mu[A] < +\infty$, $\nu[B] < +\infty$. Pour tout k on définit

$$f_k = \nu[B_k] 1_{A_k}.$$

Clairement, f_k est une fonction mesurable positive, et son intégrale vaut $\nu[B_k] \mu[A_k] = \theta[P_k]$. Pour tout $x \in A$, la fonction $1_{(x,y) \in P_k}$ est clairement mesurable sur B (si

$x \in B_k$, c'est la fonction indicatrice de A_k , sinon c'est la fonction nulle); on peut donc écrire

$$\begin{aligned} \sum_k f_k(x) &= \sum_k \int_B 1_{B_k}(y) 1_{A_k}(x) d\nu(y) = \sum_k \int_B 1_{(x,y) \in P_k} d\nu(y) \\ &= \int_B \sum_k 1_{(x,y) \in P_k} d\nu(y). \end{aligned}$$

Puisque chaque (x, y) de P appartient à un et un seul des P_k , la somme à l'intérieur de l'intégrale est identiquement égale à 1. Il s'ensuit, pour tout $x \in A$,

$$\sum_k \nu[B_k] 1_{A_k}(x) = \nu[B].$$

Pour tout $x \notin A$, cette somme est clairement nulle, d'où finalement

$$\sum_k \nu[B_k] 1_{A_k}(x) = \nu[B] 1_A(x).$$

Les fonctions f_k étant positives, on peut alors échanger somme et intégrale, et trouver

$$\sum_k \theta[P_k] = \sum_k \int_X f_k d\mu = \int_X \sum_k f_k = \int \nu[B] 1_A d\mu = \nu[B] \mu[A] = \theta[P],$$

ce qui achève la démonstration du point (i).

Pour prouver le point (ii), il suffit de remarquer que la famille $\mathcal{F} \otimes \mathcal{G}$ génère la tribu produit d'après la Proposition III-35, et que $X \times Y$ est réunion croissante d'une suite d'éléments de cette famille. On peut alors appliquer le Théorème I-71 (ii) pour conclure à l'unicité d'une mesure satisfaisant aux conditions requises. \square

REMARQUE III-44. La démonstration du point (i) n'est pas très intuitive. Voici une esquisse d'argument plus intuitif, mais qui ne marche pas! Introduisons une partition de A plus fine que tous les ensembles A_k , et une partition de B plus fine que tous les B_k . Chaque pavé $A_k \times B_k$ peut se redécouper en une union (au plus dénombrable) disjointe de pavés obtenus à partir des partitions plus fines : P est donc recouvert par une union dénombrable de pavés $A'_k \times B'_\ell$, où tous les A'_k sont disjoints, et tous les B'_ℓ sont disjoints. Tous les couples (k, ℓ) sont forcément représentés, sinon l'union de tous les $A'_k \times B'_\ell$ ne recouvrirait pas $A \times B$. On se ramène alors à montrer que

$$\sum_{k,\ell} \mu[A'_k] \nu[B'_\ell] = \mu[A] \nu[B],$$

ce qui est vrai puisque toutes deux quantités sont égales à

$$\left(\sum_k \mu[A'_k] \right) \left(\sum_\ell \nu[B'_\ell] \right).$$

L'erreur dans ce raisonnement est qu'il est impossible en général de définir une partition dénombrable qui soit plus fine qu'une famille dénombrable de partitions finies. Ainsi, sur $[0, 1]$, la seule partition qui soit plus fine que toutes les partitions $[0, q_n] \cup [q_n, 1]$, où $(q_n)_{n \in \mathbb{N}}$ est une énumération des rationnels de $[0, 1]$, est la partition triviale, non dénombrable, de tous les singletons.

REMARQUE III-45. La mesure produit a une importance considérable en théorie des probabilités, où elle est associée à la notion d'**indépendance**. C'est assez naturel : pour calculer la probabilité jointe de deux événements A et B qui n'ont rien à voir l'un avec l'autre, on considère comme normal de multiplier les probabilités respectives de ces deux événements.

EXEMPLE III-46. Soit $\lambda_1 = \lambda$ la mesure de Lebesgue sur \mathbb{R} ; on peut définir $\lambda_2 = \lambda \otimes \lambda$, c'est une mesure borélienne sur \mathbb{R}^2 , appelée mesure de Lebesgue 2-dimensionnelle. Alors que λ_1 mesure les longueurs, λ_2 mesure les *aires*. Nous reviendrons par la suite sur les propriétés de cette mesure et de ses analogues en dimension plus grande.

III-2.4. Généralisation : mesures dépendant d'un paramètre.

PROPOSITION III-47 (produit tensoriel par une famille de mesures). *Soient (X, μ) un espace mesuré, et Y un espace mesurable. Soit une famille $(\nu_x)_{x \in X}$ de mesures définies sur Y . On suppose que $x \mapsto \nu_x$ est mesurable, au sens où pour tout $B \subset Y$ l'application*

$$x \mapsto \nu_x[B]$$

est mesurable sur X . On suppose également que $Y = \cup Y_k$, où chaque Y_k est un ensemble mesurable de ν_x -mesure finie pour tout x . On peut alors définir sur la tribu produit une mesure $\mu \otimes \nu_x$ par la formule

$$(\mu \otimes \nu_x)[A] = \int_X \nu_x[A_x] d\mu(x).$$

DÉMONSTRATION. Il y a deux choses à vérifier : (i) que la fonction $x \mapsto \nu_x[A_x]$ est mesurable, et (ii) que la formule précédente définit bien une mesure. Sans perte de généralité, on peut supposer les Y_k disjoints ; alors les ensembles $A \cap Y_k$ induisent des tranches $(Y_k)_x$ disjointes, et

$$\nu_x[A_x] = \nu_x[\cup_k (A \cap Y_k)_x] = \sum_k \nu_x[(A \cap Y_k)_x].$$

Il suffit donc de vérifier que chaque application $x \mapsto \nu_x[(A \cap Y_k)_x]$ est mesurable ; on supposera donc, sans perte de généralité, que ν_x est finie pour tout x .

Pour l'assertion (i), soit \mathcal{A} l'ensemble des éléments de la tribu produit tels que $\nu_x[A_x]$ soit mesurable. Par hypothèse, \mathcal{A} contient tous les pavés. Il est facile de voir que \mathcal{A} est stable par union disjointe : deux ensembles disjoints A^1 et A^2 donnent lieu à des tranches distinctes A_x^1 et A_x^2 le long de Y , pour chaque x , d'où $\nu_x[(A_1 \cup A_2)_x] = \nu_x[A_x^1] + \nu_x[A_x^2]$. On montre de même que cet ensemble est stable par limite croissante. En utilisant la finitude de ν_x , on montre également qu'il est stable par différence : si $A, B \in \mathcal{A}$, $B \subset A$, alors $B \setminus A \in \mathcal{A}$. En particulier, \mathcal{A} contient la classe monotone engendrée par les pavés. Comme l'ensemble des pavés est stable par intersection finie, le Lemme de Classe Monotone assure que cette classe monotone coïncide avec la tribu produit tout entière.

Pour l'assertion (ii) on note que, si (A^n) est une famille d'ensembles mesurables disjoints, alors pour tout x les tranches A_x^n sont disjointes, d'où

$$\int_X \nu_x[(\cup A^n)_x] d\mu(x) = \int_X \sum_x \nu_x[A_x^n] d\mu(x) = \sum_n \int_X \nu_x[A_x^n] d\mu(x).$$

□

REMARQUE III-48. Si $\mu[A] = 0$ alors $(\mu \otimes \nu_x)[A \times Y] = 0$; Nous verrons dans la Section X-2 que sous certaines hypothèses peu contraignantes, cette propriété caractérise les mesures qui peuvent s'écrire sous la forme $\mu \otimes \nu_x$.

EXEMPLE III-49. Soit f une fonction intégrable positive sur \mathbb{R}^2 par rapport à la mesure $\lambda \otimes \lambda$ (mesure de Lebesgue 2-dimensionnelle). Alors la mesure $f(x, y) \lambda(dx) \lambda(dy)$ peut être considérée de deux manières équivalentes : soit comme la mesure de densité f par rapport à $\lambda \otimes \lambda$, soit comme le produit tensoriel $\lambda(dx) \otimes \nu_x$, où $\nu_x(dy) = f(x, y) \lambda(dy)$.

III-2.5. Théorème de Fubini-Tonelli-Lebesgue. Le théorème de Fubini est celui qui permet d'échanger les intégrales de fonctions Riemann-intégrables ; en voici la version dans le cadre de la théorie abstraite de l'intégration de Lebesgue. On lui donne parfois le nom de théorème de Tonelli quand on considère des fonctions positives. Dans la pratique, on dira "théorème de Fubini".

THÉORÈME III-50 (Théorème de Fubini-Tonelli-Lebesgue). *Soient (X, μ) et (Y, ν) deux espaces mesurés, σ -finis. On munit $X \times Y$ de la tribu produit. Alors*

(i) *Pour toute fonction mesurable f définie sur $X \times Y$, à valeurs dans $[0, +\infty]$, les fonctions*

$$x \longmapsto \int_Y f(x, y) d\nu(y), \quad y \longmapsto \int_X f(x, y) d\mu(x)$$

sont mesurables sur X et Y respectivement. En outre,

$$\begin{aligned} \int_{X \times Y} f(x, y) d(\mu \otimes \nu)(x, y) &= \int_X \left(\int_Y f(x, y) d\nu(y) \right) d\mu(x) = \\ &= \int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y). \end{aligned}$$

(ii) *Soit f une fonction mesurable définie sur $X \times Y$, à valeurs dans $\overline{\mathbb{R}}$. Si*

$$\int_{X \times Y} |f(x, y)| d(\mu \otimes \nu)(x, y) < +\infty,$$

alors, pour μ -presque tout x , la fonction $f(x, \cdot)$ est ν -sommable ; et pour ν -presque tout y , la fonction $f(\cdot, y)$ est μ -sommable. La fonction

$$\varphi : y \longmapsto \int_X f(x, y) d\mu(x)$$

est alors ν -sommable sur l'ensemble S_Y des y tels que $f(\cdot, y)$ est μ -sommable ; et la fonction

$$\psi : x \longmapsto \int_Y f(x, y) d\nu(y)$$

est μ -sommable sur l'ensemble S_X des x tels que $f(x, \cdot)$ est ν -sommable. En outre, si l'on redéfinit arbitrairement les valeurs de φ (resp. ψ) sur le complémentaire de S_X (resp. S_Y), on a l'égalité

$$\begin{aligned} (18) \quad \int_{X \times Y} f(x, y) d(\mu \otimes \nu)(x, y) &= \int_X \left(\int_Y f(x, y) d\nu(y) \right) d\mu(x) \\ &= \int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y). \end{aligned}$$

REMARQUE III-51. Si l'on applique ce théorème dans le cas particulier où $Y = \mathbb{N}$ et ν est la mesure de comptage, on retrouve les énoncés d'interversion somme-série que nous avons vus en corollaire des théorèmes de convergence monotone, et de convergence dominée.

DÉMONSTRATION DU THÉORÈME III-50. Il est assez facile de se convaincre que (ii) est une conséquence de (i). En effet, en appliquant (i) à la fonction positive $|f(x, y)|$, on constate que les fonctions

$$x \mapsto \int_Y |f(x, y)| d\nu(y), \quad y \mapsto \int_X |f(x, y)| d\mu(x)$$

sont sommables ; en particulier, elles sont finies presque partout, ce qui assure que pour μ -presque tout x , la fonction $f(x, y)$ est ν -sommable ; et de même, pour ν -presque tout y , cette fonction est μ -sommable. Les fonctions φ et ψ sont donc bien définies presque partout. La fonction

$$x \mapsto \int_X |f(x, y)| d\mu(y)$$

étant mesurable, l'ensemble S_X des x pour lesquels $f(x, \cdot)$ est non sommable est mesurable ; de même pour S_Y . On peut donc redéfinir φ et ψ en-dehors de ces ensembles, sans altérer leur mesurabilité. L'inégalité

$$\left| \int_Y f(x, y) d\nu(y) \right| \leq \int_Y |f(x, y)| d\nu(y)$$

assure alors que la fonction ψ est effectivement μ -sommable sur S_X ; par symétrie, il en est de même pour φ . Enfin, pour établir (18) on décompose f en partie positive et partie négative, et on applique (i) à chacune de ces fonctions.

Il reste à établir (i). La preuve de ce théorème est assez laborieuse et utilise des schémas que nous avons déjà rencontrés : remplacer les fonctions mesurables par des fonctions indicatrices, remplacer les ensembles mesurables par des pavés. Nous allons démontrer **en même temps** l'assertion de mesurabilité et la formule d'échange des intégrales. Soit \mathcal{G} l'ensemble des fonctions f vérifiant (i), et \mathcal{A} l'ensemble des parties mesurables de $X \times Y$ dont la fonction indicatrice appartient à \mathcal{G} . Dans un premier temps, on supposera que μ et ν sont finies.

1. \mathcal{A} contient les pavés. En effet, dans ce cas l'application $x \mapsto \int f(x, y) d\nu(y)$ est un multiple de la fonction indicatrice d'un ensemble mesurable, donc mesurable. En outre, le théorème de Fubini se réduit alors à la définition de la mesure produit sur les pavés.

2. \mathcal{A} est stable par limite croissante. Pour le montrer, on écrit, pour tout x

$$\int_Y 1_{\cup A_n}(x, y) d\nu(y) = \lim_{k \rightarrow \infty} \int_Y 1_{A_k}(x, y) d\nu(y),$$

ce qui est une conséquence du théorème de convergence monotone ; et une relation similaire en échangeant les rôles de X et Y . On applique une deuxième fois le théorème de convergence monotone pour établir la formule de Fubini.

3. \mathcal{A} est stable par soustraction. Pour le voir, on écrit simplement que $1_{B \setminus A} = 1_B - 1_A$ si $A \subset B$, et on applique les règles d'addition de l'intégrale : $\int (f - g) =$

$\int f - \int g$. On note que **la finitude de μ et ν est utilisée ici** ; sans cette hypothèse nous aurions des indéterminations du type $(+\infty) - (+\infty)$.

4. \mathcal{A} contient donc toute la tribu produit. C'est une conséquence du Lemme de Classe Monotone. En termes équivalents, \mathcal{G} contient toutes les fonctions indicatrices mesurables.

5. \mathcal{G} contient toutes les fonctions simples. C'est évident par linéarité de l'intégrale (on utilise ici la linéarité des deux intégrales, par rapport à μ et par rapport à ν).

6. \mathcal{G} contient toutes les fonctions mesurables. Pour le voir, on approche f mesurable par une suite croissante de fonctions simples, et on passe à la limite dans toutes les expressions en jeu en utilisant le Théorème de Convergence Monotone comme en 2).

Pour conclure la preuve, il ne reste plus qu'à remplacer l'hypothèse de finitude par celle de σ -finitude. Par hypothèse, X est une union d'ensembles mesurables X_k de mesure finie, et Y une union d'ensembles mesurables Y_k de mesure finie. Pour tout k , les conclusions de (i) sont donc vérifiées si l'on remplace X et X_k par Y et Y_k ; ou, de manière équivalente, si l'on remplace f par $f1_{X_k \times Y_k}$. Puisque f est la limite croissante des $f1_{X_k \times Y_k}$, on conclut par application répétée du Théorème de Convergence Monotone, comme en 2). \square

REMARQUE III-52. Dans l'énoncé, on a pris soin de définir (arbitrairement) les fonctions φ, ψ en-dehors de certains ensembles négligeables où leur valeur n'était pas définie (on ne définit pas l'intégrale d'une fonction non sommable dont le signe n'est pas constant). Une alternative classique consisterait à admettre que les fonctions φ, ψ ne soient définies qu'en-dehors d'un ensemble négligeable. Dans tout ce cours, on préfère éviter ce point de vue, et ne parler que de fonctions qui soient définies partout. Dans le même esprit, on évitera d'identifier a priori des fonctions qui coïncident hors d'un ensemble de mesure nulle. Indispensable dans l'analyse (fonctionnelle) des espaces de Lebesgue, cette identification est nuisible dans d'autres situations.

C'est l'occasion d'énoncer quelques remarques sur le Théorème de Fubini-Tonelli-Lebesgue, que nous illustrerons en utilisant la mesure de Lebesgue λ sur \mathbb{R} .

REMARQUE III-53. La σ -finitude de X et Y est une hypothèse importante dans le Théorème III-50. Un contre-exemple classique consiste à considérer la mesure de Lebesgue λ sur $[0, 1]$ d'une part, la mesure de comptage ν sur $[0, 1]$ d'autre part (clairement, la mesure de comptage n'est pas σ -finie, sinon $[0, 1]$ serait dénombrable). Si l'on intègre la diagonale $\Delta := \{(x, x); x \in [0, 1]\}$ de deux façons différentes, on trouve

$$\forall y, \quad \int 1_{\Delta}(x, y) d\lambda(x) = 0; \quad \forall x, \quad \int 1_{\Delta}(x, y) d\nu(y) = 1.$$

En particulier,

$$\int \int 1_{\Delta}(x, y) d\lambda(x) d\nu(y) = 0; \quad \int \int 1_{\Delta}(x, y) d\nu(y) d\lambda(x) = 1.$$

Noter que Δ est mesurable puisque intersection d'une famille dénombrable d'union de pavés (comme suggéré par la figure 4). Noter également que nos hypothèses ne garantissent pas que la mesure produit $\lambda \otimes \nu$ soit bien définie ; est bien définie en revanche la mesure extérieure $(\lambda \otimes \nu)^*$ associée aux recouvrements par des pavés. En l'occurrence, on se convainc facilement que $(\lambda \otimes \nu)^*[\Delta] = +\infty$.

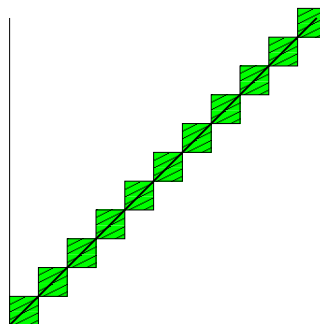


FIG. 4. La diagonale est limite d'une union de petits carrés

REMARQUE III-54. Il est également important que la fonction f soit *mesurable pour la tribu produit*! Un contre-exemple surprenant dû à Sierpinski [Rudin, p. 167] montre que les intégrales doubles

$$\int_{[0,1]} \left(\int_{[0,1]} f(x, y) d\lambda(y) \right) d\lambda(x); \quad \int_{[0,1]} \left(\int_{[0,1]} f(x, y) d\lambda(x) \right) d\lambda(y)$$

peuvent être toutes deux bien définies comme intégrales de fonctions positives mesurables, et pourtant différentes! On peut toutefois exclure ce type de pathologie par des hypothèses topologiques : par exemple, si $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est telle que les applications partielles $f(x, \cdot)$ et $f(\cdot, y)$ sont respectivement Borel-mesurables pour tout x , et continues pour tout y , alors f est automatiquement Borel-mesurable [Rudin, p. 176]. Notons également que le contre-exemple de Sierpinski utilise l'axiome du choix et l'hypothèse du continu, dont nous reparlerons dans le Chapitre IV.

En utilisant le Théorème de Fubini, nous allons maintenant démontrer la formule de sommation par tranches généralisée (sous une hypothèse de σ -finitude).

DÉMONSTRATION DU THÉORÈME III-31 QUAND X EST σ -FINI. Puisque $\Phi(f) = \nu[[0, f[$, on peut écrire, en utilisant Fubini-Tonelli-Lebesgue,

$$\begin{aligned} \int_X \Phi(f(x)) d\mu(x) &= \int_X \int_{\mathbb{R}} 1_{[0, f(x)[}(t) \nu(dt) d\mu(x) \\ &= \int_{\mathbb{R}} \left(\int_X 1_{[0, f(x)[}(t) d\mu(x) \right) \nu(dt) \\ &= \int_{\mathbb{R}} \mu[\{f > t\}] \nu(dt). \end{aligned}$$

□

III-2.6. Généralisation : intégrales multiples. Il n'y a aucune difficulté à généraliser les constructions précédentes à un produit fini d'un nombre quelconque d'espaces mesurés; on obtient ainsi le théorème ci-dessous, dont la preuve pourra être traitée en exercice.

THÉORÈME III-55 (produits multiples et intégrales multiples). (i) Soient $(X_1, \mathcal{A}_1), \dots, (X_n, \mathcal{A}_n)$ des espaces mesurés. Alors la tribu $((\mathcal{A}_1 \otimes \mathcal{A}_2) \otimes \mathcal{A}_3) \dots \otimes \mathcal{A}_n$ est la

tribu engendrée par les pavés multiples, de la forme $A_1 \times \dots \times A_n$, où $A_i \in \mathcal{A}_i$ pour tout i . On l'appelle tribu produit de $\mathcal{A}_1, \dots, \mathcal{A}_n$ et on la note

$$\mathcal{A}_1 \otimes \mathcal{A}_2 \dots \otimes \mathcal{A}_n.$$

(ii) Soient X_1, \dots, X_n des ensembles quelconques, et $\mathcal{F}_1, \dots, \mathcal{F}_n$ des familles de parties de X_1, \dots, X_n respectivement. On suppose que X_i est réunion dénombrable d'éléments de \mathcal{F}_i , pour tout $i \in \{1, \dots, n\}$. Alors la tribu produit $\sigma(\mathcal{F}_1) \otimes \sigma(\mathcal{F}_2) \dots \otimes \sigma(\mathcal{F}_n)$ est engendrée par les pavés de la forme $A_1 \times \dots \times A_n$, où $A_i \in \mathcal{F}_i$ pour tout i .

(iii) Soient X_1, \dots, X_n des espaces métriques séparables, munis de leurs tribus boréliennes respectives. Alors la tribu produit sur $X_1 \times \dots \times X_n$ coïncide avec la tribu borélienne sur $X_1 \times \dots \times X_n$.

(iv) Soient $(X_1, \mathcal{A}_1), \dots, (X_n, \mathcal{A}_n)$ des espaces mesurables, et soit $A \subset \prod X_i$ un ensemble mesurable pour la tribu produit. Alors pour tout k les $(n - k)$ -tranches

$$\{(x_1, \dots, x_{n-k}) \in X_1 \times \dots \times X_{n-k}; (x_1, \dots, x_n) \in A\}$$

sont mesurables pour la tribu produit $\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_{n-k}$.

(iv) Soient $(X_1, \mathcal{A}_1, \mu_1), \dots, (X_n, \mathcal{A}_n, \mu_n)$ des espaces mesurés σ -finis. Alors sur $X_1 \times \dots \times X_n$, muni de la tribu produit, il existe une unique mesure μ telle que pour tout pavé $P = A_1 \times \dots \times A_n$, où $A_i \in \mathcal{A}_i$ pour tout i ,

$$\mu[P] = \prod_{i=1}^n \mu_i[A_i].$$

Cette mesure coïncide avec $((\mu_1 \otimes \mu_2) \otimes \dots) \otimes \mu_n$; on l'appelle mesure produit de μ_1, \dots, μ_n et on la note

$$\mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n.$$

Si $\mathcal{F}_1, \dots, \mathcal{F}_n$ sont des familles de parties de X_1, \dots, X_n telles que pour tout i , $\mathcal{A}_i = \sigma(\mathcal{F}_i)$, \mathcal{F}_i est stable par intersection finie, et X_i est union dénombrable d'une famille croissante d'éléments de \mathcal{F}_i , alors la mesure produit est caractérisée par la propriété

$$\forall i, A_i \in \mathcal{F}_i \implies \mu[A_1 \times \dots \times A_n] = \prod_{i=1}^n \mu_i[A_i].$$

(v) Soient $(X_0, \mathcal{A}_0), \dots, (X_n, \mathcal{A}_n)_{n \in \mathbb{N}}$ des espaces mesurables. On se donne une mesure μ_0 sur X_0 ; et pour tout $j \in \{1, \dots, n-1\}$ on se donne une famille de mesures ν_{x_j} sur X_{j+1} , dépendant mesurablement de $x_j \in X_j$. On pose $X := \prod X_j$ et on le munit de la tribu produit. Alors il existe une unique mesure μ sur X telle que pour toutes parties mesurables A_i de X_i ,

$$\mu[\prod A_i] = \int_{A_0} \int_{A_1} \dots \int_{A_{n-1}} \int_{A_n} d\nu_{x_{n-1}}(x_n) d\nu_{x_{n-2}}(x_{n-1}) \dots d\nu_{x_0}(x_1) d\mu(x_0).$$

On la note $\mu_0 \otimes \nu_{x_0} \otimes \dots \otimes \nu_{x_{n-1}}$.

(vi) Soient $(X_1, \mathcal{A}_1, \mu_1), \dots, (X_n, \mathcal{A}_n, \mu_n)$ des espaces mesurés σ -finis. On munit $X^k := X_1 \times \dots \times X_k$ de la tribu produit $\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_k$. Alors, pour tout $k \in$

$\{1, \dots, n-1\}$ et toute fonction mesurable f définie sur X^n , à valeurs dans $[0, +\infty]$, la fonction

$$(x_{k+1}, \dots, x_n) \mapsto \int_{X_1 \times \dots \times X_k} f(x_1, \dots, x_n) d(\mu_1 \otimes \dots \otimes \mu_k)(x_1, \dots, x_k)$$

est mesurable sur $X_{k+1} \times \dots \times X_n$. En outre

$$\int_{X^n} f(x_1, \dots, x_n) d(\mu_1 \otimes \dots \otimes \mu_n)(x_1, \dots, x_n) = \int_{X_n} \dots \int_{X_1} f(x_1, \dots, x_n) d\mu_1(x_1) \dots d\mu_n(x_n).$$

En outre, si f est une fonction mesurable définie sur X^n , à valeurs dans $\overline{\mathbb{R}}$, telle que

$$\int_{X^n} |f(x_1, \dots, x_n)| d(\mu_1 \otimes \dots \otimes \mu_n)(x_1, \dots, x_n) < +\infty,$$

alors pour chaque k la fonction

$$(x_{k+1}, \dots, x_n) \mapsto \int_{X_1 \times \dots \times X_k} f(x_1, \dots, x_n) d(\mu_1 \otimes \dots \otimes \mu_k)(x_1, \dots, x_k)$$

est bien définie et sommable hors d'un ensemble négligeable Z_k ; quitte à la redéfinir arbitrairement sur Z_k , on a

$$\begin{aligned} \int_{X^n} f(x_1, \dots, x_n) d(\mu_1 \otimes \dots \otimes \mu_n)(x_1, \dots, x_n) &= \int_{X_n} \dots \int_{X_1} f(x_1, \dots, x_n) d\mu_1(x_1) \dots d\mu_n(x_n) \\ &= \int_{X_{\sigma(n)}} \dots \int_{X_{\sigma(1)}} f(x_1, \dots, x_n) d\mu_{\sigma(1)}(x_{\sigma(1)}) \dots d\mu_{\sigma(n)}(x_{\sigma(n)}) \end{aligned}$$

pour toute permutation σ de $\{1, \dots, n\}$.

EXEMPLE III-56. On peut définir la mesure de Lebesgue en dimension n par $\lambda_n = \lambda^{\otimes n}$.

REMARQUE III-57. Nous verrons plus loin que l'on peut, sous certaines conditions, définir des produits infinis de mesures. Cette opération n'est pas toujours permise, ainsi le produit $\lambda^{\otimes \infty}$ n'a pas de sens.

III-3. Changement de variable

Nous allons maintenant considérer des changements de variable dans un cadre mesurable. Bien sûr, dans un cadre abstrait, il n'est pas question de parler de difféomorphisme, de Jacobien, etc. Nous verrons plus tard comment faire le lien avec les notions habituelles de changement de variable. Ce théorème est basé sur la notion importante de **mesure image**.

III-3.1. Image d'une mesure par une fonction mesurable.

PROPOSITION III-58 (Tribu image). Soit (X, \mathcal{A}) un espace mesurable, Y un ensemble quelconque, et $f : X \rightarrow Y$. On peut définir une tribu, notée $f_{\#}\mathcal{A}$, ou $f\#\mathcal{A}$, ou $f_*\mathcal{A}$, ou $f\mathcal{A}$, sur Y , par

$$f_{\#}\mathcal{A} = \{B \subset Y; f^{-1}(B) \in \mathcal{A}\}.$$

Cette tribu est appelée tribu image de \mathcal{A} par f , et c'est la plus grande tribu qui rende f mesurable.

Si Y est au départ un espace mesurable, muni d'une tribu \mathcal{B} et f une application mesurable, alors $\mathcal{B} \subset f_{\#}\mathcal{A}$.

DÉFINITION III-59 (Mesure image). Soit (X, \mathcal{A}, μ) un espace mesuré, et $f : X \rightarrow Y$. Alors la formule

$$\nu[B] := \mu[f^{-1}(B)]$$

définit une mesure sur la tribu image $f_{\#}\mathcal{A}$, appelée mesure image de μ par f et notée $f_{\#}\mu$, ou $f_{\#}\mu$, ou $f_*\mu$, ou plus rarement $f\mu$.

Si (Y, \mathcal{B}) est au départ un espace mesurable, et f est une application mesurable, alors $f_{\#}\mu$ définit par restriction une mesure sur \mathcal{B} .

La preuve des assertions énoncées ci-dessus est un exercice simple de maniement des axiomes de théorie de la mesure.

III-3.2. Théorème de changement de variable.

THÉORÈME III-60. Soient X, Y deux espaces mesurables, et $\varphi : X \rightarrow Y$ une application mesurable. On suppose que X est muni d'une mesure μ . Alors,

(i) Pour toute fonction f mesurable sur Y , à valeurs dans $[0, +\infty]$,

$$(19) \quad \int f d(\varphi_{\#}\mu) = \int (f \circ \varphi) d\mu.$$

(ii) Pour toute fonction f mesurable sur Y , à valeurs dans $\overline{\mathbb{R}}$, la fonction $f \circ \varphi$ est μ -sommable si et seulement si la fonction f est $(\varphi_{\#}\mu)$ -sommable, et l'égalité ci-dessus est alors vérifiée.

DÉMONSTRATION. Il est facile de voir que (ii) est une conséquence de (i); on va donc se contenter de démontrer (i). Si f est une fonction simple, l'égalité (19) est une conséquence de la définition de $\varphi_{\#}\mu$. Dans le cas général où f est seulement supposée mesurable, on peut approcher f par une famille croissante de fonctions simples f_n ; alors $f_n \circ \varphi$ est une famille croissante de fonctions simples convergeant vers $f \circ \varphi$, et on peut passer à la limite grâce au Théorème de convergence monotone de Beppo Levi. \square

REMARQUE III-61. Il peut se produire que $f \circ \varphi$ soit mesurable sans que f le soit. C'est le cas par exemple dès que $\varphi(X)$ n'est pas mesurable, pour la fonction $f = 1_{\varphi(X)}$.

III-3.3. Morphismes d'espaces mesurés. Le théorème de changement de variables que nous avons vu ne suppose aucune régularité et pourra être intéressant dans des questions théoriques abstraites.

Soit la situation où (X, \mathcal{A}, μ) est un espace mesuré, φ une application $X \rightarrow Y$, et Y est muni de la tribu image $\varphi_{\#}\mathcal{A}$ et de la mesure image $\varphi_{\#}\mu$. Tout énoncé faisant intervenir la mesure μ et des ensembles mesurables, ou des intégrales de fonctions mesurables, se traduira en un énoncé similaire sur $(Y, \varphi_{\#}\mathcal{A}, \varphi_{\#}\mu)$. On peut dire que φ réalise un morphisme entre les espaces mesurés X et Y .

Si maintenant f est bijective, de réciproque mesurable (on dit parfois "bimesurable"), alors f^{-1} réalisera également un morphisme entre Y et X , et les énoncés de théorie de la mesure faisant intervenir (X, μ) seront équivalents aux énoncés correspondants faisant intervenir $(Y, f_{\#}\mu)$. On dit que f réalise un **isomorphisme** entre les espaces mesurés X et Y . Cette notion permet parfois de ramener des problèmes définis sur un espace en apparence compliqué, à des problèmes définis sur un espace beaucoup plus familier; il s'agit d'une abstraction de la procédure de changement de variable.

A titre d'exemple, citons un résultat de classification surprenant qui montre que dans le cadre de n'importe quel espace Polonais, on peut effectuer un changement de variable abstrait pour se ramener à l'intervalle $[0, 1]$.

THÉORÈME III-62 (représentation des espaces Polonais). *Soit (X, μ) un espace Polonais muni d'une mesure de Borel finie. Soit I l'intervalle $[0, 1]$ muni de sa tribu borélienne. Alors*

(i) *Il existe une mesure finie λ sur I , et une application mesurable $f : I \rightarrow X$ qui réalise un morphisme entre (I, λ) et (X, μ) . Autrement dit, toute mesure finie sur un Polonais est image d'une mesure finie sur $[0, 1]$.*

(ii) *Si μ est sans atome, alors on peut choisir la fonction f bijective. Autrement dit, toute mesure finie sans atome sur un Polonais est isomorphe à une mesure finie sans atome sur $[0, 1]$.*

REMARQUE III-63. L'hypothèse de non-atomicité équivaut à dire que les points sont négligeables; c'est bien sûr une condition nécessaire à l'existence d'un isomorphisme. Rappelons également (Théorème II-18) que la réciproque d'une bijection mesurable entre espaces Polonais est automatiquement mesurable.

EXEMPLE III-64. Comme nous le verrons au Chapitre IV, l'espace $[0, 1]$, muni de la mesure de Lebesgue, est isomorphe à l'espace $\{0, 1\}^{\mathbb{N}}$, muni de la mesure produit (infini) obtenue par produit tensoriel de la mesure de Bernoulli sur $\{0, 1\}$, i.e. la mesure qui attribue un poids $1/2$ à 0 et à 1. Pour autant, \mathbb{R} et $\{0, 1\}^{\mathbb{N}}$ ne sont pas topologiquement isomorphes : ainsi, le premier est connexe, alors que le second est totalement discontinu (ses composantes connexes sont des points).

Nous allons maintenant utiliser le théorème de changement de variables pour prouver le Théorème III-31 dans le cas général (rappelons que ce théorème a été démontré dans le cas où X est σ -fini à l'aide du Théorème de Fubini).

DÉMONSTRATION DU THÉORÈME III-31. Appliquons le Théorème III-29 à la fonction positive $\Phi \circ f$: on a

$$\int_X \Phi \circ f \, d\mu = \int_{\mathbb{R}_+} \mu[\{\Phi \circ f > t\}] \, d\lambda(t).$$

De par sa définition, la fonction Φ est croissante et continue à gauche (en effet, $\nu[[0, x]] = \lim_{k \rightarrow \infty} \nu[0, x - k^{-1}]$). On définit son inverse généralisé par la formule

$$\Phi^{-1}(t) := \inf\{s \geq 0; \quad \Phi(s) > t\}.$$

Il est facile de vérifier que Φ^{-1} est croissante et continue à droite. Par définition de Φ^{-1} , si $f > \Phi^{-1}(t)$ alors $\Phi(f) > t$. Si maintenant $\Phi(f) > t$, par continuité de Φ à gauche on peut trouver $\varepsilon > 0$ tel que $\Phi(f - \varepsilon) > t$, et par définition de Φ^{-1} on a $\Phi^{-1}(t) \leq f - \varepsilon < f$. On a donc

$$\Phi(f) > t \Leftrightarrow f > \Phi^{-1}(t).$$

Il s'ensuit

$$\int_{\mathbb{R}} \mu[\{\Phi \circ f > t\}] \, d\lambda(t) = \int_{\mathbb{R}} \mu[\{f > \Phi^{-1}(t)\}] \, d\lambda(t) = \int_{\mathbb{R}} \mu[\{f > s\}] \, d(\Phi^{-1})_{\#} \lambda(s).$$

Pour conclure, il suffit d'établir que

$$(\Phi^{-1})_{\#} \lambda = \nu.$$

Or la tribu borélienne sur \mathbb{R}_+ est engendrée par les intervalles de la forme $[0, s[$; il suffit donc de vérifier que

$$\lambda[\{\Phi^{-1} < s\}] = \nu[[0, s[.$$

Or la première quantité est $\lambda[[0, \Phi(s)[= \Phi(s)$, puisque $\Phi^{-1}(t) < s$ équivaut à $t < \Phi(s)$; et la deuxième quantité est par définition $\Phi(s)$. \square

III-4. Inégalités intégrales élémentaires

Pour établir des majorations sur des quantités faisant intervenir des intégrales, on utilise le plus souvent un petit nombre d'inégalités souples et puissantes, qui apparaissent dans un nombre incalculable de contextes différents. Ces inégalités, qui sont valables en toute généralité, sont les inégalités de **Chebyshev**, **Jensen**, et **Hölder**. Une quatrième inégalité d'usage très fréquent, appelée inégalité de **Minkowski** sera étudiée plus tard, au Chapitre VI.

Une certaine familiarité avec les propriétés des fonctions convexes sera utile pour lire cette section; en cas de besoin on pourra se reporter aux rappels contenus dans l'Appendice en fin de chapitre.

III-4.1. Inégalité de Chebyshev. L'inégalité de Chebyshev¹ est aussi élémentaire qu'utile, particulièrement dans le domaine des probabilités.

THÉORÈME III-65 (inégalité de Chebyshev). *(i) Soient (X, μ) un espace mesuré et $f : X \rightarrow \mathbb{R}_+$ une fonction mesurable positive. Alors, pour tout $a > 0$,*

$$(20) \quad \mu[\{x \in X; f(x) \geq a\}] \leq \frac{1}{a} \int_X f d\mu.$$

(ii) Soient (X, μ) un espace mesuré, $f : X \rightarrow \mathbb{R}_+$ une fonction mesurable positive, et $\Phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ une fonction mesurable croissante. Alors, pour tout $a \geq 0$,

$$\mu[\{x \in X; f(x) \geq a\}] \leq \frac{1}{\Phi(a)} \int_X \Phi(f(x)) d\mu(x).$$

REMARQUE III-66. Dans le cas dégénéré où $a = 0$ et f est nulle presque partout, l'inégalité (20) est a priori fautive (si l'on adopte la convention habituelle $0/0 = 0$).

REMARQUE III-67. L'énoncé (i) est souvent appelé inégalité de Markov; l'énoncé (ii) est souvent appelé inégalité de Bienaymé-Chebyshev quand $\Phi(r) = r^2$, et inégalité de Chebyshev exponentielle quand $\Phi(r) = e^{ar}$. Si l'on peut choisir des fonctions Φ croissant très vite à l'infini, mais telles que $\Phi \circ f$ soit toujours intégrable, on peut obtenir des estimations de décroissance très rapide de la mesure de $\{x; f(x) \geq a\}$ quand $a \rightarrow \infty$. En fait, dans la plupart des situations concrètes, on obtient des estimations de décroissance presque optimales par un choix convenable de Φ .

DÉMONSTRATION DU THÉORÈME III-65. Posons $A := \{x \in X; f(x) \geq a\}$. Comme f est positive, on a

$$f \geq a1_A.$$

En intégrant cette inégalité contre μ et en divisant par le nombre positif a , on obtient l'énoncé (i).

¹On ne discutera pas ici des multiples orthographes que le nom Chebyshev peut revêtir...

Pour en déduire l'énoncé (ii), il suffit d'appliquer (i) avec f remplacé par $\Phi \circ f$, et de noter que, Φ étant croissante,

$$\{x; f(x) \geq a\} \subset \{x; \Phi(f(x)) \geq \Phi(a)\}.$$

□

REMARQUE III-68. Il est facile de vérifier que l'énoncé est en général faux si f n'est pas positive! Dans la pratique, on cherchera donc toujours à s'y ramener, par exemple en prenant la valeur absolue. En utilisant des normes, on peut aussi appliquer ce théorème à des estimations de fonctions à valeurs vectorielles.

III-4.2. Inégalité de Jensen.

THÉORÈME III-69 (inégalité de Jensen dans \mathbb{R}^n). Soient X un espace mesuré, μ une mesure de probabilité sur X , $f : X \rightarrow \mathbb{R}^n$ une fonction mesurable dont chaque composante est μ -sommable, et $\Phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe. On suppose que $f \in L^1(\mu)$, et on note $\int f d\mu$ le vecteur de \mathbb{R}^n dont la composante d'ordre i est $\int f_i d\mu$. Alors

$$\Phi\left(\int f d\mu\right) \leq \int \Phi \circ f d\mu.$$

De plus, si les deux membres de l'inégalité sont finis, il y a égalité si et seulement si Φ coïncide, $f_{\#}\mu$ -presque partout, avec une fonction affine; en particulier, si Φ est strictement convexe, f doit être égale à une constante, μ -presque partout.

COROLLAIRE III-70 (inégalité de Jensen pour des puissances). Soient X un espace mesuré, μ une mesure de probabilité sur X , $f : X \rightarrow \mathbb{R}^n$ une fonction mesurable dont chaque composante est μ -sommable, et $p \in [1, +\infty[$. On note $\int f d\mu$ le vecteur de \mathbb{R}^n dont la composante d'ordre i est $\int f_i d\mu$. Alors

$$\left|\int f d\mu\right|^p \leq \int |f|^p d\mu.$$

De plus, si le membre de droite de l'inégalité est fini et $p > 1$, il y a égalité si et seulement si il existe une constante $a \in \mathbb{R}^n$ telle que, μ -presque partout, $f = a$.

REMARQUES III-71. (i) Quand $\mu = \lambda\delta_x + (1 - \lambda)\delta_y$, l'inégalité de Jensen se réduit à la définition de la convexité. Plus généralement, si l'on pose $X = \{1, \dots, N\}$, $\mu = \sum \lambda_i \delta_i$ et $f(i) = x_i$, l'inégalité de Jensen se réduit à l'inégalité

$$\Phi\left(\sum \lambda_i x_i\right) \leq \sum \lambda_i \Phi(x_i),$$

que l'on peut également adopter comme définition de la convexité. L'inégalité de Jensen n'est qu'une "limite continue" de l'inégalité ci-dessus.

(ii) L'inégalité de Jensen s'étend à n'importe quelle notion "raisonnable" d'intégrale à valeurs vectorielles, même si l'espace d'arrivée de f est de dimension infinie; voir le Théorème VII-193. En fait, compte tenu de son importance dans des contextes très divers, on pourrait ajouter l'inégalité de Jensen au cahier des charges d'une intégrale abstraite.

DÉMONSTRATION DE L'INÉGALITÉ DE JENSEN. Je vais d'abord présenter une démonstration générale, qui ne craindra pas les valeurs infinies, mais ne permettra pas de traiter les cas d'égalité.

Considérons d'abord le cas où f prend un nombre fini de valeurs $y_1, \dots, y_k \in \mathbb{R}^n$, et notons $A_k = f^{-1}(y_k)$, $\alpha_k = \mu[A_k]$. Les ensembles A_k sont mesurables et $\sum \alpha_k = 1$. Par convexité de Φ ,

$$\Phi\left(\int f d\mu\right) = \Phi\left(\sum \alpha_k y_k\right) \leq \sum \alpha_k \Phi(y_k) = \int \Phi \circ f d\mu.$$

Supposons maintenant que Φ est lipschitzienne. Par hypothèse $f \in L^1(\mu)$, donc chaque composante f_j de f peut être approchée dans $L^1(\mu)$ par une famille $(g_j^{(\ell)})_{\ell \in \mathbb{N}}$ de fonctions prenant un nombre fini de valeurs réelles. On en déduit

$$\int g^{(\ell)} d\mu \xrightarrow[\ell \rightarrow \infty]{f} \int f d\mu,$$

donc

$$\Phi\left(\int g^{(\ell)} d\mu\right) \xrightarrow[\ell \rightarrow \infty]{\Phi} \Phi\left(\int f d\mu\right);$$

et par lipschitzianité de Φ ,

$$\left| \int \Phi(f(x)) d\mu(x) - \int \Phi(g^{(\ell)}(x)) d\mu(x) \right| \leq \int |\Phi(f(x)) - \Phi(g^{(\ell)}(x))| d\mu(x) \leq \|\Phi\|_{\text{Lip}} \int |f(x) - g^{(\ell)}(x)| d\mu(x)$$

On peut donc passer à la limite dans l'inégalité de Jensen appliquée à chaque fonction $g^{(\ell)}$, et obtenir l'inégalité de Jensen pour la fonction f .

Pour conclure, on note que si Φ est convexe semi-continue inférieurement, à valeurs dans $\mathbb{R} \cup \{+\infty\}$, on peut écrire $\Phi = \sup_{k \in \mathbb{N}} \Phi_k$, où chaque Φ_k est lipschitzienne. On a donc

$$\Phi\left(\int f d\mu\right) = \sup_{k \in \mathbb{N}} \Phi_k\left(\int f d\mu\right) \leq \sup_{k \in \mathbb{N}} \int \Phi_k \circ f d\mu \leq \int \Phi \circ f d\mu.$$

Pour obtenir les cas d'égalité, nous devons travailler un peu plus. Sans perte de généralité, on peut supposer que $f(x) = x$: pour s'y ramener, il suffit de remplacer μ par $f_{\#}\mu$. L'inégalité devient alors

$$\Phi(\xi) \leq \int \Phi d\mu, \quad \xi = \int x d\mu(x).$$

(En d'autres termes, ξ est le barycentre de μ .) Supposons donc que $\Phi(\xi) = \int \Phi d\mu$. Plaçons-nous dans l'espace affine E engendré par le support de μ . Soit $\Omega = \Phi^{-1}(\mathbb{R}) \cap E$ le domaine de Φ , ou plutôt de sa restriction à E . Si $\int \Phi d\mu < +\infty$, forcément $\mu[E \setminus \Omega] = \mu[\Phi = +\infty] = 0$, autrement dit μ peut être considéré comme une mesure de probabilité sur Ω . Par le lemme III-73 ci-dessous, ξ est intérieur à Ω (dans E) ; le sous-différentiel $\partial\Phi(\xi)$ est donc non nul (on démontrera plus tard ce résultat sous des hypothèses plus générales, voir le Corollaire VII-190). Soit $y \in \partial\Phi(\xi)$; pour tout $z \in E$, on a

$$\Phi(z) - \Phi(\xi) - \langle y, z - \xi \rangle \geq 0.$$

Comme l'intégrale de cette fonction vaut

$$\int \Phi d\mu - \Phi(\xi) - \int \langle y, z - \xi \rangle d\mu(z) = \left(\int \Phi d\mu - \Phi(\xi) \right) - \int \langle y, \xi - \xi \rangle = 0 - 0 = 0,$$

elle est forcément nulle. On conclut que

$$\Phi(z) = \Phi(\xi) + \langle y, z - \xi \rangle,$$

μ -presque partout, ce qui démontre la conclusion. \square

REMARQUE III-72. Dans la preuve des cas d'égalité, on a en fait redémontré l'inégalité de Jensen, sous l'hypothèse supplémentaire que Φ est finie μ -presque partout.

LEMME III-73. Soit $\Omega \subset \mathbb{R}^n$ un convexe (non nécessairement ouvert ou fermé), et soit μ une mesure de probabilité borélienne sur Ω , telle que $\int |x| d\mu(x) < +\infty$. On note E l'espace affine euclidien engendré par le support de μ , et $\xi = \int x d\mu(x)$ le barycentre de μ . Alors ξ est intérieur à $\Omega \cap E$ dans E .

DÉMONSTRATION. On note $\Omega' = \Omega \cap E$. Supposons que $\xi \in \partial\Omega'$; par le Théorème de séparation de Hahn–Banach (dont la démonstration sera rappelée plus tard dans un cadre général, voir le Théorème VII-84) il existe une forme linéaire $\lambda \in E^*$, et $\alpha \in \mathbb{R}$, tels que $\lambda \leq \alpha$ sur Ω' et $\lambda(\xi) = \alpha$. On considère $\nu = \lambda_{\#}\mu$: c'est une mesure sur $(-\infty, \alpha]$ dont le barycentre est égal à α ; elle est donc forcément égale à δ_{α} . On en déduit que μ est concentrée sur un hyperplan de E , ce qui est en contradiction avec la définition de E . \square

Il faut prendre bien garde, quand on applique l'inégalité de Jensen, à l'hypothèse sur la mesure μ : ce doit être une mesure de probabilité. Il existe cependant un cas intéressant où cette hypothèse peut être omise : c'est celui où la fonction Φ est **homogène de degré 1**, au sens où

$$\forall \lambda \in \mathbb{R}, \forall x \in \mathbb{R}^n, \quad \Phi(\lambda x) = \lambda \Phi(x).$$

THÉORÈME III-74 (inégalité de Jensen pour des fonctions convexes homogènes). Soient X un espace mesuré, μ une mesure sur X , $f : X \rightarrow \mathbb{R}^n$ une fonction mesurable dont chaque composante est μ -sommable, et $\Phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, semi-continue inférieurement, homogène de degré 1. Alors

$$\Phi\left(\int f d\mu\right) \leq \int \Phi \circ f d\mu,$$

où l'on convient que le membre de droite vaut $+\infty$ si $\Phi \circ f$ n'est pas sommable.

REMARQUE III-75. La convention sur le membre de droite est très naturelle : on peut montrer que Φ est minorée par une fonction affine, et il s'ensuit que $\Phi \circ f$ est minorée par une fonction intégrable; $\Phi \circ f$ est donc la somme d'une fonction sommable et d'une fonction positive.

DÉMONSTRATION DU THÉORÈME III-74. La μ -sommabilité de chaque composante de f implique celle de $|f|$. Pour tout entier $k \geq 1$, notons $A_k := \{x \in X; |f(x)| \geq k^{-1}\}$. L'inégalité de Chebyshev implique

$$\mu[A_k] \leq k \int |f| d\mu.$$

En particulier, $\mu[A_k]$ est fini. Soit μ_k la mesure de probabilité définie par

$$\mu_k[B] := \frac{\mu[A_k \cap B]}{\mu[A_k]}.$$

L'inégalité de Jensen implique

$$\Phi\left(\int f d\mu_k\right) \leq \int \Phi \circ f d\mu_k.$$

En utilisant l'homogénéité de Φ , on en déduit

$$\Phi \left(\mu[A_k] \int f d\mu_k \right) = \mu[A_k] \Phi \left(\int f d\mu_k \right) \leq \mu[A_k] \int \Phi \circ f d\mu_k.$$

En résumé,

$$(21) \quad \Phi \left(\int_{A_k} f d\mu \right) \leq \int_{A_k} \Phi \circ f d\mu.$$

En distinguant les cas $f(x) = 0$ et $f(x) \neq 0$, on voit que $f1_{A_k}$ converge partout vers f quand $k \rightarrow \infty$. Par convergence dominée,

$$\int_{A_k} f d\mu \xrightarrow{k \rightarrow \infty} \int_X f d\mu,$$

et par semi-continuité inférieure de Φ ,

$$(22) \quad \Phi \left(\int f d\mu \right) \leq \liminf_{k \rightarrow \infty} \int_{A_k} \Phi \circ f d\mu.$$

D'autre part, l'homogénéité de Φ impose $\Phi(0) = 0$, ce qui permet de montrer que la fonction $(\Phi \circ f)1_{A_k}$ converge partout vers $\Phi \circ f$. Si cette fonction est intégrable, alors on a, par convergence dominée,

$$\int_{A_k} \Phi \circ f d\mu \xrightarrow{k \rightarrow \infty} \int_X \Phi \circ f d\mu.$$

Cela conclut l'argument. \square

REMARQUE III-76. Soit φ une fonction convexe sur \mathbb{R}^{n-1} ; alors la fonction définie sur $\mathbb{R}^{n-1} \times \mathbb{R}_+$ par

$$\Phi(x, z) = z\varphi\left(\frac{x}{z}\right)$$

est convexe, comme on peut le voir en revenant à la définition de la convexité (exercice).

III-4.3. Inégalités de Young intégrées. Pour toutes fonctions f et g à valeurs réelles, définies sur un espace X quelconque, et toute fonction convexe Φ sur \mathbb{R} , on peut écrire

$$f(x)g(x) \leq \Phi(f(x)) + \Phi^*(g(x)).$$

Si X est muni d'une mesure μ et que toutes ces fonctions sont intégrables, on a alors

$$\int_X fg d\mu \leq \int_X \Phi \circ f d\mu + \int_X \Phi^* \circ g d\mu.$$

Plus généralement, si f et g sont à valeurs dans \mathbb{R}^n , on peut écrire

$$\int \langle f, g \rangle d\mu \leq \int_X \Phi \circ f d\mu + \int_X \Phi^* \circ g d\mu.$$

Ces inégalités aussi élémentaires que cruciales peuvent parfois être améliorées, en particulier quand Φ présente certaines propriétés d'**homogénéité**, comme nous allons maintenant le voir.

III-4.4. Inégalité de Hölder. Très souvent, quand on majore des produits de fonctions, on cherche à utiliser des intégrabilités différentes pour les deux facteurs. Un exemple évident est l'inégalité utile

$$|f| \leq C \text{ } \mu\text{-presque partout} \implies \left| \int fg \, d\mu \right| \leq C \int |g| \, d\mu.$$

L'inégalité de Hölder affine cette inégalité grâce à l'usage de fonctions puissances.

THÉOREME III-77 (inégalité de Hölder). Soit (X, μ) un espace mesuré, soit $p \in]1, +\infty[$, et soit $p' = p/(p-1)$ l'exposant conjugué de p . Soient f et g deux fonctions mesurables sur X , à valeurs dans $\overline{\mathbb{R}}$. Alors

$$\left| \int fg \, d\mu \right| \leq \left(\int |f|^p \, d\mu \right)^{1/p} \left(\int |g|^{p'} \, d\mu \right)^{1/p'},$$

où l'on convient que le membre de gauche vaut $+\infty$ si fg n'est pas intégrable. De plus, si le membre de droite de l'inégalité est fini et non nul, il y a égalité si et seulement si il existe $\alpha > 0$, $\varepsilon \in \{\pm 1\}$ tel que, μ -presque partout, f et εg ont même signe et $|f|^p = \alpha |g|^{p'}$, ce qui revient à $g = |f|^{p-2} f$.

Les mêmes conclusions valent si f et g sont à valeurs dans \mathbb{R}^n , quitte à remplacer le produit fg par le produit scalaire $\langle f, g \rangle$, et à interpréter $|f|$ et $|g|$ comme les normes euclidiennes de f et g .

REMARQUES III-78. (i) Dans le cas où $p = p' = 2$, l'inégalité de Hölder est appelée **inégalité de Cauchy–Schwarz**; on peut alors la démontrer par un argument abstrait, valable pour n'importe quel espace de Hilbert.

(ii) Nous rencontrerons au chapitre VI une autre inégalité importante faisant intervenir des fonctions puissances et de la convexité, appelée **inégalité de Minkowski**.

(iii) L'inégalité de Hölder reste vraie pour $p = 1$ ou $p = \infty$, si l'on convient de poser

$$\left(\int |g|^\infty \, d\mu \right)^{1/\infty} := \inf \left\{ C \in \mathbb{R}; |g| \leq C \text{ } \mu\text{-presque partout} \right\}.$$

Cette dernière quantité est appelée le **supremum essentiel** de $|g|$; il s'agit de la définition habituelle du supremum, à laquelle on a ajoutée les mots " μ -presque partout".

(iv) Pour $0 < p < 1$, l'inégalité reste vraie avec f, g positives, et $q := p/(1-p)$... au signe près! Il faut en effet remplacer " \leq " par " \geq ".

DÉMONSTRATION DE L'INÉGALITÉ DE HÖLDER. Si l'une des intégrales du membre de droite est nulle, alors, μ -presque partout, $fg = 0$, et l'inégalité est satisfaite. Si l'une de ces intégrales est infinie et l'autre non nulle, alors l'inégalité est bien sûr satisfaite. Supposons donc que les deux intégrales sont strictement positives et finies. On pose $f' = f/(\int |f|^p)^{1/p}$, $g' = g/(\int |g|^{p'})^{1/p'}$, on a alors $\int |f'|^p = \int |g'|^{p'} = 1$ et on doit prouver

$$\left| \int f'g' \, d\mu \right| \leq 1.$$

On sait déjà que

$$\left| \int f'g' \, d\mu \right| \leq \int |f'g'| \, d\mu,$$

et si le membre de droite est fini, l'égalité n'est possible que si, μ -presque partout, $f'g' = |f'g'|$, c'est-à-dire si f et g ont même signe. Pour récapituler, nous nous sommes ramenés, par cet argument d'homogénéité, au cas particulier suivant : montrer que, si f et g sont deux fonctions positives,

$$\int f^p = \int g^{p'} = 1 \implies \int fg \leq 1,$$

avec égalité si et seulement si $f = g$ presque partout. On écrit alors l'inégalité du Lemme III-118 avec $a := f(x)$, $b := g(x)$, et on intègre par rapport à μ : on trouve

$$\int fg \, d\mu \leq \frac{1}{p} + \frac{1}{p'} = 1;$$

avec égalité si et seulement si $f = g$ presque partout. La preuve est complète. \square

L'inégalité de Hölder admet plusieurs avatars simples et intéressants.

THÉORÈME III-79 (variantes de l'inégalité de Hölder). *(i) Soit (X, μ) un espace mesuré, soit $p \in]1, +\infty[$ et soit $p' = p/(p-1)$ son exposant conjugué. Soient f et g deux fonctions mesurables sur X , à valeurs dans $\overline{\mathbb{R}}$. Alors, pour tout $\lambda > 0$,*

$$\left| \int fg \, d\mu \right| \leq \frac{\lambda^p}{p} \int |f|^p \, d\mu + \frac{1}{\lambda^q} \int |g|^q \, d\mu,$$

où l'on convient que le membre de gauche vaut $+\infty$ si fg n'est pas intégrable.

(ii) Soit (X, μ) un espace mesuré, soient $p_1, \dots, p_k \in]1, +\infty[$ tels que

$$\sum_{i=1}^k \frac{1}{p_i} = 1,$$

et soient f_1, \dots, f_k des fonctions mesurables sur X , à valeurs dans $\overline{\mathbb{R}}$. Alors

$$\left| \int \left(\prod_i f_i \right) d\mu \right| \leq \prod_i \left(\int |f_i|^{p_i} \, d\mu \right)^{1/p_i},$$

où l'on convient que le membre de gauche vaut $+\infty$ si $\prod f_i$ n'est pas intégrable.

(iii) Soient (X, μ) un espace mesuré σ -fini et (Y, π) un espace de probabilité. Alors, pour toute fonction F mesurable de $X \times Y$ dans $\mathbb{R}_+ \cup \{+\infty\}$, on a

$$\int_X \exp \left(\int_Y \log F(x, y) \, d\pi(y) \right) d\mu(x) \leq \exp \left(\int_Y \log \left(\int_X F(x, y) \, d\mu(x) \right) d\pi(y) \right).$$

(iv) Soient X et Y deux ensembles quelconques, et L un opérateur linéaire, défini sur un sous-espace vectoriel de l'ensemble des fonctions de X dans \mathbb{R} , à valeurs dans l'ensemble des fonctions de Y dans \mathbb{R} . On suppose que L est positif, i.e. $Lf \geq 0$ si $f \geq 0$. Soient $f, g \geq 0$ dans le domaine de L , soit $p \in]1, +\infty[$ et $p' = p/(p-1)$ son exposant conjugué. Alors

$$L(fg) \leq [L(f^p)]^{1/p} [L(g^{p'})]^{1/p'},$$

ce qui est une inégalité entre deux fonctions de Y dans \mathbb{R} .

(v) Soient (X, μ) un espace mesuré, E un espace vectoriel normé, et E^* l'espace des formes linéaires continues sur E , muni de sa norme naturelle. Soient $f : X \rightarrow E^*$ et $g : X \rightarrow E$ des fonctions mesurables, soit $p \in]1, +\infty[$ et $q := p/(p-1)$. Alors

$$\left| \int \langle f, g \rangle_{E^* \times E} d\mu \right| \leq \left(\int \|f\|_{E^*}^p d\mu \right)^{1/p} \left(\int \|g\|_E^{p'} d\mu \right)^{1/p'},$$

où l'on convient que le membre de gauche vaut $+\infty$ si $\langle f, g \rangle$ n'est pas intégrable.

(vi) Soit (X, μ) un espace mesuré, $f, g : X \rightarrow \mathbb{C}$ deux fonctions mesurables à valeurs complexes, $p \in]1, +\infty[$ et $q := p/(p-1)$. Alors

$$\left| \int fg d\mu \right| \leq \left(\int |f|^p d\mu \right)^{1/p} \left(\int |g|^{p'} d\mu \right)^{1/p'},$$

où l'on convient que le membre de gauche vaut $+\infty$ si fg n'est pas intégrable. De plus, si le membre de droite de l'inégalité est fini et non nul, il y a égalité si et seulement si il existe $\alpha > 0$ et $\theta \in \mathbb{R}$ tels que, μ -presque partout, $fg \in e^{i\theta}\mathbb{R}$ et $|f|^p = \alpha|g|^q$.

DÉMONSTRATION. (i) Il suffit de remarquer que

$$\inf_{\lambda \in [0,1]} \left\{ \frac{\lambda^p}{p} \int |f|^p d\mu + \frac{1}{\lambda^{p'} p'} \int |g|^{p'} d\mu \right\} = \left(\int |f|^p d\mu \right)^{1/p} \left(\int |g|^{p'} d\mu \right)^{1/p'}.$$

Cette façon de procéder fournit d'ailleurs la base d'une autre démonstration de l'inégalité de Hölder : on commence par appliquer l'inégalité du Lemme III-118 avec $a := \lambda f(x)$, $b := g(x)/\lambda$, où $\lambda > 0$ est arbitraire. On intègre l'inégalité obtenue par rapport à μ , puis on *optimise* par rapport au paramètre λ .

(ii) Sans perte de généralité, on peut supposer que toutes les fonctions f_i sont positives ; l'inégalité à démontrer s'obtient alors à partir de l'inégalité de Hölder par récurrence.

(iii) Par homogénéité, on peut supposer que

$$\forall y \in Y, \quad \int_X F(x, y) d\mu(x) = 1,$$

auquel cas l'inégalité à établir est

$$\int_X \exp \left(\int_Y \log F(x, y) d\pi(y) \right) d\mu(x) \leq 1.$$

C'est alors une conséquence immédiate de l'inégalité de Jensen pour la fonction convexe $-\log$ et pour la mesure de probabilité π , combinée avec le théorème de Fubini.

Les énoncés (iv), (v) et (vi) se démontrent sans difficulté en adaptant la preuve de l'inégalité de Hölder ou en s'y ramenant, par exemple en écrivant que $\langle f, g \rangle_{E^* \times E} \leq \|f\|_{E^*} \|g\|_E$. \square

REMARQUE III-80. Pour comprendre en quoi l'énoncé (iii) est une variante de l'inégalité de Hölder dans le cas où les fonctions f et g sont strictement positives, il suffit de poser $Y := \{0, 1\}$, $\pi := (1/p)\delta_0 + (1/q)\delta_1$, $F(x, 0) := f(x)^p$, $F(x, 1) := g(x)^q$. En fait, on peut facilement se convaincre que l'énoncé (iii) est une "limite continue" de l'énoncé (ii). L'énoncé (iii) n'est pas très utile en pratique, son principal intérêt pour nous est de mettre en évidence un lien étroit entre les inégalités de Hölder et

de Jensen. L'énoncé (iv) quant à lui a le mérite de montrer que l'inégalité de Hölder est vraie dans un cadre beaucoup plus général que celui de l'intégration ; noter que l'inégalité de Hölder correspond au cas où Y est réduit à un point.

III-4.5. Inégalités entropiques. L'inégalité de Hölder fait intervenir des puissances des fonctions en jeu, ce qui est largement suffisant pour la majorité des problèmes. Cependant, il arrive que l'on considère, par commodité ou par nécessité, d'autres fonctions. C'est le cas en particulier quand on étudie l'**entropie** d'une fonction positive :

DÉFINITION III-81 (entropie). Soient (X, μ) un espace mesuré et $f : X \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ une fonction positive. On appelle entropie de f par rapport à μ la quantité

$$S_\mu(f) := - \int_X f \log f \, d\mu.$$

On appelle information de Kullback de $f\mu$ par rapport à μ la quantité

$$H(f\mu|\mu) := \int_X (f \log f - f + 1) \, d\mu.$$

REMARQUE III-82. La fonction $f \mapsto f \log f - f + 1$ est positive, et donc l'information de Kullback toujours positive, alors que l'entropie peut prendre n'importe quel signe.

En physique statistique, si f est une densité de probabilité par rapport à la mesure $\mu := \mathcal{L}$, mesure de Lebesgue sur \mathbb{R}^n , on appelle $S_\mu(f)$ l'entropie de Boltzmann ; en théorie de l'information, on appelle $S_\mu(f)$ l'entropie de Shannon. Par ailleurs, l'information de Kullback coïncide avec l'opposé de l'entropie dès que $\int f \, d\mu = \int d\mu$, ce qui est très souvent le cas. Dans chacun de ces domaines, l'entropie joue un rôle fondamental. L'inégalité suivante remplace alors l'inégalité de Hölder :

THÉORÈME III-83 (inégalité de convexité pour l'entropie). Soit (X, μ) un espace de probabilité, et soient $f, g : X \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ deux fonctions mesurables positives. Alors,

$$\int f g \, d\mu \leq \int (f \log f - f + 1) \, d\mu + \log \int e^g \, d\mu.$$

DÉMONSTRATION DU THÉORÈME III-83. Par homogénéité, on peut se ramener au cas où $\int e^g \, d\mu = 1$, et il s'agit alors de prouver que

$$\int f g \, d\mu \leq \int (f \log f - f + 1) \, d\mu.$$

Pour cela on écrit l'inégalité de Young logarithmique ci-dessus avec $a := f(x)$, $b := g(x)$, et on l'intègre contre μ . Il vient

$$\int f g \, d\mu \leq \int (f \log f - f + 1) \, d\mu + \int (e^g - 1) \, d\mu,$$

et la dernière intégrale s'annule car $\int e^g = 1$ et μ est une mesure de probabilité. \square

Une autre question naturelle que l'on peut se poser est la façon dont l'entropie se compare aux fonctions définies par des puissances. On ne peut bien sûr espérer contrôler par l'entropie aucune puissance de f strictement supérieure à 1. L'inégalité suivante, dite de **Csiszár-Kullback-Pinsker**, que nous abrègerons en CKP, répond de manière assez précise à cette question.

THÉORÈME III-84 (inégalité CKP). Soit (X, μ) un espace de probabilité, et soit f une densité de probabilité, i.e. $f\mu$ est une mesure de probabilité sur X . Alors

$$\int_X |f - 1| d\mu \leq \sqrt{2 \int_X (f \log f - f + 1) d\mu}.$$

DÉMONSTRATION. Par formule de Taylor avec reste intégrale,

$$f \log f - f + 1 = (f - 1)^2 \int_0^1 \frac{(1 - t) dt}{1 + t(f - 1)}.$$

La mesure μ étant finie, on est en droit d'appliquer le théorème de Fubini, d'où

$$\int (f \log f - f + 1) d\mu = \int_0^1 (1 - t) \left[\int_X \frac{(f - 1)^2}{1 + t(f - 1)} d\mu \right] dt.$$

Par inégalité de Cauchy-Schwarz, pour tout $t \in [0, 1]$,

$$\left(\int |f - 1| d\mu \right)^2 \leq \left(\int_X \frac{(f - 1)^2}{1 + t(f - 1)} d\mu \right) \left(\int (1 + t(f - 1)) d\mu \right) = \int_X \frac{(f - 1)^2}{1 + t(f - 1)} d\mu,$$

puisque μ et $f\mu$ sont toutes deux des mesures de probabilité. On conclut que

$$\left(\int_0^1 (1 - t) dt \right) \left(\int |f - 1| d\mu \right)^2 \leq \int (f \log f - f + 1) d\mu,$$

ce qui est équivalent à la conclusion souhaitée. \square

III-5. Équi-intégrabilité et tension

Cette section pourra être omise en première lecture.

III-5.1. Équi-intégrabilité. On dit qu'un ensemble \mathcal{F} de fonctions définies sur un espace métrique est *équicontinu* s'il admet un module de continuité uniforme : pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour tous x, y distants d'au plus δ , les images $f(x)$ et $f(y)$ soient distantes d'au plus ε , et ce pour tout $f \in \mathcal{F}$ (δ est donc indépendant de f). Ce concept joue un rôle important, par exemple dans l'étude de la compacité dans des espaces de fonctions continues. Bien évidemment, si \mathcal{F} est équicontinu, alors tout $f \in \mathcal{F}$ est uniformément continu ; et la réciproque est fausse.

En théorie de la mesure, un concept analogue est l'équi-intégrabilité :

DÉFINITION III-85. Soit (X, μ) un espace mesuré.

(i) On dit qu'un ensemble \mathcal{F} de fonctions $f : X \rightarrow \mathbb{R}$ est *équi-intégrable* si pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que

$$(23) \quad \mu[A] \leq \delta \implies \forall f \in \mathcal{F}, \quad \int_A |f| d\mu \leq \varepsilon.$$

(ii) On dit que \mathcal{F} est *équi-intégrable à l'infini* si pour tout $\varepsilon > 0$ il existe une partie $Y_\varepsilon \subset X$, de mesure finie, telle que

$$(24) \quad \forall f \in \mathcal{F}, \quad \int_{X \setminus Y_\varepsilon} |f| d\mu \leq \varepsilon.$$

L'équi-intégrabilité se prouve le plus souvent grâce au critère suivant :

PROPOSITION III-86. Soient (X, μ) un espace mesuré, et \mathcal{F} un ensemble de fonctions mesurables de X dans \mathbb{R} , tel que

$$\sup_{f \in \mathcal{F}} \int |f| d\mu < +\infty.$$

Alors

(i) \mathcal{F} est équi-intégrable si et seulement si il existe une fonction $\Phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, telle que

$$(25) \quad \lim_{r \rightarrow \infty} \frac{\Phi(r)}{r} = +\infty \quad \text{et} \quad \sup_{f \in \mathcal{F}} \int_X \Phi(|f|) d\mu < +\infty.$$

En outre, sans perte de généralité, on peut choisir la fonction Φ convexe et régulière.

(ii) \mathcal{F} est équi-intégrable à l'infini s'il existe une fonction $\varphi : X \rightarrow \mathbb{R}_+$ telle que

$$\forall r > 0, \quad \mu[\{\varphi \leq r\}] < +\infty; \quad \sup_{f \in \mathcal{F}} \int_X |f| \varphi d\mu < +\infty.$$

DÉMONSTRATION. Commençons par la propriété (i). Supposons (25) réalisé, et soit $\varepsilon > 0$. On pose $I := \sup_{f \in \mathcal{F}} \int \Phi(|f|) d\mu$, et on choisit R assez grand pour que

$$r \geq R \implies \frac{\Phi(r)}{r} \geq \frac{C}{2\varepsilon}.$$

On pose ensuite $\delta := \varepsilon/(2R)$. Alors, pour tout $f \in \mathcal{F}$,

$$\int_{|f| \leq R} |f| d\mu \leq \frac{\varepsilon}{2C} \int_{|f| \geq R} \int_X \Phi(|f|) d\mu \leq \frac{\varepsilon}{2}.$$

Donc, pour tout ensemble A de mesure $\mu[A] \leq \delta$ et pour tout $f \in \mathcal{F}$,

$$\int_A |f| d\mu \leq \frac{\varepsilon}{2} + \int_{A \cap \{|f| \leq R\}} |f| d\mu \leq \frac{\varepsilon}{2} + R\mu[A] \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2};$$

et \mathcal{F} est bien équi-intégrable. (Cette implication n'utilise pas la borne sur les intégrales $\int |f| d\mu$.)

Réciproquement, supposons que \mathcal{F} est équi-intégrable, et $\int |f| d\mu \leq C$. Par l'inégalité de Chebyshev, pour tout $N \geq 1$ on a

$$\mu[\{|f| \geq N\}] \leq \frac{\int |f| d\mu}{N} \leq \frac{C}{N}.$$

Pour $\varepsilon > 0$, soit δ comme dans (23), et N tel que $C/N \leq \delta$. Alors on a, pour tout $f \in \mathcal{F}$,

$$\int_{|f| \geq N} |f| d\mu \leq \varepsilon.$$

En conséquence,

$$\forall \varepsilon > 0, \quad \exists N(\varepsilon) \geq 1; \quad \forall f \in \mathcal{F}, \quad \int_{|f| \geq N(\varepsilon)} |f| \leq \varepsilon.$$

On pose $N_0 = 0$ et on construit par récurrence une suite d'entiers $(N_k)_{k \geq 1}$ tels que pour tout $k \geq 1$,

$$N_k > 2N_{k-1}, \quad \text{et} \quad \sup_{f \in \mathcal{F}} \int_{|f| \geq N_k} \leq 2^{-k}.$$

La suite $(N_k)_{k \in \mathbb{N}}$ tend bien sûr vers l'infini, et pour tout $x \geq 0$ il existe un nombre fini d'indices k tels que $N_k \leq x$. On définit

$$\Phi(x) := x \sum_{N_k \leq x} k.$$

Il est clair que $\Phi(x)/x \rightarrow +\infty$ quand $x \rightarrow \infty$. D'autre part, pour tout $f \in \mathcal{F}$ on a

$$\int_X \Phi(|f|) d\mu = \int_X |f| \sum_{|f| \geq N_k} k d\mu = \sum_{k \in \mathbb{N}} k \int_X |f| 1_{|f| \geq N_k} d\mu \leq \sum_{k \in \mathbb{N}} k 2^{-k} < +\infty.$$

En particulier, les quantités $\int \Phi(|f|) d\mu$ sont bien majorées uniformément pour $f \in \mathcal{F}$.

La fonction Φ ainsi construite est affine par morceaux, et discontinue. On définit Φ_c comme son "enveloppe affine continue", i.e. la plus grande fonction affine par morceaux et continue qui minore f , obtenue en joignant les points $(N_k, \Phi(N_k-))$. Sur l'intervalle $[N_k, N_{k+1}]$ cette fonction varie d'une quantité $kN_{k+1} - (k-1)N_k$, sa pente est donc

$$p_k = \frac{kN_{k+1} - (k-1)N_k}{N_{k+1} - N_k} = k + \frac{N_k}{N_{k+1} - N_k}.$$

Par construction, $N_{k+1} > 2N_k$, donc $N_k/(N_{k+1} - N_k) < 1$. On a donc $k \leq p_k < k+1$, ce qui montre que la suite p_k est strictement croissante, et la fonction Φ_c est donc convexe.

Pour conclure, il suffit de vérifier qu'on peut trouver une fonction convexe positive Ψ , de classe C^∞ , telle que $\Phi_c - 1 \leq \Psi \leq \Phi$. C'est un simple exercice qui est laissé au lecteur.

On passe ensuite à la partie (ii), qui est plus simple. Cette partie n'utilise pas non plus la borne uniforme sur les $\int |f| d\mu$. Soit \mathcal{F} un ensemble de fonctions vérifiant la condition indiquée, on pose $C = \sup\{\int |f| \varphi d\mu; f \in \mathcal{F}\}$. Soient $r = C/\varepsilon$ et $A_r := \{x; \varphi > r\}$. Pour tout $f \in \mathcal{F}$, on applique l'inégalité de Chebyshev à la fonction φ et à la mesure $|f|\mu$:

$$\int_{A_r} |f| d\mu \leq \frac{C}{r} \leq \varepsilon.$$

Puisque A_r est de mesure finie, l'ensemble \mathcal{F} est bien équi-intégrable à l'infini.

Réciproquement, supposons que \mathcal{F} est équi-intégrable à l'infini. Par récurrence, on peut construire une suite croissante d'ensembles Y_k , de mesure finie, tels que $\int_{X \setminus Y_k} |f| d\mu \leq 2^{-k}$ pour tout $f \in \mathcal{F}$. Posons $Y = \cup Y_k$. Si $f \in \mathcal{F}$ est fixé, on a par convergence monotone

$$\int_{X \setminus Y} |f| d\mu = \lim_{k \rightarrow \infty} \int_{X \setminus Y_k} |f| d\mu = 0.$$

Il s'ensuit que f est nul presque partout en-dehors de Y .

On pose

$$\varphi = \sum_{k \in \mathbb{N}} k 1_{Y_k}.$$

En outre on définit $\varphi(x) = +\infty$ sur $X \setminus Y$. Si $\varphi(x) \leq k$, alors x n'appartient à aucun des Y_j pour $j > k$, et n'appartient pas non plus à $X \setminus Y$; x appartient donc à Y_k , qui

est un ensemble de mesure finie. Autrement dit, l'ensemble $\{\varphi < k\}$ est de mesure finie. D'autre part, pour tout $f \in \mathcal{F}$,

$$\int_X \varphi f d\mu = \int_Y \varphi f d\mu = \sum_{k \in \mathbb{N}} k \int_{A_k} |f| d\mu \leq \sum_k \frac{k}{2^k} < +\infty.$$

□

On sait bien qu'une fonction continue sur un espace métrique compact X est automatiquement uniformément continue, et que donc un singleton dans $C(X)$ est équicontinu (ce qui revient à dire qu'un singleton est compact!) Un théorème analogue est valable dans le cadre de l'équi-intégrabilité :

PROPOSITION III-87. *Soit f une fonction sommable dans un espace mesuré (X, μ) . Alors f est uniformément intégrable, au sens où l'ensemble $\{f\}$ est équi-intégrable, et équi-intégrable à l'infini.*

DÉMONSTRATION. Soit $f_M := \max(-M, \min(f, M))$ (on tronque f aux hauteurs $-M$ et M). Par convergence dominée,

$$\eta(M) := \int_{|f| \geq M} |f_M - f| d\mu \xrightarrow{M \rightarrow \infty} 0.$$

Il s'ensuit que, pour toute partie mesurable A avec $\mu[A] \leq \delta$,

$$\int_A |f| d\mu \leq \int_A |f_M| d\mu + \varepsilon(M) \leq M\delta + \eta(M).$$

Si $\varepsilon > 0$ est donné, on choisit donc M assez grand pour que $\eta(M) \leq \varepsilon/2$, et $\delta = \varepsilon/(2M)$. Ceci prouve l'équi-intégrabilité.

Pour obtenir l'équi-intégrabilité à l'infini, si $\varepsilon > 0$ est donné, on pose $A_k = \{x \in X; |f(x)| \geq k^{-1}\}$. L'inégalité de Chebyshev entraîne que A_k est de mesure finie pour tout k . D'autre part, la fonction $|f|1_{|f| < k^{-1}}$ converge vers 0 partout, et elle est majorée par la fonction intégrable $|f|$. Donc, par convergence dominée,

$$\int_{X \setminus A_k} |f| d\mu = \int 1_{|f| < k^{-1}} |f| d\mu \xrightarrow{k \rightarrow \infty} 0.$$

On peut donc trouver k assez grand pour que cette quantité soit majorée par ε . □

REMARQUE III-88. Ce résultat, de démonstration simple, est conceptuellement subtil!! Il entraîne que si une fonction f est intégrable, alors il existe une fonction Φ positive, avec $\Phi(r)/r \rightarrow \infty$, telle que

$$\int_X \Phi(|f|) d\mu < +\infty.$$

En un certain sens, “si une fonction est intégrable, alors elle est un petit mieux qu'intégrable”...!

REMARQUE III-89. Nous verrons au Chapitre VII que l'équi-intégrabilité est associée à un critère de compacité, ce qui est en accord avec le fait que cette propriété soit automatiquement vérifiée par les singletons, ou plus généralement par les ensembles finis.

III-5.2. Tension. La tension est l'analogue naturel de la propriété d'équi-intégrabilité quand on parle de familles de mesures sur une σ -algèbre donnée, et non plus de familles de fonctions intégrables sur un espace mesuré. Elle s'exprime en termes d'ensembles compacts et non en termes d'ensembles de mesure finie.

DÉFINITION III-90. Soient X un espace topologique, muni de sa tribu borélienne, et \mathcal{M} un ensemble de mesures de Borel sur X . On dit que \mathcal{M} est tendu si, pour tout $\varepsilon > 0$ on peut trouver un compact K_ε dans X tel que

$$\sup_{\mu \in \mathcal{M}} \mu[X \setminus K_\varepsilon] \leq \varepsilon.$$

Il est clair que cette notion n'est pas complètement sans rapport avec l'intégrabilité à l'infini : dès que les compacts sont de mesure finie, la tension d'une famille de mesures de la forme $f\nu$, où \mathcal{F} est une famille de fonctions positives ν -intégrables, implique son équi-intégrabilité à l'infini. Il y a une formulation équivalente de la tension, très semblable à celle de l'équi-intégrabilité à l'infini :

PROPOSITION III-91. Soient X un espace métrique, muni de sa tribu borélienne, et \mathcal{M} un ensemble de mesures de Borel sur X . Alors \mathcal{M} est tendu si et seulement si il existe une fonction $\varphi : X \rightarrow \mathbb{R}_+$, tendant vers l'infini à l'infini, au sens où pour tout $r > 0$ il existe un compact K_r tel que $x \notin K_r \implies \varphi(x) \geq r$, et telle que

$$\sup_{\mu \in \mathcal{M}} \int_X \varphi d\mu < +\infty.$$

En outre, si X est localement compact, on peut sans perte de généralité choisir la fonction φ continue.

DÉMONSTRATION. Supposons l'existence de φ tendant vers l'infini à l'infini, et telle que pour tout $\mu \in \mathcal{M}$, $\int \varphi d\mu \leq C$. On pose $r = C/\varepsilon$: par inégalité de Chebyshev,

$$\mu[\{\varphi \geq r\}] \leq C/r = \varepsilon.$$

Soit K_r comme dans l'énoncé ; on a alors $X \setminus K_r \subset \{\varphi \geq r\}$, d'où $\mu[K_r] \leq \varepsilon$.

Réciproquement, soit \mathcal{M} un ensemble tendu. Par récurrence, on peut construire une suite croissante de compacts K_n tels que pour tout $\mu \in \mathcal{M}$,

$$\mu[K_n] \leq 2^{-n}.$$

On pose alors $\varphi = \sum n 1_{X \setminus K_n}$, et on effectue un raisonnement similaire à celui de la démonstration de la Proposition III-86(ii).

La dernière assertion (on peut choisir φ continue si X est localement compact) est laissée en exercice. \square

III-6. Produits infinis

Cette section pourra être omise en première lecture. La théorie de la mesure dans des produits infinis a une importance considérable en théorie des probabilités. Avec le matériel de la section III-2, nous sommes en mesure de démontrer les principaux résultats de ce sujet.

III-6.1. Tribu et topologie d'un produit infini. Commençons par rappeler la définition de la tribu produit dans le cas d'un produit infini (pas forcément dénombrable) d'espaces mesurés.

DÉFINITION III-92 (cylindre). *Soit T un ensemble arbitraire et $(X_t)_{t \in T}$ une famille d'ensembles indexés par le paramètre T ; on pose $X = \prod X_t$. Pour toute partie finie $F \subset T$, on pose $X_F = \prod_{t \in F} X_t$. Pour tout sous-ensemble A_F de X_F , on définit le cylindre $C_F(A_F)$, noté abusivement $C(A_F)$, par*

$$C(A_F) = \{(x_t)_{t \in T}; (x_t)_{t \in F} \in A_F\}.$$

REMARQUES III-93. Comme dans le langage courant, un cylindre n'est pas forcément un pavé. D'autre part, le concept n'a d'intérêt que pour un ensemble d'indices infini : si T est fini, alors tout ensemble mesurable est un cylindre.

DÉFINITION III-94 (tribu produit infini). *Soit T un ensemble arbitraire et $(X_t, \mathcal{A}_t)_{t \in T}$ une famille d'ensembles mesurables, indexés par le paramètre T ; on pose $X = \prod X_t$. Pour toute partie finie $F \subset T$, on pose $X_F = \prod_{t \in F} X_t$, que l'on munit de la tribu produit. Pour toute partie mesurable A_F de X_F , on appelle $C(A_F)$ le cylindre mesurable de base A_F . Si A_F est de la forme $\prod_{t \in F} A_t$, avec $A_t \in \mathcal{A}_t$, on dit que $C(A_F)$ est un cylindre mesurable produit.*

On définit alors la tribu produit sur X comme la tribu engendrée par les cylindres mesurables, ou de manière équivalente comme la tribu engendrée par les cylindres mesurables produits.

Cette définition est formellement analogue à celle de la topologie produit.

DÉFINITION III-95 (topologie produit infini). *Soit T un ensemble arbitraire et $(X_t, \mathcal{A}_t)_{t \in T}$ une famille d'ensembles mesurables, indexés par le paramètre T ; on pose $X = \prod X_t$. Pour toute partie finie $F \subset T$, on pose $X_F = \prod_{t \in F} X_t$, que l'on munit de la topologie produit. Pour tout ouvert O_F de X_F , on appelle $C(O_F)$ le cylindre ouvert de base O_F . Si O_F est de la forme $\prod_{t \in F} O_t$, où chaque O_t est un ouvert de X_t , on dit que $C(O_F)$ est un cylindre ouvert produit.*

On définit alors la topologie produit sur X comme la topologie engendrée par les cylindres ouverts, ou de manière équivalente comme la topologie engendrée par les cylindres ouverts produits.

REMARQUES III-96. (i) Soit $C(A_F)$ un cylindre produit ; alors c'est l'intersection des cylindres $C(A_t)$ pour $t \in F$. Aussi bien les tribus que les topologies étant stables par intersection finie, on pourrait donc, dans les définitions précédentes, se limiter à des familles F ne contenant qu'un élément.

(ii) Soit $S \subset T$; pour tout $s \in S$ on se donne A_s une partie mesurable (resp. ouverte) de X_s . En général, on ne peut rien dire de $A_S = \prod_{s \in S} A_s$. Si la famille S est **dénombrable** et les A_s sont mesurables, alors A_S est mesurable. Si la famille S est **finie** et les A_s sont ouverts, alors A_S est ouvert.

(iii) Si $t \in T$, on peut définir un opérateur de projection π_t qui à $x \in X$ associe x_t ; et si F est une partie finie de T , on peut définir un opérateur de projection π_F , qui à toute partie de $(X_t)_{t \in T}$ associe par restriction une partie de $(X_t)_{t \in F}$. La tribu produit est alors la plus petite tribu qui rende mesurables toutes les applications π_F , ou, de manière équivalente, la plus petite tribu qui rende mesurable toutes les applications π_t . De même, la topologie produit est la plus grossière topologie qui rende continues toutes ces applications.

La tribu produit sera l'objet de la suite de ce chapitre.

III-6.2. Produit infini de mesures. Dans le Chapitre I, nous avons prouvé le Théorème I-75 uniquement dans le cas particulier où les espaces considérés étaient de cardinal fini. On va maintenant proposer une démonstration dans le cas général. On pourra consulter [Dudley, p. 257-259] pour une variante.

THÉORÈME III-97 (produit dénombrable de mesures). *Soient $(X_k, \mu_k)_{k \in \mathbb{N}}$ une famille d'espaces mesurés, tels que*

$$\prod_{k \geq 1} \mu_k[X_k] < +\infty.$$

Alors il existe une unique mesure μ^∞ sur $X := \prod X_k$ telle que pour tout $n \geq 1$, et pour toutes parties mesurables A_i de X_i , $1 \leq i \leq n$,

$$\mu^\infty[A_1 \times \dots \times A_n \times \prod_{i \geq n+1} X_i] = \mu_1[A_1] \times \dots \times \mu_n[A_n] \times \prod_{i \geq n+1} \mu_i[A_i].$$

On a alors, pour toutes parties mesurables A_k de X_k ,

$$\mu^\infty[\prod A_k] = \prod \mu_k[A_k].$$

EXEMPLE III-98. Soit $\gamma(dx) = (2\pi)^{-1/2} e^{-x^2/2} dx$ la mesure gaussienne standard sur \mathbb{R} . On peut définir la mesure gaussienne standard γ_n sur \mathbb{R}^n par $\gamma_n = \gamma^{\otimes n}$, mais aussi la mesure gaussienne standard $\gamma_\infty = \gamma^{\otimes \infty}$ sur \mathbb{R}^∞ (que l'on peut identifier à ℓ^2).

DÉMONSTRATION DU THÉORÈME III-97. Si A_1, \dots, A_n sont des parties mesurables de X_1, \dots, X_n respectivement. On note \mathcal{C} l'ensemble des cylindres produits :

$$\mathcal{C}(A_1 \times \dots \times A_n) = \{x \in X; \forall i \leq n, x_i \in A_i\}.$$

Il est facile de vérifier que \mathcal{C} engendre la tribu produit, est stable par intersection finie, et que le complémentaire de tout élément de \mathcal{C} est une réunion finie disjointe d'éléments de \mathcal{C} . Comme $\mu[X] < +\infty$, le Théorème I-71(i) garantit l'unicité du prolongement éventuel. Pour prouver l'existence de ce prolongement, en vertu du Théorème I-71(iii) il suffit de vérifier la σ -additivité de μ sur \mathcal{C} .

Soient donc C un cylindre produit, et $(C_k)_{k \in \mathbb{N}}$ une famille de cylindres produits ; on suppose que les C_k sont disjoints et d'union égale à C , et on cherche à montrer que $\mu^\infty[C] = \sum_k \mu^\infty[C_k]$. Quitte à restreindre chaque mesure μ_k à la k -ème composante de C , on peut supposer que le cylindre C est l'espace tout entier ; on fera cette hypothèse dans la suite.

Supposons par l'absurde que

$$(26) \quad \sum \mu^\infty[C_k] < \mu[X].$$

Sans perte de généralité, on supposera tous les C_k non vides.

Un cylindre produit (ou pavé) C_k étant donné, on définit son *ordre* comme le plus petit entier K tel que C_k s'écrive sous la forme $C(A_1, \dots, A_K)$; et sa *composante* d'ordre ℓ comme A_ℓ . On vérifie facilement, grâce à l'existence de la mesure produit $\mu_1 \otimes \dots \otimes \mu_K$, que pour tout K , μ^∞ est σ -additive sur l'ensemble des cylindres d'ordre inférieur ou égal à K . En particulier, la relation $\sum \mu^\infty[C_k] < \mu^\infty[X]$ implique qu'il y a des C_k d'ordre arbitrairement grand. Il est également clair que la mesure de la projection d'un cylindre est au moins égale à la mesure du cylindre lui-même.

La mesure de l'union des cylindres C_k d'ordre 1 est la somme des mesures de ces cylindres, strictement inférieure à $\mu[X]$. Soit Z_1 le complémentaire de l'union des bases de ces cylindres : c'est un sous-ensemble mesurable, de mesure $\varepsilon_1 > 0$. Il est recouvert par les cylindres d'ordre 2 ou plus, et la somme des mesures de ces cylindres est strictement inférieure à $M_2\varepsilon_1$, où $M_2 = \prod_{j \geq 2} \mu_j[X_j]$. (sinon (26) serait contredit).

Chaque cylindre C_k s'écrit comme un produit infini de A_k^j pour $j \in \mathbb{N}$, $A_k^j \in \mathcal{A}_j$. Soit, pour $x_1 \in Z_1$,

$$\phi_1(x_1) := \sum_{k=1}^{\infty} \prod_{j=2}^{\infty} \mu_j[A_k^j] 1_{x_1 \in A_k^1}.$$

La fonction ϕ_1 est mesurable, et son intégrale vaut

$$\int_{Z_1} \phi_1 = \sum_{k=1}^{\infty} \prod_{j=2}^{\infty} \mu_j[A_k^j] \mu_1[A_k^1 \cap Z_1].$$

Pour tout pavé C_k d'ordre 1, $A_k^1 \cap Z_1 = \emptyset$; et pour tout pavé C_k d'ordre 2 ou plus, $A_k^1 \subset Z_1$. On en déduit que $\int_{Z_1} \phi_1$ vaut la somme des mesures des pavés d'ordre 2 ou plus. En particulier,

$$\int_{Z_1} \phi_1(x_1) d\mu_1(x_1) < \mu_1[Z_1] = M_2\varepsilon_1.$$

Il existe donc au moins un x_1 dans Z_1 tel que $\phi_1(x_1) < M_2 = \prod_{j \geq 2} \mu_j[X_j]$. On décompose chaque pavé C_k dont la première projection contient x_1 , en $A_k^1 \times C'_k$. Les C'_k recouvrent alors $X' := \prod_{j=2}^{\infty} X_j$, et, si l'on note $\mu'[C'_k] = \prod_{j=2}^{\infty} \mu_j[A_k^j]$, on trouve

$$\sum_{k=1}^{\infty} \mu'[C'_k] < \mu'[X'].$$

On note que les composantes d'ordre ℓ des C'_k sont exactement les composantes d'ordre $\ell + 1$ des C_k , et que x_1 a été construit de telle sorte qu'il n'est première composante d'aucun cylindre C_k d'ordre 1. En particulier, il est équivalent de dire que (x_1, x_2) appartiennent aux deux premières composantes d'un des cylindres C_k , ou de dire que x_1 appartient à la première composante de C_k , que C_k se décompose en $A_k^1 \times C'_k$, et que x_2 appartient à la première composante de C'_k .

On peut alors recommencer le même raisonnement avec X' et les C'_k en place de X et des C_k ... Par récurrence, on construit une suite $x = (x_1, x_2, \dots)$ telle que pour tout K , (x_1, \dots, x_K) ne sont les K premières composantes d'aucun cylindre d'ordre K . Comme x appartient nécessairement à l'un des cylindres C_k , on aboutit à une contradiction. Ceci achève la preuve de l'existence de μ^∞ .

La dernière assertion de l'énoncé s'obtient sans peine par σ -additivité, si l'on note que les ensembles $A_1 \times \dots \times A_n \times \prod_{j \geq n+1} X_j$ décroissent vers $\prod A_k$ quand $n \rightarrow \infty$, et que $\mu[X] < +\infty$. \square

Pour généraliser ce résultat à un produit infini quelconque d'ensembles mesurés X_t , il est naturel d'imposer des conditions plus fortes sur les mesures des X_t ; par exemple, pour que ce produit ne soit ni 0 ni $+\infty$, il est nécessaire qu'au plus une infinité dénombrable des nombres $\mu_t[X_t]$ soit différente de 1. Il est donc naturel, dans ce cadre général, d'imposer toutes ces mesures égales à 1, autrement dit de se restreindre à des espaces de probabilités.

THÉORÈME III-99 (produit infini de mesures de probabilités). Soient T un ensemble quelconque, et $(X_t, \mathcal{A}_t, \mu_t)_{t \in T}$ une famille d'espaces mesurés de probabilités ; on pose $X := \prod X_t$, et on le munit de la tribu produit. Pour toute famille finie $F \subset T$, et pour toute famille d'ensembles mesurables $A_F = (A_t)_{t \in F}$, on définit le cylindre produit de base A_F , noté $C_F(A_F)$, ou par abus de notation $C(A_F)$, par

$$C(A_F) := \{x \in X; \forall t \in F, x_t \in A_t\}.$$

Alors il existe une unique mesure de probabilités μ sur X telle que pour toute famille finie $F \subset T$, et tout cylindre produit $C(A_F)$, on ait

$$\mu[C(A_F)] = \prod_{t \in F} \mu_t[A_t].$$

En d'autres termes, si π_F désigne la projection de X_T dans $X_F = \prod_{t \in F} X_t$, et $\mu_F = \prod_{t \in F} \mu_t$, alors pour toute famille finie $F \subset T$ on a

$$(\pi_F)_\# \mu = \mu_F.$$

En outre, si pour chaque $t \in T$ on se donne une famille \mathcal{F}_t qui engendre la tribu \mathcal{A}_t , alors μ est caractérisée par les valeurs qu'elle attribue aux cylindres produits $C(A_F)$ pour $A_F \in \prod_{t \in F} \mathcal{F}_t$.

DÉMONSTRATION. La preuve de l'unicité est facile : les cylindres dont la base est choisie parmi les produits d'éléments de \mathcal{F}_t ($t \in F$) engendrent les cylindres dont la base est choisie parmi les produits d'éléments de \mathcal{A}_t , grâce à la partie (ii) du Théorème III-55 (noter que chaque X_t est de mesure finie) ; on en déduit que ces cylindres particuliers suffisent à engendrer toute la tribu produit, et l'unicité découle du résultat d'unicité dans le Théorème I-69.

Pour démontrer l'existence, il nous suffit encore une fois de vérifier la σ -additivité sur les cylindres produits. Comme cette propriété ne concerne qu'une famille dénombrable de cylindres, dont la définition ne fait intervenir qu'une famille dénombrable d'indices t , on peut toujours, quitte à changer les notations, supposer que T est dénombrable. La σ -additivité est alors une conséquence du Théorème III-97. \square

Quand le nombre de variables d'intégration est infini, la signification même d'un énoncé du type du Théorème de Fubini n'est pas très claire ; on conserve cependant l'invariance par permutation et/ou regroupement des variables, ce que traduit l'énoncé suivant, assez abstrait. Sa preuve est une conséquence presque immédiate de l'unicité dans le Théorème III-99.

PROPOSITION III-100 (invariance de l'intégrale produit par permutation ou regroupement des variables). Soient T un ensemble arbitraire, et $(X_t, \mu_t)_{t \in T}$ une famille d'espaces de probabilités.

(i) Soient T' un ensemble en bijection avec T , et π une application bijective de T' dans T ; π induit alors par permutation des coordonnées un isomorphisme d'espaces mesurables entre $\prod_{t \in T} X_t$ et $\prod_{t' \in T'} X_{\pi(t')}$, de telle sorte que

$$\pi_\# \prod_{t' \in T'} \mu_{\pi(t')} = \prod_{t \in T} \mu_t.$$

(ii) Si $T = \prod_{s \in S} (\prod_{t \in T_s} X_t)$, alors

$$\prod_{t \in T} \mu_t = \prod_{s \in S} \left(\prod_{t \in T_s} \mu_t \right).$$

Un exemple simple d'application de la règle précédente est

$$\prod_{k \in \mathbb{N}} \mu_k = \mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_N \otimes \left(\prod_{k \geq N+1} \mu_k \right),$$

où l'on a décomposé le produit infini en $N + 1$ facteurs; on peut alors appliquer le théorème de Fubini à ces facteurs, en les permutant, etc.

III-6.3. Approximation cylindrique. Il n'est pas facile en général de se représenter les éléments de la tribu produit infini, et on cherche le plus souvent à se ramener par approximation à un nombre fini de variables.

THÉORÈME III-101 (approximation cylindrique pour la mesure produit). *Soit $(X_n)_{n \in \mathbb{N}}$ une famille d'espaces mesurés; on munit $X = \prod X_n$ de la tribu produit, et on se donne une mesure finie μ sur X . Alors, pour toute partie mesurable A de X il existe une suite $(C_n)_{n \in \mathbb{N}}$ de cylindres tels que*

$$(27) \quad \mu[C_n \setminus A] + \mu[A \setminus C_n] \xrightarrow{n \rightarrow \infty} 0.$$

DÉMONSTRATION. Soit \mathcal{C} l'ensemble des parties mesurables A de X telles qu'il existe une suite $(A_n)_{n \in \mathbb{N}}$ de cylindres satisfaisant (27). Il est clair que \mathcal{C} contient les cylindres; pour conclure il nous suffit de montrer que c'est une tribu.

Le complémentaire d'un cylindre étant un cylindre, il est évident que \mathcal{C} est stable par passage au complémentaire. De même, l'union de deux cylindres étant un cylindre, \mathcal{C} est stable par union finie. Soit maintenant $(A^k)_{k \in \mathbb{N}}$ une suite d'éléments de \mathcal{C} , et A leur union. Soit $\varepsilon > 0$, notre but est de prouver qu'il existe un cylindre C tel que

$$\mu[C \setminus A] + \mu[A \setminus C] \leq \varepsilon.$$

Puisque μ est finie, par σ -additivité il existe N tel que

$$\mu[A \setminus (\cup_{1 \leq k \leq N} A^k)] \leq \varepsilon/2.$$

D'autre part, puisque \mathcal{C} est stable par union finie, il existe un cylindre C tel que

$$\mu[C \setminus (\cup_{1 \leq k \leq N} A^k)] + \mu[(\cup_{1 \leq k \leq N} A^k) \setminus C] \leq \varepsilon/2.$$

On en déduit la conclusion souhaitée. \square

III-6.4. Lois du 0-1. Les lois du 0-1 concernent des probabilités définies sur des produits infinis, et énoncent que dans certaines circonstances, la probabilité de certains événements est forcément "triviale". Il y a deux lois du 0-1 célèbres; la plus connue est celle de Kolmogorov, l'autre est celle de Hewitt et Savage. Nous allons déduire la première de la seconde. Nous aurons besoin de quelques définitions préliminaires pour les énoncer.

DÉFINITION III-102 (permutation finie). *Soit $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ une bijection; on dit que σ est une permutation finie s'il n'y a qu'un nombre fini d'entiers j tels que $\sigma(j) \neq j$.*

DÉFINITION III-103 (ensemble symétrique). Soit X un ensemble quelconque ; une partie $A \subset X^{\mathbb{N}}$ est dite symétrique si pour toute permutation finie σ ,

$$(x_n)_{n \in \mathbb{N}} \in A \quad \Leftrightarrow \quad (x_{\sigma(n)})_{n \in \mathbb{N}} \in A.$$

DÉFINITION III-104 (tribu asymptotique). Soit $(X_n)_{n \in \mathbb{N}}$ une suite d'espaces mesurables. On définit la projection π_n , de $X := \prod X_k$ dans X_n , par restriction. On définit \mathcal{A}_m comme la plus petite tribu qui rende mesurables toutes les applications π_n pour $n \geq m$. On définit la tribu asymptotique, \mathcal{A}_∞ , comme l'intersection de toutes les tribus \mathcal{A}_m .

REMARQUE III-105. Intuitivement, la tribu asymptotique est celle qui rassemble tous les ensembles dont la définition peut s'exprimer en fonction de variables d'ordre arbitrairement grand.

EXEMPLES III-106. On pose $X = \mathbb{R}_+$, de sorte que $X^{\mathbb{N}}$ est l'ensemble des suites réelles positives. L'ensemble des suites (x_n) pour lesquelles $\sum nx_n = 1$ n'est pas symétrique. L'ensemble des suites (x_n) dont la somme vaut 1 est un ensemble symétrique, mais non mesurable pour la tribu asymptotique : sa définition fait intervenir la valeur x_1 . L'ensemble des suites qui convergent, ou l'ensemble des suites qui convergent vers 1, ou l'ensemble des suites dont la somme converge, sont des ensembles symétriques et mesurables pour la tribu asymptotique.

Dans tous les exemples précédents, les ensembles asymptotiques se trouvaient également être symétriques. La proposition suivante assure que c'est la règle.

PROPOSITION III-107 (asymptotique implique symétrique). Soit X un espace mesurable, et soit A un ensemble mesurable pour la tribu asymptotique sur $X^{\mathbb{N}}$. Alors A est symétrique.

DÉMONSTRATION. On note, comme ci-dessus, \mathcal{A}_n la plus petite tribu qui rende mesurable les projections π_m pour $m \geq n$; de manière équivalente, c'est la tribu engendrée par les cylindres

$$C(m, B) := \{x \in X^{\mathbb{N}}; x_m \in B\},$$

où $m \geq n$ et B décrit l'ensemble des parties mesurables de X .

Pour prouver la Proposition III-107, il suffit de montrer que si $A \in \mathcal{A}_n$, alors A est invariant par les permutations finies qui n'affectent que les entiers $\{0, \dots, n-1\}$. Soit donc \mathcal{C} l'ensemble des parties de \mathcal{A}_n qui sont invariantes par de telles permutations. Il est clair que \mathcal{A} laisse invariants tous les cylindres $C(m, B)$; il suffit donc de vérifier que \mathcal{C} est une algèbre.

Soit $A \in \mathcal{C}$, et soit σ une permutation n'affectant que les entiers $\{0, \dots, n\}$. Si $x \in X \setminus A$, alors $x \notin A$, d'où $\sigma(x) \notin A$, i.e. $\sigma(x) \in X \setminus A$. De même, si $x \notin X \setminus A$, alors $\sigma(x) \notin X \setminus A$. On voit donc que $X \setminus A \in \mathcal{C}$.

Soit $(A_k)_{k \in \mathbb{N}}$ une suite d'éléments de \mathcal{C} , et soit σ une permutation n'affectant que les entiers $\{0, \dots, n\}$. Si $x \in (\cup A_k)$, alors il existe k tel que $x \in A_k$, d'où $\sigma(x) \in A_k$, en particulier $\sigma(x) \in (\cup A_k)$. La réciproque est identique, et on conclut que $\cup A_k \in \mathcal{C}$. L'ensemble \mathcal{C} est bien stable par passage au complémentaire et union dénombrable, ce qui achève la preuve. \square

Nous pouvons maintenant passer aux lois du 0-1 proprement dites.

THÉORÈME III-108 (loi du 0-1 de Hewitt-Savage). *Soit (X, μ) un espace de probabilité; on munit $X^{\mathbb{N}}$ de la tribu produit, et de la probabilité produit $\mu^{\otimes \mathbb{N}}$. Soit A une partie mesurable symétrique de $X^{\mathbb{N}}$, alors $\mu^{\otimes \mathbb{N}}[A]$ vaut soit 0, soit 1.*

En combinant cet énoncé avec la Proposition III-107, on obtient le

COROLLAIRE III-109 (loi du 0-1 de Kolmogorov). *Soit (X, μ) un espace de probabilité; on munit $X^{\mathbb{N}}$ de la tribu produit, et de la probabilité produit $\mu^{\otimes \mathbb{N}}$. Soit A un élément de la tribu asymptotique de $X^{\mathbb{N}}$, alors $\mu^{\otimes \mathbb{N}}[A]$ vaut soit 0, soit 1.*

DÉMONSTRATION DU THÉORÈME III-108. Soit $\varepsilon > 0$; par le Théorème III-101, il existe un cylindre C tel que

$$(28) \quad \mu^{\otimes \mathbb{N}}[A \setminus C] + \mu^{\otimes \mathbb{N}}[C \setminus A] \leq \varepsilon.$$

Le cylindre C est de la forme $B \times \prod_{k \geq n+1} X_k$, où B est une partie mesurable de X^n ; et $\mu^{\otimes \mathbb{N}}[C] = \mu^n[B]$.

Soit σ une permutation finie qui échange $\{1, \dots, n\}$ et $\{n+1, \dots, 2n\}$. On fait agir σ sur les éléments de $X^{\otimes \mathbb{N}}$ par permutation des coordonnées. L'invariance de $\mu^{\otimes \mathbb{N}}$ par permutation (Proposition III-100) se traduit par

$$\mu^{\otimes \mathbb{N}}[\sigma(A) \setminus \sigma(C)] + \mu^{\otimes \mathbb{N}}[\sigma(C) \setminus \sigma(A)] \leq \varepsilon.$$

Par hypothèse $\sigma(A) = A$, d'où en fait

$$\mu^{\otimes \mathbb{N}}[A \setminus \sigma(C)] + \mu^{\otimes \mathbb{N}}[\sigma(C) \setminus A] \leq \varepsilon.$$

En combinant cela avec (28), on obtient

$$\mu^{\otimes \mathbb{N}}[C \setminus \sigma(C)] + \mu^{\otimes \mathbb{N}}[\sigma(C) \setminus C] \leq 2\varepsilon.$$

Autrement dit,

$$\mu^{\otimes 2n}[(B \times X^n) \setminus (X^n \times B)] + \mu^{\otimes 2n}[(X^n \times B) \setminus (B \times X^n)] \leq 2\varepsilon,$$

Chacun des deux termes du membre de gauche est égal à

$$\mu^{\otimes 2n}[B \times (X^n \setminus B)] = \mu^{\otimes n}[B] \mu^{\otimes n}[X^n \setminus B] = \mu^{\otimes n}[B](1 - \mu^{\otimes n}[B]).$$

On a donc montré que

$$\mu^{\otimes n}[B](1 - \mu^{\otimes n}[X^n \setminus B]) \leq \varepsilon.$$

Posons $b := \mu^{\otimes n}[B]$, $a := \mu^{\otimes \mathbb{N}}[A]$. On a donc

$$a(1 - a) \leq (b + \varepsilon)(1 - b + \varepsilon) = b(1 - b) + \varepsilon + \varepsilon^2 \leq 2\varepsilon + \varepsilon^2.$$

En faisant tendre ε vers 0, on trouve $a(1 - a) = 0$, ce qui conclut la preuve. \square

REMARQUE III-110. On peut aussi établir la loi du 0-1 de Kolmogorov directement, par un argument assez similaire à celui qui est présenté ci-dessus. Il est également possible, mais beaucoup plus délicat, de déduire la loi de Hewitt-Savage de celle de Kolmogorov [Dudley, p. 272].

III-6.5. Théorème de prolongement de Kolmogorov. Pour conclure cette section sur les produits infinis, démontrons le Théorème I-77. Pour le confort du lecteur, nous allons rappeler son énoncé (sous une forme un peu différente mais équivalente) et le préciser très légèrement.

THÉORÈME III-111 (théorème de prolongement de Kolmogorov). *Soit T un ensemble arbitraire, et $(X_t)_{t \in T}$ une famille d'espaces Polonais ; on définit*

$$X := \prod_{t \in T} X_t,$$

que l'on munit de la tribu produit. Pour toute partie finie $F = \{t_1, \dots, t_K\} \subset T$, on définit $X_F := X_{t_1} \times \dots \times X_{t_K}$; et pour tout Borélien A_F de X_F , on définit le cylindre $C_F(A_F)$, noté abusivement $C(A_F)$, par

$$C(A_F) := \{x \in X; (x_{t_1}, \dots, x_{t_K}) \in A_F\}.$$

Pour toute partie finie F de T on se donne une mesure de probabilité μ_F sur X_F . On suppose que les μ_F sont compatibles, au sens où pour toutes parties finies F et F' , et pour tous Boréliens A_F et $B_{F'}$ de X_F et $X_{F'}$ respectivement,

$$C(A_F) = C(B_{F'}) \implies \mu_F[A_F] = \mu_{F'}[B_{F'}].$$

Alors on peut définir une unique mesure de probabilité μ sur X , telle que pour toute partie F finie de T , et pour tout Borélien A_F de X_F ,

$$\mu[C(A_F)] = \mu_F[A_F].$$

Si pour tout t la tribu Borélienne sur X_t est engendrée par une famille \mathcal{F}_t , alors μ est caractérisée par la valeur qu'elle attribue aux ensembles élémentaires, i.e. les ensembles $C(A_F)$ avec

$$A_F = A_{t_1} \times \dots \times A_{t_K},$$

chacun des A_t appartenant à \mathcal{F}_t .

PREUVE DU THÉORÈME III-111. On vérifie facilement que l'ensemble \mathcal{A} de tous les cylindres $C(A_F)$ est une algèbre, qui contient X . La condition de compatibilité nous permet de définir μ sur cette algèbre, et de prouver qu'elle y est additive. Par le Théorème I-69, l'existence et l'unicité du prolongement de μ seront assurées si l'on prouve la σ -additivité de μ sur \mathcal{A} . Pour établir la fin de l'énoncé, il suffit de remarquer que les cylindres élémentaires engendrent la tribu complète, et que tous les espaces X_t sont de mesure finie.

Soient donc A et $(A_k)_{k \geq 1}$ des cylindres, tels que $A = \cup A_k$; on doit prouver que

$$\mu[A] = \sum_{k \geq 1} \mu[A_k].$$

Pour tout N , les cylindres A, A_1, \dots, A_N peuvent s'inclure dans un même cylindre, et comme on sait que μ est σ -additive sur les cylindres on aura

$$\sum_{1 \leq k \leq N} \mu[A_k] \leq \mu[A].$$

En passant à la limite quand $N \rightarrow \infty$, on obtient

$$\sum_{k \geq 1} \mu[A_k] \leq \mu[A].$$

Notre problème est maintenant d'établir l'inégalité inverse, beaucoup plus délicate.

La définition des cylindres A , A_k , et de leur mesure, ne fait intervenir qu'une partie dénombrable des espaces $(X_t)_{t \in T}$; quitte à changer les notations, on peut oublier les autres espaces, et supposer que T est dénombrable; sans perte de généralité $T = \{1, 2, \dots\}$. On définit alors $\mu_n = \mu_F$ pour $F = \{1, \dots, n\}$. La condition de compatibilité implique bien sûr que μ_{n+1} prolonge μ_n .

On conviendra qu'un cylindre C est d'ordre n si n est le plus petit indice tel que la définition de C ne fasse intervenir que les n premiers espaces X_j . Un tel cylindre est de la forme $B \times (\prod_{j \geq n+1} X_j)$; on dit alors que B est sa base. En particulier, $\mu[C] = \mu_n[B]$. On introduit n_0 l'ordre de A , et n_k l'ordre de A_k , pour tout $k \geq 1$; on définit également B comme la base de A , et B_k comme la base de A_k . On pose enfin $X^n = X_1 \times \dots \times X_n$.

L'espace X^{n_0} est produit fini d'espaces Polonais, donc Polonais lui-même. Par le Théorème de régularité I-54, la mesure μ_{n_0} est régulière, et en particulier on peut trouver un compact K_{n_0} contenu dans B tel que

$$\mu_{n_0}[K_{n_0}] \geq \mu_{n_0}[B] - \varepsilon = \mu[A] - \varepsilon.$$

En particulier,

$$\mu_{n_0+1}[K_{n_0} \times X_{n_0+1}] \geq \mu[A] - \varepsilon.$$

Par régularité de la mesure $\mu_{n_0+1}[K \times \cdot]$, on peut trouver un compact K_{n_0+1} de X_{n_0+1} , tel que

$$\mu_{n_0+1}[K_{n_0} \times K_{n_0+1}] \geq \mu[A] - 3\varepsilon/2.$$

Par récurrence, on construit ainsi une famille de compacts $K_n \subset X_n$, $n \geq n_0$, tels que pour tout n

$$\mu_n[K_{n_0} \times \dots \times K_n] \geq \mu[A] - 2\varepsilon.$$

Le produit infini de ces compacts est un compact K contenu dans A .

De même, l'espace X^{n_k} étant Polonais, pour tout k on peut trouver un ouvert O_k , contenant B_k et tel que

$$\mu[O_k] \leq \mu_{n_k}[B_k] + \varepsilon 2^{-k} = \mu[A_k] + \varepsilon 2^{-k}.$$

On définit alors $U_k = O_k \times \prod_{j \geq n_{k+1}} X_j$: c'est un cylindre ouvert, qui contient A_k , tel que $\mu[U_k] \leq \mu[A_k] + \varepsilon 2^{-k}$.

Puisque $K \subset A \subset \cup A_k \subset \cup U_k$, que K est compact et que les U_k sont ouverts, on peut extraire des U_k un sous-recouvrement fini, soit U_1, \dots, U_N . Leur union est un cylindre, d'ordre fini M ; comme ils recouvrent K , ils recouvrent également le cylindre d'ordre M

$$C_M(K) := K_{n_0} \times K_{n_1} \times \dots \times K_M \times X_{M+1} \times X_{M+2} \times \dots$$

En particulier, $\mu[C_M(K)] \leq \sum_{k=1}^N \mu[U_k]$. Mais, par construction,

$$\mu[C_M(K)] = \mu_M[K_{n_0} \times \dots \times K_M] \geq \mu[A] - 2\varepsilon$$

et

$$\sum_k \mu[U_k] \leq \sum_k \mu[A_k] + \varepsilon.$$

On conclut que

$$\mu[A] \leq \sum_k \mu[A_k] + 3\varepsilon.$$

Comme ε est arbitraire, il s'ensuit que $\mu[A] \leq \sum_k \mu[A_k]$, ce qui était notre but. \square

REMARQUE III-112. Comme on l'a vu, même si T est un ensemble arbitraire, on se ramène dans la preuve à ne considérer qu'une famille dénombrable.

On pourra trouver dans [Dudley, p. 441–443] une version légèrement différente de l'argument présenté ci-dessus, et des hypothèses topologiques un peu plus générales, sans que le gain de généralité soit très appréciable.

EXEMPLE III-113. Le lecteur peut vérifier que l'énoncé du Théorème de Kolmogorov implique le résultat suivant, qui permet de fonder la théorie des chaînes de Markov en probabilités :

THÉORÈME III-114 (Théorème de Ionescu Tulcea). *Soient $(X_n, \mathcal{A}_n)_{n \in \mathbb{N}}$ des espaces mesurables. On se donne une mesure de probabilité μ_0 sur X_0 ; et pour tout $n \in \mathbb{N}$ on se donne une famille de mesures de probabilités ν_{x_n} sur X_{n+1} , dépendant mesurablement de $x_n \in X_n$. On pose $X := \prod X_n$ et on le munit de la tribu produit. Alors il existe une unique mesure de probabilité μ^∞ sur X telle que pour tout cylindre $C := A_0 \times A_1 \times A_2 \times \dots \times A_n \times (\prod_{j \geq n+1} X_j)$ de X ,*

$$\mu[C] = \int_{A_0} \int_{A_1} \dots \int_{A_{n-1}} \int_{A_n} d\nu_{x_{n-1}}(x_n) d\nu_{x_{n-2}}(x_{n-1}) \dots d\nu_{x_1}(x_0) d\mu(x_0).$$

Voici maintenant un exemple où l'espace T n'est pas dénombrable.

THÉORÈME III-115 (Existence de processus stochastique). *Soient $T = \mathbb{R}_+$ et X un espace Polonais, muni de sa tribu borélienne. On se donne une probabilité μ_0 sur X , et une famille de mesures $(\gamma_{t,x})_{t \geq 0, x \in X}$, dépendant mesurablement du paramètre $(t, x) \in \mathbb{R}_+ \times X$, satisfaisant*

$$\int_X \gamma_{s,y} \gamma_{t-s,x}(dy) = \gamma_{t,x}$$

pour tous (s, t) in $\mathbb{R}_+ \times \mathbb{R}_+$, $0 \leq s \leq t$, et pour tout $x \in X$. Pour toute famille finie $F = \{t_0, \dots, t_N\}$, $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_N$, pour toutes parties mesurables A_0, A_1, \dots, A_N de X , on définit le cylindre

$$C_F(A_0, \dots, A_N) = \{x \in X^{\mathbb{R}_+}; \forall k \in \{0, \dots, N\}, x_{t_k} \in A_k\}.$$

Alors il existe une unique mesure de probabilité μ sur $X^{\mathbb{R}_+}$ telle que pour tout cylindre $C_F(A_0, \dots, A_N)$,

$$\mu[C_F(A_0, \dots, A_N)] = \int_{A_0 \times \dots \times A_N} d\mu_{t_0}(x_0) d\gamma_{t_1-t_0, x_0}(x_1) d\gamma_{t_2-t_1}(x_2) \dots d\gamma_{t_N-t_{N-1}}(x_N).$$

En outre, si \mathcal{F} est une famille de parties de X engendrant la tribu borélienne, telle que X est union dénombrable d'une suite croissante d'éléments de \mathcal{F} , alors μ est uniquement déterminée par sa valeur sur les cylindres $C_F(A_0, \dots, A_N)$, où F est une partie finie arbitraire de T , et les A_j sont des éléments arbitraires de \mathcal{F} .

EXEMPLE III-116. L'exemple le plus célèbre est certainement celui où $X = \mathbb{R}$, $\mu_0 = \delta_0$, \mathcal{F} est (par exemple) la famille des intervalles compacts, et

$$\gamma_{t,x}[[a, b]] = \int_a^b \frac{e^{-\frac{|y-x|^2}{2t}}}{\sqrt{2\pi}} dy.$$

Dans ce cas, la mesure μ est appelée **processus stochastique Brownien** sur $\mathbb{R}^{\mathbb{R}_+}$: c'est une mesure de probabilité définie sur l'ensemble de toutes les trajectoires, au sens de fonctions de \mathbb{R}_+ dans \mathbb{R} .

Le théorème précédent est cependant très loin d'impliquer l'existence de la mesure de Wiener : en effet, la probabilité que nous avons construite est définie sur $\mathbb{R}^{\mathbb{R}_+}$; il est considérablement plus délicat de prouver qu'elle est en fait concentrée sur l'espace $C(\mathbb{R}_+)$ des trajectoires **continues**.

Appendice : Rappels sur les fonctions convexes

Les fonctions convexes, qui jouent un rôle aussi important dans des espaces vectoriels généraux, que les fonctions croissantes sur \mathbb{R} . C'est l'occasion de faire quelques rappels sur la théorie des fonctions convexes dans \mathbb{R}^n . Le traité de référence en la matière est [Rockafellar]. On se limitera ici aux fonctions convexes sur \mathbb{R}^n ; plus tard dans le cours, on parlera de convexité dans des espaces plus généraux.

On dit que $C \subset \mathbb{R}^n$ est une partie convexe de \mathbb{R}^n si, pour tous $x, y \in C$, $\lambda \in [0, 1]$, la “combinaison convexe” $\lambda x + (1 - \lambda)y$ est un élément de C . On dit que $\Phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ est une fonction convexe si, pour tous $x, y \in X$, $\lambda \in [0, 1]$,

$$\Phi(\lambda x + (1 - \lambda)y) \leq \lambda \Phi(x) + (1 - \lambda)\Phi(y).$$

Du point de vue géométrique, cela veut dire que “entre deux points x et y , le graphe de Φ est situé en-dessous de la corde joignant $[x, \Phi(x)]$ et $[y, \Phi(y)]$ ”.

Il est équivalent d'imposer que pour tout $N \in \mathbb{N}$, et tous $x_1, \dots, x_N \in \mathbb{R}^n$, $\lambda_1, \dots, \lambda_N \geq 0$, $\sum \lambda_i = 1$,

$$\Phi\left(\sum \lambda_i x_i\right) \leq \sum \lambda_i \Phi(x_i).$$

L'ensemble des x tels que $\Phi(x) < +\infty$ est appelé domaine de Φ . C'est un convexe (qui peut être ouvert, fermé, ou ni l'un ni l'autre).

Il est équivalent de dire qu'une fonction est convexe dans \mathbb{R}^n , ou que sa restriction à tout segment $[x, y]$ de \mathbb{R}^n est convexe.

Une fonction convexe sur \mathbb{R}^n est automatiquement continue dans l'intérieur de son domaine. En particulier, si elle est à valeurs dans \mathbb{R} , elle est continue sur tout \mathbb{R}^n . Dans le cas où cette fonction prend la valeur $+\infty$, il est naturel d'imposer que

- Φ ne soit pas identiquement $+\infty$;
- Φ soit semi-continue inférieurement. Comme Φ est continue dans l'intérieur de son domaine Ω , et vaut $+\infty$ identiquement à l'extérieur, cette hypothèse ne concerne en fait que le bord $\partial\Omega$ du domaine. Elle revient à imposer que pour tout segment $[x_0, x_1]$, avec x_0 dans l'intérieur de Ω et $x_1 \in \partial\Omega$, on ait $\Phi(x_t) \rightarrow \Phi(x_1)$ quand $t \rightarrow 1$, où $x_t := (1 - t)x_0 + tx_1$.

Si une fonction est donnée par un supremum de fonctions affines, elle est automatiquement convexe et semi-continue inférieurement.

Une fonction convexe n'est pas forcément différentiable, mais toujours **sous-différentiable** : en tout point x de l'intérieur du domaine de Φ on peut trouver au moins un vecteur $z = z(x)$ tel que

$$(29) \quad \forall y \in \mathbb{R}^n, \quad \Phi(y) \geq \Phi(x) + \langle z, y - x \rangle.$$

Géométriquement, cette inégalité exprime le fait que le graphe de Φ est situé au-dessus de l'hyperplan (dans \mathbb{R}^{n+1}) passant par $(x, \Phi(x))$ et orthogonal au vecteur $(z, 1)$. L'ensemble de tous les z admissibles dans (29) est appelé **sous-différentiel de Φ en x** , et noté $\partial\Phi(x)$.

Soit x dans l'intérieur du domaine de Φ . Si Φ est différentiable en x , alors $\partial\Phi(x)$ contient le vecteur $\nabla\Phi(x)$ constitué des dérivées partielles de Φ en x . En d'autres termes,

$$\forall y \in \mathbb{R}^n, \quad \Phi(y) \geq \Phi(x) + \langle \nabla\Phi(x), y - x \rangle.$$

En fait, $\partial\Phi(x)$ ne contient que ce vecteur : $\partial\Phi(x) = \{\nabla\Phi(x)\}$. Réciproquement, si le sous-différentiel au point x se limite à un singleton z , alors Φ est différentiable en x , et $z = \nabla\Phi(x)$.

Soit Φ une fonction deux fois dérivable dans un ouvert convexe de \mathbb{R}^n . Alors Φ est convexe si et seulement si sa matrice Hessienne $\nabla^2\Phi$ est positive sur Ω . C'est le critère que l'on utilise le plus souvent, en pratique, pour vérifier la convexité.

Une fonction $\Phi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ étant donnée, on peut définir sa **transformée de Legendre** :

$$\Phi^*(y) := \sup_{x \in \mathbb{R}^n} (\langle x, y \rangle - \Phi(x)).$$

Comme supremum de fonctions affines, c'est une fonction convexe et semi-continue inférieurement.

Si Φ est une fonction convexe et semi-continue inférieurement, alors

$$\Phi^{**} = \Phi.$$

C'est un cas particulier de la **dualité de Fenchel-Rockafellar**. Si Φ est convexe mais pas semi-continue inférieurement, alors Φ^{**} coïncide avec Φ dans l'intérieur et l'extérieur du domaine de Φ , mais pas sur le bord de ce domaine. Si Φ n'est pas convexe, alors Φ^{**} est l'enveloppe convexe de Φ : c'est la plus grande fonction convexe (semi-continue inférieurement) qui minore Φ .

Soit Φ une fonction convexe semi-continue inférieurement. Par définition de la transformée de Legendre, on a, pour tous $x, y \in \mathbb{R}^n$,

$$\langle x, y \rangle \leq \Phi(x) + \Phi^*(y).$$

Cette inégalité est appelée **inégalité de Young**. On a l'équivalence

$$\langle x, y \rangle = \Phi(x) + \Phi^*(y) \iff y \in \partial\Phi(x).$$

En particulier, si Φ est différentiable en x , alors

$$\langle x, \nabla\Phi(x) \rangle = \Phi(x) + \Phi^*(\nabla\Phi(x)),$$

et le vecteur $\nabla\Phi(x)$ est le seul vecteur possédant cette propriété.

Les fonctions $\nabla\Phi$ et $\nabla\Phi^*$ sont inverses l'une de l'autre : si Φ est différentiable en x et Φ^* différentiable en $\nabla\Phi(x)$, alors $\nabla\Phi^*(\nabla\Phi(x)) = x$. C'est cette propriété que l'on utilise le plus souvent en pratique pour calculer les transformées de Legendre.

EXERCICE III-117. Vérifier que, pour tout $p \in [1, +\infty[$,

$$\Phi(x) = \frac{|x|^p}{p} \implies \Phi^*(y) = \frac{|y|^{p'}}{p'}, \quad p' = \frac{p}{p-1}.$$

Cette identité est vraie dans \mathbb{R} ou plus généralement dans \mathbb{R}^n . Vérifier également que, sur \mathbb{R}_+ ,

$$\Phi(x) = x \log x - x \implies \Phi^*(y) = e^y.$$

La proposition qui suit est une conséquence immédiate de l'exercice.

PROPOSITION III-118 (inégalités de convexité de Young pour les puissances). Soient $x, y \in \mathbb{R}^n$, $p \in [1, +\infty]$ et $p' := p/(p-1)$. Alors

$$x \cdot y \leq \frac{|x|^p}{p} + \frac{|y|^{p'}}{p'};$$

si $p \notin \{1, +\infty\}$, l'inégalité précédente est une égalité si et seulement si x et y sont colinéaires et $|x|^p = |y|^{p'}$, autrement dit $|y| = |x|^{p-2}x$.

Ces résultats sont d'usage constant, ce qui motive la définition suivante.

DÉFINITION III-119 (exposant conjugué). Soit $p \in [1, +\infty]$; on appelle *exposant conjugué* de p le nombre $p' = p/(p-1) \geq 1$, caractérisé par l'identité

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

REMARQUE III-120. L'exposant p' est très souvent noté q , ce qui a l'inconvénient de ne pas imposer de lien notationnel avec p ; en outre, en théorie des probabilités, la convention d'usage est $q = 1 - p$... C'est pourquoi on préfère ici la notation p' (également très courante). Bien sûr, $(p')' = p$.

La deuxième partie de l'exercice III-117 mène à l'inégalité suivante, également fort utile.

PROPOSITION III-121 (inégalité de convexité de Young logarithmique). Soient a, b deux nombres réels positifs. Alors

$$ab \leq (a \log a - a + 1) + (e^b - 1).$$

CHAPITRE IV

La mesure de Lebesgue

Jusqu'à présent, nous avons étudié la théorie de Lebesgue dans le cadre abstrait développé par Radon et ses successeurs. Dans ce chapitre et le suivant, nous allons étudier de plus près des mesures particulières dans l'espace Euclidien \mathbb{R}^n : la mesure de Lebesgue d'une part ; et ses généralisations appelées mesures de Hausdorff d'autre part. La mesure de Lebesgue est celle que l'on utilise couramment, "par défaut", dans \mathbb{R}^n , et il est vital, en analyse réelle, d'être familier avec ses principales propriétés.

Après avoir expliqué comment on peut construire la mesure de Lebesgue (i.e. prouver son existence), nous étudierons quelques-unes de ses propriétés d'invariance qui expliquent pourquoi elle est si naturelle. Nous verrons ensuite que l'intégrale qui est associée à la mesure de Lebesgue généralise le concept d'intégrale de Riemann. Puis nous passerons en revue quelques-unes des principales propriétés de l'intégrale de Lebesgue, en particulier la formule de changement de variables. Pour finir, on s'interrogera sur la mesurabilité au sens de Lebesgue, d'un point de vue "axiomatique".

IV-1. Construction de la mesure de Lebesgue

La mesure de Lebesgue est celle que l'on utilise couramment dans \mathbb{R}^n ; l'intégrale qui lui est associée prolonge le concept d'intégrale de Riemann. Voici une façon de la définir.

DÉFINITION IV-1 (mesure de Lebesgue). *Soit $n \geq 1$ un entier, et $\mathcal{B}(\mathbb{R}^n)$ la tribu borélienne sur \mathbb{R}^n .*

(i) *Il existe sur $\mathcal{B}(\mathbb{R}^n)$ une unique mesure λ_n telle que pour tout pavé $P = [a_1, b_1] \times \dots \times [a_n, b_n] \subset \mathbb{R}^n$, on ait*

$$(30) \quad \lambda_n[P] = \prod_{j=1}^n (b_j - a_j).$$

*Cette mesure est appelée **mesure de Lebesgue n -dimensionnelle** et notée λ_n , λ , \mathcal{L}_n , \mathcal{L}^n , ou \mathcal{L} . On note également*

$$\lambda_n[A] = |A|_n = |A|.$$

(ii) *Si $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ est une fonction borélienne λ -sommable, on note*

$$\int_{\mathbb{R}^n} f(x) d\lambda_n(x) = \int_{\mathbb{R}^n} f(x) d^n x = \int_{\mathbb{R}^n} f(x) dx$$

et on dit que f est Lebesgue-intégrable. Si $n = 1$, on note également

$$\int_{[a,b]} f d\lambda_1 = \int_a^b f(x) dx = \int_a^b f.$$

(iii) La complétion de λ_n est également appelée mesure de Lebesgue; elle est définie sur la tribu des **ensembles Lebesgue-mesurables**, constituée de toutes les parties E de \mathbb{R}^n telles qu'il existe des ensembles Boréliens A et B tels que

$$A \subset E \subset B; \quad \lambda_n[B \setminus A] = 0.$$

Une fonction $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$, mesurable pour cette tribu, est dite *Lebesgue-mesurable*.

La définition de la tribu complétée en (iii) suit celle du Théorème I-80; on reparlera dans la section IV-5 de la structure des ensembles Lebesgue-mesurables. Pour l'instant, commençons par vérifier que la Définition IV-1 est licite, au sens où elle définit bien la mesure de Lebesgue sans équivoque. La famille des pavés est stable par intersection finie (l'intersection de deux pavés est un pavé), et \mathbb{R}^n est l'union des pavés $[-k, k]^n$ pour $k \in \mathbb{N}$; l'unicité de la mesure de Lebesgue est donc une conséquence directe du Théorème I-71(i). En revanche, établir l'**existence** de la mesure de Lebesgue nécessite un peu plus de travail. Comme nous allons le voir, on peut le faire de plusieurs façons légèrement différentes; le résultat d'unicité assure que toutes sont équivalentes. Toutes les méthodes présentées ci-après reposent *in fine* sur le théorème de prolongement de Carathéodory.

IV-1.1. De la dimension 1 à la dimension n . Supposons construite la mesure de Lebesgue $\lambda = \lambda_1$ sur $\mathcal{B}(\mathbb{R})$. On peut alors définir la mesure produit $\lambda^{\otimes n}$ sur $\mathcal{B}(\mathbb{R})^{\otimes n}$, qui d'après la Proposition III-36 n'est autre que $\mathcal{B}(\mathbb{R}^n)$. Par définition de la mesure produit, cette mesure vérifie (30), c'est donc la mesure de Lebesgue. En conclusion, il est équivalent de construire directement la mesure de Lebesgue en dimension n , ou de l'obtenir par tensorisation successive de la mesure de Lebesgue en dimension 1. Une autre conséquence est l'identité

$$\lambda_m \otimes \lambda_n = \lambda_{m+n},$$

vue comme une égalité entre mesures définies sur $\mathcal{B}(\mathbb{R}^{m+n})$.

Dans la suite, on se bornera donc à construire λ_1 ; au demeurant il n'est pas très difficile d'adapter les preuves pour obtenir des constructions directes de λ_n .

IV-1.2. Via le théorème de prolongement de Carathéodory. C'est la démonstration qui a déjà été présentée dans la section I-6.

IV-1.3. Via le théorème d'existence de produit infini. Cette construction va utiliser un "changement de variables" mesurable. Il est bien connu que tout nombre réel dans $[0, 1]$ admet une écriture binaire,

$$x = \sum_{k \geq 1} x_k 2^{-k}, \quad x_k \in \{0, 1\}.$$

Cette écriture est unique si l'on exclut les nombres dyadiques, i.e. de la forme $x = p/2^k$, $p, k \in \mathbb{N}$ (on peut conserver 0 et 1). L'application "écriture binaire" nous permet de changer la variable $x \in [0, 1]$ en une variable $x \in \{0, 1\}^{\mathbb{N}}$.

Munissons l'ensemble $\{0, 1\}$ de la **mesure de Bernoulli**, i.e. la mesure β définie par

$$\beta[\{0\}] = \frac{1}{2}, \quad \beta[\{1\}] = \frac{1}{2}.$$

Comme c'est une mesure de probabilité, on peut considérer son produit tensoriel infini, $\beta^{\otimes \mathbb{N}}$, bien défini par le Théorème I-75. On peut alors transporter la mesure $\beta^{\otimes \mathbb{N}}$ sur l'intervalle $[0, 1]$, via l'application

$$\varphi : (x_k)_{k \in \mathbb{N}} \mapsto \sum_{k \geq 1} x_k 2^{-k}.$$

L'ensemble des nombres de $[0, 1]$ dont le développement en base 2 commence par une suite donnée (x_1, \dots, x_k) est un intervalle de $[0, 1]$ appelé "intervalle dyadique" (de la forme $[p2^{-k}, (p+1)2^{-k}]$); l'image réciproque par φ d'un intervalle dyadique est donc l'union d'un cylindre et d'un ou deux points (les nombres dyadiques admettent deux écritures différentes). Un point de $\{0, 1\}^{\mathbb{N}}$ est mesurable car intersection de cylindres, on conclut que l'image réciproque d'un intervalle dyadique est mesurable. On vérifie aisément que tout intervalle ouvert peut s'écrire comme réunion d'intervalles dyadiques; la tribu engendrée par les intervalles dyadiques est donc la tribu borélienne tout entière, et φ est bien mesurable pour la tribu borélienne.

La mesure image

$$\lambda := \varphi_{\#} \beta^{\otimes \mathbb{N}}$$

est donc bien définie sur la tribu borélienne. Et c'est la mesure de Lebesgue sur $[0, 1]$! Pour s'en convaincre, il suffit de remarquer que tous les intervalles dyadiques de la forme $[p2^{-k}, (p+1)2^{-k}]$, ont mesure 2^{-k} : en effet, l'image réciproque d'un tel intervalle est l'union disjointe d'un cylindre de mesure 2^{-k} et d'un ou deux points, de mesure nulle (correspondant aux "écritures impopres": les entiers dyadiques admettent deux écritures binaires distinctes). Par exemple, l'image réciproque de $[1/4, 3/8]$ est constituée du cylindre $(0, 1, 0) \times \{0, 1\}^{\mathbb{N}}$ (écritures propres des nombres dans $[1/4, 3/8[$ et écriture impropre de $3/8$), du point $(0, 0, 1, 1, 1, 1, \dots)$ (écriture impropre de $1/4$) et du point $(0, 1, 1, 0, 0, 0, 0, \dots)$ (écriture propre de $3/8$).

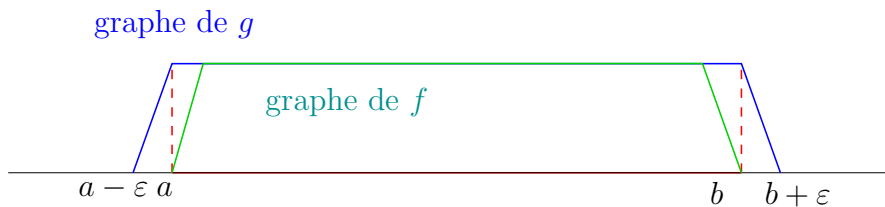
Comme les intervalles dyadiques engendrent la tribu borélienne et que la mesure que nous venons de définir coïncide avec la mesure de Lebesgue sur les intervalles dyadiques, elle coïncide avec la mesure de Lebesgue sur l'ensemble de tous les boréliens de $[0, 1]$.

IV-1.4. Via le théorème de représentation de Riesz. Une autre façon de construire la mesure de Lebesgue consiste à faire appel au Théorème de Riesz II-54, ou même à sa version simplifiée II-58. Dans ce cas, la forme linéaire positive que nous allons considérer est tout simplement **l'intégrale de Riemann** des fonctions continues à support compact dans \mathbb{R} . Le théorème de Riesz assure qu'il existe une mesure λ sur la tribu borélienne, telle que $\int f d\lambda = \int f(x) dx$, pour toute fonction f continue à support compact.

Nous allons vérifier que λ est la mesure de Lebesgue; il suffit de montrer que λ attribue à un intervalle $I = [a, b]$ la mesure $b - a$. Pour cela on se donne $\varepsilon > 0$ et on construit deux fonctions continues f et g , à support compact et à valeurs dans $[0, 1]$, telles que

$$f \leq 1_I \leq g, \quad \int f d\lambda = \int_{\text{Riem}} f = (b - a) - \varepsilon, \quad \int g d\lambda = \int_{\text{Riem}} g = (b - a) + \varepsilon,$$

où \int_{Riem} désigne bien sûr l'intégrale au sens de Riemann. On en déduit que $\lambda[I]$ est compris entre $b - a - \varepsilon$ et $b - a + \varepsilon$; en faisant tendre ε vers 0 on conclut que $\lambda[I] = b - a$. La mesure λ est donc bien la mesure de Lebesgue.

FIG. 1. Fonctions continues approchant $1_{[a,b]}$

REMARQUE IV-2. L'espace euclidien \mathbb{R}^n est non seulement localement compact, mais muni d'une structure différentiable : on a une notion de fonctions différentiables, et même indéfiniment différentiables, sur \mathbb{R}^n . Les mesures boréliennes finies sur les compacts font partie de la grande famille des **distributions**, qui sont des formes linéaires sur l'espace vectoriel des fonctions indéfiniment différentiables et à support compact, satisfaisant certaines propriétés de continuité. Un résultat remarquable stipule que les distributions positives sont exactement les mesures de Borel finies sur les compacts.

IV-2. Propriétés fondamentales de la mesure de Lebesgue

Dans cette section nous allons établir diverses propriétés importantes et intuitives de la mesure de Lebesgue, que l'on est en droit d'exiger de toute notion raisonnable de volume dans l'espace euclidien :

- a) le volume est diffus et "bien réparti" dans l'espace ;
- b) le volume est invariant par translation, et plus généralement par isométrie ;
- c) multiplier les distances par un facteur $\lambda > 0$ entraîne une multiplication du volume par un facteur λ^n ;
- d) contracter les distances diminue le volume ;
- e) le volume d'un parallélépipède coïncide avec son volume algébrique, défini grâce au déterminant des vecteurs qui l'engendrent.

Pour traduire la propriété a), nous utiliserons la notion de "mesure doublante" (Définition I-62) ; la propriété souhaitée découlera alors de c). En chemin, on reviendra sur la notion de Lebesgue-négligeabilité.

IV-2.1. Invariance par translation. Si P est un pavé de \mathbb{R}^n et $\tau : x \rightarrow x + h$ est une translation de vecteur $h \in \mathbb{R}^n$ fixé, il est évident que le produit des longueurs de P est identique au produit des longueurs de $\tau(P)$. Ceci, et la construction de la mesure de Lebesgue, implique immédiatement que λ_n est invariante par translation : $\tau_{\#}\lambda_n = \lambda_n$ pour toute translation τ .

Il est intéressant de noter que cette propriété **caractérise** la mesure de Lebesgue. Du fait de la structure d'espace affine de \mathbb{R}^n , l'invariance par translation est une autre justification du caractère naturel (voire incontournable) de la mesure de Lebesgue.

THÉORÈME IV-3 (caractérisation via l'invariance par translation). *La mesure de Lebesgue est, à multiplication scalaire près, l'unique mesure de Borel sur \mathbb{R}^n , finie sur les compacts, qui soit invariante par translation.*

DÉMONSTRATION. Soit μ une mesure vérifiant le cahier des charges ci-dessus. Puisque μ est finie sur les compacts, $\mu[C] < +\infty$. Soit $C_k := [0, 1/k]^n$; on peut

recouvrir C par une union disjointe de k^n cubes semi-ouverts de rayon $1/k$, obtenus par translation de C_k . Il s'ensuit que $\mu[C] = k^n \mu[C_k]$. Les mesures μ et $\mu[C] \lambda_n$ attribuent donc la même mesure à tous les cubes semi-ouverts de côté $1/k$, et cette famille suffit à engendrer la tribu borélienne (Exemple I-11 (ii)). Il s'ensuit que $\mu = \mu[C] \lambda_n$. \square

IV-2.2. Passage au quotient. Une conséquence presque immédiate de l'invariance par translation est la possibilité de passer au quotient par un réseau régulier, par exemple \mathbb{Z}^n . On définit le tore \mathbb{T}^n comme le quotient de \mathbb{R}^n par \mathbb{Z}^n , autrement dit par la relation d'équivalence : $x \mathcal{R} y \Leftrightarrow x - y \in \mathbb{Z}^n$. Le tore \mathbb{T}^n est aussi le produit de n copies du tore \mathbb{T}^1 . C'est un espace métrique compact, en bijection naturelle avec $C = [0, 1]^n$, puisque toute classe d'équivalence dans \mathbb{T}^n admet un unique représentant dans C . En particulier, $\mathbb{R}^n = C + \mathbb{Z}^n$: tout élément de \mathbb{R}^n est obtenu en ajoutant des coordonnées entières à un élément de C .

On utilise cette bijection f pour "identifier" \mathbb{T}^n et C en tant qu'espaces mesurés : si μ est une mesure sur C , on en déduit une mesure $f_{\#}\mu$ sur \mathbb{T}^n , et réciproquement.

PROPOSITION IV-4 (quotient de la mesure de Lebesgue). *La mesure de Lebesgue λ_n induit par restriction à $C = [0, 1]^n$ une mesure de probabilité sur \mathbb{T}^n , invariante par addition modulo \mathbb{Z}^n .*

DÉMONSTRATION. On peut écrire \mathbb{R}^n comme l'union disjointe des C_k , où $C_k = C + k$, et $k \in \mathbb{Z}^n$. Pour tout $A \subset C$, pour tout $x \in \mathbb{R}^n$, on peut décomposer $A + x$ en l'union dénombrable disjointe des $B_k = C_k \cap (A + x)$ (seul un nombre fini de ces ensembles sont non vides). Les ensembles $B_k - x$ sont disjoints : si y était un élément commun à deux tels ensembles, par différence on trouverait deux indices distincts k et ℓ tels que $\ell - k$ soit différence de deux éléments de A , ce qui est impossible puisque $A \subset C$.

On en déduit que A est l'union disjointe des $B_k - x$. Comme $A + x(\bmod \mathbb{Z}^n)$ est l'union disjointe des B_k , et que $B_k - x$ a même mesure que B_k , on conclut que A et $A + x(\bmod \mathbb{Z}^n)$ ont même mesure. \square

IV-2.3. Action des homothéties. Soit $f : x \rightarrow \alpha x$, avec $\alpha > 0$. L'application f est bijective, et l'image d'un pavé P est un pavé $f(P)$ dont toutes les longueurs ont été multipliées par α ; il s'ensuit que le volume de $f(P)$ est égal à α^n fois le volume de P . Bien sûr f établit une bijection entre pavés, donc la mesure image de λ_n par f est exactement la mesure de Lebesgue, à un facteur α^{-n} près. (Pourquoi α^{-n} et pas α^n ?) On en déduit que pour tout Borélien A de \mathbb{R}^n ,

$$\lambda_n[\alpha A] = \alpha^n \lambda_n[A].$$

Plus généralement, pour tout $\alpha \in \mathbb{R}$ et tout $A \in \mathcal{B}(\mathbb{R}^n)$,

$$\lambda_n[\alpha A] = |\alpha|^n \lambda_n[A].$$

IV-2.4. Régularité.

PROPOSITION IV-5. *La mesure de Lebesgue sur \mathbb{R}^n est régulière, et 2^n -doublante.*

DÉMONSTRATION. La mesure de Lebesgue est (bien sûr) finie sur les compacts de \mathbb{R}^n . Sa régularité découle donc du Corollaire I-56. Par ailleurs, si on se donne

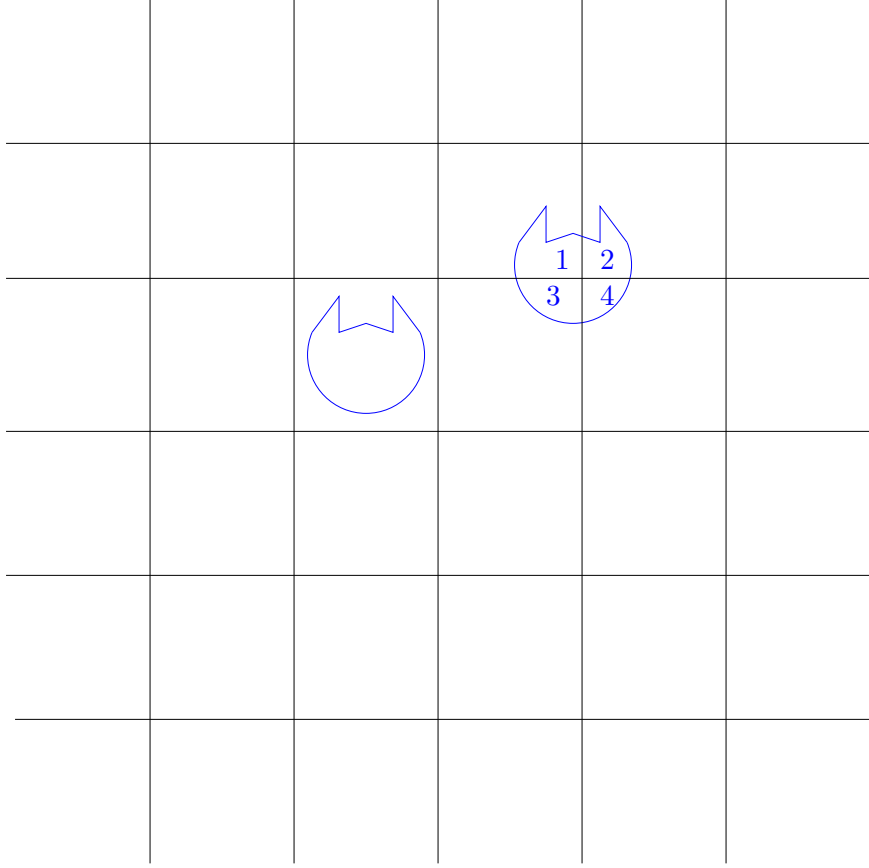


FIG. 2. Invariance de la mesure de Lebesgue par translation dans le tore. La somme des aires des quatre morceaux du chat translaté coïncide avec l'aire totale du chat initial.

$x \in \mathbb{R}^n$ et $r > 0$, la boule $B_{2r}(x)$ est obtenue à partir de $B_r(x)$ par homothétie de rapport 2, donc d'après le paragraphe précédent,

$$\lambda_n[B[x, 2r]] \leq 2^n \lambda_n[B[x, r]].$$

□

IV-2.5. Lebesgue-négligeabilité. De la définition de la mesure extérieure on déduit qu'un ensemble est Lebesgue-négligeable si et seulement si on peut l'inclure dans une famille dénombrable de pavés (ou de cubes) dont la somme des volumes est arbitrairement petite. En dimension 1, un ensemble est Lebesgue-négligeable si et seulement si on peut l'inclure dans une famille de segments dont la somme des longueurs est arbitrairement petite : c'est la définition qu'utilisait déjà Lebesgue.

EXEMPLE IV-6. Tout ensemble dénombrable est de mesure nulle (ce que l'on peut déduire d'ailleurs directement de la σ -additivité et de l'absence d'atomes). Tout sous-espace affine strict de \mathbb{R}^n est de mesure de Lebesgue (n -dimensionnelle) nulle. De même pour un ensemble inclus dans une union dénombrable d'hyperplans.

REMARQUE IV-7. Même en dimension 1, il existe des ensembles non dénombrables de mesure nulle ; par exemple l'ensemble triadique de Cantor.

Voici maintenant deux critères un peu plus sophistiqués de négligeabilité :

PROPOSITION IV-8 (les graphes mesurables sont négligeables). *Soient $D \subset \mathbb{R}^n$, et $f : D \rightarrow \mathbb{R}^m$ une application mesurable, avec $m \geq 1$. Alors le graphe de f est de mesure de Lebesgue nulle dans \mathbb{R}^{n+m} .*

PROPOSITION IV-9 (l'image lipschitzienne d'un ensemble négligeable est négligeable). *Soient A un borélien de mesure de Lebesgue nulle dans \mathbb{R}^n et $f : \Omega \rightarrow \mathbb{R}^m$ une application lipschitzienne, où Ω est un ouvert de \mathbb{R}^n contenant A . Alors $f(A)$ est Lebesgue-négligeable, au sens où il est inclus dans un borélien de mesure nulle.*

PREUVE DE LA PROPOSITION IV-8. Pour l'instant nous n'allons faire la preuve que dans le cas où f est continue ; on verra le cas général plus tard, comme conséquence du théorème de Fubini.

Comme \mathbb{R}^n est union dénombrable de cubes, on peut supposer que D est inclus dans le cube unité, auquel cas f est uniformément continue. On recouvre ce cube par N^n cubes C_k de côté $\delta = 1/N$; on en déduit un recouvrement du graphe de f par N^n pavés de la forme $C_k \times Q_k$, où Q_k est un cube de \mathbb{R}^m , de côté $2\omega(\delta)$, ω étant le module de continuité de f . La mesure totale de ces pavés est exactement $2^m \omega(\delta)^m$, qui tend vers 0 quand $\delta \rightarrow 0$. \square

PREUVE DE LA PROPOSITION IV-9. Soit $\varepsilon > 0$. Recouvrons A par une famille dénombrable de cubes C_j dont la somme des volumes est au plus ε . Chaque cube C_j , disons de côté c_j , est inclus dans une boule B_j de rayon $\sqrt{n}c_j$; $f(C_j)$ est alors inclus dans une boule de rayon $k\sqrt{n}c_j$, où k est la constante de Lipschitz de f , et donc dans un cube C'_j de côté knc_j . Le volume de C'_j est au plus $(kn)^n$ fois le volume de C_j , donc $f(A)$ est inclus dans une union de cubes dont le volume est au plus $(kn)^n \varepsilon$. Il s'ensuit que $f(A)$ est Lebesgue-négligeable. \square

La Proposition IV-9 admet un corollaire intéressant :

COROLLAIRE IV-10 (Les applications lipschitziennes préservent la Lebesgue-mesurabilité). *Soit Ω un ouvert de \mathbb{R}^n , et $f : \Omega \rightarrow \mathbb{R}^m$ une application lipschitzienne. Soit $A \subset \Omega$ un ensemble Lebesgue-mesurable de \mathbb{R}^n ; alors $f(A)$ est Lebesgue-mesurable.*

Pour apprécier cet énoncé, on notera que l'image d'un ensemble Lebesgue-mesurable par une application *continue* n'a aucune raison d'être Lebesgue-mesurable (les images continues d'ensembles Lebesgue-mesurables ont d'ailleurs un nom, ce sont les ensembles sousliniens).

DÉMONSTRATION DU COROLLAIRE IV-10. Par régularité de la mesure de Lebesgue, on peut écrire $A = (\cup K_i) \cup N$, où les K_i forment une famille dénombrable de compacts, et N est Lebesgue-négligeable. Puisque f est continue, les ensembles $f(K_i)$ sont tous compacts, donc leur union forme un ensemble borélien. Alors

$$\cup f(K_i) \subset f(A) \subset (\cup f(K_i)) \cup f(N);$$

la Proposition IV-9 implique que $f(N)$ est Lebesgue-négligeable, et il s'ensuit que $f(A)$ est Lebesgue-mesurable. \square

Terminons cette sous-section avec quelques remarques sur la négligeabilité. Les ensembles Lebesgue-négligeables peuvent être beaucoup plus complexes que ceux que nous avons vus jusqu'à présent ; un ensemble négligeable peut même être **gras** au sens de la topologie, c'est-à-dire intersection dénombrable d'ouverts denses.

EXEMPLE IV-11. Soit $I = [0, 1]$ muni de la mesure de Lebesgue, on énumère tous les rationnels de $[0, 1]$ en une suite $(q_n)_{n \in \mathbb{N}}$. Pour tout $\varepsilon > 0$ on pose

$$O_\varepsilon = \bigcup_{n \in \mathbb{N}} B_{\varepsilon/n^2}(q_n).$$

(O_ε est constitué de l'union de tous les intervalles ouverts de longueur $2\varepsilon/n^2$ centrés en q_n .) Bien sûr O_ε est ouvert dans I et dense. En outre

$$\lambda[O_\varepsilon] \leq 2\varepsilon \sum_{n \in \mathbb{N}} \frac{1}{n^2} = \left(\frac{\pi^2}{3}\right) \varepsilon.$$

Soit alors

$$A = \bigcap_{k \in \mathbb{N}} O_{1/k};$$

par σ -additivité on a bien $\lambda[A] = 0$, bien que A soit gras.

REMARQUE IV-12. Le complémentaire de A dans $[0, 1]$ est alors de mesure pleine bien que **maigre**, c'est-à-dire union dénombrable de fermés d'intérieur vide. Cette construction se généralise facilement en remplaçant I par \mathbb{R} tout entier, ou \mathbb{R}^n , ou n'importe quel ouvert de \mathbb{R}^n . C'est un débat classique de savoir si la "bonne" notion de négligeabilité est celle que fournit la théorie de la mesure, ou celle que fournit la topologie, et la plupart des mathématiciens se rangent dans un camp ou dans l'autre en fonction de leur sensibilité, de leur expérience personnelle, ou des problèmes qu'ils ont l'habitude de considérer. (Le Théorème KAM en mécanique classique est un exemple non académique pour lequel ce débat devient important.)

IV-2.6. Action des contractions. Il est intuitif que la contraction des longueurs induit une contraction du volume. Le théorème suivant précise cette idée.

THÉORÈME IV-13 (Réduire les distances réduit les volumes). *Soit Ω un ouvert de \mathbb{R}^n , et $f : \Omega \rightarrow \mathbb{R}^n$ une application 1-lipschitzienne :*

$$\forall x, y \in \mathbb{R}^n, \quad |f(x) - f(y)| \leq |x - y|.$$

Alors

$$(31) \quad \lambda_n[f(\Omega)] \leq \lambda_n[\Omega].$$

REMARQUE IV-14. Ce qui rend ce théorème non trivial est le fait que la mesure de Lebesgue est définie en termes de mesures de pavés, et que l'on ne peut pas dire grand chose de l'image d'un pavé par une application 1-lipschitzienne. Ce sont les boules qui se comportent bien vis-à-vis de l'hypothèse de lipschitzianité.

REMARQUE IV-15. Le Théorème IV-13 admet une généralisation immédiate au cas où f est seulement supposée L -lipschitzienne, avec L éventuellement différent de 1 : il suffit de remplacer (31) par

$$\lambda_n[f(\Omega)] \leq L^n \lambda_n[\Omega].$$

Le concept de mesure de Hausdorff nous permettra de démontrer un énoncé encore bien plus général : voir la Proposition V-6.

PREUVE DU THÉORÈME IV-13. Notons pour commencer que $f(\Omega)$ est mesurable en vertu de la Proposition IV-10 ; de toute façon l'argument qui suit redémontrera ce résultat.

Si A est une boule fermée $B[x, r]$, alors $f(A)$ est inclus dans la boule $B[f(x), r]$, qui a même volume que $B[x, r]$; donc

$$\lambda_n[f(A)] \leq \lambda_n[A].$$

La mesure de Lebesgue étant 2^n -doublante, on peut appliquer le Corollaire I-87 pour épuiser Ω par une union dénombrable de boules fermées disjointes B_j :

$$\Omega = \left(\bigcup_{j \in \mathbb{N}} B_j \right) \cup N,$$

où N est un borélien de mesure nulle.

Par la Proposition IV-9, $f(N)$ est de mesure nulle. Donc

$$\begin{aligned} \lambda_n[f(\Omega)] &= \lambda_n \left[f \left(\bigcup_{j \in \mathbb{N}} B_j \right) \right] = \lambda_n \left[\bigcup_{j \in \mathbb{N}} f(B_j) \right] \\ &\leq \sum_{j \in \mathbb{N}} \lambda_n[f(B_j)] \leq \sum_{j \in \mathbb{N}} \lambda_n[B_j] \\ &= \lambda_n \left[\bigcup_{j \in \mathbb{N}} B_j \right] = \lambda_n[\Omega], \end{aligned}$$

où l'on a utilisé le fait que les boules B_j sont disjointes. \square

COROLLAIRE IV-16 (les isométries préservent le volume). *Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une isométrie ; alors pour tout ensemble mesurable A de \mathbb{R}^n on a $\lambda_n[f(A)] = \lambda_n[A]$.*

DÉMONSTRATION. Il suffit d'appliquer le Théorème IV-13 à f et à f^{-1} . \square

REMARQUE IV-17. Une isométrie de \mathbb{R}^n est forcément une application affine ; on peut donc voir le corollaire précédent comme un cas particulier de l'action des applications affines sur la mesure de Lebesgue, que nous allons étudier dans la suite de ce chapitre.

REMARQUE IV-18. La classe des transformations (ou changements de variables) qui préservent la mesure de Lebesgue est infiniment plus vaste que celle des isométries. Par exemple, sur le segment $[0, 1]$, on peut permuter des sous-intervalles... La figure IV-18 représente les graphes de quelques applications simples préservant la mesure de Lebesgue sur $[0, 1]$ (les deux premières sont des bijections, la troisième non ; la première et la troisième sont continues, la deuxième non).

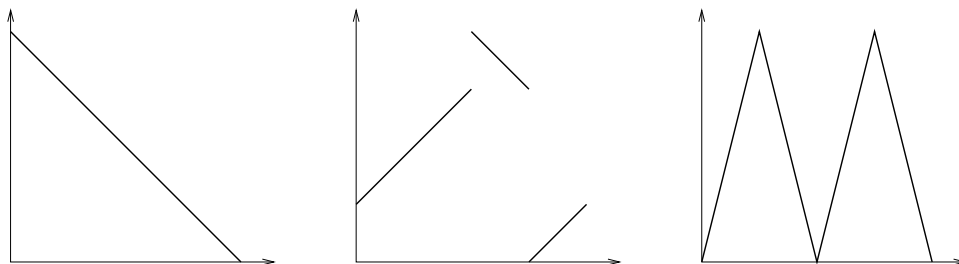


FIG. 3. Quelques graphes de transformations préservant la mesure de Lebesgue

En plusieurs dimensions, la classe des applications préservant la mesure de Lebesgue est d'une très grande importance dans de nombreux domaines des mathématiques. Par exemple, en mécanique des fluides, on utilise les bijections préservant la mesure de Lebesgue (restreinte à un ouvert de \mathbb{R}^n) pour représenter la collection des trajectoires d'un **fluide incompressible**; l'ensemble de ces bijections est un espace de dimension infinie (ce n'est pas un espace vectoriel, mais c'est un sous-ensemble d'une sphère dans un espace vectoriel normé) qui a fait l'objet de nombreuses études.

IV-2.7. Action des transformations affines. Le théorème suivant fait le lien entre deux notions naturelles de volume (l'une analytique, l'autre algébrique) pour un parallélépipède :

THÉORÈME IV-19 (mesure de Lebesgue et déterminant). *Soient $A \in M_n(\mathbb{R})$ et $b \in \mathbb{R}^n$; on note T l'application affine définie par $T(x) = Ax + b$, $C = [0, 1]^n$ le cube unité de \mathbb{R}^n , et $P = T(C)$ le parallélépipède formé des vecteurs colonnes de A . Alors,*

$$(32) \quad \lambda_n[P] = |\det A|.$$

REMARQUE IV-20. Il s'agit ici de volume non orienté.

Avant de donner la preuve, on va s'efforcer de faire sentir pourquoi ce résultat est naturel. Considérons une application linéaire de la forme $T(x_1, \dots, x_n) = (\alpha_1 x_1, \dots, \alpha_n x_n)$, où les α_i sont des nombres réels. Si $P = \prod [a_i, b_i]$ est un pavé de \mathbb{R}^n , le pavé $T(P)$ a pour côtés les nombres positifs $|\alpha_i| |b_i - a_i|$, son volume est donc égal à $(\prod |\alpha_i|) \prod |b_i - a_i|$, ce qui est le volume initial de P multiplié par le coefficient $\prod |\alpha_i| = |\det T|$. Il se peut que certaines longueurs soient allongées, d'autres raccourcies, ce qui compte pour évaluer la variation de volume c'est le *produit* des valeurs propres α_i . Comme le volume est invariant par changement de base orthonormée (Corollaire IV-16), le même résultat devrait être vrai pour toute application linéaire symétrique (diagonalisable dans une base orthonormée). L'examen de ce cas particulier suggère bien que le facteur multiplicatif du volume est la valeur absolue du déterminant.

DÉMONSTRATION. Commençons par le cas où A est non inversible. D'une part, $\det A = 0$; d'autre part, $T(\mathbb{R}^n)$ est inclus dans un hyperplan affine, donc de mesure nulle. Les deux membres de (32) sont donc nuls.

Dans le cas où A est inversible (et donc T est bijective $\mathbb{R}^n \rightarrow \mathbb{R}^n$), on va établir l'énoncé plus général

$$(33) \quad T_{\#} \lambda_n = |\det A|^{-1} \lambda_n.$$

Montrons que (32) et (33) sont équivalents. L'équation (33) s'écrit $\lambda_n[T^{-1}(B)] = |\det A|^{-1} \lambda_n[B]$ pour tout borélien $B \subset \mathbb{R}^n$; comme T est bijective cela équivaut à $\lambda_n[B] = |\det A|^{-1} \lambda_n[T(B)]$, d'où (32) par le choix $B = C$. Réciproquement, si (32) est vrai, alors la mesure $\mu = |\det A|^{-1} (T^{-1})_{\#} \lambda_n$ satisfait à $\mu[C] = 1$; et μ est invariante par translation puisque $\mu[B + h] = |\det A|^{-1} \lambda_n[T(B + h)] = |\det A|^{-1} \lambda_n[T(B) + Ah] = |\det A|^{-1} \lambda_n[T(B)]$ pour tout borélien $B \subset \mathbb{R}^n$ et tout vecteur $h \in \mathbb{R}^n$. Grâce au Théorème IV-3, on conclut que $\mu = \lambda_n$, ce qui revient à (33).

Toujours grâce à l'invariance par translation, il suffit de se restreindre au cas où $b = 0$, c'est-à-dire $T(x) = Ax$. Le résultat voulu, soit sous la forme (33), soit sous

la forme (32), peut se démontrer facilement dans un certain nombre de cas simples. Par exemple,

- (I) si A est une matrice de permutation, c'est évident puisque $AC = C$.
- (II) si A se contente de multiplier une coordonnée :

$$Ax = (\alpha x_1, x_2, \dots, x_n),$$

alors l'image de C est un pavé de côtés $|\alpha|, 1, \dots, 1$; donc de volume $|\alpha| = |\det A|$.

- (III) si A est de la forme

$$Ax = (x_1 + x_2, x_2, \dots, x_n),$$

alors $AC = P \times [0, 1]^{n-2}$, où P est le parallélogramme (2-dimensionnel) de sommets $(0, 0)$, $(0, 1)$, $(1, 1)$ et $(1, 0)$. On peut découper ce parallélogramme en deux triangles (plus un reste de mesure nulle) que l'on peut recoller en le carré $[0, 1]^2$; ce qui revient à découper AC en deux morceaux et à les recoller en le cube $[0, 1]^n$ (voir la figure). Le volume de AC est donc égal à 1, ce qui est aussi le déterminant de A .

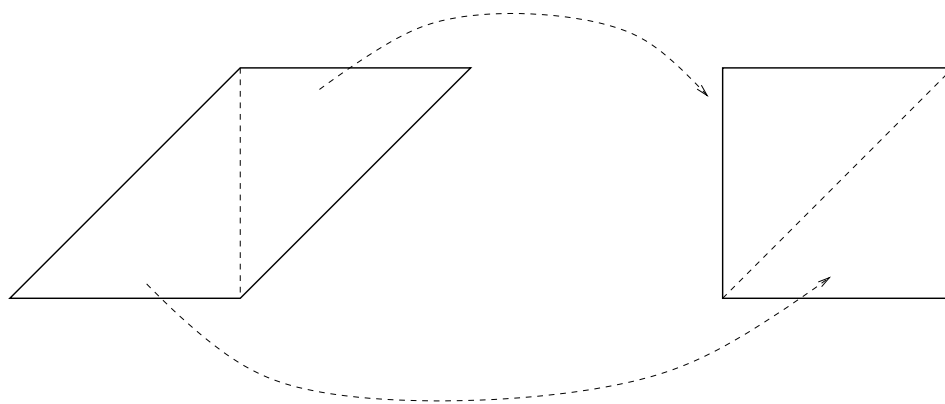


FIG. 4. Le parallélogramme a même aire que le carré

On note ensuite que la formule (33) est invariante par composition : si elle est vraie pour deux applications A_1 et A_2 , elle est aussi vraie pour $A = A_1 A_2$ puisque $|\det A_1 A_2| = |\det A_1| |\det A_2|$. Or un argument d'algèbre linéaire montre que toute matrice inversible est produit d'un nombre fini de matrices du type (I), (II) ou (III). On conclut à la validité de (33) pour n'importe quel A inversible. \square

IV-3. L'intégrale de Lebesgue généralise l'intégrale de Riemann

Nous avons déjà mentionné sans preuve que l'intégrale de Lebesgue généralisait l'intégrale de Riemann. Ce fait est important à plusieurs titres : d'une part, l'intégrale de Riemann est un objet simple, avec lequel le lecteur est sans doute familier ; d'autre part, tous les procédés habituels d'intégration numérique de fonctions (ou de calcul d'aire ou de volume) se ramènent à des variantes de l'intégrale de Riemann (méthode des rectangles, des trapèzes, etc.).

Nous limiterons notre discussion à la dimension 1, même si les démonstrations se généralisent sans autre problème que la lourdeur des notations. Pour simplifier, nous allons également nous limiter à des fonctions positives, le cas général étant conçu

pour s'y ramener. On commence par rappeler précisément le concept de Riemann-intégrabilité, en ne considérant que des fonctions **localement bornées** sur un intervalle J de \mathbb{R} , c'est-à-dire les fonctions f qui sont sommables sur tout intervalle compact $[a, b] \subset J$ (par exemple, les fonctions $x \mapsto x$ ou $x \mapsto 1/\sqrt{x}$ sont localement bornées sur $]0, +\infty[$).

DÉFINITION IV-21 (Riemann-intégrabilité). *Soit J un intervalle de \mathbb{R} , et $f : J \rightarrow \mathbb{R}_+$ une fonction localement bornée sur J . Une subdivision σ de l'intervalle $[a, b] \subset J$ en sous-intervalles I_1, \dots, I_K étant donnée, on définit $m_k(f) := \inf_{I_k} f$, $M_k(f) := \sup_{I_k} f$, et*

$$I^-(f, \sigma) := \sum_k |I_k| m_k(f), \quad I^+(f, \sigma) := \sum_k |I_k| M_k(f).$$

On pose alors

$$\mathcal{R}_{[a,b]}^-(f) := \sup_{\sigma \in \Sigma} I^-(f, \sigma), \quad \mathcal{R}_{[a,b]}^+(f) := \inf_{\sigma \in \Sigma} I^+(f, \sigma),$$

où Σ est l'ensemble de toutes les subdivisions de $[a, b]$. On dit que f est Riemann-intégrable sur $[a, b]$ si $\mathcal{R}_{[a,b]}^+(f) = \mathcal{R}_{[a,b]}^-(f)$, et dans ce cas on appelle intégrale de f sur $[a, b]$ la valeur commune de ces deux nombres. On définit enfin l'intégrale de Riemann de f sur J comme la limite de l'intégrale de f sur $[a, b]$ quand a et b tendent respectivement vers les extrémités gauche et droite de J .

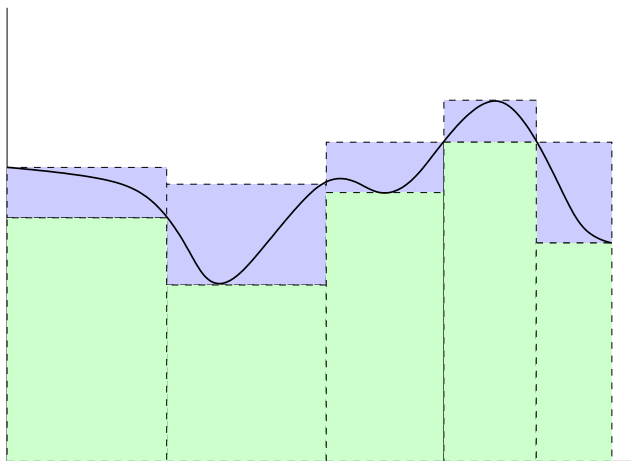


FIG. 5. L'intégrale de Riemann définie par encadrements : $I^-(f, \sigma)$ est la somme des aires des rectangles inférieurs, $I^+(f, \sigma)$ l'aire totale hachurée.

Le principal résultat de cette section est le suivant.

THÉORÈME IV-22 (L'intégrale de Lebesgue généralise celle de Riemann). *Soient J un intervalle de \mathbb{R} et $f : J \rightarrow \mathbb{R}_+$ une fonction bornée sur les compacts de J . Alors f est Riemann-intégrable si et seulement si (i) elle est Lebesgue-mesurable, et (ii) l'ensemble de ses points de discontinuité est de mesure nulle. Dans ce cas, l'intégrale de Riemann de f est égale à l'intégrale de Lebesgue de f .*

REMARQUE IV-23. La fonction $1_{[0,1] \cap \mathbb{Q}}$ est un exemple de fonction Lebesgue-mesurable bornée qui ne soit pas Riemann-intégrable.

DÉMONSTRATION. 1. On peut trouver des suites $(a_k)_{k \geq 1}$ et $(b_k)_{k \geq 1}$, respectivement croissante et décroissante, telles que $J = \cup [a_k, b_k]$. Alors on a $\int_J f = \lim_{k \rightarrow \infty} \int_{[a_k, b_k]} f$, au sens de Lebesgue, par convergence monotone. En outre, une réunion dénombrable d'ensembles négligeables est négligeable. Pour prouver le théorème dans le cas général, il suffit donc de se limiter au cas particulier où $J = [a, b]$ et f est bornée.

2. Supposons que f est Riemann-intégrable sur $[a, b]$; on note $\mathcal{R}(f)$ son intégrale au sens de Riemann. Soit $\varepsilon > 0$ et soient σ, σ' deux subdivisions de $[a, b]$ en sous-intervalles $(I_k)_{1 \leq k \leq K}$ et $(I'_\ell)_{1 \leq \ell \leq L}$ respectivement, telles que

$$\sum_k |I_k| m_k(f) \geq \mathcal{R}(f) - \varepsilon, \quad \sum_\ell |I'_\ell| M_\ell(f) \leq \mathcal{R}(f) + \varepsilon.$$

Quitte à remplacer σ et σ' par une subdivision plus fine, on peut supposer que $\sigma = \sigma'$. Pour chaque valeur de m on peut donc construire une subdivision $\sigma = \sigma_m$ de $[a, b]$ en sous-intervalles I_k , telle que

$$\sum_k |I_k| m_k(f) \geq \mathcal{R}(f) - \frac{1}{m}, \quad \sum_k |I_k| M_k(f) \leq \mathcal{R}(f) + \frac{1}{m}.$$

On définit alors les fonctions f_m^- et f_m^+ par

$$x \in I_k \implies f_m^-(x) = m_k(f), \quad f_m^+(x) = M_k(f),$$

en convenant que ces deux fonctions coïncident avec f aux extrémités des sous-intervalles.

Quitte à remplacer σ_m par une subdivision plus fine que $\sigma_1, \dots, \sigma_m$, on peut supposer que les subdivisions σ_m sont de plus en plus fines quand m augmente, auquel cas les fonctions f_m^+ forment une suite décroissante, et les fonctions f_m^- forment une suite croissante. Appelons f^+ et f^- les limites respectives de ces suites : les fonctions f^+ et f^- sont mesurables puisque limites de fonctions constantes par morceaux, et clairement $f^+ \geq f^-$.

Par convergence monotone,

$$\int f^+ = \lim \int f_m^+, \quad \int f^- = \lim \int f_m^-,$$

et par hypothèse ces deux limites sont égales à $\mathcal{R}(f)$. On en déduit que

$$\int (f^+ - f^-) = 0,$$

et il s'ensuit que $f^- = f^+$ presque partout sur $[a, b]$; en conséquence, ces fonctions coïncident presque partout avec f . En particulier, $\int f = \mathcal{R}(f)$.

Soit D l'ensemble de toutes les extrémités des sous-intervalles I_k des subdivisions σ_m ; comme D est dénombrable, il est de mesure nulle. Pour presque tout $x \in [a, b] \setminus D$, on a $f^+(x) = f^-(x)$, ce qui veut dire que x est point intérieur d'une famille d'intervalles J_m , décroissante, vérifiant

$$\lim_{m \rightarrow \infty} \left(\inf_{J_m} f \right) = \lim_{m \rightarrow \infty} \left(\sup_{J_m} f \right) = f(x).$$

Il s'ensuit que f est continue en x . Nous avons donc montré que l'ensemble des points de discontinuité de f est de mesure nulle.

3. Réciproquement, soit f une fonction bornée, Lebesgue-intégrable sur $[a, b]$, positive, dont l'ensemble des points de discontinuité est de mesure nulle ; nous allons montrer que f est Riemann-intégrable. Pour tout $m \geq 1$, on définit une subdivision σ_m en subdivisant l'intervalle $[a, b]$ en 2^m intervalles ouverts I_k^m , de longueur égale, et on définit les fonctions f_m^+ et f_m^- comme ci-dessus. Soit D l'ensemble de toutes les extrémités de ces intervalles, et soit x un point de continuité de f n'appartenant pas à D . Pour tout $m \geq 1$, il existe un k tel que $x \in I_k^m$, et l'intervalle I_k^m est de longueur $(b-a)/2^m$. Comme f est continue en x , l'oscillation de f sur I_k^m tend vers 0 quand $m \rightarrow \infty$, autrement dit

$$\sup_{I_k^m} f - \inf_{I_k^m} f \xrightarrow{m \rightarrow \infty} 0,$$

soit encore $f_m^+(x) - f_m^-(x) \rightarrow 0$. La famille $(f_m^+ - f_m^-)$ est une suite de fonctions positives, bornées sur $[a, b]$, convergeant vers 0 presque partout, par convergence dominée on a $\int f_m^+ - \int f_m^- \rightarrow 0$, ce qui signifie exactement que f est Riemann-intégrable. On a déjà vu qu'en pareil cas, la valeur de l'intégrale de Riemann et celle de l'intégrale de Lebesgue coïncident. \square

IV-4. Règles de calcul associées à l'intégrale de Lebesgue

IV-4.1. Dérivation et intégration dans \mathbb{R} . Nous avons déjà mentionné que l'une des motivations de Lebesgue était de construire une théorie dans laquelle intégration et dérivation étaient toujours des opérations inverses l'une de l'autre. Le cadre naturel de son principal résultat en la matière est celui des applications absolument continues.

DÉFINITION IV-24 (absolue continuité). Soient I un intervalle de \mathbb{R} , et $f : I \rightarrow \mathbb{R}$ une application mesurable. On dit que f est absolument continue sur I si pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour toute famille finie $[a_k, b_k]$ d'intervalles disjoints inclus dans I ($1 \leq k \leq N$),

$$\sum_{1 \leq k \leq N} |b_k - a_k| \leq \delta \implies \sum_{1 \leq k \leq N} |f(b_k) - f(a_k)| \leq \varepsilon.$$

EXEMPLE IV-25. Il est clair qu'une fonction lipschitzienne est absolument continue : dans la définition ci-dessus on peut choisir $\delta = \varepsilon/L$, où L est la constante de Lipschitz de f . En particulier, par la formule des accroissements finis, toute application dérivable, dont la dérivée est bornée, est lipschitzienne, et donc absolument continue.

Il est clair par ailleurs que l'absolue continuité implique l'uniforme continuité ; le concept d'absolue continuité est donc intermédiaire entre celui d'uniforme continuité et celui de lipschitzianité. Formellement, les applications absolument continues sur un intervalle borné sont celles dont **la dérivée est sommable** ; plus rigoureusement ce sont celles dont la dérivée (au sens des distributions) est une mesure absolument continue par rapport à la mesure de Lebesgue.

THÉORÈME IV-26 (dérivation et intégration). Soit f une fonction absolument continue sur un intervalle $I = [a, b]$ de \mathbb{R} . Alors f est dérivable presque partout dans

I , et sa dérivée f' est une application sommable sur I . En outre, pour tout $x \in [a, b]$, on a l'identité

$$f(x) - f(a) = \int_a^x f'(t) dt.$$

Nous démontrerons ce résultat plus tard, dans le Chapitre ??.

IV-4.2. Théorème de Fubini. L'espace $(\mathbb{R}^n, \lambda_n)$ est bien sûr σ -fini ; on peut donc appliquer le théorème de Fubini–Tonelli–Lebesgue dans ces espaces. En outre, comme on l'a déjà rappelé,

$$\lambda_m \otimes \lambda_n = \lambda_{m+n}.$$

REMARQUE IV-27. Il arrive souvent que l'on ait besoin de découper une intégrale en tranches “curvilignes”, pour lesquelles le théorème de Fubini ne s'applique pas. La célèbre **formule de la co-aire** permet de traiter de telles situations ; nous en parlerons dans le Chapitre ??.

À titre d'illustration, on va donner deux applications du théorème de Fubini : la démonstration générale de la Proposition IV-8 ; puis une démonstration alternative du Théorème IV-19.

PREUVE DE LA PROPOSITION IV-8. Notons d'abord que le graphe de f est mesurable, comme l'image réciproque de 0 par l'application mesurable $(x, y) \mapsto y - f(x)$. En outre, l'application indicatrice du graphe de f vaut $1_{f(x)=y}(x, y)$. Ensuite, par Fubini,

$$\int_{\mathbb{R}^{n+m}} 1_{f(x)=y}(x, y) d\lambda_n(x) d\lambda_m(y) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} 1_{y=f(x)}(x, y) d\lambda_m(y) \right) d\lambda_n(x) = \int_{\mathbb{R}^n} 0 d\lambda_n(x) = 0.$$

Ceci prouve que le graphe de f est négligeable. \square

PREUVE ALTERNATIVE DU THÉORÈME IV-19. Comme dans la preuve vue précédemment, on se ramène au cas où A est inversible, on montre l'équivalence entre (32) et (33) et l'invariance de la formule par composition. La différence est dans le choix des “cas élémentaires” : on note que toute application linéaire inversible A est produit d'applications linéaires laissant une coordonnée invariante.

Soit alors A une telle application ; sans perte de généralité (en utilisant l'invariance par permutation) on peut supposer que $Ax = (A'(x_1, \dots, x_n), x_n)$, et la sous-matrice A_{n-1} formée des $n-1$ premières lignes et colonnes de A est inversible :

$$A = \begin{pmatrix} & 0 \\ & \vdots \\ A_{n-1} & \\ * & \dots & * & 1 \end{pmatrix}$$

On note que $\det A = \det A_{n-1}$. Soit $P = Q \times [a, b]$ un pavé dans \mathbb{R}^n , on note $A(P)_y$ la tranche de P selon $x_n = y$; cette tranche vaut Q si $y \in [a, b]$, et \emptyset sinon. Par Fubini,

$$\lambda_n[A(P)] = \int_{\mathbb{R}} \lambda_{n-1}[A(P)_{x_n}] dx_n = \int_a^b \lambda_{n-1}[A_{n-1}(Q)] dx_n = (b-a) \lambda_{n-1}[A_{n-1}(Q)].$$

On en déduit que

$$(A^{-1})_{\#} \lambda_n = ((A_{n-1}^{-1})_{\#} \lambda_{n-1}) \otimes \lambda_1.$$

En particulier, si la formule (33) est vraie pour la sous-matrice A_{n-1} , elle sera vraie également pour la matrice A .

Pour conclure, on raisonne par récurrence sur la dimension. Si $n = 1$, la propriété souhaitée est évidente. Si la propriété est démontrée au rang $n - 1$, soit alors A une matrice inversible de taille n ; on peut l'écrire comme $A = A' \times A''$, où les matrices A' et A'' préservent chacune une coordonnée. Par hypothèse de récurrence, et la remarque ci-dessus, la formule (33) est vraie pour A' et A'' , elle est donc aussi vraie pour A grâce à l'invariance par produit. \square

IV-4.3. Changement de variable. Dans la Section III-3 nous avons considéré le théorème abstrait de changement de variable : $\int f d(T_{\#}\lambda) = \int (f \circ T) d\lambda$. Dans le cadre "concret" de \mathbb{R}^n muni de la mesure de Lebesgue, cette identité peut être précisée grâce à une formule qui exprime la mesure image en fonction du déterminant jacobien du changement de variable.

Il existe de nombreuses variantes de ce théorème, sous diverses hypothèses. Celle que nous allons voir maintenant semble être le meilleur compromis entre généralité et simplicité. On notera $d_x\varphi$ la différentielle de φ en x , que l'on peut l'identifier à la matrice des dérivées partielles :

$$[d_x\varphi]_{ij} = \left(\frac{\partial \varphi_i}{\partial x_j} \right).$$

La notation $\lambda_n|_A$ désignera la restriction de λ_n au borélien A .

THÉORÈME IV-28 (Changement de variable C^1 dans \mathbb{R}^n). *Soient U un ouvert de \mathbb{R}^n , $\varphi : U \rightarrow V$ un C^1 -difféomorphisme, alors*

(i) *pour tout borélien $B \subset U$,*

$$\lambda_n[\varphi(B)] = \int_B |\det d\varphi| d\lambda_n;$$

(ii) *pour toute fonction f sommable $V \rightarrow \mathbb{R}$ (ou pour toute fonction f mesurable positive $V \rightarrow [0, +\infty]$),*

$$\int_{\varphi(U)} f d\lambda_n = \int_U (f \circ \varphi) |\det d\varphi| d\lambda_n;$$

(iii) $\varphi_{\#}(\lambda_n|_U) = m \lambda_n|_{\varphi(U)}$, où m est la fonction définie par

$$m(y) = \frac{1}{|\det d_{\varphi^{-1}(y)}\varphi|}.$$

REMARQUE IV-29. Une formulation équivalente de l'énoncé (ii) ci-dessus est :
(ii') *pour toute fonction mesurable positive g ,*

$$\int_U g(x) |\det d_x\varphi| dx = \int_{\varphi(U)} g(\varphi^{-1}(y)) dy.$$

Pour s'en convaincre, il suffit de poser $g = f \circ \varphi$, $f = g \circ \varphi^{-1}$ dans (ii).

REMARQUE IV-30. Les erreurs classiques dans l'application du Théorème IV-28 sont (a) la confusion entre φ et φ^{-1} , surtout au niveau de la formule (iii); (b) l'oubli des valeurs absolues autour du déterminant; (c) l'oubli de la restriction à U et $\varphi(U)$ dans la formulation (iii); (d) la non-vérification de l'injectivité de φ .

REMARQUE IV-31. Le Théorème IV-19 est un cas particulier du Théorème IV-28, correspondant au cas où φ est affine bijective.

REMARQUE IV-32. À son tour, le Théorème IV-28 se généralise considérablement :

(a) L'hypothèse de régularité C^1 peut être relâchée en régularité lipschitz, ou même des hypothèses encore beaucoup plus générales telles que la différentiabilité presque partout (mais la Remarque IV-29 n'est plus forcément valide).

(b) L'hypothèse de bijectivité de φ peut être remplacée par l'injectivité en-dehors d'un ensemble négligeable ; en revanche sans cette propriété d'"injectivité presque partout", le théorème devient *faux*, et on a seulement, pour $f \geq 0$,

$$\int_{\varphi(U)} f \leq \int_U f \circ \varphi |\det d\varphi|$$

(pourquoi ?).

(c) On peut cependant modifier les formules pour traiter des cas où φ n'est pas injective (il convient alors d'introduire la multiplicité), et où φ est un changement de variables $\mathbb{R}^n \rightarrow \mathbb{R}^m$ avec $m > n$ (formule de l'aire) ou $m < n$ (formule de la co-aire). La mesure de Lebesgue doit alors être remplacée par une **mesure de Hausdorff**. On reparlera brièvement de ces formules dans le Chapitre V.

Ces sujets sont abordés dans le livre [Evans–Gariepy] ; et dans divers articles de recherche dont certains sont très récents.

On peut prouver le Théorème IV-28 de diverses manières. L'équivalence entre les énoncés (i)–(iii) est une conséquence facile de la définition de la mesure image (Définition III-59), du théorème abstrait de changement de variable (Théorème III-60) et de la bijectivité de φ . Il suffit donc de démontrer n'importe laquelle de ces trois formules.

Je vais présenter ici deux stratégies : la première, empruntée à [Gramain], copie l'argument utilisé à la fin de la sous-section IV-4.2 pour (re)démontrer le Théorème IV-19 ; la seconde, au contraire, considère le Théorème IV-19 comme une brique élémentaire à laquelle on peut se ramener par approximation. Cette dernière stratégie est plus intuitive, mais aussi plus élaborée puisqu'elle reposera sur le Théorème IV-13, qui lui-même fait appel au lemme de recouvrement de Vitali. Les deux preuves utilisent un argument de localisation.

PREMIÈRE PREUVE DU THÉORÈME IV-28. On va chercher à démontrer la formule (i). Par régularité de la mesure de Lebesgue (ou tout simplement parce que les compacts engendrent la tribu borélienne), il suffit de se limiter au cas où B est compact.

On raisonne par récurrence sur la dimension n . Commençons par $n = 1$. Comme les intervalles compacts engendrent la tribu borélienne, il suffit de montrer que pour tout $[a, b] \subset U$,

$$(34) \quad \lambda[\varphi(I)] = \int_a^b |\varphi'(x)| dx.$$

Étant un difféomorphisme, φ est soit strictement croissante, soit strictement décroissante sur $[a, b]$. Dans le premier cas, la formule (34) devient $\varphi(b) - \varphi(a) = \int_a^b \varphi'(x) dx$,

ce qui est évidemment vrai. Dans le deuxième cas, (34) se réécrit $\varphi(a) - \varphi(b) = \int_a^b (-\varphi'(x)) dx$, ce qui revient au même. Le théorème est donc vrai pour $n = 1$.

Supposons maintenant le théorème démontré en dimension $n-1 \geq 1$. Pour traiter la dimension n , on utilisera le

LEMME IV-33. *Soient U et V deux ouverts de \mathbb{R}^n et $\varphi : U \rightarrow V$ un difféomorphisme. Alors localement φ s'écrit comme composition de permutations des coordonnées, et de difféomorphismes préservant au moins une coordonnée.*

PREUVE DU LEMME IV-33. Soit $z \in U$. Il existe des indices j et k tels que $(\partial\varphi_k/\partial x_j)(z) \neq 0$. Quitte à permuter, on suppose $k = j = 1$. On pose alors $\psi(x_1, \dots, x_n) = (\varphi_1(x), x_2, \dots, x_n)$. Il est clair que $\det_z d\psi = (\partial\psi_1/\partial x_1)(z)$, donc par le théorème d'inversion locale ψ définit un difféomorphisme d'un voisinage O de z dans $\psi(O)$. Dans l'ouvert O on peut alors écrire $\varphi = (\varphi \circ \psi^{-1}) \circ \psi$, où ψ préserve les $n-1$ dernières coordonnées, et $\varphi \circ \psi^{-1}$ préserve la première. \square

Retournons à la preuve du Théorème IV-28. Montrons que la conclusion est vraie si φ préserve la dernière coordonnée : $\varphi(x_1, \dots, x_n) = (\varphi_1(x), \dots, \varphi_{n-1}(x), x_n)$. Notons $x' = (x_1, \dots, x_{n-1})$. Pour tout z fixé, l'application φ^z , obtenue en gelant la variable x_n à la valeur z et en ne conservant que les $n-1$ premières coordonnées de φ , définit un difféomorphisme de $U^z = U \cap \{x_n = z\}$ (vu comme un ouvert de l'hyperplan $(x_n = z)$) sur son image $\varphi^z(U^z)$. L'hypothèse de récurrence s'applique à φ^z :

$$\begin{aligned} (f^z)_\#(\lambda_{n-1}|_{U^z}) &= |\det d_{x'} f^z|^{-1} (\lambda_{n-1}|_{\varphi^z(U^z)}) \\ &= |\det d_{(x', z)} f|^{-1} (\lambda_{n-1}|_{\varphi^z(U^z)}). \end{aligned}$$

En appliquant Fubini, on en déduit, exactement comme dans la démonstration en fin de sous-section IV-4.2, que $f_\#(\lambda_n 1_U) = |\det d_x f|^{-1} (\lambda_n 1_V)$. Le théorème est donc vrai dans le cas où φ préserve l'une des coordonnées.

Le théorème est également (bien sûr) vrai quand φ est une permutation de coordonnées. Or le Lemme IV-33 montre que φ s'écrit localement comme composée de permutations et de difféomorphismes préservant une coordonnée. Grâce à l'invariance de la formule (iii) par composition, on conclut que le théorème est vrai en dimension n , pour peu que l'on remplace U par un petit voisinage de x (et φ par sa restriction à U).

Pour boucler la récurrence, il reste à "recoller les morceaux", c'est-à-dire établir la formule globale (disons pour tout borélien $B \subset U$) à partir de la formule locale (valable pour tout borélien B inclus dans un petit voisinage d'un point x fixé).

Fixons donc U , V et φ , et supposons que tout $x \in U$ est contenu dans un voisinage $U_x \subset U$ où la conclusion du théorème est vraie (avec U remplacé par U_x , φ remplacé par sa restriction à U_x); on va montrer qu'alors le théorème est vrai pour l'ouvert U tout entier. Comme on l'a déjà remarqué, il suffit de traiter le cas où B est un compact K de U . De la famille des $\{U_x, x \in K\}$ on extrait un sous-recouvrement fini $\{U_{x_1}, \dots, U_{x_N}\}$. On pose alors $C_1 = K \cap U_{x_1}$, $C_2 = K \cap (U_{x_2} \setminus U_{x_1})$, etc. de manière à définir des ensembles boréliens C_1, \dots, C_N deux à deux disjoints dont l'union est K . Puisque le théorème est vrai dans chaque U_{x_i} , on a

$$\forall j \in \{1, \dots, N\}, \quad \lambda_n[\varphi(C_j)] = \int_{C_j} |\det d_x \varphi| dx.$$

Comme φ est bijective, les $\varphi(C_j)$ sont disjoints et leur union est $\varphi(K)$. La sommation en j donne donc

$$\lambda_n[\varphi(K)] = \int_K |\det d_x \varphi| dx,$$

ce qui conclut la preuve du théorème. \square

SECONDE PREUVE DU THÉORÈME IV-28. L'idée est de comparer, au voisinage d'un point x , la mesure image $\varphi_{\#}\lambda_n$ à la mesure image $(T_x\varphi)_{\#}\lambda_n$, où $T_x\varphi$ désigne l'application affine tangente à φ en x :

$$T_x\varphi(y) = \varphi(x) + (d_x\varphi)(y - x).$$

Pour cela, on utilise le lemme suivant :

LEMME IV-34. Soient U et V deux ouverts de \mathbb{R}^n et $\varphi : U \rightarrow V$ un C^1 difféomorphisme. Alors pour tout $\delta \in]0, 1[$ et pour tout $x \in U$ il existe $r > 0$ tel que $B_r(x) \subset U$ et

(a) pour tout $y \in B_r(x)$,

$$(1 - \delta) |\det d_y \varphi| \leq |\det d_x \varphi| \leq (1 + \delta) |\det d_y \varphi|;$$

(b) pour tous $y, z \in B_r(x)$,

$$(1 - \delta) |T_x\varphi(y) - T_x\varphi(z)| \leq |\varphi(y) - \varphi(z)| \leq (1 + \delta) |T_x\varphi(y) - T_x\varphi(z)|.$$

PREUVE DU LEMME IV-34. Fixons x ; puisque φ est un difféomorphisme on a $|\det d_x \varphi| > 0$. Par continuité de $\det d\varphi$, il existe $\varepsilon > 0$ tel que

$$|x - y| \leq \varepsilon \implies |\det d_y \varphi - \det d_x \varphi| \leq \delta |\det d_x \varphi|,$$

ce qui implique la propriété (a).

Ensuite, soit $\eta \in]0, 1[$. Par continuité de $d\varphi$ (fonction à valeurs dans les applications linéaires), il existe $r > 0$ tel que $\|d_y \varphi - d_x \varphi\| \leq \eta$ pour $|x - y| \leq r$. Étant donnés y, z dans $B_r(x)$, on définit $x(t) = (1 - t)y + tz$, alors

$$\varphi(y) - \varphi(z) = \left(\int_0^1 (d_{x(t)} \varphi) dt \right) \cdot (y - z);$$

$$T_x\varphi(y) - T_x\varphi(z) = d_x\varphi \cdot (y - z);$$

d'où

$$\left| (\varphi(y) - \varphi(z)) - (T_x\varphi(y) - T_x\varphi(z)) \right| \leq \left(\int_0^1 \|d_{x(t)} \varphi - d_x \varphi\| dt \right) |y - z| \leq \eta |y - z|.$$

Par ailleurs $d_x\varphi$ est inversible, donc il existe une constante $K > 0$, ne dépendant que de x , telle que $|d_x\varphi \cdot (y - z)| \geq K |y - z|$ (il suffit de choisir $K = \|(d_x\varphi)^{-1}\|^{-1}$). On obtient ainsi

$$\left| |\varphi(y) - \varphi(z)| - |T_x\varphi(y) - T_x\varphi(z)| \right| \leq (\varepsilon K^{-1}) |T_x\varphi(y) - T_x\varphi(z)|.$$

La conclusion découle du choix $\eta = K\delta$. \square

Revenons maintenant à la preuve du Théorème IV-28. Soit $\delta \in]0, 1[$; soit $x \in U$ et soit $r > 0$ tels que les énoncés (a) et (b) du Lemme IV-34 soient vrais. On pose $U_x = B(x, r) \cap U$.

Les applications φ et $T_x\varphi$ sont bijectives, on peut donc définir $f = (1 + \delta)^{-1}\varphi \circ (T_x\varphi)^{-1}$, et l'inégalité de droite dans l'énoncé (b) montre que f est 1-lipschitzienne

sur $T_x\varphi(U_x)$. Par le Théorème IV-13 (une application contractante réduit les volumes), on sait que pour tout borélien $B' \subset T_x\varphi(U_x)$, on a $\lambda_n[f(B')] \leq \lambda_n[B']$. Si B est un borélien quelconque de U_x , on peut appliquer la relation précédente à $B' = T_x\varphi(B)$ et on trouve, $T_x\varphi$ étant bijective, $\lambda_n[(1+\delta)^{-1}\varphi(B)] \leq \lambda_n[T_x\varphi(B)]$. Puisque l'homothétie de rapport $(1+\delta)^{-1}$ contracte les volumes par un facteur $(1+\delta)^{-n}$, on conclut que

$$\lambda_n[\varphi(B)] \leq (1+\delta)^n \lambda_n[T_x\varphi(B)].$$

(Le lecteur familier avec la notion de mesure de Hausdorff pourra voir que cette conclusion découle immédiatement de l'inégalité de droite dans (b); ici j'ai utilisé un chemin légèrement détourné pour me ramener au Théorème IV-13.)

Un raisonnement identique à partir de l'inégalité de gauche dans (b) mène finalement à

$$(35) \quad (1-\delta)^n \lambda_n[T_x\varphi(B)] \leq \lambda_n[\varphi(B)] \leq (1+\delta)^n \lambda_n[T_x\varphi(B)];$$

cette inégalité est valable pour tout borélien B inclus dans U_x .

Soit maintenant K un compact de U . Par un raisonnement simple rappelé au cours de la première preuve du Théorème IV-28, on peut partitionner K en boréliens disjoints C_1, \dots, C_N tels que chaque C_i est inclus dans un U_{x_i} . On peut alors appliquer (35) à chaque C_i :

$$(36) \quad (1-\delta)^n \lambda_n[T_{x_i}\varphi(C_i)] \leq \lambda_n[\varphi(C_i)] \leq (1+\delta)^n \lambda_n[T_{x_i}\varphi(C_i)].$$

Pour chaque i , on peut appliquer le Théorème IV-19 pour calculer le volume de $T_{x_i}\varphi(C_i)$:

$$\lambda_n[T_{x_i}\varphi(C_i)] = |\det d_{x_i}\varphi| \lambda_n[C_i].$$

De la condition (a) on déduit alors

$$(1-\delta) \int_{C_i} |\det d_y\varphi| dy \leq \lambda_n[T_{x_i}\varphi(C_i)] \leq (1+\delta) \int_{C_i} |\det d_y\varphi| dy.$$

En reportant cette inégalité dans (36) on obtient

$$(1-\delta)^{n+1} \int_{C_i} |\det d_y\varphi| dy \leq \lambda_n[\varphi(C_i)] \leq (1+\delta)^{n+1} \int_{C_i} |\det d_y\varphi| dy.$$

Comme les $\varphi(C_i)$ forment une partition de $\varphi(K)$, en sommant cette double inégalité par rapport à l'indice i on trouve

$$\frac{(1-\delta)^n}{1+\delta} \int_K |\det d_y\varphi| dy \leq \lambda_n[\varphi(K)] \leq \frac{(1+\delta)^n}{1-\delta} \int_K |\det d_y\varphi| dy.$$

On conclut la preuve en faisant tendre δ vers 0. \square

IV-5. Mesurabilité, non-mesurabilité, et paradoxes de Banach–Tarski

IV-5.1. Ensembles boréliens et Lebesgue-mesurables. Pour passer de la tribu borélienne à la tribu des ensembles Lebesgue-mesurables, nous nous sommes contentés d'appliquer l'opération de complétion. Les ensembles Lebesgue-mesurables sont donc toutes les parties E de \mathbb{R}^n telles qu'il existe des boréliens A et B vérifiant

$$A \subset E \subset B, \quad |B \setminus A| = 0.$$

En particulier, tout sous-ensemble d'un ensemble négligeable est lui-même négligeable, ce qui semble naturel. De manière équivalente, les ensembles Lebesgue-mesurables sont ceux pour lesquels

$$\lambda^*[X \cap E] + \lambda^*[X \setminus E] = \lambda^*[X]$$

pour toute partie $X \subset \mathbb{R}$ (on peut se limiter au cas où X décrit l'ensemble des pavés). C'est cette dernière définition qu'adoptait Lebesgue.

La régularité de la mesure de Lebesgue nous permet de donner une autre caractérisation, en apparence un peu plus précise.

PROPOSITION IV-35 (Lebesgue-mesurabilité, F_σ et G_δ). *Tout ensemble Lebesgue-mesurable E peut s'écrire sous la forme $A \cup N$, où A est une union dénombrable de fermés (un F_σ) et N un ensemble négligeable; s'il est de mesure finie, il peut également s'écrire sous la forme $B \setminus N$, où B est une intersection dénombrable d'ouverts (un G_δ) et N un ensemble négligeable.*

DÉMONSTRATION. Si E est de mesure finie, c'est une conséquence de la Proposition I-49. Dans le cas où E n'est pas de mesure finie, on s'y ramène en considérant son intersection avec une suite croissante de pavés. \square

Ces descriptions sont bien sûr très grossières. Une branche de la **théorie géométrique de la mesure** [Federer] s'attache à décrire géométriquement les ensembles Lebesgue-mesurables.

IV-5.2. Fonctions boréliennes. On peut se demander à quoi ressemble une fonction borélienne, disons de \mathbb{R}^n dans \mathbb{R} , et si la classe des fonctions boréliennes est vraiment beaucoup plus large que la classe des fonctions continues, ou semi-continues... La question est bien sûr formulée ici de manière trop vague; cependant, le théorème de Lusin (Théorème II-60) implique que toute fonction borélienne $f : \mathbb{R}^n \rightarrow \mathbb{R}$, nulle en-dehors d'un ensemble de mesure de Lebesgue finie, coïncide avec une fonction continue en-dehors d'un ensemble de mesure arbitrairement petite; et que f est, en-dehors d'un ensemble de mesure nulle, limite simple de fonctions continues. Son corollaire II-61 indique que pour toute fonction f sommable de \mathbb{R}^n dans \mathbb{R} on peut trouver une famille $(f_k)_{k \in \mathbb{N}}$ de fonctions continues à support compact, telles que

$$\int |f_k(x) - f(x)| dx \xrightarrow[k \rightarrow \infty]{} 0.$$

Le théorème de Vitali-Carathéodory (Théorème II-63) s'applique aussi, sans restriction, à la mesure de Lebesgue sur \mathbb{R}^n , et permet d'encadrer une fonction sommable f par des fonctions semi-continues, au prix d'une erreur arbitrairement petite sur les intégrales.

IV-5.3. Existe-t-il des ensembles non mesurables ? Nous allons maintenant aborder quelques-unes des questions très subtiles liées au concept de non-mesurabilité, dont certaines touchent à rien moins que les fondations logiques des mathématiques.

Commençons par essayer de résoudre le problème de l'existence de parties non mesurables par un argument de cardinalité. On peut montrer [Rudin, p. 53] que la famille des Boréliens de \mathbb{R} a même cardinalité que \mathbb{R} . Si l'on admet que l'ensemble des parties de \mathbb{R} est de cardinal strictement supérieur à \mathbb{R} lui-même, il en découle que "la grande majorité" des sous-ensembles de \mathbb{R} ne sont pas des Boréliens.

Le raisonnement précédent n'est pas forcément très convaincant car il présuppose des résultats subtils de théorie des cardinaux. En tous les cas, il ne fonctionne plus pour la famille des ensembles Lebesgue-mesurables, qui a même cardinalité que l'ensemble de toutes les parties de \mathbb{R} : pour s'en convaincre, on peut noter que l'ensemble triadique de Cantor C sur $[0, 1]$ a même cardinal que \mathbb{R} , et que toutes ses parties sont mesurables puisqu'il est de mesure nulle. Si l'on admet que l'ensemble des parties de \mathbb{R} est "strictement plus grand" que \mathbb{R} lui-même, il en résulte que "la plupart" des parties de \mathbb{R} ne sont pas des boréliens.

Se pose alors la question de savoir s'il existe des parties qui ne soient pas Lebesgue-mesurables ! La réponse à cette question est un des résultats les plus frappants de la logique moderne : on peut effectivement construire de telles parties, mais leur construction nécessite l'**axiome du choix**. Si en revanche on ne postule pas cet axiome, la mesurabilité de toutes les parties de \mathbb{R} est un problème *indécidable*. On peut alors poser comme axiome que toutes les parties de \mathbb{R} sont mesurables, ou au contraire qu'il existe (au moins) une partie non mesurable ; les mathématiques que l'on pourra développer dans l'un et l'autre cas seront incompatibles, mais chacune aura a priori sa cohérence propre. Cette découverte majeure est due au célèbre logicien Solovay, aidé des techniques introduites par Cohen pour démontrer l'indécidabilité de l'"hypothèse du continu" (qui énonce, essentiellement, que le plus petit cardinal non dénombrable est celui de \mathbb{R}). Le théorème d'incomplétude de Gödel, l'indécidabilité de l'hypothèse du continu, et le théorème de Solovay, peuvent être considérés comme les résultats de logique les plus marquants du vingtième siècle.

Comment construire un ensemble non mesurable ? On se souvient que d'après la Proposition IV-4, la mesure de Lebesgue induit sur le tore \mathbb{T}^n une mesure invariante par addition modulo \mathbb{Z}^n . Si l'on construit un ensemble non mesurable E dans $\mathbb{T} = \mathbb{T}^1$, alors $E \times \mathbb{T}^{n-1}$ constituera un ensemble non mesurable de \mathbb{T}^n . C'est précisément ce que montre le résultat suivant.

THÉORÈME IV-36 (Paradoxe de Vitali). *Il existe une partie E de \mathbb{T} telle que \mathbb{T} puisse s'écrire comme réunion dénombrable disjointe de translatés de E . En particulier, E n'est pas Lebesgue-mesurable.*

DÉMONSTRATION. Notons que la seconde partie du théorème découle de la première : si E était mesurable, de mesure positive, alors la mesure de \mathbb{T} serait infinie, ce qui est faux ; et s'il était de mesure nulle, alors la mesure de E serait nulle, par σ -additivité.

Pour construire E , on introduit une relation d'équivalence \mathcal{R} dans \mathbb{T} comme suit :

$$x\mathcal{R}y \iff x - y \in \mathbb{Q}.$$

L'ensemble \mathbb{T} est alors partagé en une infinité de classes d'équivalence, et on choisit un représentant dans chaque classe ; on note E l'ensemble des représentants ainsi sélectionnés.

Toute classe d'équivalence s'obtient à partir de son représentant, par addition (modulo 1) de rationnels ; si l'on se limite à des rationnels de $[0, 1[$ on obtient des éléments distincts de \mathbb{T} . La conclusion est que \mathbb{T} est la réunion disjointe dénombrable des $q_k + E$, où les q_k sont les rationnels de $[0, 1[$ et l'addition est considérée modulo 1. \square

Dans le raisonnement précédent, nous avons utilisé l'axiome du choix pour *choisir* arbitrairement un représentant dans chaque classe d'équivalence. Au vu du théorème de Solovay, cela est inévitable. Notons que le contre-exemple de Sierpinski, mentionné dans la Remarque III-54, reposait également sur l'axiome du choix.

Mentionnons pour conclure deux autres “paradoxes” menant à l'existence d'ensembles non mesurables ; bien évidemment, tous deux reposent encore sur l'axiome du choix :

- Bernstein a construit un sous-ensemble $B \subset \mathbb{R}$ tel que B et $\mathbb{R} \setminus B$ intersectent tout sous-ensemble fermé non dénombrable de \mathbb{R} . En particulier, tout compact inclus dans B est au plus dénombrable, donc de mesure nulle, et la régularité de la mesure de Lebesgue impliquerait $|B| = 0$ si B était mesurable. De même on aurait $|\mathbb{R} \setminus B| = 0$... Ce paradoxe est étudié dans [Oxtoby, pp.22–23]
- Sierpinski a construit un ensemble $E \subset \mathbb{R}^2$ tel que (i) E intersecte tout ensemble fermé mesurable de \mathbb{R}^2 de mesure positive, et (ii) on ne peut pas trouver trois points de E alignés. Il s'ensuit que E n'est pas mesurable, sinon le théorème de Fubini impliquerait que $|E| = 0$ (puisque l'intersection de E avec toute droite verticale est réduite à au plus deux points), et on pourrait trouver un ensemble de mesure positive dans $\mathbb{R}^2 \setminus E$... On trouvera dans [Oxtoby, p.54–55] une preuve simplifiée (utilisant, outre l'axiome du choix, l'hypothèse du continu).

À ce stade, on pourrait encore conserver l'espoir de définir une mesure sur toutes les parties de \mathbb{R} , même si elle ne pourrait vérifier ni l'invariance par translation (à cause du paradoxe de Vitali), ni la régularité (à cause du paradoxe de Bernstein), ni le théorème de Fubini (à cause du paradoxe de Sierpinski). Au vu de toutes ces restrictions, on peut douter de l'intérêt qu'aurait une telle mesure ; quoi qu'il en soit, ce dernier espoir est ruiné par le théorème suivant, dû à Banach et Kuratowski, généralisé par Ulam [Dudley, pp.526–527 ; Billingsley, p.37].

THÉORÈME IV-37. *Soit μ une mesure finie sur la σ -algèbre de toutes les parties de $[0, 1]$, telle que $\mu[\{x\}] = 0$ pour tout $x \in [0, 1]$. Alors μ est identiquement nulle.*

Ce théorème également utilise l'axiome du choix et l'hypothèse du continu. Ici l'obstruction a trait à la théorie des cardinaux, comme le montre une généralisation abstraite due à Ulam (voir [Oxtoby, p. 25–26]).

IV-5.4. Contre-exemple de Hausdorff. L'existence d'ensembles non mesurables pourrait être considérée comme un défaut majeur de la théorie de la mesure telle que nous l'avons évoquée. Après tout, en mécanique classique, tous les objets ont une masse, et la théorie de la mesure peut être vue comme une tentative de formaliser le concept de masse dans un cadre abstrait. Un coupable tout désigné serait le fameux axiome de σ -additivité, imposé pour des raisons mathématiques. Se pose alors la question naturelle de savoir si l'on peut remplacer la mesure de Lebesgue par une fonction additive d'ensembles, ou “**mesure finiment additive**”, à savoir une fonction μ vérifiant $\mu[\emptyset] = 0$, et $\mu[A \cup B] = \mu[A] + \mu[B]$ quand A et B sont disjoints, qui serait définie sur l'ensemble de *toutes* les parties de \mathbb{R}^n . On souhaite en outre que cette fonction additive conserve la propriété naturelle d'invariance par isométrie affine. Nous voici donc face à la

Question : *Existe-t-il des fonctions additives d'ensembles définies sur toutes les parties de \mathbb{R}^n , non triviales, finies sur les compacts et invariantes par isométries ?*

La réponse à cette question est encore négative. Pour l'expliquer, introduisons le concept d'**ensemble paradoxal**, qui généralise à un cadre finiment additif l'idée utilisée dans le contre-exemple de Vitali. Dans la suite, nous emploierons le terme de "mesure" pour des fonctions additives qui ne sont pas forcément σ -additives, aucune confusion n'étant possible.

DÉFINITION IV-38 (découpage et recollement). *Soient A et B deux ensembles de \mathbb{R}^n . On dit que A peut être découpé et recollé en B s'il existe une partition finie de A en morceaux A_1, \dots, A_k , et des déplacements (i.e. des isométries affines de déterminant 1) g_1, \dots, g_k tels que les morceaux $g_1(A_1), \dots, g_k(A_k)$ forment une partition de B .*

Remarquons que cette définition ne correspond pas de très près au concept intuitif de "découpage et recollement" : il se peut que les parties A_i soient imbriquées de manière très complexe, de sorte que leur séparation physique soit impossible. Cependant, c'est une approximation naturelle de ce concept.

DÉFINITION IV-39 (ensemble paradoxal). *Un ensemble de \mathbb{R}^n est dit paradoxal si on peut le découper et le recoller en deux copies disjointes de lui-même.*

Il est naturel de penser que de tels ensembles sont forcément de mesure nulle, sinon on aurait une contradiction apparente avec le fait que les deux copies disjointes doivent avoir deux fois le volume de l'objet originel. L'existence même d'ensembles paradoxaux semble douteuse. Le théorème suivant, dû à Hausdorff [Wagon, p. 18], répond à cette question.

THÉORÈME IV-40 (paradoxe de Hausdorff). *Il existe un sous-ensemble dénombrable de la sphère S^2 tel que $S^2 \setminus D$ soit paradoxal.*

Ce paradoxe, qui utilise lui aussi l'axiome du choix, repose sur la constatation suivante, d'une grande importance en théorie des groupes : le groupe SO_3 des déplacements linéaires de \mathbb{R}^3 , qui laisse la sphère S^2 invariante, admet pour sous-groupe une copie du groupe libre à deux éléments. Un système explicite de générateurs peut d'ailleurs être construit, voir [Wagon, p. 15]. Voyons maintenant ce que l'on peut déduire de ce paradoxe.

COROLLAIRE IV-41 (non-existence de mesures finiment additives). *Soit μ une mesure finiment additive définie sur toutes les parties de \mathbb{R}^3 , invariante par déplacement et finie sur les compacts. Alors $\mu[K] = 0$ pour tout compact de \mathbb{R}^3 .*

DÉMONSTRATION. Soit μ une mesure vérifiant les hypothèses, non nulle sur les compacts ; il existe donc $R > 0$ tel que $\mu[RB^3] > 0$, où B^3 est la boule unité dans \mathbb{R}^3 . Quitte à remplacer μ par $(m_{1/R})_\# \mu$, où $m_a(x) = ax$, on peut supposer que $\mu[B^3] > 0$. Tous les singletons ont même mesure pour μ , forcément nulle, sinon $\mu[A]$ serait infini pour toute partie A infinie. La restriction de μ à la boule privée de son centre est donc une mesure finiment additive, bien définie et invariante par l'action de SO_3 .

L'application $x \mapsto x/|x|$ envoie la boule privée de son centre sur la sphère S^2 , et transporte donc la mesure μ en une mesure non nulle sur S^2 , qui reste finiment additive et invariante par l'action de SO_3 ; on la notera toujours μ . En particulier, si D est la partie dénombrable apparaissant dans le paradoxe de Hausdorff, on a

$$\mu[S^2 \setminus D] = 2\mu[S^2 \setminus D],$$

ce qui montre que $\mu[S^2 \setminus D] = 0$.

Il reste à vérifier que D est de mesure nulle pour aboutir à une contradiction. Soit ℓ une ligne issue de l'origine, n'intersectant pas D ; on définit R_θ comme la rotation d'angle θ autour de ℓ . L'ensemble des angles θ tels que R_θ envoie au moins un élément de D dans D est dénombrable; en conséquence, il existe au moins un angle θ pour lequel D et $R_\theta(D)$ sont disjoints. On en déduit que $\mu[D'] = 2\mu[D] \leq \mu[S^2]$, où $D' = D \cup R_\theta(D)$; en particulier $\mu[D] \leq \mu[S^2]/2$. Comme D' lui-même est dénombrable, le même raisonnement montre que $\mu[D'] \leq \mu[S^2]/2$, en particulier $\mu[D] \leq \mu[S^2]/4$. Par récurrence, on montre que $\mu[D] \leq \mu[S^2]/2^m$ pour tout $m \geq 1$, d'où finalement $\mu[D] = 0$. \square

IV-5.5. Paradoxe de Banach–Tarski. Du paradoxe de Hausdorff on pourrait déduire qu'il est trop ambitieux de chercher à “mesurer” **toutes** les parties de l'espace Euclidien, au moins dans \mathbb{R}^3 . Les paradoxes dits de Banach–Tarski vont plus loin et mènent à s'interroger sur le bien-fondé de l'axiomatique mathématique habituelle.

THÉORÈME IV-42 (paradoxes de Banach–Tarski). *(i) Pour tout $n \geq 3$, la boule unité de \mathbb{R}^n est paradoxale : on peut la découper et la recoller en deux boules disjointes de rayon 1 (pour cela, cinq morceaux sont nécessaires et suffisants).*

(ii) Soient A et B deux parties de \mathbb{R}^n d'intérieur non vide, $n \geq 3$. Alors on peut découper A et le recoller en B .

Ces paradoxes sont discutés en détail dans [Wagon]. Le premier énoncé, considéré par certains comme “le théorème le plus surprenant de toutes les mathématiques”, est suffisamment incroyable pour qu'on l'énonce sous une forme encore plus explicite : *Il est possible de découper une boule de rayon 1 en 5 morceaux A_1, \dots, A_5 et de trouver des isométries g_1, \dots, g_5 telles que l'union des $g_i(A_i)$ forme deux boules disjointes, chacune étant de rayon 1.*

En corollaire, au moins pour $n \geq 3$, il est impossible de définir une mesure finiment additive sur l'ensemble de toutes les parties de \mathbb{R}^n , au moins pour $n \geq 3$.

Il se trouve que la conclusion est différente pour $n \leq 2$!

THÉORÈME IV-43 (existence de mesure finiment additive dans \mathbb{R}^2). *Pour $n = 1$ ou $n = 2$, il est possible de définir une mesure finiment additive μ sur toutes les parties de \mathbb{R}^n , qui soit invariante par déplacement et coïncide avec la mesure de Lebesgue sur tous les pavés.*

Tous ces résultats, et bien d'autres qui leur sont liés, sont démontrés et commentés dans [Wagon]. Le théorème IV-43 ne prétend pas que la mesure μ coïncide avec la mesure de Lebesgue sur tous les Boréliens, ce qui laisse un doute quant à la possibilité de l'utiliser. Quoi qu'il en soit, la différence de comportement entre les dimensions inférieures ou égales à 2 d'une part, et supérieures ou égales à 3 d'autre part, reflète des différences fondamentales dans la structure des groupes d'isométries correspondants, qui a donné naissance à la notion de groupe moyennable, aujourd'hui une branche importante de la théorie géométrique des groupes.

La conclusion du paradoxe de Banach–Tarski, qui ne fait pas intervenir le concept de mesure, semble si choquante pour le sens commun, que l'on peut se demander s'il ne faut pas revoir l'ensemble des axiomes qui a permis de l'établir. C'est l'occasion de discuter un peu plus en détail de l'axiome du choix...

IV-5.6. Axiome du choix. L'écrasante majorité des démonstrations mathématiques, en-dehors du domaine de la logique, repose sur un ensemble d'axiomes "incontestables", appelé couramment théorie des ensembles ZF (Zermelo–Fraenkel). Cette théorie permet de construire les entiers, les rationnels, les réels, etc. On peut ajouter, ou pas, à cette théorie l'axiome dit axiome du choix, qui énonce, essentiellement, que le produit d'ensembles non vides est toujours non vide. En clair, étant donnée une collection d'ensembles non vides, on peut **choisir** un élément dans chacun d'entre eux, et rassembler tous les éléments ainsi choisis en un ensemble. Cet axiome peut paraître inoffensif, mais il est suffisant à aboutir à des paradoxes tels que ceux de Vitali, Hausdorff ou Banach–Tarski. En outre il n'est pas très intuitif, car dans le cas où la famille d'ensembles est infinie on ne pourra jamais construire explicitement leurs représentants.

D'un autre côté, si l'on supprime complètement l'axiome du choix, on tombe vite sur des paradoxes qui heurtent le sens mathématique commun. Ainsi, ZF est compatible avec l'assertion selon laquelle \mathbb{R} est union dénombrable d'ensembles dénombrables... Cela montre bien que nous utilisons sans nous en rendre compte des versions de l'axiome du choix. L'argument diagonal de Cantor en fournit un exemple !

Une variante de l'axiome du choix qui permet d'éviter ce genre de paradoxe, et en même temps n'est pas assez forte pour impliquer l'existence d'ensembles non mesurables, est l'**axiome du choix dénombrable**, qui énonce qu'un produit dénombrable d'ensembles non vides est non vide.

Une variante légèrement plus forte est l'**axiome du choix dépendant** : étant donné une famille dénombrable d'ensembles E_n , pour tout n on peut choisir $x_n \in E_n$, dépendant de x_{n-1} d'une manière que l'on spécifie. La théorie obtenue en ajoutant cet axiome à ZF permet d'effectuer (presque) tous les raisonnements habituels en analyse, et reste compatible avec la non-existence d'ensembles mesurables. En particulier, la démonstration des paradoxes de Vitali, Hausdorff et Banach–Tarski est impossible dans ce contexte.

Notons pour finir qu'il existe un autre axiome célèbre que l'on peut ou pas imposer, et que nous avons déjà rencontré dans certains paradoxes, c'est l'**hypothèse du continu**, à savoir que \mathbb{R} est le plus petit ensemble infini non dénombrable.

IV-5.7. Quelle attitude adopter ? Au vu de la discussion précédente, il y a (au moins) trois attitudes possibles :

- *l'attitude classique* : accepter l'axiome du choix, se résoudre à ce que certains ensembles ne soient pas mesurables, et vérifier la mesurabilité des objets avec lesquels on travaille, quand il y a besoin ;
- *l'attitude iconoclaste* : n'accepter que l'axiome du choix dépendant, postuler la mesurabilité de toutes les parties de \mathbb{R} , et se résoudre à ce que l'axiome du choix ne soit pas vrai ;
- *l'attitude sceptique* : s'interdire l'usage de l'axiome du choix, mais sans supposer a priori la mesurabilité de toutes les parties de \mathbb{R}^n .

L'attitude classique est le choix le plus répandu dans les mathématiques des dernières décennies. L'axiome du choix mène parfois à des démonstrations relativement élégantes et à des énoncés synthétiques ; nous le verrons par exemple quand nous parlerons de mesure de Haar au Chapitre ???. L'axiome du choix apparaît sous de nombreux avatars tels que le Lemme de Zorn, ou le Théorème de Tychonov

dans sa version la plus générale (un produit de compacts est compact). Il a bien sûr le défaut considérable de mener aux paradoxes de Banach–Tarski évoqués plus haut ; mais on pourra objecter que les mathématiques contiennent d’autres énoncés paradoxaux qui ne nécessitent pas l’axiome du choix.

L’attitude iconoclaste est quelque peu dangereuse, car cela réclame une discipline importante que de savoir quels sont les résultats qui nécessitent l’axiome du choix et quels sont ceux qui ne le réclament pas. Ainsi, le théorème de Hahn–Banach, ou le théorème de Banach–Alaoglu, dans leurs versions les plus générales, nécessitent l’axiome du choix. En outre, l’axiome que l’on adopte (la non-existence de parties non mesurables) est un axiome très fort, et probablement inutile.

C’est finalement l’attitude sceptique que je recommande vivement. Elle n’exclut pas, bien sûr, de recourir à l’axiome du choix dans une phase prospective de recherche de preuve.

IV-5.8. Justification pratique de la mesurabilité. Comme nous l’avons vu, on ne peut construire une application non mesurable que si on cherche absolument à le faire ; justifier la mesurabilité d’une application est donc en général une opération de routine. En pratique, il n’y a guère qu’une situation à laquelle il faut prendre garde : si $f(x, y)$ est une fonction de deux variables, disons réelles, il n’y a pas de raison a priori pour que la fonction

$$f : x \longmapsto \sup_y f(x, y)$$

soit mesurable. Dans la pratique, on cherche toujours, en présence d’une telle situation, à se ramener à un supremum pris sur une famille dénombrable.

EXEMPLE IV-44. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction localement λ_n -sommable ; on définit la **fonction maximale** de f par

$$Mf(x) := \sup_{r>0} \frac{1}{|B_r(x)|} \int_{B_r(x)} f,$$

où $B_r(x)$ désigne la boule Euclidienne de rayon r , centrée en x . La famille des $r > 0$ n’est pas dénombrable, la mesurabilité de Mf n’est donc pas évidente a priori. Cependant, étant donnés deux rationnels q et q' tels que $q \leq r \leq q'$, on a

$$|B_q(x)| \leq |B_r(x)| \leq |B_{q'}(x)|; \quad \int_{B_q(x)} f \leq \int_{B_r(x)} f \leq \int_{B_{q'}(x)} f.$$

Le supremum sur tous les nombres réels positifs peut donc être remplacé par un supremum sur tous les nombres rationnels positifs. La fonction maximale est donc effectivement mesurable !

Une situation du même type, dont on ne peut malheureusement pas se tirer si facilement, survient quand on considère des sommes de Minkowski :

$$A + B := \{x \in \mathbb{R}^n; \exists a \in A; \exists b \in B; x = a + b\}.$$

Dans un tel cas, la mesurabilité de A et B n’implique pas forcément celle de $A + B$. Pour y remédier, on fera par exemple l’hypothèse que A et B sont compacts, auquel cas $A + B$ l’est également ; ou bien on utilisera des mesures extérieures.

REMARQUE IV-45. On éprouve parfois le besoin, en théorie géométrique de la mesure, de définir les ensembles sousliniens comme les images des ensembles Lebesgue-mesurables par des applications continues ; on peut donner une description plus ou moins explicite de tels ensembles, en termes d'unions transfinies d'intersections de fermés, voir [Falconer1, p. 6]. Il est clair que la somme de Minkowski de deux ensembles mesurables fait partie de cette catégorie...

IV-5.9. Subtilités liées au produit. Plus graves que les exemples précédents sont les subtilités liées à la structure de produit tensoriel. En effet, si l'on note $\mathcal{L}(\mathbb{R}^n)$ la tribu des ensembles Lebesgue-mesurables en dimension n , et si l'on admet l'existence de parties non mesurables, alors

$$\mathcal{L}(\mathbb{R}^m) \otimes \mathcal{L}(\mathbb{R}^n) \neq \mathcal{L}(\mathbb{R}^{m+n})!$$

Pour s'en convaincre, il suffit de choisir un ensemble non Lebesgue-mesurable X dans \mathbb{R} , et de l'envoyer sur la diagonale ($y = x$) dans \mathbb{R}^2 , via l'application $\phi : x \mapsto (x, x)$, ou sur l'axe horizontal ($y = 0$) dans \mathbb{R}^2 , via l'application $\phi : x \mapsto (x, 0)$. Comme la diagonale (ou l'axe horizontal) est de mesure nulle dans \mathbb{R}^2 , $\phi(X)$ est négligeable, et en particulier mesurable ; mais son image réciproque par ϕ n'est pas mesurable. La conclusion est que la tribu $\mathcal{L}(\mathbb{R}) \otimes \mathcal{L}(\mathbb{R})$ ne contient pas $\phi(X)$; cette tribu est en fait strictement plus petite que la tribu $\mathcal{L}(\mathbb{R}^2)$.

Comme corollaire particulièrement déplaisant de ce contre-exemple, la formule de **découpage en tranches**,

$$\lambda_2[A] = \int_{\mathbb{R}} \lambda_1[A_x] dx,$$

où $A_x := \{y \in \mathbb{R}; (x, y) \in A\}$ *n'est pas valide* pour un ensemble Lebesgue-mesurable quelconque (sauf si l'on postule que toutes les parties sont mesurables...) Cette formule redevient valide si l'on définit les tranches après avoir modifié A sur un ensemble de mesure nulle “bien choisi” (voir [Rudin, pp.168-169]).

IV-5.10. Quelle tribu utiliser ? La tribu de Lebesgue est obtenue à partir de la tribu borélienne “simplement” par complétion. A priori, cette opération devrait rendre les raisonnements plus simples. Cependant, le paragraphe précédent a bien montré qu'elle implique des subtilités importantes. C'est pourquoi je recommande de **travailler uniquement avec la tribu borélienne** chaque fois que cela est possible, c'est-à-dire dans la grande majorité des problèmes d'analyse. L'ouvrage [Lieb–Loss] est un exemple de traité d'analyse réelle entièrement basée sur la tribu borélienne.

CHAPITRE V

Les mesures de Hausdorff

Dans ce chapitre, nous étudions les mesures de Hausdorff, qui généralisent la mesure de Lebesgue. Il est utile, dans de nombreux domaines des mathématiques, d'être un tant soit peu familier avec ces mesures, même si elles n'ont pas la même importance pratique que la mesure de Lebesgue.

V-1. Motivations

La théorie des mesures de Hausdorff est née une quinzaine d'années après celle de la mesure de Lebesgue, et fut développée principalement par Besicovich pendant les quarante années qui ont suivi. Elle répondait à plusieurs motivations.

V-1.1. Mesures d'objets de dimension inférieures. Plaçons-nous en dimension 3 pour simplifier la discussion. La mesure de Lebesgue λ_3 permet d'attribuer à toutes les parties (mesurables) de \mathbb{R}^3 un "volume" ; mais dans de nombreux problèmes on a besoin de définir l'aire d'une surface, ou la longueur d'une courbe tracée dans \mathbb{R}^3 . La mesure de Lebesgue de tels objets est bien sûr nulle, ce qui suggère l'introduction de nouvelles mesures pour définir les concepts d'"aire" ou de "longueur" de parties de \mathbb{R}^3 . Bien sûr, on s'attend à ce que l'aire d'un objet soit infinie si son volume est non nul, de sorte que ces nouvelles mesures seraient intéressantes uniquement quand on les appliquerait à des ensembles Lebesgue-négligeables.

C'est dans cette perspective que Carathéodory construisit, vers 1914, des mesures de dimension k dans \mathbb{R}^n , avec $1 \leq k < n$, grâce à la notion de mesure extérieure qu'il venait de développer.

V-1.2. Changements de variables. Nous avons vu au chapitre précédent des formules faisant intervenir un changement de variables T entre sous-ensembles de \mathbb{R}^n , et noté l'apparition du déterminant Jacobien $|\det \nabla T|$. Qu'advient-il si notre changement de variables fait intervenir des fonctions $\mathbb{R}^m \rightarrow \mathbb{R}^n$, avec, par exemple, $m > n$? Nous avons déjà rencontré un exemple très simple : le théorème de Fubini peut être considéré comme un changement de variables $z \in \mathbb{R}^{m+n} \rightarrow (x, y) \in \mathbb{R}^m \times \mathbb{R}^n$ avec $z = (x, y)$, et on peut écrire

$$\int_{\mathbb{R}^{m+n}} f(z) d\lambda_{m+n}(z) = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} f(x, y) d\lambda_n(y) \right) d\lambda_m(x).$$

Bien sûr, dans ce cas il n'y a aucun problème car le changement de variables correspond à un produit cartésien ; mais que se passe-t-il quand les choses sont plus complexes ?

Un exemple familier et très utile est le **changement de variables polaire** (aussi appelé changement de variables sphérique), dans lequel on troque la variable $x \in \mathbb{R}^n \setminus \{0\}$ pour le couple $(r, \sigma) \in \mathbb{R}_+ \times S^{n-1}$, avec $r := |x|$ et $\sigma := x/|x|$. Comment écrire la formule de changement de variables correspondante ? Voici comment pourrait raisonner un physicien ou un ingénieur : *Faisons varier r dans un "intervalle*

infinitésimal” $[r - dr, r + dr]$ et σ à l’intérieur d’un “disque infiniment petit” tracé sur la sphère S^{n-1} , de centre σ et de surface $d\sigma$. La région ainsi visitée par le point x est, à des infiniment petits d’ordre supérieur près, un cylindre centré en $r\sigma$, dont la hauteur est $2dr$ et la section a (par homogénéité) une “surface” $r^{n-1}d\sigma$. On en déduit la formule de changement de variable

$$dx = r^{n-1} dr d\sigma.$$

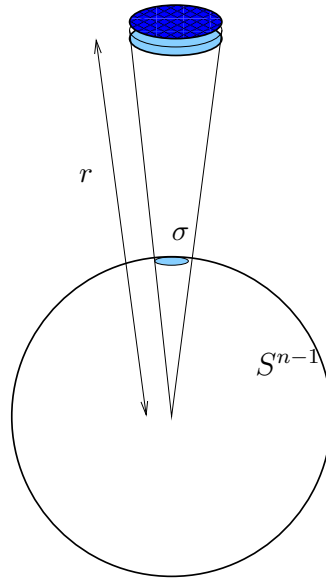


FIG. 1. Élément de volume autour de x dans les variables sphériques

Quelle est la signification de ce symbole $d\sigma$, que l’on peut interpréter comme une “mesure d’aire infinitésimale sur la sphère S^{n-1} ”? Il s’agit bien de l’élément d’intégration par rapport à une mesure σ sur S^{n-1} , que l’on peut introduire comme la restriction à S^{n-1} de la mesure de Hausdorff de dimension $n - 1$ dans \mathbb{R}^n . Il est donc parfaitement licite d’écrire

$$\int_{\mathbb{R}^n} f(x) dx = \int_{\mathbb{R}_+} \left(\int_{S^{n-1}} f(r\sigma) d\sigma \right) r^{n-1} dr.$$

Il y a bien d’autres façons de définir σ : par exemple, comme l’élément de volume sur S^{n-1} , vu comme une variété Riemannienne. On peut également, en basse dimension la définir au moyen de coordonnées explicites, ce qui est commode pour effectuer des calculs : ainsi, quand $n = 2$, on peut identifier σ à un angle dans $[0, \pi[$ et écrire $dx = r dr d\theta$; quand $n = 3$ on introduit traditionnellement deux “angles solides” $\theta \in [0, \pi]$ et $\phi \in [0, 2\pi[$, tels que (par exemple) $\sigma = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$, et alors la formule correspondante est $dx = r^{n-1} dr \sin \theta d\theta d\phi$... Mais c’est l’interprétation en termes de restriction de mesure de Hausdorff qui s’avère conceptuellement la plus naturelle pour généraliser la formule de changement de variables.

V-1.3. Notion de dimension. Nous avons l’habitude de penser qu’une courbe “régulière” est de dimension 1, car on peut localement la déformer “continûment” en un morceau de droite ; de manière plus générale, il est naturel de penser à un

ensemble comme étant de dimension k si on peut le décrire localement au moyen de k fonctions indépendantes ; en particulier l'image par une application régulière d'un ensemble de dimension k devrait être de dimension au plus k .

De tels énoncés sont effectivement vrais quand on travaille avec des fonctions régulières ; mais il est possible de construire des courbes continues surjectives de $[0, 1]$ dans $[0, 1]^2$, dites “courbes de Peano”. L'image d'une telle courbe est incontestablement de dimension 2... Cet exemple montre bien qu'il est impossible de définir une notion de dimension qui soit basée sur les dimensions d'espaces de départ et d'arrivée. Comment déterminer si l'image d'une fonction continue $[0, 1] \rightarrow [0, 1]^2$, non surjective, doit être considérée comme étant de dimension 0, 1 ou 2 ?

Le point de vue adopté en théorie de la mesure est le suivant : pour définir la dimension d'un objet, on essaie de le mesurer par toute une famille de mesures, qui sont associés à des objets de dimension déterminée. Ainsi, si un objet a une longueur positive non nulle, il est naturel de penser qu'il est de dimension 1 ; s'il a une surface positive non nulle, de dimension 2. Ce sont les mesures de Hausdorff qui vont jouer ce rôle en définissant rigoureusement les notions de longueur, surface, etc.

Comme l'a découvert Hausdorff vers 1919, il est en fait possible de définir ces mesures pour des dimensions non entières, et d'en déduire une notion de dimension qui peut elle aussi être non entière. C'est cette contribution, techniquement simple mais conceptuellement remarquable, qui a valu à son nom de rester attaché aux mesures de Hausdorff et à la dimension ainsi définie, dite **dimension de Hausdorff**.

La dimension de Hausdorff permet de définir un ordre dans la notion de négligeabilité : plutôt que de dire qu'un objet est de mesure de Lebesgue nulle, on pourra souvent dire plus précisément qu'il est de telle ou telle dimension de Hausdorff. Comme on s'y attend, un point sera de dimension 0, un segment de droite de dimension 1, etc.

On note dès à présent qu'il existe une autre notion de dimension, antérieure à celle de dimension de Hausdorff, qui est souvent plus simple à manipuler, même si son usage est moins courant que celui de la dimension de Hausdorff : c'est la **dimension de Minkowski**. Cependant, l'un des grands avantages du formalisme de Hausdorff, c'est qu'il fournit à la fois une notion de dimension et une notion de mesure.

Dans les dernières décennies, l'étude des objets fractals s'est développée considérablement, motivée par les progrès de l'informatique, les suggestions visionnaires de Mandelbrot, et la découverte de formalismes fractals dans des domaines aussi variés que la mécanique des fluides, les systèmes dynamiques chaotiques, la théorie du signal, etc. Les mesures de Hausdorff, déjà très utilisées par les spécialistes de théorie géométrique de la mesure, en particulier dans le domaine du calcul des variations, se sont alors imposées comme l'un des outils-clés dans l'étude des objets fractals [Falconer].

V-1.4. Mesures de références abstraites. Comme nous l'avons vu au chapitre précédent, ses propriétés d'invariance font de la mesure de Lebesgue une mesure de référence naturelle dans \mathbb{R}^n . Si maintenant on se donne un espace métrique abstrait (X, d) , peut-on y définir une mesure borélienne de référence “naturelle” ? Les mesures de Hausdorff sont de bons candidats pour cela. En effet, pour toute dimension N , on peut définir a priori sur (X, d) une “mesure de Hausdorff de dimension N ” (qui malheureusement sera souvent triviale). Ainsi, si X est une variété de dimension n , muni de sa distance géodésique, alors la mesure de Hausdorff N -dimensionnelle

sur X coïncidera avec la mesure de volume quand $N = n$, et avec la mesure nulle quand $N > n$.

V-2. Construction des mesures de Hausdorff

V-2.1. Définition. La mesure de Lebesgue, ou longueur, d'une partie A de \mathbb{R} est définie comme l'infimum des sommes des longueurs des intervalles recouvrant A :

$$|A| := \inf \left\{ \sum \ell(I_k); I_k \text{ intervalle}; A \subset \bigcup I_k \right\}.$$

C'est cette définition que l'on a envie de généraliser. En dimension plus grande que 1, les candidats naturels pour jouer le rôle d'intervalles sont les boules. Un calcul assez simple (basé sur un changement de variable polaire!) montre que le volume de la boule de rayon r en dimension $d \in \mathbb{N}$ est

$$|B_r|_d = \alpha(d)r^d, \quad \alpha(d) := \frac{\pi^{d/2}}{\Gamma(\frac{d}{2} + 1)},$$

où $\Gamma(x) := \int_0^\infty e^{-t} t^{x-1} ds$ est la fonction Γ habituelle.

Si l'on cherche à définir des dimensions d non entières, il est naturel (mais non obligatoire) d'utiliser la même formule pour $\alpha(d)$, ce qui revient à prolonger par analyticit   la fonction "volume d'une boule de rayon r en dimension d ".

Voici maintenant une premi  re tentative de construction de la "mesure d -dimensionnelle", copiant la d  finition de la mesure de Lebesgue :

$$\mu^d[A] := \inf \left\{ \sum_{k=1}^{\infty} \alpha(d)r_k^d; A \subset \bigcup_{k=1}^{\infty} B_{r_k}(x_k) \right\} \quad ?$$

On se rend compte tout de suite que cette d  finition est absurde : le volume 1-dimensionnel d'une boule de \mathbb{R}^2 serait fini ! Le probl  me vient de ce que la notion de dimension doit d  pendre uniquement de la structure locale d'un objet, et que donc on doit forcer le recouvrement par des boules    "  pouser les d  tails" de l'ensemble A ; autrement dit, il faut d  finir la mesure d -dimensionnelle en fonction de *recouvrements par des petites boules*. Avec la mesure de Lebesgue sur \mathbb{R} , cette propri  t     tait superflue : un "gros" intervalle de longueur L peut se partager en L/ε "petits" intervalles de longueur ε , et les deux recouvrements ainsi obtenus sont   quivalents en termes de mesure.

Nous arrivons ainsi    une deuxi  me tentative de d  finition de "mesure d -dimensionnelle" :

$$\mu^d[A] := \liminf_{\varepsilon \rightarrow 0} \left\{ \sum_{k=1}^{\infty} \alpha(d)r_k^d; A \subset \bigcup_{k=1}^{\infty} B_{r_k}(x_k); r_k \leq \varepsilon \right\}.$$

La mesure ainsi d  finie est dite **mesure de Hausdorff sph  rique** [Falconer1, p. 7]. Elle a le d  faut de reposer sur la notion de boule, qui n'est pas invariante par restriction : si $A \subset \mathbb{R}^n$, et $B \subset \mathbb{R}^n$ est une boule de rayon r , alors $A \cap B$ n'est pas forc  ment une boule dans A (il suffit que le centre de la boule n'appartienne pas    A ...). Ce qui est vrai en revanche, c'est que le diam  tre de $A \cap B$ est inf  rieur ou   gal    $2r$.

Pour avoir une notion aussi intrins  que que possible, et   tre s  r que la mesure d'un objet ne d  pend pas de la taille de l'espace dans lequel on le plonge, on souhaiterait donc d  finir les mesures de Hausdorff en fonction des diam  tres, sans r  f  rence

à la notion de boule. Nous arrivons ainsi à la définition finalement retenue pour les mesures de Hausdorff :

DÉFINITION V-1 (mesure de Hausdorff). *Soient $A \subset \mathbb{R}^n$, et $d \in \mathbb{R}_+$. On définit la mesure de Hausdorff d -dimensionnelle de A par*

$$(37) \quad \mathcal{H}^d[A] = \liminf_{\varepsilon \rightarrow 0} \left\{ \sum_{k=1}^{\infty} \alpha(d) r(C_k)^d; \quad A \subset \bigcup_{k=1}^{\infty} C_k \quad \text{diam}(C_k) \leq \varepsilon \right\},$$

où les C_k sont des parties arbitraires de \mathbb{R}^n , $r(C_k) := \text{diam}(C_k)/2$ est le demi-diamètre de C_k , et

$$\alpha(d) := \frac{\pi^{d/2}}{\Gamma(\frac{d}{2} + 1)}$$

est le “volume de la boule unité de dimension d ”.

REMARQUES V-2. (i) Posons

$$\mathcal{H}_\varepsilon^d[A] = \inf \left\{ \sum_{k=1}^{\infty} \alpha(d) r(C_k)^d; \quad A \subset \bigcup_{k=1}^{\infty} C_k \quad \text{diam}(C_k) \leq \varepsilon \right\}.$$

Comme $\mathcal{H}_\varepsilon^d[A]$ est clairement une fonction décroissante de ε , l'existence de $\lim_{\varepsilon \rightarrow 0} \mathcal{H}_\varepsilon^d[A]$ est assurée, et cette limite est un supremum.

(ii) Soit A tel que $\mathcal{H}^d[A] < +\infty$, alors pour tout $\varepsilon > 0$ on a $\mathcal{H}_\varepsilon^d[A] < +\infty$, et pour tout $\delta > 0$ on peut trouver une famille dénombrable (C_k) d'ensembles de diamètre au plus ε , recouvrant A , telle que

$$\sum_{k=1}^{\infty} \alpha(d) r(C_k)^d \leq \mathcal{H}_\varepsilon^d[A] + \delta.$$

Pour tout $\varepsilon_k > 0$, l'ensemble C'_k des points dont la distance à C_k est strictement inférieure à ε_k est un ouvert contenant A ; en choisissant ε_k suffisamment petit, on peut faire en sorte que les quantités $\sum r(C_k)^d$ et $\sum r(C'_k)^d$ ne diffèrent pas de plus que δ . On a donc l'énoncé suivant : Pour tous $\delta > 0$, $\varepsilon > 0$ et $\varepsilon' > \varepsilon$ on peut trouver une famille dénombrable (C'_k) d'ouverts, de diamètre au plus ε' , recouvrant A , telle que

$$\sum_{k=1}^{\infty} \alpha(d) r(C'_k)^d \leq \mathcal{H}_\varepsilon^d[A] + \delta.$$

En remplaçant les C_k par les $\overline{C_k}$, on voit également que le mot “ouverts” dans l'énoncé précédent peut être remplacé par “fermés”. En faisant ensuite tendre ε et ε' vers 0, on vérifie facilement que la définition de la mesure de Hausdorff est inchangée si l'on impose au recouvrement d'être constitué d'ensembles ouverts (resp. fermés).

L'énoncé suivant justifie la terminologie “mesure de Hausdorff”.

PROPOSITION V-3 (la mesure de Hausdorff est une mesure de Borel). *Pour tout $d \geq 0$, la fonction $A \mapsto \mathcal{H}^d[A]$ est une mesure extérieure sur \mathbb{R}^n , et définit une mesure sur la tribu Borélienne $\mathcal{B}(\mathbb{R}^n)$.*

DÉMONSTRATION. Il est clair que $\mathcal{H}^d[\emptyset] = 0$ et que \mathcal{H}^d est une fonction croissante d'ensembles. On vérifie facilement que

$$\mathcal{H}_\varepsilon^d[\cup A_k] \leq \sum_{k \in \mathbb{N}} \mathcal{H}_\varepsilon^d[A_k].$$

En passant à la limite $\varepsilon \rightarrow 0$ dans le terme de gauche, et en utilisant l'inégalité $\mathcal{H}_\varepsilon^d \leq \mathcal{H}^d$ dans le terme de droite, on trouve

$$\mathcal{H}^d[\cup A_k] \leq \sum_{k \in \mathbb{N}} \mathcal{H}^d[A_k].$$

La fonction \mathcal{H}^d est donc sous-additive : c'est bien une mesure extérieure, définie sur l'ensemble de toutes les parties de \mathbb{R}^n .

Soit \mathcal{M} la tribu des ensembles \mathcal{H}^d -mesurables, au sens de l'énoncé du Théorème I-71 ; on sait que \mathcal{H} définit une mesure sur \mathcal{M} . Pour vérifier que \mathcal{M} contient toutes les parties boréliennes, on utilise le critère de Carathéodory présenté au Théorème I-79. Soient donc A et B deux parties de \mathbb{R}^n vérifiant $d(A, B) > 0$, on cherche à montrer que

$$\mathcal{H}^d[A \cup B] = \mathcal{H}^d[A] + \mathcal{H}^d[B].$$

Pour tout $\varepsilon < d(A, B)/2$, un ensemble de diamètre ε ne peut intersecter à la fois A et B ; si l'on se donne un recouvrement de $A \cup B$ par des ensembles de diamètre au plus ε on pourra donc en extraire des sous-recouvrements disjoints de A et B en considérant d'une part les ensembles qui intersectent A , d'autre part ceux qui intersectent B . On en déduit que $\mathcal{H}_\varepsilon^d[A \cup B] = \mathcal{H}_\varepsilon^d[A] + \mathcal{H}_\varepsilon^d[B]$, et la conclusion en découle par passage à la limite. \square

EXEMPLES V-4. (i) Soit $A = \{x_0\}$ un singleton. Il est clair que l'on peut recouvrir A par une boule de rayon nul, ce qui est de volume d -dimensionnel nul pour tout $d > 0$. Il s'ensuit que $\mathcal{H}^0[A] = 1$, $\mathcal{H}^d[A] = 0$ pour tout $d > 0$. Par σ -additivité, pour tout A dénombrable, $\mathcal{H}^0[A]$ n'est autre que le cardinal de A ; et cette identité reste valable si A n'est pas dénombrable. On conclut que \mathcal{H}^0 n'est autre que la mesure de comptage.

(ii) Il est facile de vérifier que la mesure de Hausdorff \mathcal{H}^1 dans \mathbb{R} n'est autre que la mesure de Lebesgue. Le caractère intrinsèque de la définition de \mathcal{H}^1 garantit que la mesure \mathcal{H}^1 restreinte à un segment de droite de \mathbb{R}^2 est également la mesure de Lebesgue sur ce segment de droite (vu comme sous-ensemble d'une copie de \mathbb{R}).

(iii) Soit μ la mesure définie sur \mathbb{R}^2 par

$$\int f d\mu = \int_0^1 f(0, t) dt.$$

Un peu de réflexion montre que $\mu = \delta_0 \otimes \mathcal{H}^1|_{[0,1]}$, où le symbole $|$ signifie "restriction".

(iv) On pourra montrer en exercice que si $I = [x, y]$ est un segment de droite dans \mathbb{R}^2 (non réduit à un point), alors $\mathcal{H}^d[I]$ vaut $+\infty$ si $d < 1$, $|x - y|$ si $d = 1$ et 0 si $d > 1$.

V-2.2. Propriétés élémentaires. Commençons par un critère de négligeabilité, conséquence immédiate de la structure de mesure extérieure :

PROPOSITION V-5 (critère pratique de Hausdorff-négligeabilité). *Soit $A \subset \mathbb{R}^n$. Alors $\mathcal{H}^d[A] = 0$ si et seulement si on peut inclure A dans une union d'ensembles B_k tels que $\sum_{k=1}^{\infty} \text{diam}(B_k)^d$ est arbitrairement petit.*

Cet énoncé généralise le critère de négligeabilité habituel pour la mesure de Lebesgue dans \mathbb{R} : un sous-ensemble de \mathbb{R} est de mesure nulle si et seulement si on peut l'inclure dans une union dénombrable d'intervalles dont la somme des longueurs est arbitrairement petite.

Voici maintenant une propriété bien commode des mesures de Hausdorff, qui explique en partie le rôle privilégié des fonctions Lipschitziennes dans ce contexte :

PROPOSITION V-6 (borne sous l'action des fonctions Lipschitziennes). *(i) Soit f une fonction k -Lipschitzienne définie sur un borélien de \mathbb{R}^n , à valeurs dans \mathbb{R}^m , alors pour tout ensemble borélien $B \subset A$ et pour tout $d \in [0, n]$ on a*

$$\mathcal{H}^d[f(A)] \leq k^d \mathcal{H}^d[A].$$

(ii) Plus généralement, si f et g sont deux fonctions définies sur un borélien de \mathbb{R}^n , à valeurs dans \mathbb{R}^m , telles que pour tous $x, y \in A$ on ait

$$|f(x) - f(y)| \leq |g(x) - g(y)|,$$

alors pour tout borélien $B \subset A$ et pour tout $d \in [0, n]$ on a

$$\mathcal{H}^d[f(A)] \leq \mathcal{H}^d[g(A)].$$

DÉMONSTRATION. Si f est k -Lipschitzienne, en passant au supremum pour $(x, y) \in B \times B$ dans l'inégalité $|f(x) - f(y)| \leq k|x - y|$, on voit que pour tout ensemble $C \subset A$, $\text{diam}(f(C)) \leq k \text{diam}(C)$. L'énoncé (i) en découle immédiatement.

La démonstration de (ii) est similaire : l'hypothèse implique $\text{diam}(f(C)) \leq \text{diam}(g(C))$. \square

Passons maintenant à des propriétés d'invariances, qui elles aussi découlent directement de la définition :

PROPOSITION V-7 (invariance par isométrie-multiplication). *Soit T une application affine de la forme*

$$T(x) = \alpha Ax + b,$$

où A est une isométrie, $\alpha > 0$ et $b \in \mathbb{R}^n$. Alors

$$T\#\mathcal{H}^d = \alpha^{-d}\mathcal{H}^d.$$

En particulier, pour tout Borélien C , on a $\mathcal{H}^d[C+b] = \mathcal{H}^d[C]$ et $\mathcal{H}^d[\alpha C] = \alpha^d \mathcal{H}^d[C]$, et \mathcal{H}^d est 2^d -doublante.

REMARQUE V-8. Lors d'un changement de variables dans une intégrale faisant intervenir des mesures de Hausdorff, ce n'est donc pas le déterminant Jacobien qui apparaît, même pour des opérations de multiplication scalaire.

La propriété V-7 peut sembler étrange si l'on se souvient de la caractérisation de la mesure de Lebesgue par son invariance sous l'action des translations : les mesures de Hausdorff vérifient la même propriété d'invariance ! Il n'y a pas de contradiction car les mesures de Hausdorff, malgré leur propriété de doublement, sont souvent très singulières (ou triviales), comme le montre la propriété suivante.

PROPOSITION V-9. Soit $C_n := [0, 1]^n \subset \mathbb{R}^n$; alors $\mathcal{H}^d[C_n] = +\infty$ pour tout $d < n$ et 0 pour tout $d > n$. En particulier,

- si $d < n$, alors $\mathcal{H}^d[O] = +\infty$ pour tout ouvert (non vide) de \mathbb{R}^n ;
- si $d > n$, alors \mathcal{H}^d est la mesure nulle sur \mathbb{R}^n .

DÉMONSTRATION. Pour tout $k \geq 1$, on peut partager C_n en 2^{nk} “petits” cubes semi-ouverts de côté 2^{-k} , qui sont tous de mesure $2^{-dk}\mathcal{H}^d[C_n]$, par la Proposition V-7. La σ -additivité implique donc

$$\mathcal{H}^d[C_n] = 2^{nk}2^{-dk}\mathcal{H}^d[C_n].$$

Si $\mathcal{H}^d[C_n] \notin \{0, +\infty\}$ on a donc forcément $n = k$.

Dans le cas où $d > n$, on peut appliquer la Proposition V-5 : les petits cubes sont de diamètre $\sqrt{n}2^{-k}$, et la somme de leurs diamètres à la puissance d vaut donc

$$2^{nk}n^{d/2}2^{-dk} \xrightarrow{k \rightarrow \infty} 0.$$

Il s’ensuit que $\mathcal{H}^d[C_n] = 0$. Comme \mathbb{R}^n est union dénombrable de copies de C_n , il s’ensuit que $\mathcal{H}^d[\mathbb{R}^n] = 0$.

Dans le cas où $d < n$, pour montrer que $\mathcal{H}^d[C_n] = +\infty$ il suffit de montrer que $\mathcal{H}^d[C_n] > 0$. On peut raisonner comme suit : si B_k est un ensemble de diamètre $2r_k$, alors on peut l’inclure dans une boule euclidienne de rayon $2r_k$, et $\lambda_n[B_k] \leq 2^n\alpha(n)r_k^n$. On a donc, pour tout recouvrement de C_n par des ensembles B_k de demi-diamètre $r_k \leq 1$,

$$1 = \lambda_n[C_n] \leq \sum_k \lambda_n[B_k] \leq 2^n\alpha(n) \sum_k r_k^n \leq 2^n\alpha(n) \sum_k r_k^d,$$

et en passant à l’infimum on voit que $\mathcal{H}^d[C_n] \geq 1/(2^n\alpha(n))$. Il s’ensuit que $\mathcal{H}^d[C_n] = +\infty$, et donc $\mathcal{H}^d[C] = +\infty$ pour tout cube semi-ouvert de \mathbb{R}^n . On conclut en notant que tout ouvert contient un cube semi-ouvert. \square

V-2.3. Régularité. Il résulte de la Proposition V-9 que la mesure de Hausdorff \mathcal{H}^d en dimension $n > d$ n’est ni σ -finie, ni régulière au sens de la Définition I-48. Cependant, la propriété I-49 (que de nombreux auteurs appellent aussi régularité) reste vraie :

THÉORÈME V-10 (“régularité” de la mesure de Hausdorff). Soit $d \geq 0$, et soit $A \subset \mathbb{R}^n$ une partie quelconque. Alors

- (i) Il existe G , intersection dénombrable d’ouverts contenant A , telle que

$$\mathcal{H}^d[G] = \mathcal{H}^d[A];$$

- (ii) Si A est \mathcal{H}^d -mesurable et $\mathcal{H}^d[A] < +\infty$, alors il existe F , union dénombrable de fermés contenus dans A , telle que

$$\mathcal{H}^d[F] = \mathcal{H}^d[A].$$

En particulier,

$$\mathcal{H}^d[A] = \sup\{\mathcal{H}^d[K]; K \text{ compact}; K \subset A\}.$$

REMARQUE V-11. L’énoncé (i) peut surprendre, puisque G est l’intersection décroissante des U_k , où chaque U_k est une intersection finie d’ouverts, donc un ouvert ; si $\mathcal{H}^d[A] < +\infty$ on a donc

$$\mathcal{H}^d[A] = \mathcal{H}^d[G] < \lim \mathcal{H}^d[U_k] = +\infty.$$

Pourquoi cela n'est-il pas en contradiction avec la σ -additivité de \mathcal{H}^d ?

DÉMONSTRATION DU THÉORÈME V-10. (i) Sans perte de généralité, on suppose que $\mathcal{H}^d[A] < +\infty$. Pour tout k , on a $\mathcal{H}_{1/k}^d[A] \leq \mathcal{H}^d[A] < +\infty$, et par la remarque V-2(ii), on peut trouver un recouvrement de A par des ouverts $C'_{k,j}$ de diamètre au plus $2/k$, tel que

$$\sum_j \alpha(d) r(C'_{k,j})^d \leq \mathcal{H}_{1/k}^d[A] + \frac{1}{k}.$$

On pose alors

$$O_k := \bigcup_{j \in \mathbb{N}} C'_{k,j}, \quad G := \bigcap_{k \geq 1} O_k.$$

Il est clair que G contient A , et d'autre part pour tout k on a

$$\mathcal{H}_{2/k}^d[G] \leq \sum_j \alpha(d) r(C'_{k,j})^d \leq \mathcal{H}_{1/k}^d[A] + \frac{1}{k}.$$

Il s'ensuit que $\mathcal{H}^d[G] \leq \mathcal{H}^d[A]$, d'où la conclusion.

(ii) Chacun des ouverts O_k peut s'écrire comme union dénombrable croissante de fermés $F_{k,j}$ ($j \in \mathbb{N}$). Par σ -additivité,

$$\lim_{j \rightarrow \infty} \mathcal{H}^d[A \cap F_{k,j}] = \mathcal{H}^d[A \cap O_k] = \mathcal{H}^d[A].$$

Pour tout $\delta > 0$ et $k \geq 1$ on choisit j_k tel que

$$\mathcal{H}^d[A \cap F_{k,j_k}] \geq \mathcal{H}^d[A] - 2^{-k}\delta.$$

On pose

$$F' := \bigcap_{k \geq 1} F_{k,j_k};$$

on a alors $\mathcal{H}^d[A \cap F'] \geq \mathcal{H}^d[A] - \delta$. Attention, rien ne garantit que F' soit inclus dans A ! Cependant, F' est inclus dans $\bigcap O_k = G$, et $\mathcal{H}^d[G \setminus A] = 0$, il s'ensuit que $\mathcal{H}^d[F' \setminus A] = 0$. Par la partie (i) du théorème, il existe un ensemble G' , intersection dénombrable d'ouverts, de mesure nulle, tel que $F' \setminus A \subset G'$. Alors $F_\delta := F' \setminus G'$ est contenu dans A , c'est une intersection dénombrable de fermés, et

$$\mathcal{H}^d[F_\delta] \geq \mathcal{H}^d[F'] - \mathcal{H}^d[G'] = \mathcal{H}^d[F'] \geq \mathcal{H}^d[A] - \delta.$$

On conclut en posant $F := \bigcap_{k \geq 1} F_{1/k}$. □

Les mesures de Hausdorff vérifient certaines des propriétés de densité au sens de Lebesgue. On établit ainsi le théorème suivant [Evans-Gariepy, pp. 72-75]

THÉORÈME V-12 (densité au sens de Hausdorff). *Soit $A \subset \mathbb{R}^n$ un ensemble \mathcal{H}^d -mesurable, avec $\mathcal{H}^d[A] < +\infty$, $0 < d < n$. Alors pour \mathcal{H}^d -presque tout $x \in A$,*

$$2^{-d} \leq \limsup_{r \rightarrow 0} \frac{\mathcal{H}^d[B_r(x) \cap A]}{\alpha(d)r^d} \leq 1$$

et pour \mathcal{H}^d -presque tout $x \in \mathbb{R}^n \setminus A$,

$$\lim_{r \rightarrow 0} \frac{\mathcal{H}^d[B_r(x) \cap A]}{\alpha(d)r^d} = 0.$$

REMARQUE V-13. Il ne faut pas être surpris par la dissymétrie des deux énoncés : les ensembles \mathcal{H}^d -mesurables de mesure finie sont “très petits”, en particulier leur complémentaire est toujours de mesure infinie. En tous les cas, il n’est pas toujours vrai que \mathcal{H}^d -presque tout point x de A soit régulier, au sens où on aurait

$$\lim_{r \rightarrow 0} \frac{\mathcal{H}^d[B(x, r) \cap A]}{\alpha(d)r^d} = 1.$$

Un travail considérable a été accompli dans la deuxième moitié du vingtième siècle pour préciser l’énoncé ci-dessus et décrire les ensembles \mathcal{H}^d -mesurables de manière plus précise. De manière générale, on peut décomposer un ensemble \mathcal{H}^d -mesurable en une “partie régulière”, dont \mathcal{H}^d -presque tous les points sont réguliers, et une partie “totalement irrégulière”, dont \mathcal{H}^d -presque aucun point n’est régulier. Les propriétés de ces ensembles et leur description géométrique (existence de tangentes, etc.) occupent une bonne partie de [Falconer1], et constituent encore un domaine de recherche en activité.

V-2.4. Généralisation abstraite. Il est facile de généraliser la notion de mesure de Hausdorff à un espace métrique X arbitraire : il suffit d’utiliser la formule (37) pour $A \subset X$, en prenant l’infimum sur tous les recouvrements de A par des parties B_k de X , de diamètre au plus ε .

V-3. Identification des mesures de Hausdorff

Quand d n’est pas un entier, il est difficile d’interpréter la mesure de Hausdorff \mathcal{H}^d d’une manière intuitive ; elle définit une sorte de volume en dimension fractionnaire, qu’il vaut sans doute mieux considérer de manière purement formelle. En revanche, quand d est un entier, la question se pose de savoir si on retrouve des concepts familiers de longueur, surface, volume, etc.

Juste après avoir défini la notion de mesure de Hausdorff, nous avons remarqué que la mesure de Hausdorff 0-dimensionnelle coïncide avec la mesure de comptage. Nous allons maintenant voir qu’il y a bien identité entre les deux notions naturelles de “volume n -dimensionnel dans \mathbb{R}^n ”, données respectivement par la mesure de Lebesgue et par la mesure de Hausdorff n -dimensionnelle. Nous verrons aussi que la mesure de Hausdorff 1-dimensionnelle prolonge une définition courante de la longueur.

V-3.1. Inégalité isodiamétrique. Soit $A \subset \mathbb{R}^n$, de demi-diamètre r . Il est clair que le volume de A est égal à $\alpha(n)r^n$ si A est une boule, mais que peut-on dire dans le cas général ? On est tenté de penser que A est inclus dans une boule de rayon r , ou $r + \varepsilon$ avec $\varepsilon > 0$ arbitrairement petit, mais ce n’est pas forcément le cas, comme le montre l’exemple d’un triangle de côté 1 dans \mathbb{R}^2 est de diamètre 1. Cependant, l’inégalité isodiamétrique assure que le volume d’un tel ensemble est inférieur ou égal à celui d’une boule de même rayon.

THÉORÈME V-14 (inégalité isodiamétrique). *Soit $A \subset \mathbb{R}^n$ un ensemble Lebesgue-mesurable, et r son demi-diamètre. Alors*

$$\lambda^n[A] \leq \alpha(n)r^n.$$

En d’autres termes, à diamètre fixé, les boules maximisent le volume.

- REMARQUES V-15. (i) On pourra comparer cet énoncé à celui de l'inégalité **isopérimétrique**, qui stipule qu'à surface fixée, les boules maximisent le volume.
- (ii) L'inégalité isodiamétrique peut paraître évidente à première vue, mais elle ne l'est pas, car un ensemble de diamètre $2r$ ne peut pas, en général, s'inclure dans une boule de rayon r .

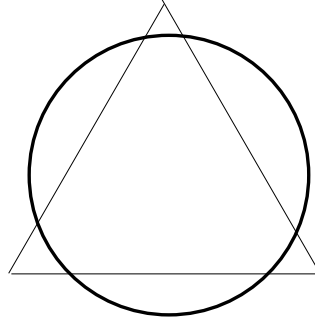


FIG. 2. Le triangle équilatéral ne rentre pas dans le disque de même diamètre.

Pour démontrer le Théorème V-14, nous utiliserons le concept utile de **symétrisation de Steiner**.

DÉFINITION V-16 (symétrisation de Steiner). Soit $A \subset \mathbb{R}^n$, et soit $a \in \mathbb{R}^n$ un vecteur de norme 1. Soit P_a l'hyperplan passant par 0, orthogonal à a . On peut écrire A comme l'union disjointe des $L_{a,z} \cap A$, où $L_{a,z}$ est la ligne dirigée par a , passant par $z \in P_a$. Pour chaque $z \in P_a$, on construit le segment A'_z centré en z , tel que $\mathcal{H}^1[A'_z] = \mathcal{H}^1[L_{a,z} \cap A]$. La réunion disjointe des segments A'_z ainsi obtenus est appelé *symétrisé de Steiner de A par rapport à l'hyperplan P_a* .

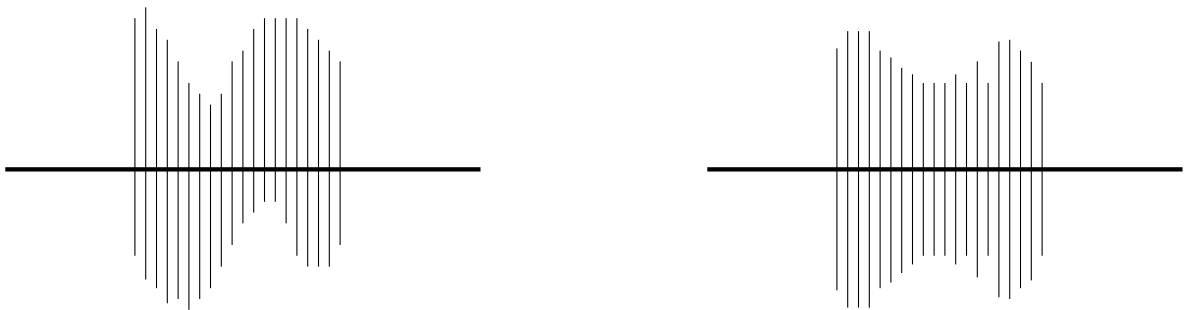


FIG. 3. Représentation schématique de la symétrisation de Steiner

Nous admettrons le lemme suivant [Evans-Gariepy pp. 67-68], que l'on pourra chercher à démontrer de manière non rigoureuse en exercice.

LEMME V-17 (propriétés de la symétrisation de Steiner). *La symétrisation de Steiner réduit le diamètre et préserve la mesure de Lebesgue.*

DÉMONSTRATION DE L'INÉGALITÉ ISODIAMÉTRIQUE. Soit (e_1, \dots, e_n) une base euclidienne de \mathbb{R}^n . On note S_a la symétrisation de Steiner par rapport à P_a , et $A^* := S_{e_n} S_{e_{n-1}} \dots S_{e_1} A$. Le diamètre de A^* est alors inférieur ou égal à celui de A , tandis que la mesure de Lebesgue de A^* est égale à celle de A ; il suffit donc de montrer le résultat pour A^* .

Par récurrence, et en utilisant le fait que la réflexion autour de P_{e_k} laisse e_j invariant pour tout $j \neq k$, on montre que A^* est symétrique par rapport à P_{e_1}, \dots, P_{e_n} , et donc symétrique par rapport à l'origine. Il s'ensuit que A^* est contenu dans une boule de centre 0 et de rayon $\text{diam}(A^*)/2$. Le résultat en découle. \square

V-3.2. Dimension n : le volume. La mesure de Hausdorff n -dimensionnelle en dimension n coïncide avec la mesure de Lebesgue λ_n :

THÉORÈME V-18 ($\mathcal{H}^n = \lambda_n$). *Soit $A \subset \mathbb{R}^n$ un ensemble Borélien. Alors*

$$\mathcal{H}^n[A] = \lambda_n[A].$$

En particulier, si E_k est un sous-espace affine de \mathbb{R}^n , de dimension k , alors la restriction de \mathcal{H}^k à E_k coïncide avec la mesure de Lebesgue sur E_k .

Je vais commencer par présenter une démonstration simple d'un énoncé plus faible selon lequel \mathcal{H}^n est *proportionnelle* à λ_n . La démonstration complète du Théorème V-18 est plus subtile et utilisera l'inégalité isodiamétrique.

DÉMONSTRATION PARTIELLE DU THÉORÈME V-18. Il est clair que \mathcal{H}^n est invariante par translation (de même que toutes les mesures de Hausdorff sur \mathbb{R}^n). Pour montrer que \mathcal{H}^n et λ_n sont proportionnelles (il existe $c(n) > 0$ tel que $\mathcal{H}^n = c(n) \lambda_n$), il suffit donc de montrer que $\mathcal{H}^n[C_n] \in (0, +\infty)$, où $C_n = [0, 1]^n$.

Soit $\varepsilon > 0$, et $k \in \mathbb{N}$ tel que $2^{-k} \leq \varepsilon/\sqrt{n} \leq 2^{-k+1}$. On peut recouvrir C_n par 2^{nk} cubes de côté 2^{-k} , dont chacun aura un diamètre $\sqrt{n}2^{-k} \leq \varepsilon$. Il s'ensuit que $\mathcal{H}_\varepsilon^n[C_n] \leq C(n) 2^{nk} 2^{-nk} = C(n)$, où $C(n)$ est une constante ne dépendant que de n . En prenant la limite quand $\varepsilon \rightarrow 0$ on conclut que

$$\mathcal{H}^n[C_n] < +\infty.$$

Par ailleurs, si A est un ensemble quelconque, sa mesure de Lebesgue extérieure est majorée par $C'(n) \text{diam}(A)^n$, où $C'(n)$ est le volume de la boule de rayon 2 dans \mathbb{R}^n . Si l'on a un recouvrement de C_n par des ensembles A_j , la somme de toutes les mesures extérieures de ces ensembles est au moins égale à celle du cube, d'où $\sum C'(n) (\text{diam}(A_j))^n \geq 1$. On en déduit que $\mathcal{H}_\varepsilon^n[C_n]$ est minoré par une constante positive indépendante de ε , et en faisant tendre ε vers 0 on conclut que

$$\mathcal{H}^n[C_n] > 0.$$

\square

DÉMONSTRATION COMPLÈTE DU THÉORÈME V-18. La deuxième partie de ce théorème se déduit de la première grâce au caractère intrinsèque de la définition de mesure de Hausdorff : la restriction de la mesure de Hausdorff \mathcal{H}^k à E_k est exactement la mesure de Hausdorff \mathcal{H}^k définie sur E_k , qui est une copie de \mathbb{R}^k .

Soit $(C_k)_{k \geq 1}$ un recouvrement de A par des ensembles de diamètre inférieur ou égal à ε . Grâce à l'inégalité isodiamétrique, on a

$$\lambda_n[A] \leq \sum_k \lambda_n[C_k] \leq \sum_k \alpha(n) r(C_k)^n.$$

En passant à l'infimum, on voit que $\lambda_n[A] \leq \mathcal{H}_\varepsilon^n[A]$, et donc $\lambda_n[A] \leq \mathcal{H}^n[A]$. Il nous reste à montrer l'inégalité inverse.

Il est facile de montrer, en utilisant des cubes dyadiques, que

$$\lambda_n[A] = \inf \left\{ \sum_{k=1}^{\infty} \lambda_n[Q_k]; \quad A \subset \bigcup Q_k, \quad r(Q_k) \leq \varepsilon \right\},$$

où les Q_k sont des cubes dyadiques de côtés parallèles aux axes. Pour de tels cubes, on peut trouver une constante c_n , dépendant uniquement de n , telle que

$$\alpha(n)r(Q_k)^n = c_n \lambda_n[Q_k].$$

On en déduit que $\mathcal{H}^n \leq c_n \lambda_n$.

Pour conclure, on admet le lemme suivant : étant donné un cube Q et $\varepsilon > 0$, on peut écrire

$$Q = \bigcup_{j \leq 1} B_j \cup N,$$

où les B_j sont des boules fermées de rayon au plus ε , disjointes, et N est un ensemble Lebesgue-négligeable.

Soit maintenant A un ensemble Lebesgue-mesurable, on choisit une famille (C_k) de cubes Q_k recouvrant A , telle que

$$\sum_k \lambda_n[Q_k] \leq \lambda_n[A] + \delta,$$

où $\delta > 0$ est arbitrairement petit. Pour chaque Q_k on introduit une famille de boules $(B_{k,j})_{j \geq 1}$ et un ensemble négligeable N_k vérifiant les conclusions du lemme admis ci-dessus ; en particulier, $\mathcal{H}[N_k] \leq c_n \cdot 0 = 0$. On donc

$$\begin{aligned} \mathcal{H}_\varepsilon^n[A] &\leq \sum_{k \geq 1} \mathcal{H}^n[Q_k] = \sum_{k \geq 1} \left(\sum_{j \geq 1} \mathcal{H}[B_{k,j}] + \mathcal{H}(N_k) \right) \\ &\leq \sum_{k \geq 1} \sum_{j \geq 1} \lambda_n[B_{k,j}] = \sum_{k \geq 1} \lambda_n[\bigcup B_{k,j}] = \sum_{k \geq 1} \lambda_n[Q_k] \leq \lambda_n[A] + \delta. \end{aligned}$$

Ceci conclut l'argument. \square

V-3.3. Dimension 1 : la longueur. On pourrait convenir a priori de choisir \mathcal{H}^1 comme définition de la longueur d'une partie de \mathbb{R}^n . Cependant, il existe une autre notion simple et populaire de longueur, bâtie sur le concept de **rectifiabilité**. Commençons par en rappeler les propriétés principales.

DÉFINITION V-19 (rectifiabilité). Soient I un intervalle de \mathbb{R} et $\gamma : I \rightarrow \mathbb{R}^n$ une courbe continue injective. On dit que γ est rectifiable sur I si pour tout intervalle compact $[a, b] \subset I$, la quantité

$$L_{[a,b]}(\gamma) := \sup \left\{ \sum_{k=0}^N |\gamma(t_{k+1}) - \gamma(t_k)|; \quad a = t_0 \leq t_1 \leq \dots \leq t_N \leq t_{N+1} = b, \quad N \in \mathbb{N} \right\} < +\infty$$

où le supremum est pris sur toutes les subdivisions finies ($a = t_0, t_1, \dots, t_N, t_{N+1} = b$) de $[a, b]$. On appelle alors

$$L(\gamma) := \sup_{[a,b] \subset I} L_{[a,b]}(\gamma)$$

la longueur de γ .

En d'autres termes, la longueur d'une courbe est le supremum de toutes les longueurs des "approximations polygonales" de cette courbe.

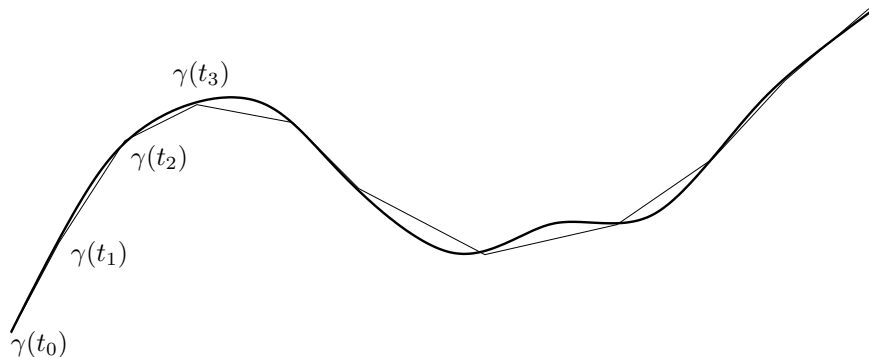


FIG. 4. Approximation polygonale d'une courbe

REMARQUES V-20. (i) Par définition, $L_{[a,b]}(\gamma)$ est toujours supérieur ou égal à $|\gamma(b) - \gamma(a)|$, et on peut vérifier qu'il y a égalité quand la courbe est une fonction affine : la ligne droite est bien le plus court chemin entre deux points !

(ii) On généralise sans difficulté cette notion à un espace métrique abstrait.

Noter l'hypothèse d'injectivité faite dans la définition : des maux de tête s'en suivraient si on devait prendre en compte la multiplicité ; ou alors il faudrait bien prendre garde à définir la longueur de la courbe γ , et non simplement de son image $\gamma([a, b])$. Dans cette partie, pour simplifier, nous ne travaillerons qu'avec des courbes injectives.

Une courbe γ étant donnée, on appelle **reparamétrage** de γ toute courbe (injective) $\tilde{\gamma}$, dont l'image est la même que celle de γ . On note que la longueur est invariante par reparamétrage.

Si γ est une courbe rectifiable définie sur un intervalle I , et x_0 est un point arbitraire de I , alors on peut définir un reparamétrage privilégié de γ , dit **paramétrage par longueur d'arc** : on définit la longueur orientée à partir de x_0 par

$$\ell_{x_0}(x) = \begin{cases} L_{[x_0, x]}(\gamma) & (x \geq x_0) \\ -L_{[x, x_0]}(\gamma) & (x < x_0) \end{cases} ;$$

on vérifie que la fonction ℓ_{x_0} est continue et strictement croissante, en particulier inversible sur son image. On définit alors le reparamétrage $\tilde{\gamma}$ par

$$\tilde{\gamma}(x_0 + \ell_{x_0}(x)) = \gamma(x), \quad \ell(x) = L_{[x_0, x]}(\gamma).$$

Les propriétés suivantes découlent presque immédiatement de la définition.

PROPOSITION V-21 (propriétés du paramétrage par longueur d'arc). *Soit $\gamma : I \rightarrow \mathbb{R}$ une courbe paramétrée par longueur d'arc. Alors pour tout $[a, b] \subset I$,*

$$L_{[a, b]}(\gamma) = b - a;$$

En particulier,

$$(38) \quad |\gamma(b) - \gamma(a)| \leq b - a,$$

et

$$L(\gamma) = |I|.$$

Le théorème suivant montre que la dimension de Hausdorff de dimension 1 est une généralisation du concept de rectifiabilité.

THÉORÈME V-22 ($L = \mathcal{H}^1$). *Soient I un intervalle de \mathbb{R} , et $\gamma : I \rightarrow \mathbb{R}$ une courbe injective rectifiable. Alors*

$$\mathcal{H}^1[\gamma(I)] = L(\gamma).$$

DÉMONSTRATION. Sans perte de généralité, on supposera que γ est paramétrée par longueur d'arc. Si $(A_k)_{k \in \mathbb{N}}$ est un recouvrement de $\gamma(I)$, on définit un recouvrement $(B_k)_{k \in \mathbb{N}}$ de I en définissant $B_k := \gamma^{-1}(A_k)$. L'inégalité (38) implique alors que $\text{diam}(B_k) \geq \text{diam}(A_k)$. En utilisant les définitions des mesures de Hausdorff, on en déduit

$$\mathcal{H}^1(\gamma(I)) \geq \mathcal{H}^1(I) = |I| = L(\gamma).$$

Pour établir l'inégalité inverse, commençons par remarquer que $\mathcal{H}^1(\gamma([a, b])) \geq |\gamma(b) - \gamma(a)|$. En effet, si π est la projection orthogonale de $\gamma([a, b])$ sur la ligne droite joignant $\gamma(a)$ et $\gamma(b)$, alors π réduit les distances, donc, par définition des mesures de Hausdorff, $\mathcal{H}^1(\gamma([a, b])) \geq \mathcal{H}^1(\pi(\gamma([a, b]))) = \mathcal{H}^1([\gamma(a), \gamma(b)])$. On peut identifier la droite passant par $\gamma(a)$ et $\gamma(b)$ à \mathbb{R} ; en utilisant alors l'identité $\mathcal{H}^1 = \lambda_1$ en dimension 1, on constate que $\mathcal{H}^1([\gamma(a), \gamma(b)])$ n'est autre que la longueur du segment $[\gamma(a), \gamma(b)]$, i.e. $|\gamma(b) - \gamma(a)|$.

Enfin, soit $[a, b] \subset I$ et soit $a = t_0 \leq t_1 \leq \dots \leq t_N \leq t_{N+1} = b$ une subdivision de $[a, b]$; cette subdivision découpe l'intervalle I en sous-intervalles ouverts $I_0, I_1, \dots, I_{N+1}, I_{N+2}$. Les points étant de mesure de Hausdorff \mathcal{H}^1 nulle, on a

$$\mathcal{H}^1(\gamma(I)) = \sum_{k=0}^{N+2} \mathcal{H}^1(\gamma(I_k)) \geq \sum_{k=1}^{N+1} \mathcal{H}^1(\gamma(I_k)) \geq \sum_{k=0}^N |\gamma(t_{k+1}) - \gamma(t_k)|.$$

En prenant le supremum sur toutes les subdivisions possibles, puis sur $[a, b] \subset I$, on conclut que

$$\mathcal{H}^1(\gamma(I)) \geq L(\gamma),$$

ce qui achève la preuve. \square

V-3.4. Autres dimensions entières. Nous venons de voir que la mesure de Hausdorff n -dimensionnelle s'identifie à la mesure de Lebesgue, i.e. au volume n -dimensionnel, et que la mesure de Hausdorff 1-dimensionnelle s'identifie à une notion de longueur, au moins dans le cas des courbes rectifiables. Il convient d'être plus prudent en ce qui concerne les autres dimensions entières! Appliquées à des objets suffisamment "réguliers", les mesures de Hausdorff donneront les résultats attendus : par exemple, la mesure \mathcal{H}^2 définit une notion de surface, etc. Cependant, pour des objets irréguliers, ces notions peuvent ne pas recouper les autres notions en vigueur... Cette remarque vaut aussi pour la dimension 1, dans le cas d'objets peu réguliers.

Le cas le plus frappant est celui où $d = n - 1$. Soit $A \subset \mathbb{R}^n$ une partie compacte (pour simplifier), comment définir la "surface" (ou volume $n - 1$ -dimensionnel) $S(\partial A)$ de son bord ∂A ? Il existe trois définitions, plus ou moins naturelles selon les

contextes. La première fait intervenir les mesures de Hausdorff, la deuxième est une définition possible de ce que l'on appelle parfois “contenu de Minkowski”, et la troisième est naturelle en théorie des distributions, ou en physique mathématique.

$$(i) S(A) := \mathcal{H}^{n-1}(\partial A);$$

$$(ii) S(A) := \liminf_{\varepsilon \rightarrow 0} \frac{\lambda_n[A_\varepsilon] - \lambda_n[A]}{\varepsilon};$$

(iii) $S(A) := \sup \left\{ \int_A \nabla \cdot J; J \in C_c^\infty(\mathbb{R}^n; \mathbb{R}^n), |J| \leq 1 \right\}$, où le supremum est pris sur l'ensemble des fonctions J de \mathbb{R}^n dans \mathbb{R}^n , de classe C^∞ et à support compact, bornées par 1 en norme, et on a noté

$$\nabla \cdot J = \sum_{k=1}^n \frac{\partial J_k}{\partial x_k}$$

la divergence de J .

C'est probablement la formule (ii) qui est la plus intuitive, et la plus simple à se représenter visuellement. D'autre part, le lecteur qui se souvient de la formule de Green-Ostrogradski ne sera pas surpris par l'apparition de l'opérateur divergence dans la formule (iii); en effet, cette formule énonce que, sous des conditions de régularité suffisante,

$$\int_A \nabla \cdot J = \int_{\partial A} J(x) \cdot N(x) d\sigma(x),$$

où $N(x)$ désigne la normale à ∂A en x et $\sigma \dots$ la mesure de surface sur ∂A .

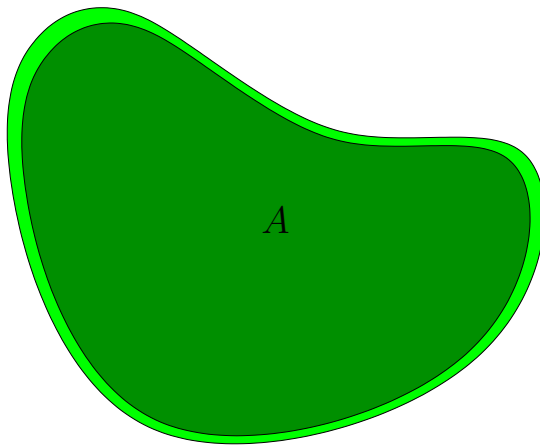


FIG. 5. Surface au sens de Minkowski : l'accroissement infinitésimal du volume est donné par le produit de la surface par la largeur d'épaississement

Les trois définitions précédentes de la surface de A peuvent donner des résultats différents pour des ensembles A “pathologiques”. Il convient donc d'être prudent quand on parle de surface k -dimensionnelle dans un contexte peu régulier, et préciser le cas échéant de quelle surface il s'agit.

V-4. Dimension

V-4.1. Echelle des mesures de Hausdorff. La proposition suivante établit le fait intuitif que si une dimension convient pour évaluer la taille d'un objet, les dimensions supérieures sont trop grossières (ainsi, si une courbe a une surface positive, sa longueur doit être infinie ; si elle a une longueur finie, sa surface doit être nulle).

PROPOSITION V-23 (au plus une dimension donne une mesure non triviale). *Soit $A \subset \mathbb{R}^n$; alors*

- (i) *si $\mathcal{H}^d[A] < +\infty$ pour un certain $d \geq 0$, alors $\mathcal{H}^{d'}[A] = 0$ pour tout $d' > d$;*
- (ii) *si $\mathcal{H}^d[A] > 0$ pour un certain $d > 0$, alors $\mathcal{H}^{d'}[A] = +\infty$ pour tout $d' < d$;*
- (iii) *pour tout $d > n$, on a $\mathcal{H}^d[A] = 0$;*

DÉMONSTRATION DE LA PROPOSITION V-23. Soient $A \subset \mathbb{R}^n$, $d_1 < d_2$, et soit $(C_k)_{k \in \mathbb{N}}$ un recouvrement de A par des ensembles de demi-diamètre respectif $r_k \leq \varepsilon/2$. Alors

$$\sum_k r_k^{d_2} \leq \varepsilon^{d_2-d_1} \sum_k r_k^{d_1}.$$

Si maintenant on a $\mathcal{H}^d[A] < +\infty$ pour un certain $d > 0$, alors pour tout $\varepsilon > 0$ on a $\mathcal{H}_\varepsilon^d[A] < +\infty$, et il existe donc un recouvrement dénombrable de A par des ensembles de demi-diamètre $r_k \leq \varepsilon/2$, tel que

$$\sum_k r_k^d \leq C < +\infty.$$

Pour ce même recouvrement, on a alors $\sum_k r_k^{d'} \leq C \varepsilon^{d'-d} \rightarrow 0$ dès que $d' > d$. Cela prouve que $\mathcal{H}_\varepsilon^{d'}[A] = O(\varepsilon^{d'-d})$, et en particulier $\mathcal{H}^{d'}[A] = 0$.

Si d'autre part on a $\mathcal{H}^d[A] > 0$ pour un certain $d > 0$, alors pour tout $\varepsilon > 0$ assez petit on a $\mathcal{H}_\varepsilon^d[A] \geq \delta > 0$; en particulier, tout recouvrement dénombrable de A par des ensembles de demi-diamètre $r_k \leq \varepsilon/2$,

$$\sum_k r_k^d \geq \frac{\delta}{\alpha(d)} > 0,$$

d'où, pour tout $d' < d$,

$$\sum_k \alpha(d') r_k^{d'} \geq \varepsilon^{d-d'} \delta \frac{\alpha(d')}{\alpha(d)} \xrightarrow{\varepsilon \rightarrow 0} +\infty,$$

et finalement $\mathcal{H}^{d'}[A] = +\infty$.

L'assertion (iii) a déjà été établie ; nous allons reproduire brièvement le raisonnement. Comme \mathbb{R}^n est union dénombrable de pavés, il suffit de prouver qu'un pavé de \mathbb{R}^n est de mesure d -dimensionnelle nulle pour $d > n$. Puisque ce pavé est de mesure de Lebesgue finie donc de mesure d -dimensionnelle finie, (iii) découle de (i). \square

V-4.2. Dimension de Hausdorff. Au vu de la Proposition V-23, la fonction $d \mapsto \mathcal{H}^d[A]$ est très particulière : on se convainc facilement qu'elle vaut $+\infty$ quand d est strictement plus petit qu'un certain d_0 , et 0 quand d est strictement supérieur à d_0 .

Ceci mène naturellement à la définition de la dimension de Hausdorff.

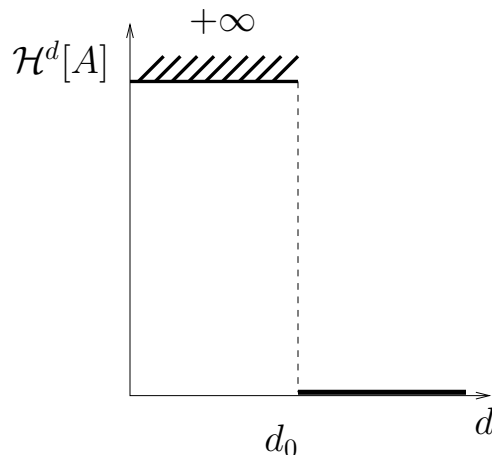


FIG. 6. Graphe de $\mathcal{H}^d[A]$; $\mathcal{H}^{d_0}[A]$ peut se situer n'importe où sur la ligne pointillée.

DÉFINITION V-24. Soit $A \subset \mathbb{R}^n$. On définit sa dimension de Hausdorff, que l'on note $\dim(A)$ ou $\dim_{\mathcal{H}}(A)$, par

$$\dim(A) := \inf\{d; \mathcal{H}^d[A] = 0\} \in [0, n]$$

De manière équivalente, $\dim(A)$ est l'unique d_0 tel que $\mathcal{H}^d[A] = +\infty$ pour tout $d < d_0$, et $\mathcal{H}^d[A] = 0$ pour tout $d > d_0$.

La dimension de Hausdorff se prête bien à de nombreux énoncés théoriques, car elle est associée naturellement aux mesures de Hausdorff ; en revanche elle est parfois difficile à calculer. Le théorème suivant se déduit facilement de la Proposition V-6 :

THÉORÈME V-25 (dimension des graphes et images). Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une fonction Lipschitzienne. Soit A une partie mesurable de \mathbb{R}^n , on note $G(f, A) = \{(x, f(x)); x \in A\}$ le graphe de f sur A . Alors

- (i) $\dim_{\mathcal{H}}(f(A)) \leq \dim_{\mathcal{H}}(A) \leq n$;
- (ii) Si $\lambda_n[A] > 0$, alors $\dim_{\mathcal{H}}(G(f, A)) = n$.

REMARQUES V-26. (i) On se souvient que le graphe d'une fonction continue est de mesure de Lebesgue nulle ; nous voyons ici que le graphe d'une application Lipschitzienne a la dimension attendue. De manière générale, la dimension de Hausdorff d'un graphe est **supérieure ou égale** à la dimension de l'espace de départ ; elle peut être strictement supérieure pour des applications qui sont seulement Hölderiennes (ou encore moins régulières) et pas Lipschitz.

(ii) L'application de Peano montre que l'image du segment $[0, 1]$ par une application continue peut être de dimension 2 (bien sûr, cette application n'est pas Lipschitzienne !). Une trajectoire typique du mouvement Brownien plan pour les temps $t \in [0, 1]$ fournit un autre exemple de courbe dont l'image est de dimension 2, cependant la mesure 2-dimensionnelle de cette image est nulle ! Les trajectoires du mouvement Brownien ne sont bien sûr pas Lipschitz, mais elles sont Hölder- α pour tout $\alpha < 1/2$ (il est naturel d'imaginer que l'exposant $1/2$ est critique pour de tels contre-exemples). En revanche, l'image d'une courbe

Lipschitz est toujours de dimension inférieure ou égale à 1. Si on considère une fonction Lipschitz définie sur un segment $[0, 1]$, à valeurs dans \mathbb{R}^n , son image sera soit réduite à un point, soit de dimension 1.

- (iii) En corollaire de ce théorème, on voit que **les applications bilipschitziennes préservent la dimension de Hausdorff** (par bilipschitzienne, on entend que la fonction f , supposée bijective, et sa réciproque, sont toutes deux Lipschitz). C'est une des raisons pour lesquelles les applications bilipschitziennes constituent une notion naturelle d'“isomorphisme” dans l'étude des objets fractals.

V-4.3. Dimension de Minkowski. Expliquons maintenant une autre notion populaire, souvent plus simple à calculer et antérieure à celle de Hausdorff, dite dimension de Minkowski.

Commençons par nous interroger sur le moyen de faire la différence entre un objet monodimensionnel et un objet bidimensionnel? Intuitivement, le second est beaucoup plus “recouvrant”; on peut formaliser cela en considérant l'ensemble des points qui leur sont proches. Plaçons-nous dans le carré $[0, 1]^2$ pour simplifier. On quadrille ce carré en petits sous-carrés de côté $\varepsilon = 1/K$, $K \gg 1$. On s'attend à ce qu'un objet monodimensionnel X rencontre environ L/ε tels sous-carrés, où L désigne la longueur de X , tandis qu'un objet bidimensionnel Y en rencontrera environ S/ε^2 , où S désigne l'aire de Y . Et si l'on considère la réunion de tous les sous-carrés rencontrés par ces objets, sa surface est environ $L\varepsilon$ dans le premier cas, S dans le deuxième. Par extrapolation, on a envie de dire qu'un objet est de dimension d si le nombre de petits carrés nécessaire à son recouvrement est de l'ordre de ε^{-d} .

Cette idée conduit à la **dimension de Minkowski** d'un sous-ensemble A de \mathbb{R}^n : on pose

$$\dim_{\mathcal{M}}(A) = \lim_{\varepsilon \rightarrow 0} \frac{\log N_{\varepsilon}(A)}{|\log \varepsilon|},$$

où $N_{\varepsilon}(A)$ est, au choix : le nombre minimal de boules de diamètre ε (resp. de cubes de côté ε , resp. de cubes pris parmi un réseau de côté ε , resp. d'ensembles de diamètre ε) par lequel on peut recouvrir A ; ou encore le nombre maximal de points que l'on peut placer dans A de telle sorte qu'ils soient tous à une distance supérieure ou égale à ε les uns des autres. Toutes ces équivalences sont passées en revue dans [Falconer2, Chapitre 3] où la dimension de Minkowski est appelée dimension de [comptage de] boîtes (“box dimension”). Notons que dans le cas où la limite quand $\varepsilon \rightarrow 0$ n'existe pas, on peut toujours définir une dimension supérieure (resp. inférieure) en remplaçant la limite par une \limsup (resp. \liminf). Enfin, il existe encore une autre façon équivalente de définir cette dimension :

$$\dim_{\mathcal{M}}(A) := n + \lim_{\varepsilon \rightarrow 0} \frac{\log \lambda_n[A_{\varepsilon}]}{|\log \varepsilon|}$$

où A_{ε} est le ε -voisinage de A , i.e.

$$A_{\varepsilon} := \{x \in \mathbb{R}^n; d(x, A) \leq \varepsilon\}.$$

La définition de la dimension de Minkowski est assez intuitive, et elle est souvent relativement facile à calculer ou estimer; mais elle a quelques défauts troublants. Par exemple, l'ensemble $([0, 1] \cap \mathbb{Q})^2$ est dense dans $[0, 1]^2$, et la définition précédente lui attribue une dimension 2; pourtant, un point est de dimension 0, et dans un cadre de mesures σ -additives, on trouverait naturel qu'une union dénombrable d'objets de

dimension donnée d soit également un objet de dimension d . La conclusion est que la dimension de Minkowski n'est pas associée à une notion naturelle de mesure.

De manière générale, la dimension de Minkowski est toujours supérieure ou égale à la dimension de Hausdorff; l'inégalité peut être stricte, puisque $[0, 1] \cap \mathbb{Q}$ est de dimension de Hausdorff 0 et de dimension de Minkowski 1... Par ailleurs, la dimension de Minkowski vérifie l'identité

$$\dim_{\mathcal{M}}(A \times B) = \dim_{\mathcal{M}}(A) + \dim_{\mathcal{M}}(B),$$

ce qui n'est pas toujours vrai de la dimension de Hausdorff.

On trouvera dans [Falconer2, Chapitre 3] d'autres définitions en usage de la notion de dimension, et une discussion des liens qui existent entre ces notions.

V-4.4. Ensembles de Cantor. Commençons par l'exemple utilisé par Hausdorff lui-même pour illustrer sa notion de dimension : l'ensemble triadique de Cantor, défini comme la limite des ensembles fermés C_k , où $C_0 = [0, 1]$ et C_k est obtenu à partir de C_{k-1} en supprimant le tiers (ouvert) central de chacune des composantes connexes de C_{k-1} . L'ensemble résultant est clairement de mesure de Lebesgue nulle, on peut se demander quelle est sa dimension.

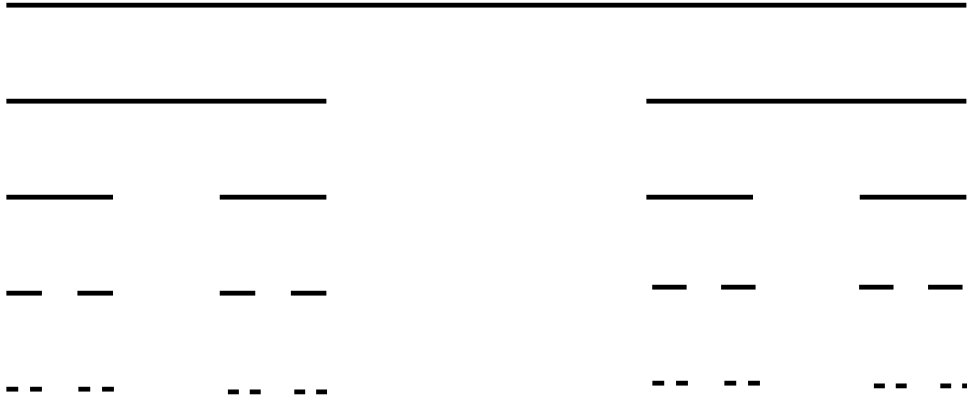


FIG. 7. Premières étapes de la construction de l'ensemble triadique de Cantor

Si l'on prend $\varepsilon = 3^{-k}$, on voit que l'ensemble triadique de Cantor C dans $[0, 1]$ peut se recouvrir par $2^k = \varepsilon^{-d}$ segments de longueur ε (soit des boules de rayon $\varepsilon/2$), avec $d = \log 2 / \log 3$, et que ce recouvrement est le plus économique que l'on puisse réaliser. Il est facile d'en déduire que

$$\dim_{\mathcal{M}}(C) = \frac{\log 2}{\log 3}.$$

La dimension de Hausdorff est déjà plus difficile à calculer. On sait qu'elle n'est pas plus grande que la dimension de Minkowski, soit $\log 2 / \log 3$. Par ailleurs, on peut faire un calcul heuristique simple en tirant parti de la construction auto-similaire de l'ensemble C et de l'identité $\mathcal{H}^d[\lambda A] = \lambda^d \mathcal{H}^d[A]$, facile à vérifier. S'il existe un exposant d tel que $\mathcal{H}^d[C] \in]0, +\infty[$, alors, comme C est l'union de deux copies de $C/3$, on aura

$$\mathcal{H}^d[C] = 2\mathcal{H}^d[C/3] = \frac{2}{3^d} \mathcal{H}^d[C],$$

ce qui impose $3^d = 2$, i.e. $d = \log 2 / \log 3$.

On est donc tenté de conclure que la dimension de Hausdorff de C est égale à la dimension de Minkowski ! C'est effectivement le cas : le raisonnement esquissé ci-après prouve en effet que pour tout recouvrement de C par une famille dénombrable d'intervalles ouverts $(I_k)_{k \in \mathbb{N}}$, on a

$$\sum_k |I_k|^d \geq \frac{1}{2},$$

et il s'ensuit que $\mathcal{H}^d[C] > 0$. Pour établir cette inégalité, on remarque d'abord que par compacité on peut se limiter à une famille finie d'intervalles ouverts, dont chacun a une longueur comprise entre $3^{-(\ell+1)}$ et (strictement) $3^{-\ell}$, pour un unique $\ell = \ell(k)$. L'intervalle I_k peut alors intersecter au plus une des composantes connexes de C_ℓ , et donc pour $j \geq \ell$ il ne peut intersecter plus de $2^{j-\ell} \leq 2^j 3^d |I_k|^d$ composantes connexes de C_j . On choisit j suffisamment grand pour que 3^{-j} soit plus petit que toutes les longueurs $|I_k|$; alors toutes les composantes connexes de C_j doivent être intersectées par les I_k , il y en a 2^j , et on a donc

$$\begin{aligned} 2^j &\leq \sum_k (\text{nombre de composantes connexes intersectées par } I_k) \\ &\leq \sum_k 2^j 3^d |I_k|^d, \end{aligned}$$

d'où $\sum_k |I_k|^2 \geq 3^{-d} = 1/2$.

Avec un peu plus d'efforts, on peut montrer que $\mathcal{H}^d[C] = 2^{1-d}$, qui constitue une sorte de mesure de la taille de C en dimension d . Notons que l'on ne peut utiliser la σ -additivité pour cela : pour tout k , on a $\mathcal{H}^d[C_k] = +\infty$...

Le **contenu de Minkowski** permet, ici encore, de prédire le résultat de manière très simple : par définition, le contenu de Minkowski d'un sous-ensemble de dimension d de \mathbb{R}^n est le produit de $\alpha(n)$ par le coefficient dominant de N_ε quand $\varepsilon \rightarrow 0$, et fournit une sorte de volume d -dimensionnel qui cadre bien avec l'intuition que l'on se fait des notions de longueur, surface, etc. Ici on a $\alpha(1) = 2$ et $N_\varepsilon \simeq 2^{-d} \varepsilon^{-d}$, de sorte que le contenu de Minkowski coïncide bien $\mathcal{H}^d[C]$. Malheureusement, cette égalité n'est pas la règle...

On note que du point de vue topologique, l'ensemble triadique de Cantor est "totalement discontinu" : bien qu'il ne soit pas dénombrable, il ne contient aucun segment, et toutes ses composantes connexes sont donc des points. Du point de vue topologique, il est naturel de lui attribuer une dimension nulle ! On peut montrer d'ailleurs que c'est le cas de toute partie dont la dimension de Hausdorff est strictement inférieure à 1 [Falconer2, Proposition 2.5]. On peut mettre cette remarque en regard de la suggestion de Mandelbrot, selon laquelle on pourrait définir un objet fractal comme un objet dont la dimension de Hausdorff est strictement supérieure à la dimension topologique.

De manière générale, on appelle *ensemble de Cantor* un espace topologique compact totalement discontinu (dont les composantes connexes sont des points) et sans point isolé (un point x_0 d'un espace X est dit isolé s'il existe un voisinage V de x_0 qui ne rencontre X qu'en x_0). Ces ensembles jouent un rôle important dans diverses branches des mathématiques ; on peut en construire de nombreux exemples par des variantes du procédé de construction diadique de Cantor. Voici quelques exemples intéressants :

- On coupe le segment $[0, 1]$ en k segments ($k \geq 3$, supposons k impair pour simplifier), on élimine les $k-2$ intervalles centraux pour ne garder que les deux segments extrêmes. On coupe chacun des segments ainsi obtenus en k parties égales, et sur ces k parties on élimine les $k-2$ parties centrales. Et ainsi de suite ! On construit de la sorte un ensemble de Cantor “ k -adique fin” de dimension $\log 2 / \log k$ (arbitrairement petite). Si au contraire à chaque étape on choisit d’éliminer seulement le segment central, l’ensemble limite C est un ensemble de Cantor “ k -adique gras” de dimension $\log 2 / \log c(k)$, où $c(k) = (2k+1)/k$ est le coefficient de proportionnalité permettant de passer de l’ensemble à sa “composante gauche” ($C = c(k)(C \cap [0, 1/2])$) ; comme $c(k) \rightarrow 1$ pour $k \rightarrow \infty$, l’ensemble ainsi construit est de dimension arbitrairement proche de 1.

- On coupe le segment $[0, 1]$ en trois tiers, on élimine le tiers central. On coupe chacun des segments ainsi obtenus en cinq parties égales, et sur ces cinq parties on élimine les trois parties centrales. On coupe chacun des segments ainsi obtenus en sept parties égales, et sur ces sept parties on élimine les cinq parties centrales. Et ainsi de suite ! On construit de la sorte un ensemble de Cantor non dénombrable mais “extrêmement fin”, en fait de dimension 0.

- On construit un Cantor triadique sur $[0, 1/2]$, un Cantor 5-adique gras sur $[1/2, 3/4]$, un Cantor 7-adique gras sur $[3/4, 7/8]$, un Cantor 9-adique gras sur $[7/8, 15/16]$, etc. L’ensemble ainsi obtenu est de mesure de Lebesgue nulle, comme union dénombrable d’ensembles de mesure nulle ; mais il sera de dimension 1, puisque la mesure d -dimensionnelle d’un Cantor k -adique gras est $+\infty$ pour $d > \log 2 / \log c(k)$.

V-4.5. Autres exemples. Le flocon de von Koch dans \mathbb{R}^2 est l’un des fractals les plus simples et les plus célèbres : partant d’un triangle équilatéral, on construit sur chaque côté un triangle équilatéral plus petit d’un facteur $1/3$, pointant vers l’extérieur. Puis on recommence.... La frontière de la figure limite est appelée flocon de von Koch (voir [Falconer2], p.xv). Il n’est pas très difficile de montrer que sa dimension fractale est $\log 4 / \log 3$, ce qui correspond au fait qu’à chaque étape on remplace chaque segment de longueur ℓ par quatre segments de longueur $\ell/3$ (comparer au Cantor triadique, dans lequel on remplaçait chaque segment de longueur ℓ par deux segments de longueur $\ell/3$).

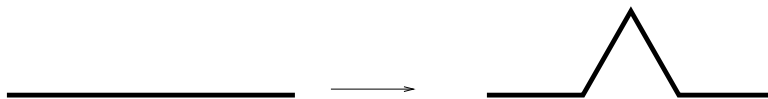


FIG. 8. Brique élémentaire de la construction du flocon de von Koch

Ici encore, la dimension de Hausdorff est strictement supérieure à la dimension topologique “naturelle” qui est 1. En particulier, le flocon de von Koch est de “longueur” infinie, et de “surface” nulle. Selon une argumentation célèbre de Mandelbrot, avec une bonne approximation on peut considérer qu’un objet tel que la côte de la Bretagne présente le même comportement : sauf à aller à des échelles ridiculement précises (de l’ordre du rocher), il est impossible de mesurer sa longueur ; des estimations de la dimension de cette côte ont même été proposées. D’autres fractals célèbres se trouvent dans [Falconer2], comme les ensembles de Julia, de Mandelbrot, ainsi que de nombreux fractals aléatoires.

Le calcul de la dimension des fractals a motivé le développement de méthodes de calcul de la dimension de Hausdorff, passées en revue dans [Falconer2]. On mentionnera en particulier la puissante et élégante technique de la **distribution de masse** (pp. 64–66) : étant donné une partie A de \mathbb{R}^n , si l'on peut trouver une mesure de probabilité μ sur A telle que

$$\iint_{A \times A} \frac{d\mu(x) d\mu(y)}{|x - y|^s} < +\infty,$$

alors $\dim_{\mathcal{H}}(A) \geq s$.

Malgré ces méthodes, le calcul de la dimension de Hausdorff est parfois un casse-tête insoluble. Voici un exemple “simple” discuté en pp. 148–153 de ce même ouvrage : sur le segment $[0, 1]$ définissons, pour $\lambda > 1$ et $s \in]1, 2[$, la fonction

$$f_{s,\lambda} : t \longmapsto \sum_{k=1}^{\infty} \lambda^{(s-2)k} \sin(\lambda^k t).$$

Cette fonction, dite “fonction de Weierstrass”, est continue (elle est donnée par un développement en série absolument convergent) mais différentiable nulle part sur $[0, 1]$ (noter que la série des dérivées est violemment divergente ; cela ne constitue bien sûr pas une preuve, mais rend plausible la non-différentiabilité). Il est prouvé dans [Falconer2] que pour λ assez grand, la dimension de Minkowski du graphe de $f_{s,\lambda}$ est exactement s . On conjecture que la dimension de Hausdorff a la même valeur, mais cela n'est toujours pas démontré (il est connu cependant que $\lim_{\lambda \rightarrow \infty} \dim_{\mathcal{H}}(G(f_{s,\lambda}, [0, 1])) = s$).

V-5. Utilisation dans les changements de variables

Les mesures de Hausdorff sont particulièrement utiles pour énoncer des changements de variables de \mathbb{R}^m dans \mathbb{R}^n de manière unifiée. Ce sont les fameuses formules de l'**aire** et de la **co-aire**. On les donne ici sans preuve ; le chapitre 3 de [Evans-Gariepy] leur est entièrement consacré.

THÉORÈME V-27 (formule de l'aire). *Soit $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une application Lipschitzienne, avec $m \geq n$, et soit $A \subset \mathbb{R}^n$ un ensemble Lebesgue-mesurable. Alors*

$$\int_A |\det \nabla T| = \int_{\mathbb{R}^m} \mathcal{H}^0[A \cap T^{-1}\{y\}] d\mathcal{H}^n(y).$$

THÉORÈME V-28 (formule de la co-aire). *Soit $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ une application Lipschitzienne, avec $m \leq n$, et soit $A \subset \mathbb{R}^n$ un ensemble Lebesgue-mesurable. Alors*

$$\int_A |\det \nabla T| = \int_{\mathbb{R}^m} \mathcal{H}^{n-m}[A \cap T^{-1}\{y\}] dy.$$

EXEMPLE V-29. Soit $f : [0, 1] \rightarrow \mathbb{R}^n$ une courbe Lipschitzienne simple ; alors on a, par la formule de l'aire,

$$\int_0^1 |f'(s)| ds = \int_{\mathbb{R}^m} 1_{f([0,1])} d\mathcal{H}^1(y) = \mathcal{H}^1[f([0,1])],$$

ce qui identifie encore \mathcal{H}^1 avec l'une des notions naturelles de longueur d'une courbe.

CHAPITRE VI

Espaces de Lebesgue et mesures signées

Jusqu'ici, nous avons considéré des fonctions “individuellement”. Dans ce chapitre et le suivant, nous porterons notre attention sur des familles entières de fonctions : des “espaces de fonctions”, ou **espaces fonctionnels**. Nous munirons ces espaces de structures géométriques (par exemple une norme pour mesurer la taille des fonctions, ou un produit scalaire pour définir l'orthogonalité) et topologiques (par exemple, la topologie induite par une norme). En caricaturant un peu, on peut dire que l'étude des propriétés géométriques et topologiques des espaces fonctionnels constitue l'**analyse fonctionnelle**.

Le but premier de l'analyse fonctionnelle est de mettre en place des schémas de démonstrations intuitifs ou simples, similaires aux arguments géométriques ou topologiques que l'on fait dans un espace Euclidien (usage de coordonnées, orthogonalité, construction de limites, etc.) On peut sans doute faire remonter ce point de vue à Fourier lui-même.

Dans ce chapitre, nous introduirons deux types d'espaces fonctionnels. Dans un premier temps, nous fixerons une mesure, et nous construirons des espaces de **fonctions** mesurables, définis par leur “degré d'intégrabilité” : intégrabilité de la puissance p pour les espaces L^p de Lebesgue. Après avoir étudié les relations entre ces espaces (ce qui nous mènera à quelques considérations subtiles telles que l'interpolation entre espaces de Lebesgue), nous élargirons le cadre pour considérer l'espace de toutes les fonctions mesurables. Enfin nous généraliserons encore en incluant les fonctions mesurables dans un espace plus grand, dont les éléments ne sont pas à proprement parler des fonctions, mais plutôt des fonctions d'ensembles : c'est l'espace des **mesures** signées ; nous verrons pourquoi on peut les considérer comme des “fonctions généralisées”.

Les premières questions que l'on se pose sur les espaces fonctionnels concernent la complétude, la séparabilité, la réflexivité, l'uniforme convexité, l'existence de systèmes de coordonnées commodés. Nous répondrons à certaines de ces questions dans le présent chapitre, et poursuivrons cette étude dans le chapitre suivant.

Dans ce chapitre, la section VI-1, qui introduit les espaces de Lebesgue et leurs propriétés élémentaires, est la plus importante. Les autres sections pourront être omises en première lecture, ou consultées en cas de besoin.

VI-1. Espaces L^p de Lebesgue

VI-1.1. Définitions.

DÉFINITION VI-1 (espaces L^p). Soit (X, μ) un espace mesuré.

- Pour tout $p \in]0, +\infty[$ on définit l'espace de Lebesgue d'ordre p comme l'ensemble des fonctions mesurables de X dans $\overline{\mathbb{R}}$ telles que $|f|^p$ soit intégrable.

- On définit l'espace de Lebesgue d'ordre ∞ comme l'ensemble des fonctions mesurables de X dans $\overline{\mathbb{R}}$ telles qu'il existe $C < \infty$ tel que $|f| \leq C$ en-dehors d'un ensemble de mesure nulle.

- On définit l'espace de Lebesgue d'ordre 0 comme l'ensemble des fonctions mesurables de X dans $\overline{\mathbb{R}}$ qui sont nulles en-dehors d'un ensemble de mesure finie.

L'espace de Lebesgue d'ordre p sur (X, μ) est noté $L^p(X, \mu)$ (ou $L^p(\mu)$ ou $L^p(X, d\mu)$ ou $L^p(d\mu)$ ou $L_p(X, \mu)$ ou $\mathcal{L}^p(X, \mu)$, etc.). S'il n'y a pas de confusion possible sur la mesure μ , on pourra le noter $L^p(X)$ (ou $\mathcal{L}^p(X)$ ou $L_p(X)$); s'il n'y a pas de confusion possible sur le couple (X, μ) on pourra le noter tout simplement L^p (ou \mathcal{L}^p ou L_p).

REMARQUE VI-2. L'espace $L^p(\mu)$ dépend de μ ; c'est évident pour $p < \infty$, mais on se laisse parfois piéger dans le cas $p = \infty$. Penser par exemple que $\|x \rightarrow x^2\|_{L^\infty(\delta_{1/2})} = 1/4 \dots$

EXEMPLES VI-3. (i) Si $X = \mathbb{N}$ et μ est la mesure de comptage, l'espace $L^p(X, \mu)$ pour $0 \leq p < +\infty$ est l'ensemble des suites réelles $(u_n)_{n \in \mathbb{N}}$ telles que (avec la convention $0^0 = 0$)

$$\sum_{n \in \mathbb{N}} |u_n|^p < +\infty.$$

Pour $p = \infty$ c'est l'ensemble des suites réelles bornées.

Dans ce cas, on utilise traditionnellement les notations ℓ^p ou $\ell^p(\mathbb{N})$ pour l'espace de Lebesgue d'ordre p .

(ii) Soient $B_1 = B_1(0) \subset \mathbb{R}^n$, et λ_n la mesure de Lebesgue dans \mathbb{R}^n . Alors, la fonction $f_\alpha : x \mapsto |x|^{-\alpha}$ appartient à $L^p(B_1, \lambda_n)$ si et seulement si $p < n/\alpha$, et à $L^p(\mathbb{R}^n \setminus B_1, \lambda_n)$ si et seulement si $p > n/\alpha$. Elle n'appartient à aucun espace $L^p(\mathbb{R}^n, \lambda_n)$. En pratique, pour vérifier l'appartenance d'une fonction à un espace de Lebesgue, on est souvent amené à étudier séparément l'intégrabilité L^p "locale" et l'intégrabilité L^p "à l'infini".

REMARQUES VI-4. (i) On rencontre exceptionnellement des espaces de Lebesgue d'ordre négatif. La définition ne fait pas de mystère : $f \in L^p$ ($p < 0$) si et seulement si $1/|f| \in L^{-p}$. Cette notion n'a guère d'intérêt que si μ est finie.

(ii) Les espaces de Lebesgue constituent en général une très bonne "échelle" pour quantifier l'intégrabilité des fonctions mesurables; mais parfois cette échelle n'est pas assez précise. On ne peut, par exemple, en termes d'appartenance à des espaces L^p , faire la différence entre des fonctions de référence telles que

$$h_{\alpha, \beta} : x \mapsto \frac{[\log(1/|x|)]^\beta}{|x|^\alpha}$$

pour des valeurs différentes de β . D'autres espaces fonctionnels plus "fins" permettent de distinguer ces fonctions : par exemple, les espaces de Lorentz $L^{p,q}$. Si $X = B_1(0) \subset \mathbb{R}^n$ est muni de la mesure de Lebesgue, alors la fonction $h_{\alpha, \beta}$ pour $\beta > 0$ appartient à L^p si et seulement si $p < n/\alpha$; et à $L^{p,q}$ si et seulement si $p < n/\alpha$ ou $p = n/\alpha$ et $q < 1/\beta$. Pour $\beta = 0$, cette fonction appartient à $L^{p, \infty}$, que l'on appelle aussi espace de Marcinkiewicz M^p . La

définition des espaces de Lorentz est un peu compliquée et nous n'en parlerons qu'après avoir discuté du concept de réarrangement. A ce stade, on donnera seulement la définition des espaces de Marcinkiewicz, qui est plus simple : *Par définition, $f \in M^p(X)$ ($p \geq 1$) s'il existe une constante C telle que*

$$\forall t \geq 0, \quad \mu[\{x; |f(x)| \geq t\}] \leq \left(\frac{C}{t}\right)^p.$$

On notera alors $\|f\|_{L^{p,\infty}(X)}$ l'infimum des constantes C admissibles (c'est un abus de notation, car il ne s'agit pas d'une norme).

EXERCICE VI-5. En utilisant l'inégalité de Chebyshev, montrer que pour tout $p \geq 1$, $L^p(X) \subset L^{p,\infty}(X)$, avec injection continue au sens où $\|f\|_{L^{p,\infty}} \leq \|f\|_{L^p}$. Montrer, en considérant des puissances inverses, que cette inclusion est stricte dans \mathbb{R}^n . L'espace $L^{p,\infty}$ est donc "un peu plus grand" que l'espace L^p .

Remarquons enfin que l'on peut étendre facilement la définition des espaces L^p à des espaces de fonctions à **valeurs vectorielles**, plus précisément à valeurs dans un espace muni d'une distance invariante par translation.

DÉFINITION VI-6. Soit E un espace vectoriel; on dit qu'une distance sur d est invariante par translation si pour tous $x, y, z \in E$ on a

$$d(x+z, y+z) = d(x, y).$$

Il est clair qu'une norme définit une distance invariante par translation. Mais le concept de distance invariante par translation est beaucoup plus général : par exemple, si N est une norme, alors $N/(1+N)$ est une telle distance.

DÉFINITION VI-7 (espaces L^p à valeurs vectorielles). Soient (X, μ) un espace mesuré, et E un espace vectoriel muni d'une distance d invariante par translation. Pour tout $p \in [0, +\infty]$, on définit alors l'espace $L^p(X; E) = L^p(X, \mu; E)$ comme l'espace des fonctions mesurables $f : X \rightarrow E$ telles que $d(0, f) \in L^p(X, \mu)$.

VI-1.2. Inégalité de Minkowski. Le point de départ de l'analyse fonctionnelle des espaces de Lebesgue est l'inégalité suivante.

THÉORÈME VI-8 (inégalité de Minkowski). Soit (X, μ) un espace mesuré, et soit $p \in [1, +\infty[$. Alors, pour toutes fonctions f, g mesurables $X \rightarrow \overline{\mathbb{R}}$,

$$\left(\int |f+g|^p d\mu\right)^{1/p} \leq \left(\int |f|^p d\mu\right)^{1/p} + \left(\int |g|^p d\mu\right)^{1/p},$$

où par convention $|(+\infty) + (-\infty)| = +\infty$. De plus, si $p > 1$ et si les deux intégrales apparaissant au membre de droite sont finies et non nulles, il y a égalité si et seulement si il existe $\alpha > 0$ tel que $f = \alpha g$ presque partout.

Il existe diverses démonstrations de l'inégalité de Minkowski ; on peut par exemple la déduire de l'inégalité de Hölder [Rudin p. 64 ; Lieb-Loss p. 48]. Nous allons donner ci-après une preuve légèrement différente, qui n'utilise pas explicitement l'inégalité de Hölder. Nous allons démontrer en même temps quelques variantes de l'inégalité de Minkowski, selon un point de vue parallèle à notre présentation de l'inégalité de Hölder au paragraphe VI-2.1.

THÉORÈME VI-9 (variantes de l'inégalité de Minkowski). Soit $p \in [1, +\infty[$.

(i) Soient (X, μ) un espace mesuré, f et g deux fonctions mesurables sur X , à valeurs dans $\overline{\mathbb{R}}$. Alors, pour tout $\lambda \in]0, 1[$,

$$\int |f + g|^p d\mu \leq \frac{1}{\lambda^{p-1}} \int |f|^p d\mu + \frac{1}{(1-\lambda)^{p-1}} \int |g|^p d\mu.$$

(ii) Soient (X, μ) un espace mesuré, f_1, \dots, f_k des fonctions mesurables sur X , à valeurs dans $\overline{\mathbb{R}}$. Alors

$$\left(\int \left| \sum_i f_i \right|^p d\mu \right)^{1/p} \leq \sum_i \left(\int |f_i|^p d\mu \right)^{1/p}.$$

(iii) Soient (X, μ) et (Y, π) deux espaces mesurés σ -finis. Alors, pour toute fonction F mesurable de $X \times Y$ dans $\mathbb{R}_+ \cup \{+\infty\}$, on a

$$\left(\int_X \left(\int_Y F(x, y) d\pi(y) \right)^p d\mu(x) \right)^{1/p} \leq \int_Y \left(\int_X F(x, y)^p d\mu(x) \right)^{1/p} d\pi(y).$$

(iv) Soient X et Y deux ensembles quelconques, et L un opérateur linéaire, défini sur un sous-espace vectoriel de l'ensemble des fonctions de X dans \mathbb{R} , à valeurs dans l'ensemble des fonctions de Y dans \mathbb{R} . On suppose que L est positif, i.e. $Lf \geq 0$ si $f \geq 0$. Soient $f, g \geq 0$ dans le domaine de L . Alors

$$L((f + g)^p)^{1/p} \leq [L(f^p)]^{1/p} + [L(g^p)]^{1/p},$$

ce qui est une inégalité entre deux fonctions de Y dans \mathbb{R} .

(v) Soient (X, μ) un espace mesuré, $(E, \|\cdot\|)$ un espace vectoriel normé, et $f, g : X \rightarrow E$ des fonctions mesurables. Alors

$$\left(\int \|f + g\|^p d\mu \right)^{1/p} \leq \left(\int \|f\|^p d\mu \right)^{1/p} + \left(\int \|g\|^p d\mu \right)^{1/p}.$$

(vi) Soient (X, μ) un espace mesuré, $f, g : X \rightarrow \mathbb{C}$ deux fonctions mesurables à valeurs complexes. Alors

$$\left(\int |f + g|^p d\mu \right)^{1/p} \leq \left(\int |f|^p d\mu \right)^{1/p} + \left(\int |g|^p d\mu \right)^{1/p}.$$

De plus, si $p > 1$ et si les deux intégrales apparaissant au membre de droite sont finies et non nulles, il y a égalité si et seulement si il existe $\alpha > 0$ tel que $f = \alpha g$ presque partout.

DÉMONSTRATION. Nous allons nous contenter de démontrer l'inégalité dans (i) et d'en déduire l'inégalité de Minkowski du Théorème VI-8. Le reste (discussion des cas d'égalité dans le Théorème VI-8, énoncés (ii) à (vi) du Théorème VI-9) est laissé en exercice.

Pour démontrer (i), on écrit d'abord, par convexité de la fonction $t \mapsto |t|^p$,

$$|f(x) + g(x)|^p = |\lambda(f(x)/\lambda) + (1-\lambda)(g(x)/(1-\lambda))|^p \leq \lambda|f(x)/\lambda|^p + (1-\lambda)|g(x)/(1-\lambda)|^p.$$

On intègre ensuite contre μ , pour trouver

$$\int |f + g|^p d\mu \leq \frac{1}{\lambda^{p-1}} \int |f|^p + \frac{1}{(1-\lambda)^{p-1}} \int |g|^p.$$

On optimise alors en λ (on minimise le membre de droite). L'inégalité de Minkowski est obtenue en utilisant l'identité élémentaire

$$\inf_{0 \leq \lambda \leq 1} \left(\frac{a}{\lambda^{p-1}} + \frac{b}{(1-\lambda)^{p-1}} \right) = (a^{\frac{1}{p}} + b^{\frac{1}{p}})^p.$$

De manière équivalente, on obtient l'inégalité souhaitée en posant

$$\lambda := \frac{(\int |f|^p)^{\frac{1}{p}}}{(\int |f|^p)^{\frac{1}{p}} + (\int |g|^p)^{\frac{1}{p}}}.$$

□

REMARQUE VI-10. Cette méthode de preuve (démonstration d'une inégalité auxiliaire dépendant d'un paramètre, puis optimisation sur ce paramètre) est très répandue en analyse.

VI-1.3. Distances L^p . Jusqu'à présent nous avons seulement défini l'ensemble des fonctions L^p ; maintenant nous allons munir cet ensemble d'une structure qui, selon les cas, sera soit une "semi-distance", soit une semi-norme.

THÉORÈME VI-11 (semi-distances L^p). *Soit (X, μ) un espace mesuré. Pour tout $p \in [0, +\infty]$, on définit sur $L^p(X, \mu)$ une application N_p , à valeurs dans $[0, \infty]$, par les formules*

$$N_p(f) = \left(\int |f|^p \right)^{\min(1, 1/p)} \quad (0 < p < \infty);$$

$$N_\infty(f) = \inf \left\{ C; \mu[\{x; |f(x)| > C\}] = 0 \right\};$$

$$N_0(f) = \mu[\{x; f(x) \neq 0\}].$$

La quantité $N_\infty(f)$ est appelée *supremum essentiel* de $|f|$, ce que l'on note $\text{esssup } |f|$. Les quantités $N_p(f)$ sont également notées $\|f\|_{L^p(X, \mu)}$ ou $\|f\|_{L^p(X)}$ ou $\|f\|_{L^p}$, etc., voire $\|f\|_p$.

L'application N_p définit alors sur L^p

(i) pour $1 \leq p \leq \infty$: une semi-norme, i.e.

$$N_p(f) \geq 0; \quad N_p(f + g) \leq N_p(f) + N_p(g); \quad N_p(\lambda f) = |\lambda| N_p(f);$$

(ii) pour $0 \leq p < 1$: une application positive, homogène de degré p , vérifiant l'inégalité triangulaire, i.e.

$$N_p(f) \geq 0; \quad N_p(f) \leq N_p(f) + N_p(g); \quad N_p(\lambda f) = |\lambda|^p N_p(f).$$

En outre, pour tout $p \in [0, +\infty]$, une fonction f dans $L^p(X)$ vérifie $N_p(f) = 0$ si et seulement si elle est nulle μ -presque partout.

DÉMONSTRATION. Les assertions d'homogénéité sont évidentes, ainsi que le traitement des cas d'égalité. Les inégalités triangulaires sont donc le coeur de cette proposition. Pour $p = 1$, $p = 0$ ou $p = \infty$, on les vérifie aisément ; pour $1 < p < \infty$ c'est l'inégalité de Minkowski ; pour $0 < p < 1$ c'est une conséquence immédiate de l'inégalité élémentaire

$$(a + b)^p \leq a^p + b^p.$$

□

Nous pouvons maintenant définir les espaces fonctionnels de Lebesgue. Pour ce faire, on va transformer les semi-distances L^p en distances, en quotientant l'espace par le noyau de N_p .

DÉFINITION VI-12 (espaces de Lebesgue). *Soit (X, μ) un espace mesuré, et soit $p \in [0, +\infty]$. On appelle espace de Lebesgue (quotienté) d'ordre p , et on note $L^p(X, \mu)$ ou $L^p(\mu)$ ou $L^p(X, d\mu)$ ou $L^p(d\mu)$ ou $L_p(X, \mu)$ ou $L^p(X)$ ou $L_p(X)$, ou simplement L^p ou L_p , l'espace vectoriel de toutes les classes d'équivalence de fonctions dans $L^p(X)$, pour la relation d'équivalence définie par l'égalité μ -presque partout. Si une classe d'équivalence f est donnée, N_p attribue la même valeur à tous ses représentants ; on note cette quantité $N_p(f)$, ou $\|f\|_{L^p}$, ou $\|f\|_p$, etc.*

L'espace $(L^p(X, \mu), N_p)$ ainsi défini est un espace vectoriel qui est

- normé pour $1 \leq p \leq \infty$;
- muni d'une distance invariante par translation pour $0 \leq p < 1$.

REMARQUES VI-13. (i) En clair, il y a deux espaces de Lebesgue L^p . Le premier est l'espace vectoriel des fonctions mesurables dont la puissance p est intégrable ; a priori, ce n'est pas un espace normé. Le deuxième est obtenu à partir du premier en identifiant des fonctions qui coïncident presque partout ; c'est un espace normé. Cette identification nous mène dans un univers peu rassurant où les "fonctions" ne sont pas définies partout, mais seulement presque partout, et où la valeur d'une fonction en un point donné n'est jamais déterminée. Cependant, si on ne procède pas à cette identification, on ne peut aller bien loin dans l'analyse fonctionnelle. Nous n'essaierons pas de distinguer les deux espaces par des notations différentes ; dans la suite, la dénomination "espace L^p " sera réservée au second. Certains auteurs tentent de différencier les deux espaces en utilisant le symbole \mathcal{L}^p pour l'espace non quotienté, mais la confusion est si répandue que nous n'essaierons pas d'en faire autant. Il est recommandé de ne pas utiliser l'espace quotienté si ce n'est pas nécessaire.

(ii) Si une fonction appartient à L^p , l'ensemble des points où elle est infinie est de mesure nulle. Quand on passe aux classes d'équivalence par la relation d'égalité presque partout, on peut donc supposer que les "fonctions" considérées sont à valeurs dans \mathbb{R} plutôt que $\overline{\mathbb{R}}$.

(iii) Pour $p \in [0, 1[$ l'espace L^p , quotienté par la relation d'égalité presque partout, est un espace vectoriel muni d'une distance invariante, mais ce n'est pas un espace vectoriel normé : l'inégalité de Minkowski telle que nous l'avons énoncée n'a plus lieu. Il existe en fait une inégalité de Minkowski dans ce cas, mais seulement pour des fonctions positives, et elle est renversée par rapport à celle que nous avons vue. On peut aller plus loin et montrer que L^p n'est pas normable. L'existence de la distance N_p dans ce cas ne suffit pas à en faire des espaces fonctionnels agréables, de sorte qu'on ne les utilise presque jamais.

Pour conclure ce paragraphe, nous allons définir les semi-distances L^p sur les espaces de Lebesgue à valeurs vectorielles $L^p(X; E)$.

PROPOSITION VI-14 (espaces de Lebesgue à valeurs vectorielles). *Soit (X, μ) un espace mesuré, et soit E un espace vectoriel muni d'une distance invariante par translation : pour tout $p \in [0, \infty]$ on définit*

$$N_p(f) := N_p(d(0, f)).$$

L'application N_p est alors une application positive, vérifiant l'inégalité triangulaire. Si E est un espace vectoriel normé et d la distance associée à la norme, N_p est homogène de degré $\min(p, 1)$, et en particulier définit une semi-norme pour $p \geq 1$.

Si l'on quotiente $L^p(X; E)$ par la relation d'égalité presque partout, on obtient un espace vectoriel sur lequel N_p définit une distance invariante par translation. Si E est un espace vectoriel normé, et $p \geq 1$, alors l'espace $L^p(X; E)$ ainsi obtenu est un espace vectoriel normé.

Beaucoup des propriétés que nous verrons par la suite se généralisent sans difficulté à ce cadre à valeurs vectorielles; nous nous contenterons d'en mentionner certaines, sans démonstration. Le seul point un tant soit peu délicat dans le manie-ment des espaces de Lebesgue à valeurs vectorielles ne concerne pas les opérations dans les espaces L^p , mais la construction de l'intégrale.

VI-1.4. Théorème de convergence dominée L^p . Avant d'aller plus loin, nous allons examiner une variante simple et utile du théorème de convergence dominée, adaptée aux espaces de Lebesgue.

THÉORÈME VI-15 (convergence dominée dans les L^p). *Soient (X, μ) un espace mesuré et $p \in]0, +\infty[$. Soit $(f_k)_{k \in \mathbb{N}}$ une suite de fonctions mesurables de X dans $\overline{\mathbb{R}}$, convergeant presque partout vers une fonction f . On suppose qu'il existe une fonction $g \in L^p(X)$ telle que $|f_k| \leq g$ presque partout, pour tout k . Alors, $f \in L^p(X, \mu)$ et*

$$\int |f_k - f|^p d\mu \xrightarrow[k \rightarrow \infty]{} 0,$$

c'est-à-dire que f_k converge vers f dans L^p .

DÉMONSTRATION. Il est clair que $|f(x)| \leq g(x)$ pour presque tout x , et donc que $f \in L^p$. Pour obtenir le reste de l'énoncé, on applique le théorème de convergence dominée à la famille $|f_k - f|^p$: cette famille est dominée par la fonction intégrable $(2g)^p$, et converge presque partout vers 0, son intégrale converge donc vers 0. \square

VI-1.5. Théorème de Riesz–Fischer. L'analyse étant basée pour une grande part sur des procédés de limite et d'approximation, on n'étudie d'ordinaire les espaces vectoriels normés que s'ils sont **complets**, i.e. toute suite de Cauchy converge. Le théorème suivant assure la complétude des espaces de Lebesgue.

THÉORÈME VI-16 (théorème de complétude de Riesz–Fischer). *Soit (X, μ) un espace mesuré, et soit $p \in [0, +\infty]$. Soit $(f_k)_{k \in \mathbb{N}}$ une suite de Cauchy dans $L^p(X, \mu)$. Alors*

(i) il existe $f \in L^p$ tel que $f_k \rightarrow f$ dans L^p ;

(ii) il existe une suite extraite de (f_k) , notée $(f_{k'})$, et une fonction g fixée dans L^p , telle que

$$|f_{k'}| \leq g \quad \mu\text{-presque partout;}$$

$$f_{k'}(x) \xrightarrow[k' \rightarrow \infty]{} f(x) \quad \text{pour } \mu\text{-presque tout } x.$$

COROLLAIRE VI-17 (statut des espaces L^p). Soit (X, μ) un espace mesuré. Alors,

(i) pour tout $p \in [1, +\infty]$, l'espace $L^p(X, \mu)$, muni de la norme L^p , est un **espace de Banach**, i.e. un espace vectoriel normé complet.

(ii) l'espace $L^2(X, \mu)$, muni de la forme bilinéaire symétrique

$$(f, g) \longrightarrow \int fg \, d\mu$$

est en outre un **espace de Hilbert**, i.e. un espace vectoriel complet muni d'une forme bilinéaire symétrique définie positive.

(iii) pour tout $p \in [0, 1[$, l'espace $L^p(X, \mu)$, muni de la distance L^p , est un **espace de Fréchet**, i.e. un espace vectoriel muni d'une distance invariante par translation, complet.

REMARQUE VI-18. La complétude éventuelle de l'espace X ne joue aucun rôle ; ce qui est utilisé en revanche de manière cruciale, c'est la complétude de l'espace d'arrivée, ici \mathbb{R} . Ces résultats se généralisent aux espaces de Lebesgue à valeurs vectorielles, $L^p(X; E)$, si E est

- un espace de Banach dans le cas (i) ;
- un espace de Hilbert dans le cas (ii) ;
- un espace de Fréchet dans le cas (iii).

L'analyse de Banach et l'analyse de Hilbert sont les branches les plus développées de l'analyse fonctionnelle. On en recensera dans le chapitre suivant les résultats les plus fondamentaux, et on les appliquera aux espaces de Lebesgue. On ne parlera pas du tout d'analyse dans les espaces de Fréchet, qui est plus délicate et d'usage beaucoup moins répandu.

DÉMONSTRATION DU THÉORÈME DE RIESZ–FISCHER. Soit (f_k) une suite de Cauchy dans L^p ; sa convergence sera assurée si on démontre l'existence d'une sous-suite convergente. Par récurrence, on construit une suite extraite, toujours notée (f_k) , telle que

$$N_p(f_{k+1} - f_k) \leq 2^{-k}.$$

Le problème est de construire une limite à cette suite. Pour cela nous distinguerons plusieurs cas.

1. Supposons d'abord $1 \leq p < \infty$. On pose $f_0 = 0$, et

$$g_k(x) := \sum_{j=1}^k |f_j(x) - f_{j-1}(x)|, \quad g(x) := \sum_{j=1}^{\infty} |f_j(x) - f_{j-1}(x)|$$

Par convergence monotone,

$$\int g(x)^p = \lim_{k \rightarrow \infty} \int g_k(x)^p \, d\mu;$$

et par inégalité de Minkowski,

$$\|g_k\|_{L^p} \leq \|f_0\|_{L^p} + \sum_{j=1}^k \|f_j - f_{j-1}\|_{L^p} \leq \|f_0\|_{L^p} + 2,$$

on en déduit que $g \in L^p(X)$. En particulier, il existe un ensemble négligeable N tel que $g(x) < +\infty$ pour tout $x \notin N$. Pour de tels x , la série $\sum (f_j(x) - f_{j-1}(x))$ est absolument convergente (par complétude de \mathbb{R}), et on pose

$$f(x) := \sum_{j=1}^{\infty} (f_j(x) - f_{j-1}(x)) = \lim_{n \rightarrow \infty} f_n(x).$$

On définit ensuite f arbitrairement (par exemple $f = 0$) sur N . la suite (f_n) est alors dominée par une fonction L^p et converge presque partout vers f , on en déduit qu'elle converge vers f dans L^p .

2. Pour $p = \infty$, on sait qu'en-dehors d'un ensemble de mesure nulle on a

$$|f(x)| \leq \|f_0\|_{L^\infty} + \sum_{k=0}^{\infty} \|f_k - f_{k+1}\|_{L^\infty},$$

ce qui montre que $f \in L^\infty$; en outre,

$$|f_n(x) - f(x)| \leq \sum_{k=n}^{\infty} \|f_k - f_{k+1}\|_{L^\infty} \xrightarrow{n \rightarrow \infty} 0,$$

ce qui prouve la convergence de f_k vers f dans L^∞ .

3. Dans le cas où $0 < p < 1$, on pose

$$g_k(x) := \left(|f_0(x)|^p + \sum_{j=1}^k |f_j(x) - f_{j-1}(x)|^p \right)^{\frac{1}{p}}, \quad g(x) := \left(|f_0(x)|^p + \sum_{j=1}^{\infty} |f_j(x) - f_{j-1}(x)|^p \right)^{\frac{1}{p}}.$$

Par convergence monotone, on a toujours

$$\int |g(x)|^p = \lim_{k \rightarrow \infty} \int |g_k(x)|^p d\mu;$$

et c'est cette fois l'inégalité triangulaire qui assure que

$$N_p(g_k) \leq N_p(f_0) + \sum_{j=1}^k N_p(f_j - f_{j-1}) \leq N_p(f_0) + 2,$$

on en déduit que $|g|^p \in L^1(X)$ et on conclut comme dans le cas $1 \leq p < \infty$.

4. Enfin, pour $p = 0$ on peut écrire

$$\int_X \sum_{k \geq 1} 1_{f_k \neq f_{k-1}} d\mu = \sum_{k \geq 1} \int_X 1_{f_k \neq f_{k-1}} d\mu < +\infty;$$

en particulier, l'ensemble N des $x \in X$ tels que $f_k(x) \neq f_{k-1}(x)$ pour une infinité de k est de mesure nulle. Pour tout $x \notin N$ on sait que la suite $(f_k(x))$ est constante à partir d'un certain rang, et en particulier converge vers une fonction que l'on note $f(x)$. On redéfinit $f = 0$ sur N . Comme

$$\lim_{k_0 \rightarrow \infty} \sum_{k \geq k_0} \int_X 1_{f_k \neq f_{k-1}} d\mu = 0,$$

on voit que pour k_0 assez grand la mesure de l'ensemble des x tels qu'il existe un $k \geq k_0$ pour lequel $f_k(x) \neq f_{k-1}(x)$ est arbitrairement petite. On conclut que $N_0(f_{k_0} - f)$ est arbitrairement petit pour k_0 assez grand. La preuve est donc (c'est le cas de le dire) complète. \square

REMARQUE VI-19. Dans le cas où $p = 1$, on a retrouvé une variante de la réciproque du théorème de convergence dominée (Théorème III-21).

VI-1.6. Produit tensoriel d'espaces de Lebesgue. Soient (X_1, μ_1) et (X_2, μ_2) des espaces mesurés. Nous avons accès aux espaces $L^p(X_1, \mu_1)$ et $L^p(X_2, \mu_2)$, faits de fonctions p -intégrables dans la variable x_1 ou dans la variable x_2 . Quand on manipule des fonctions L^p intégrables dans les deux variables x_1 et x_2 , elles peuvent être

- données intrinsèquement comme fonctions de x_1 et x_2 ;
- construites à partir de fonctions de x_1 et de fonctions de x_2 , et d'opérations élémentaires ou de passages à la limite.

Pour produire une fonction p -intégrable dans les deux variables à partir de fonctions p -intégrables d'une variable, une opération élémentaire particulièrement simple et naturelle consiste à multiplier de telles fonctions. Soient donc $f_1 \in L^p(X_1, \mu_1)$ et $f_2 \in L^p(X_2, \mu_2)$, on note

$$(f_1 \otimes f_2)(x_1, x_2) = f_1(x_1) f_2(x_2).$$

Cette fonction est appelée produit tensoriel de f_1 par f_2 .

Nous avons donc deux espaces a priori intéressants :

- l'espace $L^p(X_1 \times X_2, \mu_1 \otimes \mu_2)$;
- l'espace $L^p(X_1, \mu_1) \otimes L^p(X_2, \mu_2)$, qui par définition est l'adhérence dans $L^p(X_1 \times X_2, \mu_1 \otimes \mu_2)$ de l'espace vectoriel engendré par les produits tensoriels ; c'est donc l'ensemble de toutes les limites de combinaisons linéaires finies de produits tensoriels.

Le théorème suivant, que nous démontrerons au chapitre suivant, donne des conditions suffisantes pour qu'il y ait identité entre ces deux notions, et pour que toute fonction L^p -intégrable dans les deux variables puisse être approchée par des combinaisons linéaires de produits tensoriels :

THÉORÈME VI-20. Soient (X_1, d_1) et (X_2, d_2) des espaces métriques séparables, équipés de mesures de Borel μ_1 et μ_2 , régulières et σ -finies. On munit $X_1 \times X_2$ de la topologie produit ; alors pour tout $p \in [1, +\infty[$, $L^p(X_1, \mu_1) \otimes L^p(X_2, \mu_2) = L^p(X_1 \times X_2, \mu_1 \otimes \mu_2)$.

On trouvera la démonstration en p. 258.

VI-1.7. Espaces de Lebesgue locaux. Il est souvent utile de travailler avec des fonctions qui sont intégrables, ou L^p -intégrables, sur des ensembles bornés (par exemple) sans être nécessairement intégrables sur tout l'espace. Dans ce cours, on adoptera la définition suivante.

DÉFINITION VI-21 (espaces de Lebesgue locaux). Soit (X, d) un espace métrique. On note $L^p_{\text{loc}}(X)$ l'ensemble des fonctions mesurables $X \rightarrow \mathbb{R}$ qui sont L^p -intégrables sur toutes les boules de X .

Bien noter que cette définition dépend fortement de la métrique, pas seulement de la topologie. On pourrait bien sûr définir les espaces locaux en utilisant des ensembles compacts ; mais en pratique c'est le concept précédent qui nous sera utile.

VI-2. Inégalités et relations entre espaces de Lebesgue

Dans cette section on va passer en revue des inégalités précieuses qui lient les normes de Lebesgue L^p pour des exposants p différents.

VI-2.1. Inégalité de Hölder L^p et dualité des normes L^p .

THÉORÈME VI-22 (inégalité de Hölder dans les espaces L^p). Soit (X, μ) un espace mesuré, et soient f, g deux fonctions mesurables de X dans \mathbb{R} , $p \in [1, +\infty]$, $p' := p/(p-1)$. Alors

$$\left| \int_X fg \, d\mu \right| \leq \int_X |fg| \, d\mu \leq \|f\|_{L^p} \|g\|_{L^{p'}}$$

(où le membre de gauche est par convention $+\infty$ si fg n'est pas intégrable).

Plus généralement, soient f_1, \dots, f_k des fonctions mesurables, et p_1, \dots, p_k des exposants dans $[1, +\infty]$, tels que $\sum p_k^{-1} \geq 1$. Alors

$$\left\| \prod_j f_j \right\|_{L^r} \leq \prod_j \|f_j\|_{L^{p_j}}, \quad \frac{1}{r} = \sum_{j=1}^k \frac{1}{p_j}.$$

DÉMONSTRATION. Si $r = \infty$, nécessairement $p = q = \infty$ et l'inégalité est évidente. Dans le cas contraire, il suffit d'appliquer l'inégalité de Hölder habituelle aux fonctions $|f|^r$ et $|g|^r$, avec les exposants conjugués p/r et q/r (en effet, $(r/p) + (r/q) = 1$). \square

COROLLAIRE VI-23 (convergence de produits). Soient (X, μ) un espace mesuré, $p \in [1, \infty]$ et $p' := p/(p-1)$. Soient (f_n) et (g_n) des suites de fonctions mesurables, telles que $f_n \rightarrow f$ dans $L^p(X, \mu)$ et $g_n \rightarrow g$ dans $L^{p'}(X, \mu)$. Alors $f_n g_n$ converge vers fg dans L^1 , et en particulier

$$\int f_n g_n \, d\mu \rightarrow \int fg \, d\mu.$$

Plus généralement, si on se donne k suites de fonctions mesurables $(f_{1,n}), \dots, (f_{k,n})$ telles que

$$\forall j, \quad f_{j,n} \rightarrow f_j \quad \text{dans } L^{p_j}(\mu),$$

avec $\sum p_j^{-1} \leq 1$. Alors

$$\prod_j f_{j,n} \xrightarrow{L^r(\mu)} \prod_j f_j, \quad \frac{1}{r} = \sum_{j=1}^k \frac{1}{p_j}.$$

DÉMONSTRATION. Il est clair que le deuxième énoncé implique le premier, et que par récurrence, il suffit de traiter le cas $k = 2$. On se donne donc deux exposants p et q , et $f_n \rightarrow f$ dans L^p , $g_n \rightarrow g$ dans L^q , et on cherche à montrer que $f_n g_n \rightarrow fg$ dans L^r , avec $1/r = (1/p) + (1/q)$. Pour cela on écrit

$$\|f_n g_n - fg\|_{L^r} \leq \|f_n(g_n - g)\|_{L^r} + \|g(f_n - f)\|_{L^r} \leq \|f_n\|_{L^p} \|g_n - g\|_{L^q} + \|g\|_{L^p} \|f_n - f\|_{L^q}.$$

Puisque la suite (f_n) converge dans L^p , elle est bornée dans cet espace; on en déduit que l'expression précédente converge vers 0 quand $n \rightarrow \infty$. \square

Nous verrons au chapitre suivant que l'espace L^p peut être identifié à l'espace des formes linéaires continues sur $L^{p'}$, $p' := p/(p-1)$, sous certaines restrictions sur p ($1 < p \leq \infty$, X σ -fini pour $p = \infty$); on dit qu'il y a *dualité* entre les espaces L^p et $L^{p'}$. Indépendamment de ce théorème non trivial, on peut démontrer simplement certains liens très utiles entre norme L^p et norme $L^{p'}$, valables pour tous $p \in [1, \infty]$.

THÉORÈME VI-24 (représentation duale des normes L^p). Soient (X, μ) un espace mesuré, et $p \in [1, \infty[$. Alors, pour tout $f \in L^p(\mu)$,

$$\|f\|_{L^p(\mu)} = \sup \left\{ \int f g d\mu; \quad \|g\|_{L^{p'}(\mu)} = 1 \right\} = \sup_{\|g\|_{L^{p'}(\mu)} \neq 0} \frac{\int f g d\mu}{\|g\|_{L^{p'}(\mu)}},$$

où $p' := p/(p-1) \in]1, \infty]$. En outre, le supremum peut être restreint à l'ensembles des fonctions g qui s'écrivent comme combinaisons linéaires (finies) de fonctions indicatrices d'ensembles mesurables de mesure finie.

Si (X, μ) est σ -fini, cet énoncé est également valable pour $p = \infty$.

DÉMONSTRATION. Si $f = 0$ (presque partout), l'identité est évidente; on se limite donc au cas où $f \neq 0$. L'égalité entre les deux suprema est une conséquence de ce que $\|\cdot\|_{L^p}$ est une norme. L'inégalité de Hölder se réécrit

$$\frac{\int f g d\mu}{\|g\|_{L^{p'}(\mu)}} \leq \|f\|_{L^p(\mu)},$$

pour tout $g \in L^{p'}(\mu)$, ce qui implique

$$\sup_{\|g\|_{L^{p'}(\mu)} \neq 0} \frac{\int f g d\mu}{\|g\|_{L^{p'}(\mu)}} \leq \|f\|_{L^p(\mu)}.$$

Il suffit donc de montrer que

$$(39) \quad \|f\|_{L^p(\mu)} \leq \sup_{\|g\|_{L^{p'}(\mu)} \neq 0} \frac{\int f g d\mu}{\|g\|_{L^{p'}(\mu)}}.$$

Commençons par le cas où $p < \infty$; pour montrer (39) il suffit de choisir $g := |f|^{p-2}f \in L^{p'}(\mu)$.

Montrons maintenant, toujours dans le cas $p < \infty$, que le supremum peut être restreint à des fonctions "très simples". On sait que la partie positive f_+ de f est limite d'une suite croissante de fonctions simples h_k ; puisque $f_+ \in L^p(X, \mu)$, ces fonctions sont L^p -intégrables, et par convergence monotone, h_k converge vers f_+ dans L^p . En appliquant le même raisonnement à f_- , on voit que f est limite dans L^p d'une suite f_k de fonctions simples L^p , qui s'écrivent forcément comme combinaisons linéaires de fonctions indicatrices d'ensembles mesurables de mesure finie. Il en est de même de $g_k := |f_k|^{p-2}f_k$. On a alors $\|f_k\|_{L^p} = \left(\int f_k g_k d\mu \right)^{1/p}$. On écrit, en utilisant l'inégalité de Hölder,

$$\int f g_k d\mu = \int f_k g_k d\mu + \int (f - f_k) g_k d\mu \geq \|f_k\|_{L^p} - \|f - f_k\|_{L^p} \|g_k\|_{L^{p'}}.$$

Le premier terme du membre de droite tend vers $\|f\|_{L^p}$, tandis que le second tend vers 0 puisque $f_k \rightarrow f$ dans L^p . On en déduit que

$$\liminf \int f g_k d\mu \geq \|f\|_{L^p},$$

ce qui conclut la preuve.

Passons maintenant au cas où $p = \infty$. Par définition de $\|f\|_{L^\infty}$, pour tout $\varepsilon > 0$, l'ensemble $Y_\varepsilon := \{x; |f(x)| > \|f\|_{L^\infty} - \varepsilon\}$ est de mesure strictement positive. Comme X est σ -fini, on peut trouver dans Y_ε un sous-ensemble Z_ε de mesure finie

et strictement positive (il s'agit là d'une manière très faible d'utiliser la σ -finitude). On pose alors

$$g := \frac{1_{Z_\varepsilon} \operatorname{sign}(f)}{\mu[Z_\varepsilon]};$$

c'est une fonction intégrable, $\|g\|_{L^1} = 1$ et

$$\int_X fg \, d\mu = \frac{1}{\mu[Z_\varepsilon]} \int_{Z_\varepsilon} |f| \, d\mu \geq \frac{1}{\mu[Z_\varepsilon]} \int_{Z_\varepsilon} (\|f\|_{L^\infty} - \varepsilon) \, d\mu \geq \|f\|_{L^\infty} - \varepsilon.$$

On conclut en faisant tendre ε vers 0. \square

VI-2.2. Relations d'inclusion. Il est souvent crucial de garder en tête les relations d'inclusion entre espaces de Lebesgue. Nous limiterons la discussion au cas où $p \geq 1$. Si l'espace X est de mesure finie, alors les espaces de Lebesgue $L^p(X)$ sont emboîtés :

THÉORÈME VI-25 (emboîtement décroissant des espaces de Lebesgue). *Soit (X, μ) un espace mesuré fini : $\mu[X] < +\infty$. Alors, dès que $q \geq p \geq 1$ on a, pour tout f mesurable de X dans $\overline{\mathbb{R}}$,*

$$\|f\|_{L^p} \leq \|f\|_{L^q} \mu[X]^{\frac{1}{p} - \frac{1}{q}}.$$

En particulier, les espaces de Lebesgue $L^p(X, \mu)$ ($p \geq 1$) sont emboîtés dans le sens décroissant :

$$q \geq p \geq 1 \implies L^q \subset L^p,$$

et cette injection est continue.

DÉMONSTRATION. C'est une conséquence de l'inégalité de Hölder : on écrit

$$\int |f|^p \times 1 \leq \left(\int (|f|^p)^{q/p} \right)^{p/q} \left(\int 1 \right)^{1-p/q}$$

et on élève les deux membres de l'inégalité à la puissance $1/q$. \square

Dans le cas général, les espaces de Lebesgue ne sont pas emboîtés, et il n'y a pas de règle générale. On peut d'ailleurs trouver des situations où l'emboîtement a lieu, mais dans le sens opposé à celui que nous venons de décrire.

THÉORÈME VI-26 (emboîtement croissant des espaces de Lebesgue). *Soit (X, μ) un espace mesuré tel que*

$$\exists \varepsilon > 0; \quad \forall x \in X, \quad \mu[\{x\}] \geq \varepsilon.$$

Alors, dès que $q \geq p \geq 1$, pour tout f mesurable de X dans $\overline{\mathbb{R}}$ on a

$$\|f\|_{L^q} \leq \frac{1}{\varepsilon^{\frac{1}{p} - \frac{1}{q}}} \|f\|_{L^p}.$$

En particulier, les espaces de Lebesgue $L^p(X, \mu)$ ($p \geq 1$) sont emboîtés dans le sens croissant :

$$q \geq p \geq 1 \implies L^p(X) \subset L^q(X),$$

et l'injection est continue.

DÉMONSTRATION. Sans perte de généralité, supposons f positive. Pour tout $x_0 \in X$, on a

$$\int f^p d\mu \geq \mu[\{x_0\}] f(x_0)^p \geq \varepsilon f(x_0)^p.$$

En passant au supremum essentiel, on obtient

$$\|f\|_{L^p} \geq \varepsilon^{1/p} \|f\|_{L^\infty}.$$

En reportant cette information dans l'inégalité

$$\int f^q \leq \|f\|_{L^\infty}^{q-p} \int f^p,$$

on trouve

$$\int f^q \leq \frac{1}{\varepsilon^{\frac{q-p}{p}}} \|f\|_{L^p}^p \|f\|_{L^\infty}^{q-p},$$

d'où l'on déduit facilement le résultat. \square

EXEMPLE VI-27. Soit $X = B_1(0) \subset \mathbb{R}^n$, muni de la mesure de Lebesgue λ_n : alors les espaces $L^p(B_1)$ sont emboîtés dans le sens décroissant. En revanche, les espaces $\ell^p(\mathbb{N})$ sont emboîtés dans le sens croissant. Les espaces $L^p(\mathbb{R})$ en revanche ne sont emboîtés ni dans le sens croissant, ni dans le sens décroissant.

Cependant, même s'ils ne sont pas emboîtés, les espaces de Lebesgue sont “*en interpolation*” :

THÉORÈME VI-28 (interpolation des espaces de Lebesgue). *Soit X un espace mesuré. Alors, dès que $1 \leq p \leq q \leq r \leq \infty$, on a, pour toute fonction mesurable $f : X \rightarrow \overline{\mathbb{R}}$,*

$$(40) \quad \|f\|_{L^q} \leq \|f\|_{L^p}^\theta \|f\|_{L^r}^{1-\theta},$$

où θ est choisi de sorte que

$$\frac{1}{q} = \frac{\theta}{p} + \frac{1-\theta}{r}.$$

En particulier,

$$L^p \cap L^r \subset L^q,$$

et cette injection est continue. En outre, $L^p \cap L^r$ est dense dans L^q .

DÉMONSTRATION. La preuve de l'inégalité (40) consiste à écrire $f^q = f^a f^b$, où $q = a + b$, et à appliquer l'inégalité de Hölder avec des exposants bien choisis ; il s'agit d'un excellent exercice, vivement recommandé au lecteur. On en déduit bien sûr que $L^p \cap L^r$ est inclus dans L^q . L'injection est continue si l'on munit $L^p \cap L^r$ de sa norme “naturelle” $\|f\|_{L^p} + \|f\|_{L^r}$. Reste à prouver la densité : au vu des relations d'inclusion, on a $L^1 \cap L^\infty \subset L^p \cap L^r \subset L^q$, il suffit donc de montrer que $L^1 \cap L^\infty$ est dense dans L^q . Soit donc $f \in L^q$, on pose $f_k := f 1_{|f| \leq k}$. Alors f_k est borné par construction, et intégrable puisque $\int |f|^q \geq k^{q-1} \int |f_k|$. En appliquant le Théorème VI-15, on vérifie facilement que $\|f_k - f\|_{L^q} \rightarrow 0$ quand $k \rightarrow \infty$. \square

REMARQUE VI-29. Nous examinerons en fin de chapitre des théorèmes plus généraux, dits d'interpolation, qui vont dans la même direction.

Voici un corollaire simple et utile du théorème précédent.

COROLLAIRE VI-30 (convergence via interpolation). *Soit X un espace mesuré, et soient p, q deux exposants compris entre 1 et ∞ . Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables convergeant vers f dans L^p , et bornée dans L^q . Alors, pour tout exposant r compris entre p et q (exclus), la suite (f_n) converge vers f dans L^r .*

VI-2.3. Continuité de la norme en p . Nous avons défini une famille de normes L^p pour un paramètre p variant continûment entre 0 et ∞ . Une question très naturelle est la continuité de cette norme en le paramètre p .

THÉOREME VI-31 (continuité de la norme L^p en p). *Soient (X, μ) un espace mesuré, et $f : X \rightarrow \mathbb{R}$ une fonction mesurable. Soit*

$$J := \left\{ p \in [0, \infty]; N_p(f) < +\infty \right\}.$$

Alors J est un intervalle (éventuellement vide) et $p \mapsto N_p(f)$ est continue sur l'adhérence de J (à valeurs dans $[0, +\infty]$).

REMARQUES VI-32. (i) Remarquons que $p \mapsto N_p(f)$ n'est en général pas continue sur $[0, +\infty]$ tout entier; c'est pourquoi nous nous restreignons à l'adhérence de J . Par exemple, si l'on considère $X = \mathbb{R}$, muni de la mesure de Lebesgue, alors la fonction identiquement égale à 1 n'appartient à aucun autre espace de Lebesgue que L^∞ , donc $N_p(f) = +\infty$ pour tout $p < +\infty$; mais $N_\infty(f) = 1$, il n'y a donc pas continuité quand $p \rightarrow \infty$. En fait, pour tout $p_0 > 0$ on peut trouver, en jouant sur la décroissance à l'infini et une singularité en 0, une fonction qui appartienne à $L^p(\mathbb{R})$ uniquement si $p = p_0$ (exercice).

(ii) L'intérêt principal de ce théorème est sans doute la continuité en $+\infty$. En fait, d'après la démonstration qui suit, dès qu'il existe $q \geq 1$ tel que $f \in L^q(X)$, alors

$$N_p(f) \xrightarrow[p \rightarrow \infty]{} \|f\|_{L^\infty}.$$

Cet énoncé est parfois utile dans des problèmes de recherche très concrets (par exemple le "schéma d'itération de Moser" en théorie des équations aux dérivées partielles).

(iii) L'inégalité (40) entraîne que $\log N_p(f)$ est une fonction convexe de $1/p$, et on peut montrer que cette fonction est semi-continue inférieurement. Ces propriétés impliquent que $N_p(f)$ est continue sur \overline{J} (la convexité implique seulement la continuité dans l'intérieur de J). Cependant, nous allons donner une démonstration qui n'utilise pas explicitement cet argument.

DÉMONSTRATION. 1. Le fait que l'ensemble des valeurs de p où $N_p(f) < +\infty$ est un intervalle découle facilement du Théorème VI-28. En fait on peut montrer que la fonction $\log N_p(f)$ est une fonction convexe de $1/p$, ce qui implique aussi le résultat.

2. Considérons d'abord la continuité en $p \neq \{0, \infty\}$. Par continuité de l'application $p \mapsto X^{\min(1, 1/p)}$ pour $p \in]0, +\infty[$, il nous suffit de prouver que pour toute suite p_k convergeant vers p , $f \in L^{p_k}$,

$$\int |f|^{p_k} d\mu \xrightarrow[k \rightarrow \infty]{} \int |f|^p d\mu.$$

Notons bien que l'hypothèse $f \in L^p$ n'est pas faite, de sorte que p pourrait être au bord de l'intervalle J . On supposera par exemple que p_k tend vers p en croissant. Alors on a convergence monotone (l'une croissante, l'autre décroissante) de $|f|^{p_k} 1_{|f| \geq 1}$ et $|f|^{p_k} 1_{|f| < 1}$ vers $|f|^p 1_{|f| \leq 1}$ et $|f|^p 1_{|f| < 1}$ respectivement. Le passage à la limite croissante ne pose pas de problème; et puisque $|f|^{p_k} \in L^1$, on peut passer aussi à la limite décroissante. On conclut que

$$\int_{|f| \geq 1} |f|^{p_k} \longrightarrow \int_{|f| \geq 1} |f|^p; \quad \int_{|f| < 1} |f|^{p_k} \longrightarrow \int_{|f| < 1} |f|^p.$$

Le théorème en découle.

3. Un raisonnement du même type permet de traiter le cas $p = 0$, en séparant les trois cas $|f| = 0$, $0 < |f| \leq 1$, $|f| > 1$.

4. Passons maintenant au cas où $p = \infty$. Soit $q \geq 1$ tel que $f \in L^q$. En utilisant les identités élémentaires

$$\|f\|_{L^q} = \| |f|^q \|_{L^1}^{1/q}; \quad \| |f|^q \|_{L^\infty} = \|f\|_{L^\infty}^q,$$

on voit que l'on peut remplacer le problème sur f par le problème sur $|f|^q$, et que nous pouvons donc supposer sans perte de généralité

$$f \in L^1; \quad f \geq 0.$$

On supposera également que f n'est pas identiquement nulle, auquel cas la solution est triviale; donc $\|f\|_{L^\infty} \neq 0$.

5. Soit $K > 0$ tel que $K < \|f\|_{L^\infty}$. Par définition du supremum essentiel, on a

$$a(K) := \mu[\{f \geq K\}] > 0;$$

il s'ensuit, par inégalité de Chebyshev,

$$\|f\|_{L^p} \geq [a(K)M^p]^{1/p} = Ma(K)^{1/p} \xrightarrow{p \rightarrow \infty} M.$$

En faisant tendre M vers $\|f\|_{L^\infty}$, on en déduit (que $\|f\|_{L^\infty}$ soit fini ou non)

$$\liminf_{p \rightarrow \infty} \|f\|_{L^p} \geq \|f\|_{L^\infty}.$$

Si $\|f\|_{L^\infty} = +\infty$, ceci achève la preuve.

6. Supposons maintenant que $\|f\|_{L^\infty} < +\infty$. Comme $f \in L^1$, nous pouvons utiliser, pour tout $p \geq 1$, l'inégalité d'interpolation

$$\|f\|_{L^p} \leq \|f\|_{L^1}^{1/p} \|f\|_{L^\infty}^{1-1/p}$$

(qui se démontre très simplement, sans même que l'on ait besoin de recourir à l'inégalité de Hölder). En faisant tendre p vers l'infini dans cette inégalité, on obtient

$$\limsup_{p \rightarrow \infty} \|f\|_{L^p} \leq \|f\|_{L^\infty},$$

ce qui conclut la preuve. □

VI-2.4. Interpolation entre espaces de Lebesgue. Le Théorème VI-28 montre comment, à partir d'informations dans des espaces de Lebesgue L^p et L^q , on peut parfois obtenir des informations dans des espaces de Lebesgue L^r pour tout r compris entre p et q . Nous allons maintenant voir des théorèmes plus généraux qui rendent ce point de vue systématique. Dans la suite, on note $L^p(X) + L^q(X)$ l'espace vectoriel de toutes les fonctions mesurables de la forme $f + g$, où $f \in L^p(X)$ et $g \in L^q(X)$. En outre, si T est un opérateur linéaire d'un espace vectoriel normé E dans un espace vectoriel normé F , on pose

$$\|T\|_{E \rightarrow F} := \sup_{\|x\|_E \neq 0} \frac{\|Tx\|_F}{\|x\|_E}.$$

Les deux théorèmes qui suivent sont les deux principaux théorèmes d'interpolation entre espaces de Lebesgue. Ils reposent sur des techniques très différentes, et ne sont pas comparables. Le premier a donné naissance à la théorie de l'**interpolation complexe**, et le second à la théorie de l'**interpolation réelle**, techniques d'une grande importance en analyse.

THÉORÈME VI-33 (théorème d'interpolation de Riesz–Thorin). *Soient X et Y deux espaces mesurés et p_0, p_1, q_0, q_1 des exposants compris entre 1 et ∞ au sens large. Soit T un opérateur linéaire continu de $L^{p_0}(X)$ dans $L^{q_0}(Y)$, et de $L^{p_1}(X)$ dans $L^{q_1}(Y)$. Alors, pour tout $\theta \in]0, 1[$, l'opérateur T admet un unique prolongement continu de $L^{p_\theta}(X)$ dans $L^{q_\theta}(Y)$, où*

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \quad \frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

En outre, si l'on pose $M_\theta = \|T\|_{L^{p_\theta} \rightarrow L^{q_\theta}}$, alors

$$M_\theta \leq M_0^{1-\theta} M_1^\theta.$$

Cas particulier important : Si $1 \leq p \leq q \leq \infty$, et T est un opérateur linéaire, borné de L^p dans L^p et de L^q dans L^q , alors T se prolonge uniquement en un opérateur borné de L^r dans L^r , pour tout $r \in [p, q]$.

Le Théorème de Riesz–Thorin peut se reformuler comme suit : l'ensemble des couples $(1/p, 1/q)$ tels que T soit continu de L^p dans L^q est un ensemble *convexe*, et $\log \|T\|_{L^p \rightarrow L^q}$ est une fonction convexe du couple $(1/p, 1/q)$. Ce théorème a pour avantage de donner des bornes très précises, qui sont optimales dans le cas général (ce qui n'exclut pas qu'on ne puisse les améliorer quand on considère un opérateur T particulier). Le Théorème qui suit ne donne pas de bornes aussi bonnes, mais permet d'inclure dans la discussion les espaces de Marcinkiewicz, dont nous avons vu qu'ils sont “légèrement” plus gros que les espaces de Lebesgue ; ce raffinement s'avère parfois précieux.

THÉORÈME VI-34 (Théorème d'interpolation de Marcinkiewicz). *Soient (X, μ) et (Y, ν) des espaces mesurés, et soient $p_0, q_0, p_1, q_1 \in [1, +\infty]$ avec $q_0 \neq q_1$, $p_0 \leq q_0$, $p_1 \leq q_1$. Si T est linéaire continu de $L^{p_0}(X)$ dans $L^{q_0, \infty}(Y)$ et de $L^{p_1}(X)$ dans $L^{q_1, \infty}(Y)$, alors pour tout $\theta \in]0, 1[$, l'opérateur T admet un unique prolongement continu de $L^{p_\theta}(X)$ dans $L^{q_\theta}(Y)$, où*

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \quad \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

En outre, si l'on note $M_0 = \|T\|_{L^{p_0} \rightarrow L^{q_0, \infty}}$, $M_1 = \|T\|_{L^{p_1} \rightarrow L^{q_1, \infty}}$, $M_\theta = \|T\|_{L^{p_\theta} \rightarrow L^{q_\theta}}$, alors il existe une constante C_θ , ne dépendant que de θ , p_0 , p_1 , q_0 , q_1 , telle que

$$M_\theta \leq C_\theta M_0^{1-\theta} M_1^\theta.$$

Cas particulier important : Soient (X, μ) et (Y, ν) des espaces mesurés σ -finis, et soit T un opérateur linéaire continu de $L^1(X)$ dans $L^{1, \infty}(Y)$ et de $L^\infty(X)$ dans $L^\infty(Y)$. Alors, pour tout $p \in]1, +\infty]$, il existe un unique prolongement de T en un opérateur continu de $L^p(X)$ dans $L^p(Y)$. Plus précisément, il existe une constante numérique C ($C = e^{1/e} \leq 2$ convient) telle que pour tout $p \in [1, +\infty]$,

$$\|T\|_{L^p \rightarrow L^p} \leq \frac{Cp}{p-1} \|T\|_{L^1 \rightarrow L^{1, \infty}}^{1/p} \|T\|_{L^\infty \rightarrow L^\infty}^{1-1/p}.$$

Parlons maintenant des démonstrations de ces théorèmes. C'est Riesz qui eut le premier l'idée, vers 1926, de la technique d'interpolation entre espaces de Lebesgue, et prouva le théorème maintenant appelé théorème de Riesz–Thorin. Vers la fin des années 1930, Thorin mit au point la preuve que nous allons esquisser ci-après, basée sur l'analyse complexe ; à peu près au même moment, Marcinkiewicz démontrait le théorème qui porte son nom par des méthodes très différentes. Un outil-clé dans le théorème de Riesz–Thorin est le lemme suivant, qui est bien sûr une variante du principe du maximum pour les fonctions holomorphes (voir [Rudin] par exemple) :

LEMME VI-35 (Lemme des trois lignes). Soit $S := \{x+iy; x \in [0, 1]; y \in \mathbb{R}\} \subset \mathbb{C}$ une bande du plan complexe, et soit $f : S \rightarrow \mathbb{C}$ une fonction continue bornée, holomorphe dans l'intérieur de S . Alors,

$$(i) \sup_S |f| = \sup_{\partial S} |f|;$$

$$(ii) \text{ soit } M_\theta := \sup_{y \in \mathbb{R}} |f(\theta + iy)|; \text{ alors}$$

$$M_\theta \leq M_1^\theta M_0^{1-\theta}.$$

DÉMONSTRATION. 1. Supposons d'abord que f a pour limite 0 à l'infini, et soit $\varepsilon < \|f\|_\infty$; puisque f tend vers 0 à l'infini, il existe $M \in \mathbb{R}$ tel que $|f|$ (vu comme une fonction sur \mathbb{R}^2) atteint son maximum sur $[0, 1] \times [-M, M]$. On conclut la preuve de (i) en appliquant le principe du maximum pour les fonctions holomorphes définies sur des ouverts bornés.

2. Dans le cas général où f ne converge pas forcément vers 0, on s'y ramène en considérant z_0 tel que $|f(z_0)| \geq (1 - \delta)\|f\|_\infty$ et en posant $g(z) = e^{-\lambda(z-z_0)^2} f(z)$, $\lambda > 0$. En appliquant le résultat précédent, on voit que $|g(z)|$ atteint son maximum sur le bord ; or ce maximum est au moins $|g(z_0)| \geq (1 - \delta)\|f\|_\infty$. En particulier,

$$\sup_{\partial S} |f| \geq (1 - \delta)\|f\|_\infty,$$

et on conclut (i) en faisant tendre δ vers 0.

3. L'énoncé (ii) est obtenu à partir de (i) en posant $h(z) = e^{-\lambda z} f(z)$, $\lambda \in \mathbb{R}$. Alors

$$M_\theta \leq e^{\lambda\theta} \sup_S |h| \leq e^{\lambda\theta} \sup_{\partial S} |h| \leq e^{\lambda\theta} \max(M_0, e^{-\lambda} M_1).$$

On choisit λ de sorte que

$$M_0 = e^{-\lambda} M_1,$$

i.e. $e^\lambda = M_1/M_0$. L'estimation ci-dessus devient alors

$$M_\theta \leq M_1^\theta M_0^{1-\theta}.$$

□

DÉMONSTRATION DU THÉORÈME DE RIESZ–THORIN. On note $p = p_\theta$, $q = q_\theta$; et $M_j = \|T\|_{L^{p_j} \rightarrow L^{q_j}}$. On va utiliser le Théorème VI-24, sous la forme

$$\|f\|_{L^q(\mu)} = \sup_{\|g\|_{L^{q'}(\mu)} \neq 0} \frac{\int f g d\mu}{\|g\|_{L^{q'}(\mu)}},$$

où le supremum est pris sur toutes les fonctions g qui sont combinaisons linéaires de fonctions indicatrices d'ensembles de mesure finie; nous appellerons “fonctions simples” de telles fonctions.

Montrer que T est borné $L^p \rightarrow L^q$ avec norme au plus $M_1^\theta M_0^{1-\theta}$ revient à prouver que

$$(41) \quad \|Tf\|_{L^q} \leq M_1^\theta M_0^{1-\theta} \|f\|_{L^p}$$

pour toute fonction $f \in L^p$, ou, de manière équivalente, pour toute fonction f simple. Encore une fois, par densité et en traitant à part le cas $p = \infty$, on voit qu'il suffit d'établir (41) dans le cas où f est une fonction simple. Notre but est donc

$$(42) \quad \left| \int (Tf)g \right| \leq M_1^\theta M_0^{1-\theta} \|f\|_{L^p} \|g\|_{L^{q'}}.$$

Nous allons maintenant introduire un paramètre d'interpolation $z \in S$, et faire varier toutes les quantités ci-dessus en fonction de z . Etant données deux fonctions simples f et g , on pose donc

$$f_z(x) = |f(x)|^{p\left(\frac{1-z}{p_0} + \frac{z}{p_1}\right)} \frac{f(x)}{|f(x)|},$$

$$g_z(y) = |g(y)|^{q'\left(\frac{1-z}{q_0} + \frac{z}{q_1}\right)} \frac{g(y)}{|g(y)|},$$

avec la convention $0/0 = 0$. Ces fonctions f_z sont simples, en particulier dans tous les espaces L^r , et il s'ensuit que $Tf_z \in L^{q_0} \cap L^{q_1}$ pour tout z ; la fonction

$$\varphi : z \mapsto \int (Tf_z)g_z$$

est donc bien définie. En décomposant f_z et g_z en combinaison linéaire de fonctions indicatrices, on voit qu'en fait on peut écrire φ sous la forme

$$\varphi(z) = \sum_{1 \leq k \leq K} a_k^{\lambda_k z + \mu_k}, \quad \lambda \in \mathbb{R}, \mu \in \mathbb{R};$$

en particulier φ est holomorphe et bornée dans S , et on peut appliquer le lemme des trois lignes :

$$|\varphi(\theta)| \leq \left(\sup_{t \in \mathbb{R}} |\varphi(it)|^{1-\theta} \right) \left(\sup_{t \in \mathbb{R}} |\varphi(1+it)|^{1-\theta} \right).$$

Mais $\varphi(\theta)$ n'est autre que $\int Tfg$. Par ailleurs,

$$\left| \int T f_{it} g_{it} \right| \leq \|T f_{it}\|_{L^{q_0}} \|g_{it}\|_{L^{q'_0}} \leq \|T\|_{L^{p_0} \rightarrow L^{q_0}} \|f\|_{L^p}^{p/p_0} \|g\|_{L^{q'}}^{q'/q'_0},$$

et l'on peut faire une majoration similaire pour les $z = 1 + it$. La conclusion en découle facilement. \square

DÉMONSTRATION DU THÉORÈME DE MARCINKIEWICZ. Nous nous contenterons de démontrer le “cas particulier”, qui est utile dans de nombreuses situations. Le lecteur pourra essayer de reconstituer la démonstration générale en adaptant la technique utilisée ci-dessous ; ou consulter [Zygmund, tome II, chapitre XII, théorème 4.6].

Cette fois nous allons démontrer le théorème directement, sans passer par des fonctions simples. La preuve fait intervenir deux idées principales :

- représenter les normes des fonctions en jeu au moyen de la taille de leurs “ensembles de sur-niveau”, i.e. le lieu des points où ces fonctions sont plus grandes qu'un certain paramètre t ,

- décomposer la fonction en jeu en la somme de deux fonctions appartenant aux espaces que l'on interpole, où les deux fonctions sont choisies indépendamment *pour chaque valeur du paramètre*.

Ecrivons donc

$$\|Tf\|_{L^\infty} \leq M_1\|f\|_{L^\infty}, \quad \|Tf\|_{L^{1,\infty}} \leq M_0\|f\|_{L^1}.$$

La deuxième inégalité se réécrit

$$\forall t > 0, \quad t\mu[\{|Tf| > t\}] \leq M_0\|f\|_{L^1}.$$

Sans perte de généralité on supposera que $M_0^{1-\theta} M_1^\theta = 1$; on peut toujours se ramener à ce cas en multipliant T par une constante convenable.

On se souvient de la formule 17 :

$$\int |f|^p = p \int_0^{+\infty} \mu[\{|f| > t\}] t^{p-1} dt.$$

De même,

$$\int |Tf|^p = p \int_0^{+\infty} \mu[\{|Tf| > t\}] t^{p-1} dt.$$

Pour tout $t \geq 0$ on écrit alors

$$f = f_1^{(t)} + f_2^{(t)}, \quad f_1^{(t)} = f1_{|f| \leq At}, \quad f_2^{(t)} = f2_{|f| > At}.$$

La borne $L^\infty \rightarrow L^\infty$ entraîne que pour tout $t \geq 0$,

$$|Tf_1^{(t)}| \leq M_1 At.$$

En particulier,

$$\mu[\{|Tf| > t\}] \leq \mu[\{|Tf_2^{(t)}| > (1 - M_1 A)t\}].$$

En reportant cette inégalité dans la représentation de $\int |Tf|^p$, on trouve

$$\begin{aligned} \int |Tf|^p &\leq p \int_0^{+\infty} \mu[\{|Tf_2^{(t)}| > (1 - M_1 A)t\}] t^{p-1} dt \\ &= (1 - M_1 A)^{-1} p \int_0^{+\infty} \left((1 - M_1 A) t \mu[\{|Tf_2^{(t)}| > (1 - M_1 A)t\}] \right) t^{p-2} dt \\ &\leq (1 - M_1 A)^{-1} p M_0 \int_0^{+\infty} \|f_2^{(t)}\|_{L^1} t^{p-2} dt = (1 - M_1 A)^{-1} p M_0 \int_0^{+\infty} \int |f| 1_{|f| > At} t^{p-2} dt. \end{aligned}$$

On applique alors Fubini et un changement de variable évident pour réécrire le dernier terme sous la forme

$$\begin{aligned} & (1 - M_1 A)^{-1} p M_0 \int |f| \left(\int_0^{|f|/A} t^{p-2} dt \right) \\ &= \frac{p M_0}{(p-1)(1 - M_1 A) A^{p-1}} \int |f|^p. \end{aligned}$$

On pose $M_1 A = \lambda$, la constante apparaissant en facteur de $\int |f|^p$ est minimale pour $\lambda = 1/p'$, et vaut $c_p M_0 M_1^{p-1}$, avec

$$c_p = \frac{p^{p+1}}{(p-1)^p} = \left(p^{1/p} \frac{p}{p-1} \right)^p,$$

que l'on majore en utilisant $p^{1/p} \leq e^{1/e}$. La preuve est complète. \square

Pour conclure cette section, mentionnons une variante intéressante du théorème de Riesz–Thorin, où l'on s'autorise une dépendance de l'opérateur, est la suivante. Convenons qu'une famille (T_z) définit une famille holomorphe d'opérateurs si la fonction $z \mapsto T_z f$ est holomorphe pour tout f simple. On peut alors changer, dans l'énoncé du Théorème de Riesz–Thorin, l'opérateur T en une famille holomorphe d'opérateurs T_z ; l'hypothèse de bornes $L^{p_0} \rightarrow L^{q_0}$ et $L^{p_1} \rightarrow L^{q_1}$ sur T est alors remplacée par une hypothèse similaire sur T_0 et T_1 respectivement.

THÉORÈME VI-36 (théorème d'interpolation de Stein). *Soient X et Y deux espaces mesurés σ -finis, et p_0, p_1, q_0, q_1 des exposants compris entre 1 et ∞ au sens large. Soit $(T_z)_{z \in D}$ une famille holomorphe d'opérateurs linéaires définis sur une partie D du plan complexe incluant la bande S des nombres complexes dont la partie réelle est comprise entre 0 et 1. On suppose que T_0 est borné de $L^{p_0}(X) + L^{q_0}(Y)$ dans $L^{p_1}(X) + L^{q_1}(Y)$, tel que T_1 est borné $L^{p_0}(X) \rightarrow L^{p_1}(Y)$, et $L^{q_0}(X) \rightarrow L^{q_1}(Y)$. Alors,*

$$T_\theta \text{ est borné } L^p(X) \rightarrow L^q(Y),$$

En outre, si on pose $M_\theta = \|T_z\|_{L^{p_\theta} \rightarrow L^{q_\theta}}$, alors

$$M_\theta \leq M_0^{1-\theta} M_1^\theta.$$

La démonstration est similaire à celle du théorème de Riesz–Thorin.

EXEMPLE VI-37. Soit μ une mesure et w une fonction positive; la famille d'opérateurs

$$T_z : f \mapsto w^z f$$

satisfait aux hypothèses du théorème. Le théorème d'interpolation de Stein devient alors un théorème d'interpolation **entre espaces de Lebesgue à poids**. Par exemple, si v est une fonction positive et si l'on définit

$$\|f\|_{L_\kappa^p} = \|f v^\kappa\|_{L^p},$$

alors on a, pour tout opérateur linéaire S ,

$$\|S\|_{L_{\kappa_\theta}^{p_\theta} \rightarrow L_{\lambda_\theta}^{q_\theta}} \leq \|S\|_{L_{\kappa_0}^{p_0} \rightarrow L_{\lambda_0}^{q_0}}^{1-\theta} \|S\|_{L_{\kappa_1}^{p_1} \rightarrow L_{\lambda_1}^{q_1}}^\theta.$$

VI-3. Espace des fonctions mesurables

Nous allons maintenant introduire une notion naturelle de convergence des fonctions mesurables, ne présupposant aucune intégrabilité, et étudier ses liens avec la convergence L^p .

VI-3.1. Convergence dans L . Une première idée qui vient à l'esprit consiste à utiliser la convergence presque partout, comme naturellement associée au cadre de la théorie de la mesure. Cependant, cette notion présente de graves défauts : en particulier, la convergence L^p ($1 \leq p < \infty$) n'implique pas la convergence presque partout. En outre, la convergence presque partout n'est pas associée à une métrique : en effet, on démontre facilement que dans un espace métrique, si une suite (f_n) a la propriété que toute sous-suite extraite admet une sous-sous-suite convergeant vers un certain f , alors la suite f_n entière tend vers f . Or nous avons vu (Exemple III-20(i)) que cet énoncé n'est pas vrai pour la convergence presque partout.

La notion naturelle de convergence est en fait celle que nous venons d'invoquer implicitement.

DÉFINITION VI-38 (convergence au sens des fonctions mesurables). *Soient X et Y deux espace mesurés. On dit une famille $(f_n)_{n \geq 1}$ de fonctions mesurables de X dans Y converge vers f si de toute sous-suite extraite $(f_{n'})$ de (f_n) on peut extraire une sous-sous-suite extraite $(f_{n''})$ qui converge presque partout vers f .*

Pour abréger, on pourra dire que f_n converge “presque partout à extraction près”. Cette notion a en commun avec la notion de convergence presque partout la propriété de stabilité par composition : si $f_n : X \rightarrow Y$ converge vers f et Φ est n'importe quelle fonction mesurable de Y dans un autre espace mesurable Z , alors $\Phi \circ f_n$ converge vers $\Phi \circ f$.

Contrairement à la convergence presque partout, la convergence presque partout à extraction près est en général associée à une métrique. Pour se souvenir que cette notion de convergence est plus faible que toutes les convergences L^p , nous l'appellerons “convergence dans L ”.

PROPOSITION VI-39 (convergence dans L et convergence en mesure). *Soient (X, μ) un espace mesuré σ -fini, (Y, d) un espace métrique, et soit φ une fonction strictement positive partout sur X , d'intégrale convergente. Alors la formule*

$$\Delta(f, g) := \int_X \frac{d(f(x), g(x))}{1 + d(f(x), g(x))} \varphi(x) d\mu(x)$$

définit une distance sur l'espace $L(X, \mu; Y)$ des fonctions mesurables de X dans Y , quotienté par la relation d'égalité μ -presque partout. On note cet espace $L(X, \mu)$ dans le cas où Y est \mathbb{R} muni de la distance euclidienne. Les trois assertions suivantes sont équivalentes :

- (i) $\Delta(f_n, f) \longrightarrow 0$;
- (ii) de toute suite extraite $(f_{n'})$ on peut extraire une suite extraite $(f_{n''})$ qui converge presque partout vers f ;
- (iii) f_n converge vers f en mesure sur les parties finies, i.e. pour toute partie A de mesure finie on a

$$\forall \varepsilon > 0, \quad \mu[\{x \in A; d(f_n(x), f(x)) \geq \varepsilon\}] \xrightarrow{n \rightarrow \infty} 0.$$

Si Y est complet, l'espace L ainsi défini est un espace métrique complet. Si $Y = \mathbb{R}$, alors $L^1 \cap L^\infty$ est dense dans L .

REMARQUES VI-40. (i) L'existence d'une fonction φ intégrable et strictement positive est garantie par l'hypothèse de σ -additivité : soit $(A_k)_{k \geq 1}$ une famille d'ensembles mesurables disjoints, de mesure finie, dont la réunion est X , on peut poser

$$\varphi_k = \sum_{k \geq 1} \frac{1_{A_k}}{k^2 \mu[A_k]}.$$

(ii) En théorie des probabilités ($\mu[X] = 1$), la convergence en mesure est appelée convergence en probabilité.

DÉMONSTRATION. Nous donnerons la preuve uniquement dans le cas où $Y = \mathbb{R}$. Supposons que l'assertion (i) du théorème est vérifiée, et soit $(f_{n'})$ une suite extraite de (f_n) . La fonction positive intégrable $\varphi(x)|f_{n'}(x) - f(x)|/(1 + |f_{n'}(x) - f(x)|)$ converge vers 0 dans $L^1(X)$, on peut donc extraire une sous suite n'' pour laquelle cette expression converge vers 0 presque partout. Comme φ est strictement positive partout, on en déduit que $f_{n''}$ converge presque partout vers f . L'assertion (ii) est donc vraie.

Pour montrer que (ii) implique (i), on extrait une sous-suite n' quelconque, et de cette sous-suite on extrait une sous-sous-suite pour laquelle la convergence a lieu presque partout, et on applique le théorème de convergence dominée à la famille $\varphi|f_{n'} - f|/(1 + |f_{n'} - f|)$, dominée par φ . On montre ainsi que $\Delta(f_{n''}, f) \rightarrow 0$. Comme la sous-suite extraite $f_{n'}$ était arbitraire, et que Δ définit une métrique, on en déduit que $\Delta(f_n, f) \rightarrow 0$.

Supposons de nouveau que l'assertion (i) du théorème soit vérifiée, et soit $B_{n,\varepsilon}$ l'ensemble des $x \in X$ tels que $|f_n(x) - f(x)| \geq \varepsilon$: alors

$$\Delta(f_n, f) \geq \frac{\varepsilon}{1 + \varepsilon} \int_{B_{n,\varepsilon}} \varphi d\mu,$$

et donc

$$\int_{B_{n,\varepsilon}} \varphi d\mu \xrightarrow{n \rightarrow \infty} 0.$$

Soit maintenant A une partie de mesure finie. Comme X est la réunion dénombrable croissante des $\{\varphi \geq 1/k\}$, on peut trouver $K = K(\eta)$ tel que

$$\mu[\{\varphi \geq 1/K\} \cap A] \geq \mu[A] - \eta,$$

où η est arbitrairement petit. On a alors

$$\mu[B_{n,\varepsilon} \cap A] \leq \mu[B_{n,\varepsilon} \cap A \cap \{\varphi \geq 1/K\}] + \eta \leq K(\eta) \int_{B_{n,\varepsilon}} \varphi d\mu + \eta.$$

A η et ε fixés, le premier terme du membre de droite tend vers 0 quand $n \rightarrow \infty$; comme η est arbitrairement petit, on conclut que

$$\mu[B_{n,\varepsilon} \cap A] \xrightarrow{n \rightarrow \infty} 0,$$

ce qui veut dire qu'il y a bien convergence en mesure sur toutes les parties de mesure finie.

Finalement, supposons l'assertion (iii) du théorème vérifiée, et prouvons l'assertion (i). Pour tout $\varepsilon > 0$ on peut écrire

$$\Delta(f_n, f) \leq \frac{\varepsilon}{1 + \varepsilon} \int_{B_{n,\varepsilon}^c} \varphi \, d\mu + \mu[B_{n,\varepsilon}] \leq \varepsilon \int \varphi \, d\mu + \mu[B_{n,\varepsilon}].$$

Le premier terme est arbitrairement petit quand $\varepsilon \rightarrow 0$, et le deuxième tend vers 0 quand $n \rightarrow \infty$, ε étant fixé. On en déduit que $\Delta(f_n, f) \rightarrow 0$.

Montrons maintenant la complétude de l'espace (L, Δ) . Soit $(f_n)_{n \in \mathbb{N}}$ une suite de Cauchy pour Δ ; Pour tout $k \in \mathbb{N}$, on pose

$$A_k := \{x; (k+1)^{-2} \leq \varphi(x) < k^{-2}\}.$$

La famille $(1_{A_k} f_n)_{n \in \mathbb{N}}$ est alors une suite de Cauchy. On pose $\delta(f, g) := |f - g|/(1 + |f - g|)$. Quitte à extraire une sous-suite, on peut supposer que

$$\int_{A_k} \sum_{\ell=1}^{\infty} \delta(f_n(x), f_{n-1}(x)) \, d\mu(x) < +\infty.$$

Pour presque tout $x \in A_k$ on a donc convergence de la série $\sum \delta(f_n(x), f_{n-1}(x))$, et la suite $(f_n(x))$ converge donc vers un nombre noté $f(x)$ (on utilise ici la complétude de (\mathbb{R}, δ)). Par convergence dominée, on montre alors que

$$\int_{A_k} \delta(f_m(x), f_n(x)) \, d\mu(x) \xrightarrow{n \rightarrow \infty} \int_{A_k} \delta(f_m(x), f(x)) \, d\mu(x).$$

Comme la suite (f_n) est de Cauchy, le membre de gauche est arbitrairement petit quand m est grand et $n \geq m$. On conclut finalement que

$$\int_{A_k} \delta(f_m(x), f(x)) \, d\mu(x) \xrightarrow{m \rightarrow \infty} 0,$$

ce qui est bien sûr équivalent à

$$\int_{A_k} \varphi \delta(f_m(x), f(x)) \, d\mu(x) \xrightarrow{m \rightarrow \infty} 0.$$

On a donc, pour tout k_0 ,

$$\sum_{k \leq k_0} \int \varphi \delta(f_m(x), f(x)) \, d\mu(x) \xrightarrow{m \rightarrow \infty} 0;$$

et d'autre part, puisque $\varphi \in L^1(d\mu)$,

$$\sum_{k > k_0} \int \varphi \delta(f_m(x), f(x)) \, d\mu(x) \leq \sum_{k > k_0} \int \varphi \xrightarrow{k_0 \rightarrow \infty} 0.$$

On conclut que

$$\int \varphi \delta(f_m(x), f(x)) \, d\mu(x) \xrightarrow{m \rightarrow \infty} 0.$$

Enfin, dans le cas $Y = \mathbb{R}$, montrons que $L^1 \cap L^\infty$ est dense dans L . Soit f une fonction mesurable à valeurs réelles, on pose

$$A_k := \{x \in X; |f(x)| \leq k \text{ et } \varphi(x) \geq k^{-1}\}.$$

Puisque f est à valeurs réelles et φ strictement positive, les A_k forment une famille croissante dont l'union est égale à X tout entier, donc $\nu[X \setminus A_k] \rightarrow 0$. Soit $\varepsilon > 0$

arbitrairement petit, on choisit k tel que $\nu[X \setminus A_k] < \varepsilon$. Alors, pour toute fonction $g \in L^1(d\mu)$,

$$\int_{X \setminus A_k} \delta(f, g) \varphi d\mu \leq \nu[A_k] \leq \varepsilon.$$

En particulier, $\Delta(f, f1_{A_k}) \leq \varepsilon$. La fonction $f1_{A_k}$ est bornée par construction, et elle est également intégrable puisque A_k est de mesure finie (à cause de l'intégrabilité de φ). Ceci conclut l'argument. \square

VI-3.2. Lien avec les autres notions de convergence. La convergence dans L est une notion plus faible que la convergence au sens L^p , mais elle lui est intimement liée, comme le montre le théorème suivant, dans lequel nous nous limiterons aux fonctions à valeurs réelles.

THÉORÈME VI-41 (convergence dans L et dans L^p). *Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables à valeurs réelles sur un espace mesuré (X, μ) ; soit également f une fonction mesurable à valeurs réelles. Alors*

- (i) *si $f_n \rightarrow f$ dans $L^p(X, \mu)$ ($0 \leq p \leq \infty$), alors $f_n \rightarrow f$ dans $L(X, \mu)$.*
- (ii) *si $f_n \rightarrow f$ dans $L(X, \mu)$ et il existe $g \in L^p(\mu)$ ($0 < p < \infty$) tel que $|f_n| \leq g$ pour tout n , alors $f_n \rightarrow f$ dans $L^p(X, \mu)$.*

DÉMONSTRATION. L'assertion (i) est facile : pour toute sous-suite extraite n' , on a $f_{n'} \rightarrow f$ dans L^p , et on peut donc trouver une sous-sous-suite pour laquelle il y ait convergence presque partout.

Pour prouver l'assertion (ii), il suffit de montrer que pour toute sous-suite arbitraire n' , on a convergence d'une sous-sous-suite $f_{n''}$ vers f dans L^p . On peut supposer que $f_{n''}$ converge presque partout vers f . La conclusion découle alors du théorème de convergence dominée, appliqué à la suite $|f_{n''} - f|^p$, que l'on peut majorer par la fonction intégrable $\max(2, 2^p)|g|^p$. \square

La condition de domination peut être remplacée par une condition plus faible qui suppose seulement certaines bornes en moyenne. On va utiliser ici la notion d'équi-intégrabilité, étudiée dans la section III-5.

THÉORÈME VI-42. *Soient (X, μ) un espace mesuré σ -fini, $p \in]0, +\infty[$, et $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions dans $L^p(X, \mu)$, convergeant dans $L(X, \mu)$ vers une fonction mesurable f . On suppose que $(|f_n|^p)$ est équi-intégrable et équi-intégrable à l'infini. Alors f_n converge vers f dans $L^p(X, \mu)$.*

REMARQUE VI-43. Dans le chapitre suivant, nous retrouverons le cas particulier $p = 1$ de ce théorème comme une conséquence du théorème de Schur.

DÉMONSTRATION DU THÉORÈME VI-42. Soit $\varepsilon > 0$; on sait par hypothèse qu'il existe $M_1 > 0$ et un ensemble A_1 de mesure finie, tels que pour tout n ,

$$\int_{|f_n| > M_1} |f_n|^p d\mu + \int_{X \setminus A_1} |f_n|^p d\mu \leq \varepsilon.$$

Notons que cela impose bien sûr

$$\int |f_n|^p d\mu \leq M_1 \mu[A_1] + \varepsilon;$$

et par le lemme de Fatou on en déduit

$$\int |f|^p d\mu \leq \liminf_{n \rightarrow \infty} \int |f_n|^p < +\infty.$$

L'espace X étant σ -fini, on peut donc trouver $M_2 > 0$ et un ensemble A_2 de mesure finie, tels que

$$\int_{|f| > M_2} |f|^p d\mu + \int_{X \setminus A_2} |f|^p d\mu \leq \varepsilon.$$

On pose $A := A_1 \cup A_2$, $M := \max(M_1, M_2)$.

On a alors

$$\int_{X \setminus A} |f_n - f|^p d\mu \leq \max(2, 2^p) \left(\int_{X \setminus A_1} |f_n|^p d\mu + \int_{X \setminus A_2} |f|^p d\mu \right) \leq \max(2, 2^p) \varepsilon.$$

La même majoration est valable sur l'ensemble des x pour lesquels $|f_n(x)| \geq M$ ou $|f(x)| \geq M$; donc en particulier pour l'ensembles des x tels que $|f_n(x) - f(x)| \geq 2M$. On conclut que

$$\int_X |f_n - f|^p d\mu \leq \int_{A \cap \{|f_n - f| \leq 2M\}} |f_n - f|^p d\mu + 2 \max(2, 2^p) \varepsilon.$$

On distingue alors deux cas.

Si $p \geq 1$, on écrit

$$\int_{A \cap \{|f_n - f| \leq 2M\}} |f_n - f|^p d\mu \leq (2M)^{p-1} (1 + 2M) \int_{A \cap \{|f_n - f| \leq 2M\}} \frac{|f_n - f|}{1 + |f_n - f|} d\mu,$$

et on conclut que

$$\int_X |f_n - f|^p d\mu \leq (2M)^{p-1} (1 + 2M) \Delta(f_n, f) + 2 \max(2, 2^p) \varepsilon;$$

comme par hypothèse $\Delta(f_n, f) \rightarrow 0$, on a bien la convergence de f_n vers f dans L^p .

Si en revanche $0 < p < 1$, on écrit

$$\begin{aligned} \int_{A \cap \{|f_n - f| \leq 2M\}} |f_n - f|^p d\mu &\leq \left(\int_{A \cap \{|f_n - f| \leq 2M\}} |f_n - f| d\mu \right)^p \mu[A \cap \{|f_n - f| \leq 2M\}]^{1-p} \\ &\leq (1 + 2M)^{1/p} \left(\int_{A \cap \{|f_n - f| \leq 2M\}} \frac{|f_n - f|}{1 + |f_n - f|} d\mu \right)^p \mu[A]^{1-p}. \end{aligned}$$

On conclut que

$$\int_X |f_n - f|^p d\mu \leq (2M)^{p-1} (1 + 2M)^{1/p} \mu[A]^{1-p} \Delta(f_n, f) + 2 \max(2, 2^p) \varepsilon;$$

ce qui entraîne encore la convergence de f_n vers f dans L^p . \square

VI-4. Espaces de mesures

Dans cette dernière section, nous allons étudier les **mesures signées**, qui constituent une généralisation des fonctions mesurables, et les **mesures signées finies**, qui constituent une généralisation des fonctions sommables.

VI-4.1. Mesures signées. Si (X, \mathcal{A}, μ) est un espace mesuré, et f une fonction positive mesurable sur X , on peut définir sur X une nouvelle mesure, notée $f\mu$, par la formule

$$(f\mu)[A] = \int_A f d\mu.$$

Le théorème de convergence monotone montre que $f\mu$ est bien σ -additive. La mesure $f\mu$ détermine f uniquement, à un ensemble μ -négligeable près (en effet, si $f > g$ sur un ensemble non négligeable A , on a $(f\mu)[A] > (g\mu)[A]$). On voit donc que, dès que l'on a fixé une mesure de référence μ , l'ensemble des fonctions mesurables positives (modulo l'égalité μ -presque partout) s'identifie à une partie de l'ensemble des mesures; en ce sens, les mesures constituent une généralisation des fonctions mesurables positives.

Une fonction mesurable quelconque peut toujours s'écrire comme différence de deux fonctions positives : $f = f_+ - f_-$; en outre, f_+ et f_- sont "étrangères", au sens où elles ne sont jamais simultanément non nulles. Il est facile d'étendre cette notion à des mesures :

DÉFINITION VI-44 (mesures étrangères). Soit X un espace mesurable; on dit que deux mesures μ et ν sur X sont étrangères si elles sont concentrées sur des ensembles mesurables disjoints; en d'autres termes, s'il existe deux ensembles mesurables A et B tels que $A \cap B = \emptyset$, $\mu[X \setminus A] = 0$, $\nu[X \setminus B] = 0$.

Nous pouvons maintenant définir la notion de mesure signée, comme une généralisation du concept de fonction mesurable :

DÉFINITION VI-45 (mesure signée). Soit X un espace mesurable. On appelle mesure signée sur X un couple $\mu = (\mu_+, \mu_-)$ de mesures étrangères sur X , appelées respectivement partie positive et partie négative de μ . On notera formellement $\mu = \mu_+ - \mu_-$. On note alors $|\mu| = \mu_+ + \mu_-$.

On dira que μ est finie (ou bornée) si μ_+ et μ_- sont finies. On dira que μ est σ -finie si μ_+ et μ_- le sont. On dira que μ est de Borel si μ_+ et μ_- le sont. On dira que μ est régulière si μ_+ et μ_- le sont.

REMARQUE VI-46. Si A est mesurable et $(\mu_+[A], \mu_-[A]) \neq (+\infty, +\infty)$, on peut définir sans ambiguïté la quantité

$$\mu[A] := \mu_+[A] - \mu_-[A] \in \overline{\mathbb{R}};$$

mais si $\mu_+[A] = \mu_-[A] = +\infty$, la valeur de $\mu[A]$ n'est pas définie a priori. C'est pourquoi l'écriture $\mu_+ - \mu_-$ doit être considérée comme formelle.

EXEMPLES VI-47. Sur \mathbb{R} , δ_0 est une mesure qui n'est pas une fonction; $\delta_0 - \delta_1$ est une mesure signée; $\delta_0 - \delta_0$ ne constitue pas une mesure signée au sens de la définition précédente (les deux mesures ne sont pas étrangères); $\mu := \sum_{k \geq 0} \delta_{2k} - \sum_{k \geq 0} \delta_{2k+1}$ est une mesure signée, mais on ne peut attribuer aucune valeur à $\mu[\mathbb{R}]$.

VI-4.2. Décomposition de Hahn. Soit $\mu = (\mu_+, \mu_-)$ une mesure signée sur un ensemble mesurable (X, \mathcal{A}) . Comme nous l'avons remarqué, il est en général impossible de définir μ comme une fonction $\mathcal{A} \rightarrow [-\infty, +\infty]$, sauf si μ_+ ou μ_- est finie. À partir de maintenant nous allons concentrer notre attention sur les mesures signées finies. Le remarquable théorème de décomposition de Hahn montre que de telles mesures sont caractérisées par la propriété de σ -additivité.

DÉFINITION VI-48 (σ -additivité à valeurs réelles). Soit \mathcal{A} une σ -algèbre, et $\mu : \mathcal{A} \rightarrow \mathbb{R}$ une fonction. On dit que μ est σ -additive si, pour toute suite $(A_k)_{k \in \mathbb{N}}$ d'éléments deux à deux disjoints de \mathcal{A} , on a

$$(43) \quad \mu[\cup A_k] = \sum_{k \in \mathbb{N}} \mu[A_k],$$

où le membre de droite est défini comme la limite des sommes partielles $\sum_{1 \leq k \leq \ell} \mu[A_k]$.

REMARQUE VI-49. Le membre de gauche de (43) est invariant par permutation des A_k , donc le membre de droite aussi, ce qui veut dire que la série $\sum \mu[A_k]$ est commutativement convergente. Par un résultat classique d'analyse réelle, cette série est forcément absolument convergente : $\sum |\mu[A_k]| < +\infty$.

THÉORÈME VI-50 (théorème de décomposition de Hahn). Soit (X, \mathcal{A}) un espace mesurable ; alors on peut identifier

- d'une part, les fonctions $\mu : \mathcal{A} \rightarrow \mathbb{R}$, σ -additives ;
- d'autre part, les mesures signées finies (μ_+, μ_-) sur \mathcal{A} ;

via la formule $\mu[A] = \mu_+[A] - \mu_-[A]$.

En outre, on a alors, pour tout $A \in \mathcal{A}$,

$$(44) \quad |\mu|[A] = \mu_+[A] + \mu_-[A] = \sup \left\{ \sum_{j \in \mathbb{N}} |\mu[A_j]| ; \quad A_j \in \mathcal{A}, (A_j)_{j \in \mathbb{N}} \text{ partition de } A \right\}.$$

REMARQUE VI-51. Tout le travail dans ce théorème consiste à décomposer μ en sa partie positive et sa partie négative, d'où l'appellation "théorème de décomposition". Il faut bien noter que le résultat contient l'unicité de cette décomposition.

REMARQUE VI-52. La conclusion du Théorème VI-50 est bien sûr en défaut pour des fonctions σ -additives $\mathcal{A} \rightarrow \overline{\mathbb{R}}$ ou même $\mathcal{A} \rightarrow [0, +\infty]$ (une mesure σ -additive n'est pas pour autant finie!).

DÉMONSTRATION. 1. Il est clair qu'une mesure signée finie définit une fonction σ -additive d'ensembles ; c'est bien sûr la réciproque qui présente un intérêt.

2. Montrons maintenant l'unicité de la décomposition éventuelle. Soient $\mu_+, \mu_-, \nu_+, \nu_-$ des mesures finies vérifiant, au sens des fonctions σ -additives,

$$\mu_+ - \mu_- = \nu_+ - \nu_-,$$

et telles que (μ_+, μ_-) d'une part, (ν_+, ν_-) d'autre part, forment des couples étrangers. Introduisons $S(\mu_+)$ et $S(\mu_-)$ des ensembles mesurables disjoints tels que $\mu_+[X \setminus S(\mu_+)] = 0$, $\mu_-[X \setminus S(\mu_-)] = 0$, $S(\mu_+) \cap S(\mu_-) = \emptyset$; et de même, des ensembles $S(\nu_+)$ et $S(\nu_-)$ avec des propriétés similaires vis-à-vis de ν_{\pm} . L'ensemble $A := S(\mu_+) \cap S(\nu_-)$ vérifie

$$\mu[A] = \mu_+[S(\nu_-)] = -\nu_-[S(\mu_+)];$$

la quantité $\mu[A]$ est donc à la fois positive et négative, et donc nulle. On en déduit que $\mu_+[S(\nu_-)] = 0 = \nu_-[S(\mu_+)]$; et de même, $\mu_-[S(\nu_+)] = \nu_+[S(\mu_-)] = 0$. Pour tout $A \subset S(\nu_+)$ on a donc $\mu[A] = \nu_+[A]$, mais aussi $\mu[A] = \mu_+[A] - \mu_-[A] = \mu_+[A]$; on conclut que μ_+ et ν_+ coïncident sur $S(\nu_+)$, et donc en fait $\mu_+ = \nu_+$. De même, $\mu_- = \nu_-$.

3. Définissons provisoirement $|\mu|$ par la formule de droite dans (44) : $|\mu|[A] = \sup \sum |\mu[A_i]|$, où le supremum est pris sur toutes les partitions, finies ou dénombrable,

de A en parties mesurables A_i . Vérifions que $|\mu|$ ainsi définie est une mesure. Si $(A^k)_{k \in \mathbb{N}}$ est une famille de parties mesurables disjointes, et si on se donne des partitions $(A_j^k)_{j \in \mathbb{N}}$ de chaque A^k , on définit automatiquement une partition $(A_j^k)_{k, j \in \mathbb{N}}$ de $A = \cup A^k$. Donc

$$|\mu|[A] \geq \sum_{k \in \mathbb{N}} |\mu|[A^k].$$

Soit maintenant $(A_j)_{j \in \mathbb{N}}$ une partition de A , et $A_j^k = A^k \cap A_j$, de sorte que $(A_j^k)_{j \in \mathbb{N}}$ constitue une partition de A^k . Par σ -additivité de μ ,

$$\sum_{j \in \mathbb{N}} |\mu[A_j]| = \sum_{j \in \mathbb{N}} \left| \sum_{k \in \mathbb{N}} \mu[A_j^k] \right| \leq \sum_{j \in \mathbb{N}} \sum_{k \in \mathbb{N}} |\mu[A_j^k]| = \sum_{k \in \mathbb{N}} \sum_{j \in \mathbb{N}} |\mu[A_j^k]| \leq \sum_{k \in \mathbb{N}} |\mu|[A^k];$$

en passant au supremum on obtient

$$|\mu|[A] \leq \sum_{k \in \mathbb{N}} |\mu|[A^k],$$

et on a bien la σ -additivité de $|\mu|$.

4. L'étape suivante consiste à montrer que $|\mu|$ est une mesure finie. Si A est un ensemble mesurable tel que $|\mu|[A] = +\infty$, alors on peut trouver une partition $(A_j)_{j \in \mathbb{N}}$ de A telle que

$$\sum_{j \in \mathbb{N}} |\mu[A_j]| \geq 2|\mu[A]| + 3;$$

et donc on peut trouver J fini tel que

$$\sum_{1 \leq j \leq J} |\mu[A_j]| \geq 2(|\mu[A]| + 1).$$

En distinguant selon le signe des $\mu[A_j]$, on peut trouver une famille finie \mathcal{J} d'indices j tels que

$$\left| \sum_{j \in \mathcal{J}} \mu[A_j] \right| \geq |\mu[A]| + 1.$$

Soit $E = \cap \{A_j; j \in \mathcal{J}\}$; on a donc $E \subset A$ et $|\mu[E]| \geq |\mu[A]| + 1$. Il s'ensuit

$$|\mu[A \setminus E]| = |\mu[A] - \mu[E]| \geq |\mu[E]| - |\mu[A]| \geq 1.$$

Par ailleurs, $|\mu|$ étant σ -additive, l'une au moins des deux quantités $|\mu|[E]$ et $|\mu|[A \setminus E]$ vaut $+\infty$. Conclusion : on peut séparer A en deux parties, disons A' et B , telles que $|\mu|[A'] = +\infty$ et $|\mu[B]| \geq 1$.

Supposant par l'absurde que $|\mu|[X] = +\infty$, on peut appliquer ce résultat avec $A = X$, et séparer X en deux parties disjointes A_1 et B_1 telles que $|\mu|[A_1] = +\infty$ et $|\mu[B_1]| \geq 1$; puis réappliquer le résultat avec $A = A_1$, et ainsi de suite. On construit ainsi une famille de parties disjointes $(B_k)_{k \in \mathbb{N}}$ telle que $|\mu[B_k]| \geq 1$. Ceci contredit la σ -additivité puisque la série $\sum \mu[B_k]$ ne converge pas. On conclut que $|\mu|[X] < +\infty$.

La définition de $|\mu|$ entraîne alors

$$\mu[B] \leq |\mu|[B] \leq |\mu|[X] < +\infty :$$

la fonction μ est *bornée*.

5. Une fois cette propriété acquise, il est facile de vérifier que μ vérifie des propriétés de passage à la limite similaires à celles des mesures : pour toute famille

(A_k) croissante, en appliquant la relation de σ -additivité à la famille $(A_k \setminus A_{k-1})$, on obtient

$$\mu\left[\bigcup A_k\right] = \lim_{k \rightarrow \infty} \mu[A_k].$$

Enfin, en passant au complémentaire et en utilisant $|\mu[X]| < +\infty$, on voit que pour toute famille (A_k) décroissante,

$$\mu\left[\bigcap A_k\right] = \lim_{k \rightarrow \infty} \mu[A_k].$$

6. Soit maintenant

$$M := \sup_{A \in \mathcal{A}} \mu[A] \geq 0.$$

D'après l'étape 4, $M < +\infty$. Le but est de montrer que ce supremum est atteint par un ensemble mesurable A , et de montrer que μ_+ est concentrée sur A . Si $M = 0$, il suffit de poser $S_+ = \emptyset$; nous supposons donc $M > 0$. Soit $(A_k)_{k \in \mathbb{N}}$ une suite de parties mesurables vérifiant

$$\mu[A_k] \geq \left(1 - \frac{1}{2^k}\right) M.$$

Posons

$$A := \limsup A_k = \bigcap_{\ell \in \mathbb{N}} \bigcup_{k \geq \ell} A_k.$$

La famille $C_\ell := \bigcup_{k \geq \ell} A_k$ étant décroissante, on sait que $\mu[A] = \lim_{\ell \rightarrow \infty} \mu[C_\ell]$. D'autre part, en appliquant de manière répétée l'inégalité

$$\mu[A_k \cup B] = \mu[A_k] + \mu[B] - \mu[A_k \cap B] \geq \mu[A_k] + \mu[B] - M \geq \mu[B] - 2^{-k}M,$$

on voit que, pour tout $m \geq \ell$,

$$\mu[A_\ell \cup \dots \cup A_m] \geq \mu[A_\ell] - \sum_{k=\ell}^m 2^{-k}M \geq \mu[A_\ell] - 2^{-(\ell-1)}M.$$

En passant à la limite quand $m \rightarrow \infty$, on obtient

$$\mu[C_\ell] \geq \mu[A_\ell] - 2^{-(\ell-1)}M \geq (1 - 3 \cdot 2^{-\ell})M.$$

Il ne reste plus qu'à faire tendre ℓ vers l'infini pour obtenir

$$\mu[A] \geq M;$$

d'où $\mu[A] = M$.

7. Posons $S_+ = A$, $S_- = X \setminus A$, de sorte que (S_+, S_-) réalise une partition de X ; on va montrer que pour tout $C \subset S_+$ on a $\mu[C] \geq 0$. Dans le cas contraire, on aurait

$$\mu[S_+ \setminus C] = \mu[S_+] - \mu[C] > \mu[S_+] = M,$$

ce qui contredirait la définition de M . De même, s'il existait $C \subset S_-$ tel que $\mu[C] > 0$, alors on aurait

$$\mu[S_+ \cup C] = \mu[S_+] + \mu[C] > \mu[S_+] = M,$$

ce qui est tout aussi impossible. On conclut que la restriction de μ aux parties mesurables de S_+ est positive, tandis que la restriction de μ aux parties mesurables de S_- est négative. Il s'ensuit que (S_+, μ) et $(S_-, -\mu)$ sont deux espaces mesurés; on peut alors écrire μ comme différence de deux mesures :

$$\mu[A] = \mu[A \cap S_+] - (-\mu[A \cap S_-]).$$

Les mesures $\mu_+ := \mu[\cdot \cap S_+]$ et $\mu_- := -\mu[\cdot \cap S_-]$ sont finies et étrangères, ce qui achève la preuve de la décomposition.

8. Il reste seulement à montrer l'équivalence des formules de variation totale ; pour le moment la notation $|\mu|$ désigne la formule de droite de (44), le problème est de montrer que cela coïncide avec $\mu_+ + \mu_-$. Pour cela on note d'abord que pour tout $A \in \mathcal{A}$, $|\mu[A]| = |\mu[A \cap S_+] + \mu[A \cap S_-]| \leq |\mu[A \cap S_+]| + |\mu[A \cap S_-]| = \mu_+[A] + \mu_-[A]$; en reportant cette inégalité dans la définition de $|\mu|$ on obtient $|\mu| \leq \mu_+ + \mu_-$. Pour prouver l'inégalité inverse, on note que $(A \cap S_+, A \cap S_-)$ réalise une partition de A , de sorte que

$$\mu_+[A] + \mu_-[A] = |\mu[A \cap S_+]| + |\mu[A \cap S_-]| \leq |\mu[A]|.$$

□

REMARQUE VI-53. La fin de la preuve montre a posteriori que dans le membre de droite de la formule (44), on peut se limiter aux partitions à deux éléments.

Il sera utile dans la suite de traiter des mesures régulières. Pour faire cela, nous utiliserons la Proposition suivante :

PROPOSITION VI-54 (Reformulation de la régularité des mesures signées). *Soit (X, \mathcal{A}) un espace mesurable et μ une mesure signée finie sur X ; soit (μ_+, μ_-) la décomposition de Hahn de μ . Alors les trois propositions suivantes sont équivalentes :*

- (i) μ_+ et μ_- sont régulières ;
- (ii) $|\mu|$ est régulière ;
- (iii) pour tout $A \in \mathcal{A}$ et pour tout $\varepsilon > 0$ il existe un compact K et un ouvert O tels que $K \subset A \subset O$ et

$$(45) \quad |\mu[A] - \mu[K]| \leq \varepsilon, \quad |\mu[O] - \mu[A]| \leq \varepsilon.$$

DÉMONSTRATION. L'implication (i) \Rightarrow (ii) est (presque) triviale, il suffit donc de montrer (ii) \Rightarrow (iii) \Rightarrow (i). On notera S_+ et S_- des parties disjointes sur lesquelles μ_+ et μ_- sont concentrées.

Supposons que (ii) est vérifiée, soit A un ensemble mesurable quelconque. Par régularité de $|\mu|$, on peut trouver une suite croissante de compacts $(K_n)_{n \in \mathbb{N}}$ inclus dans A , telle que $|\mu|[A \setminus K_n] \rightarrow 0$; et une suite décroissante d'ouverts $(O_n)_{n \in \mathbb{N}}$ contenant A , telle que $|\mu|[O_n \setminus A] \rightarrow 0$. Alors

$$|\mu[O_n] - \mu[A]| = |\mu[O_n \setminus A]| \leq |\mu|[O_n \setminus A] \rightarrow 0,$$

donc $\mu[O_n] \rightarrow \mu[A]$, et de même $\mu[K_n] \rightarrow \mu[A]$. La propriété (iii) est donc vérifiée.

Supposons maintenant (iii), et prouvons que (par exemple) μ_+ est régulière. Soit A un ensemble mesurable quelconque, et $A' = A \cap S_+$. Par (45) on peut trouver une suite de compacts $(K_n)_{n \in \mathbb{N}}$, inclus dans A' et donc dans A , tels que

$$\mu_+[K_n] = \mu[K_n] \rightarrow \mu[A'] = \mu_+[A'] = \mu_+[A];$$

la mesure μ_+ est donc intérieurement régulière.

Pour montrer la régularité extérieure, on applique la régularité intérieure à $B = S_+ \setminus A$: on trouve ainsi une famille $(L_n)_{n \in \mathbb{N}}$ de compacts inclus dans B , tels que $\mu[L_n] = \mu_+[L_n] \rightarrow \mu_+[B] = \mu[B]$. L'ouvert $O_n = X \setminus L_n$ contient alors A , et on a $\mu_+[O_n] = \mu_+[X \setminus L_n] \rightarrow \mu_+[X \setminus B] = \mu_+[A \cup S_-] = \mu_+[A]$.

□

Le corollaire suivant est une conséquence immédiate du Théorème VI-50 et de la Proposition VI-54 :

COROLLAIRE VI-55 (Théorème de Hahn pour les mesures régulières). *Soit (X, \mathcal{A}) un espace mesurable ; alors on peut identifier*

- *d'une part, les fonctions $\mu : \mathcal{A} \rightarrow \mathbb{R}$, σ -additives, telles que pour tout $A \in \mathcal{A}$ il existe des suites de compacts $(K_n)_{n \in \mathbb{N}}$ et d'ouverts $(O_n)_{n \in \mathbb{N}}$ vérifiant $K_n \subset A \subset O_n$ et*

$$\lim_{n \rightarrow \infty} \mu[K_n] = \lim_{n \rightarrow \infty} \mu[O_n] = \mu[A];$$

- *d'autre part, les mesures signées finies régulières (μ_+, μ_-) sur \mathcal{A} ;*

via la formule $\mu[A] = \mu_+[A] - \mu_-[A]$.

En outre, μ est régulière si et seulement si $|\mu|$ est régulière.

VI-4.3. Espace des mesures signées finies. Comme nous l'avons vu, le théorème de Hahn identifie les mesures signées finies avec les fonctions σ -additives d'ensembles à valeurs réelles. Il est clair que ce dernier espace est un **espace vectoriel**, ce qui n'était pas évident a priori pour les mesures signées finies. On peut donc munir les mesures signées finies d'une structure naturelle d'espace vectoriel : il devient possible d'ajouter ou de soustraire des mesures signées, ou de les multiplier par des nombres réels. L'écriture $\mu = \mu_+ - \mu_-$, qui jusqu'ici était purement formelle, peut maintenant s'interpréter, dans le cas où μ_+ et μ_- sont finies, comme une soustraction au sens usuel dans un espace vectoriel.

Le Corollaire VI-55 montre de même que les mesures signées finies régulières constituent un sous-espace vectoriel de l'espace des mesures signées finies.

Ces résultats ouvrent la voie à un traitement "fonctionnel" des mesures signées. La proposition suivante se démontre sans difficulté :

PROPOSITION VI-56 (inégalités élémentaires pour les mesures signées). *Soient (X, \mathcal{A}) un espace mesurable, μ et ν deux mesures signées finies sur X , identifiées à des fonctions σ -additives d'ensembles, à valeurs réelles ; alors*

$$\begin{aligned} \mu \leq \nu &\implies \mu_+ \leq \nu_+, \mu_- \geq \nu_-; \\ \forall \alpha \geq 0, (\alpha\mu)_\pm &= \alpha\mu_\pm; \quad \forall \alpha < 0, (\alpha\mu)_\pm = |\alpha|\mu_\mp; \\ (-\mu)_+ &= \mu_-; \quad |-\mu| = |\mu|; \\ (\mu + \nu)_\pm &\leq \mu_\pm + \nu_\pm; \quad |\mu + \nu| \leq |\mu| + |\nu|. \end{aligned}$$

Pour mesurer la taille d'une mesure signée, un concept naturel est fourni par la **variation totale** :

DÉFINITION VI-57 (variation totale). *Soient X un espace mesurable, et μ une mesure signée sur X ; soient μ_+ et μ_- les parties positive et négative de μ . On appelle variation totale de μ , et on note $\|\mu\|_{VT(X)}$ ou simplement $\|\mu\|_{VT}$, la quantité positive*

$$|\mu|[X] = \mu_+[X] + \mu_-[X].$$

Si A est une partie mesurable de X , on notera $\|\mu\|_{VT(A)} = |\mu|[A]$.

PROPOSITION VI-58 (propriétés de la variation totale). *Soient X un espace mesurable et μ une mesure signée sur X . Alors*

(i) $A \mapsto \|\mu\|_{VT(A)}$ est une fonction σ -additive d'ensembles (qui coïncide avec $|\mu|$) ;

(ii) Pour toute partie mesurable A de X ,

$$|\mu[A]| \leq \|\mu\|_{VT}.$$

Plus généralement, pour toutes parties disjointes $(A_k)_{k \in \mathbb{N}}$,

$$\sum_{k \in \mathbb{N}} |\mu[A_k]| \leq \|\mu\|_{VT}.$$

(iii) $\|\mu\|_{VT} = \sup_{|h| \leq 1} \int h d\mu$, où le supremum est pris sur toutes les fonctions mesurables sur X (majorées en valeur absolue par 1); on peut également restreindre le supremum aux fonctions mesurables valant ± 1 .

(iv) $\|\mu\|_{VT} = \inf \left\{ \nu_+[X] + \nu_-[X]; \mu = \nu^+ - \nu^- \right\}$, où l'infimum est pris sur tous les couples de mesures (ν^+, ν^-) , non nécessairement étrangères, telles que $\mu = \nu^+ - \nu^-$; en outre il y a égalité si et seulement si $\mu_{\pm} = \nu^{\pm}$.

DÉMONSTRATION. L'énoncé (i) est évident. Pour obtenir (ii), il suffit d'écrire

$$|\mu[A]| = |\mu_+[A] - \mu_-[A]| \leq \mu_+[A] + \mu_-[A] \leq \mu_+[X] + \mu_-[X].$$

Pour démontrer (iii), introduisons des ensembles disjoints S_+ et S_- tels que μ_{\pm} soit supportée par S_{\pm} . Il est alors clair que, dès que $|h| \leq 1$, on a

$$\int_{S_+} h d\mu = \int_{S_+} h d\mu_+ \leq \mu_+[S_+] = \mu_+[X];$$

et de même

$$\int_{S_-} h d\mu \leq \mu_-[X].$$

On conclut que $\int h d\mu \leq \|\mu\|_{VT}$. L'égalité est obtenue pour $h = 1_{S_+} - 1_{S_-}$, ce qui achève la preuve de (iii). Enfin, pour démontrer (iv) il suffit de démontrer que

$$\mu = \nu^+ - \nu^- \implies \nu^+[S_{\pm}] + \nu^-[S_{\pm}] \geq \mu_{\pm}[S_{\pm}].$$

Démontrons par exemple $\nu^+[S_+] + \nu^-[S_+] \geq \mu_+[S_+]$. Puisque $\mu_+[S_+] = \mu[S_+] = \nu^+[S_+] - \nu^-[S_+]$, cette inégalité se réduit à $\nu^-[S_+] \geq -\nu^+[S_+]$, ce qui est évident. Le traitement des cas d'égalité ne s'effectue sans difficulté. \square

Nous allons maintenant décrire de manière un peu plus précise l'espace des mesures signées finies :

THÉORÈME VI-59 (espace des mesures signées). *Soit X un espace mesurable. L'ensemble des mesures signées finies sur X , muni de la variation totale, constitue un espace de Banach, que l'on note $M(X)$. Pour toute mesure finie ν sur X , l'espace $L^1(\nu)$ s'identifie isométriquement à un sous-espace de $M(X)$ via l'injection $f \mapsto f\nu$: en particulier,*

$$\|f\|_{L^1(\nu)} = \|f\nu\|_{VT}.$$

L'espace des mesures signées finies régulières sur X , muni de la variation totale, est un sous-espace de Banach de $M(X)$, que l'on notera $M_{\text{reg}}(X)$.

REMARQUE VI-60. Si X est polonais, $M_{\text{reg}}(X) = M(X)$ en vertu du Théorème I-54.

DÉMONSTRATION. 1. Il est facile de vérifier que la variation totale définit bien une norme, en utilisant la Proposition VI-56.

2. Montrons maintenant que $M(X)$ est complet : soit $(\mu_k)_{k \in \mathbb{N}}$ une famille de mesures signées finies telles que

$$\|\mu_k - \mu_\ell\|_{VT} \xrightarrow{k, \ell \rightarrow \infty} 0.$$

Pour toute partie A mesurable, on a, d'après la Proposition VI-58(i),

$$|\mu_k[A] - \mu_\ell[A]| \leq \|\mu_k - \mu_\ell\|_{VT} \xrightarrow{k, \ell \rightarrow \infty} 0.$$

Il s'ensuit que la suite $(\mu_k[A])_{k \in \mathbb{N}}$ est de Cauchy, et elle converge donc (par complétude de \mathbb{R} !) vers un nombre réel que nous noterons $\mu[A]$.

Montrons que l'application μ ainsi définie est une mesure signée. Par le théorème de Hahn, il suffit de vérifier que c'est une fonction σ -additive ; pour cela on se donne une famille dénombrable d'ensembles A_j disjoints, et on écrit la relation de σ -additivité pour μ_k :

$$\mu_k\left[\bigcup_j A_j\right] = \sum_j \mu_k[A_j].$$

On peut passer à la limite quand $\ell \rightarrow \infty$ dans le premier terme ; pour passer à la limite dans le deuxième, et donc prouver la σ -additivité de μ , il suffit d'établir

$$\sum_j |\mu_\ell[A_j] - \mu[A_j]| \xrightarrow{\ell \rightarrow \infty} 0.$$

Mais, les A_j étant disjoints, on a, pour tout $k \geq \ell$, grâce à la Proposition VI-58(ii),

$$\sum_j |\mu_\ell[A_j] - \mu_k[A_j]| \leq \|\mu_\ell - \mu_k\|_{VT},$$

et le membre de droite converge vers 0 quand $\ell \rightarrow \infty$, uniformément en k . En faisant tendre d'abord k vers l'infini, puis ℓ , on obtient le résultat souhaité.

À ce stade nous savons qu'il existe une mesure signée μ telle que pour tout A mesurable, $\mu_k[A]$ converge vers $\mu[A]$ quand $k \rightarrow \infty$. Pour prouver la complétude, il reste à montrer que $\|\mu_k - \mu\|_{VT}$ tend vers 0. Soit h une fonction mesurable valant ± 1 sur X , et $\varepsilon(k) := \sup_{\ell \geq k} \|\mu_k - \mu_\ell\|_{VT}$. On a, d'après la Proposition VI-58(ii),

$$\int h d\mu_k - \int h d\mu_\ell \leq \varepsilon(k).$$

La fonction h est de la forme $1_A - 1_B$; on peut donc passer à la limite dans $\int h d\mu_\ell$ quand $\ell \rightarrow \infty$, et on trouve

$$\int h d\mu_k - \int h d\mu \leq \varepsilon(k).$$

En prenant le supremum sur h et en appliquant la Proposition VI-58(ii) encore, on conclut que $\|\mu_k - \mu\|_{VT} \leq \varepsilon(k)$, ce qui conclut l'argument.

3. Vérifions maintenant l'identité

$$\|f\|_{L^1(d\nu)} = \|f\nu\|_{VT}$$

pour toute mesure finie ν . Pour cela il suffit de noter que $f_+\nu$ et $f_-\nu$ constituent la décomposition de Hahn de la mesure signée $f\nu$; en utilisant la définition de la variation totale on trouve donc

$$\|f\nu\|_{VT} = f_+\nu[X] + f_-\nu[X] = \int f_+ d\nu + \int f_- d\nu = \int (f_+ + f_-) d\nu = \int |f| d\nu.$$

4. Comme on l'a déjà remarqué, le Corollaire VI-55 montre que $M_{\text{reg}}(X)$ soit un sous-espace vectoriel de $M(X)$. Supposons maintenant que μ_k est une suite de mesures régulières finies, convergeant vers μ en variation totale, et montrons que μ est régulière. Soit $\varepsilon > 0$, et soit k tel que $\|\mu_k - \mu\| \leq \varepsilon/2$. Comme μ_k est régulière, on peut trouver un ouvert O contenant A , et un compact K inclus dans A , tels que $|\mu_k[O \setminus A]| \leq \varepsilon/2$, $|\mu_k[A \setminus K]| \leq \varepsilon/2$. On écrit alors

$$|\mu[O \setminus A]| \leq \|\mu_k - \mu\|_{VT} + |\mu_k[O \setminus A]| \leq \varepsilon,$$

et de même

$$|\mu[A \setminus K]| \leq \|\mu_k - \mu\|_{VT} + |\mu_k[A \setminus K]| \leq \varepsilon.$$

La propriété (iii) de la Proposition VI-54 est donc satisfaite, ce qui prouve la régularité de μ . \square

VI-4.4. Théorème de Riesz pour les mesures signées. Comme nous l'avons vu au chapitre II, les mesures peuvent être introduites soit à partir du concept de σ -additivité, soit comme formes linéaires sur des espaces de fonctions continues, le théorème de Riesz garantissant l'équivalence de ces deux points de vue dans le cas localement compact. Il en va de même des mesures signées : nous les avons introduites comme différence de deux mesures, mais on aurait aussi pu les introduire à partir du point de vue des formes linéaires. C'est le contenu de l'énoncé suivant.

THÉORÈME VI-61 (théorème de Riesz pour des mesures signées). *Soit X un espace topologique séparé, localement compact. Alors on peut identifier (mettre en correspondance bijective et isométrique)*

- d'une part, les formes linéaires Λ continues sur l'espace $C_c(X)$ des fonctions continues sur X à support compact, muni de la norme de la convergence uniforme ;

- ou, de manière équivalente, les formes linéaires Λ continues sur l'espace $C_0(X)$ des fonctions continues sur X tendant vers 0 à l'infini, muni de la norme de la convergence uniforme ;

- d'autre part, les mesures de Borel signées, régulières et finies μ sur X ; c'est-à-dire de la forme $\mu_+ - \mu_-$, où μ_+ et μ_- sont des mesures de Borel régulières finies étrangères sur X ;

via la formule

$$\Lambda f = \int f d\mu := \int f d\mu_+ - \int f d\mu_-.$$

En bref,

$$C_0(X)^* = M_{\text{reg}}(X),$$

où $C_0(X)$ est muni de la norme uniforme et $M_{\text{reg}}(X)$ de la norme de la variation totale. En particulier

$$\begin{aligned}\|\mu\|_{VT} &= \sup \left\{ \int_X f d\mu; \quad |f| \leq 1, f \in C_c(X) \right\} \\ &= \sup \left\{ \int_X f d\mu; \quad |f| \leq 1, f \in C_0(X) \right\}.\end{aligned}$$

REMARQUE VI-62. Si X est un espace topologique séparé **compact**, alors on peut bien sûr remplacer l'espace $C_c(X)$ dans l'énoncé ci-dessus par $C(X)$. En revanche, si X n'est pas compact, le théorème n'affirme rien sur le dual de $C_b(X)$. Sous certaines hypothèses axiomatiques, on peut identifier $C_b(X)^*$ à l'espace des fonctions d'ensemble *finiment additives*, et montrer que cet espace est strictement plus grand que $M(X)$.

EXEMPLE VI-63. Fixons un espace localement compact non compact, vérifiant aux hypothèses du Théorème ?? (de sorte que toute mesure borélienne finie sur les compacts est automatiquement régulière); par exemple $X = \mathbb{R}^n$. Notons $C_\ell(X)$ l'espace de Banach des fonctions continues admettant une limite à l'infini. La fonctionnelle \lim_∞ (limite en l'infini) est linéaire continue, et (si l'on admet l'axiome du choix) se prolonge par Hahn–Banach en une application L , linéaire continue sur $C_b(X)$, et non nulle. L'application L ne peut être représentée par aucune mesure de Borel : comme $L(f) = 0$ pour tout $f \in C_c(X)$, cette mesure ne pourrait être que la mesure nulle. En fait, L est représentée par une fonction (finiment) additive d'ensembles; noter que cette fonction viole de manière évidente les hypothèses de la Proposition I-52, en fait L est “concentrée à l'infini”.

DÉMONSTRATION. 1. Soit d'abord $\mu = \mu_+ - \mu_-$ une mesure signée finie sur X ; alors, pour toute fonction $f \in C_c(X)$,

$$\left| \int f d\mu \right| \leq \int |f| d\mu_+ + \int |f| d\mu_- \leq C \|f\|_\infty,$$

où $C = \mu_+[X] + \mu_-[X]$. La fonctionnelle $f \mapsto \int f d\mu$ est donc bien une forme linéaire continue sur $C_c(X)$.

2. Réciproquement, soit Λ une forme linéaire continue sur $C_c(X)$; pour tout $f \in C_c(X)$, $f \geq 0$ on pose

$$\Phi(f) := \sup \left\{ \langle \Lambda, h \rangle; h \in C_c(X); 0 \leq h \leq f \right\}.$$

La fonctionnelle Φ , définie sur l'ensemble des fonctions continues positives à support compact, est positive et croissante ($f \leq g \implies \Phi(f) \leq \Phi(g)$); montrons qu'elle est sur-additive. Soient f_1 et f_2 deux fonctions continues positives à support compact, soit $\varepsilon > 0$ et soient h_1, h_2 deux fonctions continues à support compact telles que pour $i = 1, 2$,

$$0 \leq h_i \leq f_i, \quad \Lambda h_i \geq \Phi(f_i) - \varepsilon.$$

Alors $h := h_1 + h_2$ est une fonction continue à support compact telle que $0 \leq h \leq f_1 + f_2$, et on a

$$\Lambda h = \Lambda h_1 + \Lambda h_2 \geq \Phi(f_1) + \Phi(f_2) - 2\varepsilon.$$

En passant au supremum sur tous les h admissibles, on obtient

$$\Phi(f_1 + f_2) \geq \Phi(f_1) + \Phi(f_2) - 2\varepsilon.$$

En faisant finalement tendre ε vers 0, on conclut à la sur-additivité de Φ . On peut alors appliquer la Remarque II-55 (v) suivant l'énoncé du Théorème de Riesz II-54 pour conclure que Φ se représente par une mesure (positive) de Borel presque régulière, que nous noterons μ_+ .

3. Montrons maintenant que μ_+ est finie. Si $0 \leq h \leq f$, alors bien sûr $\|h\|_\infty \leq \|f\|_\infty$, et par continuité de Λ il existe $C > 0$, indépendant de f et h , tel que $\Lambda h \leq C\|f\|_\infty$. En passant au supremum, on obtient $\Phi(f) \leq C\|f\|_\infty$, soit

$$\int f d\mu_+ \leq C\|f\|_\infty.$$

Pour tout compact $K \subset X$, on peut trouver une fonction f , continue à support compact, qui soit comprise entre 0 et 1, identiquement égale à 1 sur K ; en appliquant l'inégalité précédente à une telle fonction, on obtient

$$\mu_+[K] \leq C.$$

Par ailleurs, X étant ouvert et μ_+ étant presque régulière, on a

$$\mu_+[X] = \sup \left\{ \mu_+[K]; K \text{ compact} \right\};$$

ce qui prouve $\mu_+[X] \leq C$. On conclut que μ_+ est finie. Par la Remarque II-55 (iii) suivant l'énoncé du Théorème II-54 (de Riesz), μ_+ est régulière.

4. Il est maintenant facile de conclure la preuve : la mesure μ_+ construite précédemment définit une forme linéaire continue sur $C_c(X)$, et il est clair que

$$\Lambda \leq \mu_+.$$

La forme linéaire $\mu_+ - \Lambda$ est donc une forme linéaire positive sur $C_c(X)$, et une nouvelle application du Théorème de Riesz nous permet de la représenter par une mesure de Borel presque régulière, que nous noterons μ_- . On montre, de même que précédemment, que μ_- est finie et régulière. La forme linéaire Λ peut donc s'écrire sous la forme $\mu_+ - \mu_-$, où μ_+ et μ_- sont des mesures de Borel finies régulières.

5. Il est évident que

$$\begin{aligned} \|\mu\|_{VT} &\geq \sup \left\{ \int_X f d\mu; \quad |f| \leq 1, f \in C_0(X) \right\} \\ &\geq \sup \left\{ \int_X f d\mu; \quad |f| \leq 1, f \in C_c(X) \right\}. \end{aligned}$$

Pour conclure la démonstration, il suffit donc d'établir

$$\|\mu\|_{VT} \leq \sup \left\{ \int_X f d\mu; \quad |f| \leq 1, f \in C_c(X) \right\}.$$

Pour cela, on décompose μ en parties positive et négative, et on utilise la régularité de μ_+ et μ_- pour trouver des ensembles compacts K_+ et K_- avec $\mu_\pm[K_\pm] \geq \mu_\pm[X] - \varepsilon$. Les ensembles compacts K_+ et K_- étant disjoints, on peut trouver des ouverts O_+ et O_- tels que $K_\pm \subset O_\pm$ et $O_+ \cap O_- = \emptyset$ (Cf. paragraphe I-2.3). Par le lemme d'Urysohn, on peut trouver φ_+ continue à valeurs dans $[0, 1]$, identiquement égale à 1 sur K_+ et à support compact dans O_+ ; et de même φ_- continue à valeurs dans

$[0, 1]$, identiquement égale à 1 sur K_- et à support compact dans O_- . On pose alors $f = \varphi_+ - \varphi_-$, de sorte que

$$\begin{aligned} \int f d\mu &\geq \mu_+[K_+] + \mu_-[K_-] - \|\mu\|_{VT}(X \setminus (K_+ \cup K_-)) \\ &\geq (\mu_+[X] - \varepsilon) + (\mu_-[X] - \varepsilon) - 2\varepsilon = \|\mu\|_{VT} - 4\varepsilon. \end{aligned}$$

On conclut en faisant tendre ε vers 0. \square

VI-4.5. Représentation duale de la variation totale. Comme corollaire du Théorème VI-61 (Théorème de Riesz pour les mesures signées), nous avons obtenu une représentation duale de la variation totale :

$$(46) \quad \|\mu\|_{VT} = \sup \left\{ \int_X f d\mu; \quad |f| \leq 1, f \in C_0(X) \right\}$$

$$(47) \quad = \sup \left\{ \int_X f d\mu; \quad |f| \leq 1, f \in C_c(X) \right\}.$$

Mais cette formule n'a été établie que dans le cas où X est localement compact, et en fait elle peut facilement être en défaut dans des espaces non localement compacts. Pourtant, sous des hypothèses très générales elle demeure vraie, pourvu que l'on remplace $C_0(X)$ par $C(X)$ (ou de manière équivalente par $C_b(X)$, puisqu'on impose $|f| \leq 1$ de toute façon).

PROPOSITION VI-64. *Soit X un espace métrique et μ une mesure (de Borel) signée sur X , régulière. Alors*

$$\|\mu\|_{VT} = \sup \left\{ \int_X f d\mu; \quad |f| \leq 1, f \in C_b(X) \right\}.$$

En particulier, cette formule est automatiquement vérifiée si μ est finie et X est un espace polonais.

REMARQUE VI-65. En dépit de cette proposition, le dual de $C_b(X)$ est a priori plus gros que $M_{\text{reg}}(X)$.

DÉMONSTRATION. La première partie de l'énoncé implique la deuxième puisque toute mesure finie sur un espace polonais est régulière (Théorème I-54). D'autre part il est clair que $\int \varphi d\mu \leq \|\mu\|_{VT}$ pour tout φ continu à valeurs dans $[-1, 1]$; il suffit donc de prouver que $\|\mu\|_{VT} \leq \sup\{\int \varphi d\mu\}$, où le supremum est pris sur les fonctions continues à valeurs dans $[-1, 1]$.

Soient S_+ et S_- des ensembles disjoints tels que $\|\mu\|_{VT} = \mu_+[S_+] + \mu_-[S_-]$. Comme μ est régulière, pour tout $\varepsilon > 0$ on peut trouver des compacts $K_+ \subset S_+$ et $K_- \subset S_-$ (bien sûr disjoints) tels que

$$\|\mu\|_{VT} \leq \mu_+[K_+] + \mu_-[K_-] + \varepsilon;$$

en particulier, la variation totale de μ sur le complémentaire de $K_+ \cup K_-$ est au plus ε .

Sur chaque compact K_{\pm} , on peut appliquer le théorème de Riesz : par exemple

$$\mu_+[K_+] = \|\mu_+\|_{VT(K_+)} = \sup \left\{ \int \varphi d\mu_+; \varphi \in C(K_+), \|\varphi\| \leq 1 \right\}.$$

On peut donc trouver φ_{\pm} (continue sur K_{\pm} et à valeurs dans $[-1, 1]$) tels que

$$\mu_{\pm}[K_{\pm}] \leq \int_{K_{\pm}} \varphi_{\pm} d\mu_{\pm} + \varepsilon.$$

(Quitte à remplacer φ_{\pm} par sa partie positive, on peut supposer que ces fonctions sont positives, donc à valeurs dans $[0, 1]$.) Soit φ définie sur $K_+ \cup K_-$, qui vaut φ_+ sur K_+ et $-\varphi_-$ sur K_- : on a alors

$$\|\mu\|_{VT} \leq \int_{K_+ \cup K_-} \varphi d\mu + 3\varepsilon.$$

Par le Théorème d'extension de Tietze–Urysohn (rappelé dans la sous-section I-2.3), on peut prolonger φ en une fonction continue sur X , toujours notée φ , à valeurs dans $[-1, 1]$. On a alors

$$\int_{K_+ \cup K_-} \varphi d\mu \leq \int_X \varphi d\mu + \|\mu\|_{VT(X \setminus (K_+ \cup K_-))} \leq \int_X \varphi d\mu + \varepsilon.$$

On conclut que

$$\|\mu\|_{VT} \leq \int_X \varphi d\mu + 4\varepsilon,$$

et on achève l'argument en faisant tendre ε vers 0. \square

VI-4.6. Espace des mesures de Radon. Auparavant nous avons concentré notre attention sur les mesures signées finies. Les mesures de Radon constituent une classe particulière de mesures non signées, d'usage courant en analyse, en relation avec la théorie des distributions. Avant de les introduire, notons que leur définition même varie de manière assez importante d'un auteur à l'autre.

DÉFINITION VI-66 (mesures de Radon). *Soient X un espace localement compact, muni de sa tribu borélienne, et Ω un ouvert de X ; on appelle mesure de Radon sur Ω une mesure signée, localement finie (i.e. finie sur tout compact de Ω) et régulière. On notera $M_{\text{loc}}(\Omega)$ l'espace de ces mesures.*

Autrement dit, les mesures de Radon sont “localement” des mesures finies régulières, mais leur variation totale peut être infinie. Ces mesures sont assez naturelles en analyse ; si l'on munit $C_c(\Omega)$ d'une topologie adéquate, dite topologie inductive, qui en fait un espace complet, il s'avère que $(C_c(\Omega))^* = M_{\text{loc}}(\Omega)$ (c'est bien sûr un avatar du théorème de Riesz). En d'autres termes, les mesures de Radon s'identifient donc alors au dual de l'espace des fonctions continues à support compact. C'est ce que traduit l'énoncé suivant (non démontré dans ce cours) :

THÉORÈME VI-67 (mesures de Radon comme formes linéaires). *Soit X un espace topologique séparé, localement compact, dans lequel tout ouvert est union dénombrable de compacts. Pour tout compact $K \subset X$, on note $C_K(X)$ l'espace des fonctions continues dans X , dont le support est contenu dans K . Alors l'espace des mesures de Radon s'identifie à l'espace des formes linéaires sur $C_c(X)$ dont la restriction à $C_K(X)$ est continue, pour tout compact $K \subset X$.*

VI-4.7. Convergence dans $M(X)$. On reviendra sur ce sujet dans le chapitre suivant, mais il pourra être bon de dresser dès maintenant la liste des trois notions de convergence dans $M(X)$ couramment utilisées : on distingue

- la **convergence en variation totale** :

$$\mu_k \xrightarrow{VT} \mu \quad \text{si} \quad \|\mu_k - \mu\|_{VT} \xrightarrow{k \rightarrow \infty} 0.$$

Cette notion est très rigide : par exemple, δ_{x_k} converge vers δ_x en variation totale seulement si x_k est égal à x pour k assez grand !

- la **convergence faible-étoile** :

$$\mu_k \xrightarrow[w-*]{\mu} \quad \text{si} \quad \forall \varphi \in C_c(X), \quad \int \varphi d\mu_k \xrightarrow{k \rightarrow \infty} \int \varphi d\mu.$$

La terminologie de convergence faible-étoile n'est licite que dans le cas où X est un espace localement compact et si l'on se restreint à des mesures régulières, de sorte que l'on peut appliquer le Théorème de Riesz. Notons que l'on peut remplacer l'espace $C_c(X)$ par $C_0(X)$, et que cette notion est en général sans intérêt dans un espace non localement compact (il se peut que $C_c(X) = \{0\}$, auquel cas la définition devient vide...).

- la **convergence faible**, ou convergence étroite :

$$\mu_k \xrightarrow[w-*]{\mu} \quad \text{si} \quad \forall \varphi \in C_b(X), \quad \int \varphi d\mu_k \xrightarrow{k \rightarrow \infty} \int \varphi d\mu.$$

Cette notion, plus faible de la convergence en variation totale mais plus forte que la convergence faible-étoile, est très populaire parmi les probabilistes (qui l'utilisent d'ordinaire dans le contexte des mesures de probabilité).

CHAPITRE VII

Analyse de Banach - À COMPLÉTER

C'est vers le début des années 1920 que s'élabore la théorie des espaces vectoriels normés, grâce en particulier aux travaux de F. Riesz et de Banach, dont le célèbre ouvrage, *Théorie des Opérations linéaires* (1932) fonde la théorie moderne des espaces maintenant appelés espaces de Banach :

DÉFINITION VII-1. *On appelle espace de Banach un espace vectoriel normé complet.*

Dans tout ce chapitre je me limiterai aux espaces de Banach réels.

Les espaces de Banach les plus élémentaires sont les espaces de fonctions continues : $C_b(X)$ et $C_0(X)$, où X est un espace topologique, $C_b(X)$ est l'ensemble des fonctions continues bornées $X \rightarrow \mathbb{R}$, $C_0(X)$ l'ensemble des fonctions continues $X \rightarrow \mathbb{R}$ tendant vers 0 à l'infini, et tous deux sont munis de la norme de la convergence uniforme. On peut y ajouter les espaces introduits au Chapitre VI :

- les espaces $L^p(X, \mu)$ de Lebesgue, où (X, μ) est un espace mesuré ;
- l'espace $M(X)$ des mesures (signées) finies, où X est un espace mesurable (la norme étant la variation totale).

Les propriétés les plus utiles de ces espaces seront passées en revue dans ce chapitre ; ce sera également l'occasion de développer les fondements de la théorie générale des espaces de Banach.

Parmi les thèmes les plus importants que nous rencontrerons, on peut mentionner

- l'utilisation de la propriété de complétude ;
- l'étude de la géométrie induite par la norme ;
- l'identification de l'espace dual (théorèmes de représentations) ;
- les critères de compacité pour différentes topologies.

D'autre part, un recours systématique sera fait à la propriété de **séparabilité** pour éviter tout usage de **l'axiome du choix** dans sa version la plus forte. Tous les résultats de ce chapitre reposent donc uniquement sur l'axiomatique classique comprenant l'"axiome du choix dépendant", dont l'usage ne prête guère à controverse.

VII-1. La complétude et ses conséquences

Comme on l'a déjà rappelé à plusieurs reprises, un espace métrique complet est un espace dans lequel toute suite de Cauchy est convergente. Cette propriété de complétude a des conséquences remarquables, comme nous allons le voir dans cette section.

VII-1.1. Briques élémentaires. La plupart des démonstrations faisant intervenir la complétude sont basées sur les trois énoncés simples suivants (on note d la distance et $\|\cdot\|$ la norme) :

- (i) le **théorème des fermés emboîtés** : Dans un espace métrique complet E , soit $(F_k)_{k \in \mathbb{N}}$ une suite décroissante de fermés non vides, dont le diamètre tend vers 0 ; alors leur intersection est un singleton.
- (ii) le **théorème de Baire** : Soit $(O_k)_{k \in \mathbb{N}}$ une suite d'ouverts denses dans un espace métrique complet E ; alors leur intersection est dense dans E .
- (iii) la **convergence des séries absolument convergentes** : si E est un espace vectoriel normé complet, et $(u_k)_{k \in \mathbb{N}}$ une suite à valeurs dans E , avec $\sum \|u_k\| < +\infty$, alors on peut définir la somme $\sum u_k = \lim_{n \rightarrow \infty} (\sum_{k \leq n} u_k) \in E$.

Rappelons brièvement la démonstration de ces énoncés.

• Pour prouver (i), on choisit $u_k \in F_k$ pour tout $k \in \mathbb{N}$; comme $u_{k+p} \in F_{k+p} \subset F_k$, la distance entre u_k et u_{k+p} est majorée par le diamètre de F_k , qui tend vers 0 quand $k \rightarrow \infty$; la suite (u_k) est donc de Cauchy ; par complétude, elle admet une limite $u \in E$, qui appartient forcément à tous les F_k puisque ceux-ci sont fermés. Si maintenant v est un autre élément de l'intersection des F_k , alors $d(u, v) \leq \text{diam}(F_k) \rightarrow 0$ quand $k \rightarrow \infty$; donc $d(u, v) = 0$, donc $u = v$, ce qui montre bien que l'intersection des F_k est un singleton.

• Soit maintenant $(O_k)_{k \in \mathbb{N}}$ une famille d'ouverts denses dans un espace métrique complet. Puisque O_1 est ouvert, on peut y inclure une boule fermée $B_1 = B[r_1, x_1]$. Puisque O_2 est dense, il intersecte la boule ouverte $B(x_1, r_1)$: on peut trouver $x_2 \in B(x_1, r_1) \cap O_2$. Cette intersection de deux ouverts est un ouvert, on peut donc y inclure une boule fermée $B_2 = B[r_2, x_2]$, et sans perte de généralité on peut supposer que $r_2 \leq r_1/2$. En répétant la construction, on construit une suite de boules fermées emboîtées B_k , dont le diamètre tend vers 0, telles que $B_k \subset O_1 \cap O_2 \cap \dots \cap O_k$. Par le théorème des fermés emboîtés, l'intersection des B_k est non vide ; et par construction cette intersection est incluse dans tous les O_k . Ceci montre que l'intersection des O_k est non vide. En appliquant cet énoncé avec E remplacé par une boule fermée $B[x_0, r_0]$ on conclut que l'intersection des O_k est en fait dense.

• Soit $(u_k)_{k \in \mathbb{N}}$ une série absolument convergente à valeurs dans un espace vectoriel normé complet E ; on définit $s_n := \sum_{k \leq n} u_k$. Pour tous $n, p \geq 0$,

$$\|s_{n+p} - s_n\| = \|u_{n+1} + \dots + u_{n+p}\| \leq \|u_{n+1}\| + \dots + \|u_{n+p}\| \leq \sum_{k \geq n+1} \|u_k\| \xrightarrow{n \rightarrow \infty} 0.$$

La suite $(s_n)_{n \in \mathbb{N}}$ est donc de Cauchy, et par complétude elle converge vers une limite $s \in E$. Ceci conclut la preuve de (iii).

REMARQUES VII-2. (i) Attention : dans le théorème de Baire, l'intersection des O_k n'est pas en général un ouvert.

(ii) Le théorème de Baire est souvent utilisé sous la forme contraposée que voici : Si E complet est réunion de fermés, alors au moins l'un d'entre eux est d'intérieur non vide.

VII-1.2. Construction d'espaces de Banach. Commençons par un critère utile permettant de vérifier en pratique qu'un espace donné est un espace de Banach. Il s'agit de la réciproque du théorème de convergence des séries absolument convergentes.

PROPOSITION VII-3 (critère de complétude d'un espace vectoriel normé). *Un espace vectoriel normé $(E, \|\cdot\|)$ est complet si et seulement si toute série absolument convergente à valeurs dans E est convergente.*

DÉMONSTRATION DE LA PROPOSITION VII-3. Nous avons déjà vu que si E est complet et si $(u_k)_{k \in \mathbb{N}}$ est une suite à valeurs dans E telle que $\sum \|u_k\| < +\infty$, alors la somme $\sum u_k$ existe dans E . Il suffit donc d'établir la réciproque : soit E un espace vectoriel normé dans lequel toute série absolument convergente est convergente, nous allons montrer que E est complet.

Soit $(u_k)_{k \in \mathbb{N}}$ une suite de Cauchy, et soit $\varepsilon_1 = 1/2$; on sait qu'il existe n_1 tel que pour tous $m, n \geq n_1$, $\|u_m - u_n\| \leq \varepsilon_1$; on pose $u'_1 = u_{n_1}$. Soit ensuite $\varepsilon_2 = 1/4$, on fixe $n_2 \geq n_1$ tel que pour tous $m, n \geq n_2$, $\|u_m - u_n\| \leq \varepsilon_2$, et on pose $u'_2 = u_{n_2}$; par construction, on a bien sûr $\|u'_2 - u'_1\| \leq \varepsilon_1$. En continuant de même, on construit par récurrence une suite $(u'_\ell)_{\ell \in \mathbb{N}}$, extraite de $(u_k)_{k \in \mathbb{N}}$, telle que pour tout $\ell \geq 1$,

$$\|u'_{\ell+1} - u'_\ell\| \leq 2^{-\ell}.$$

La série de terme général $(u'_{\ell+1} - u'_\ell)$ est absolument convergente, donc convergente, et sa somme partielle u'_ℓ converge vers une limite $u \in E$. La suite $(u_n)_{n \in \mathbb{N}}$ est de Cauchy et admet une sous-suite convergente, elle converge donc. (En effet, soit $\varepsilon > 0$ arbitraire, et soit N tel que pour tous $m, n \geq N$, $\|u_m - u_n\| \leq \varepsilon$. On peut choisir ℓ assez grand pour que $\|u'_\ell - u\| \leq \varepsilon$, et pour que u'_ℓ soit égal à u_{n_ℓ} pour un certain $n_\ell \geq N$. Il s'ensuit que $\|u_n - u\| \leq 2\varepsilon$.) \square

Notons que c'est la Proposition VII-3 qui a servi à démontrer la complétude des espaces L^p (Théorème VI-16).

Passons maintenant en revue une liste de critères simples permettant de construire des espaces vectoriels normés complets. La majorité des espaces de Banach usuels sont obtenus par utilisation combinée de quelques-uns de ces critères.

PROPOSITION VII-4. *Soient $(E, \|\cdot\|)$ un espace vectoriel normé complet, et F un sous-espace vectoriel fermé de E . Alors F , muni de la norme induite par $\|\cdot\|$, est un espace vectoriel normé complet.*

PROPOSITION VII-5. *Soit $(E_k, N_k)_{k \in \mathbb{N}}$ une suite (finie ou infinie) d'espaces vectoriels normés complets, et soit $p \in [1, \infty]$. Alors l'espace produit $\prod E_k$ est un espace vectoriel normé quand on le munit de la norme produit ℓ^p définie par*

$$\begin{cases} \|(x_k)_{k \in \mathbb{N}}\|_{\ell^p} = (\sum_{k \in \mathbb{N}} N_k(x_k)^p)^{1/p} & (p < \infty); \\ \|(x_k)_{k \in \mathbb{N}}\|_{\ell^\infty} = \sup_{k \in \mathbb{N}} N_k(x_k). \end{cases}$$

EXEMPLE VII-6. Si E et F sont des espaces de Banach, alors $E \times F$ est un espace de Banach quand on le munit de la norme $\|(x, y)\| = \|x\| + \|y\|$.

PROPOSITION VII-7. *Soient $(F, \|\cdot\|)$ un espace vectoriel normé complet, et X un espace topologique quelconque. On note $C_b(X; F)$ l'espace des fonctions continues bornées de X dans F , et $C_0(X; F)$ l'espace des fonctions continues de X dans F , tendant vers 0 à l'infini. Alors $C_b(X; F)$ et $C_0(X; F)$ sont des espaces vectoriels normés complets quand on les munit de la norme du supremum : $N(f) = \sup_{x \in X} \|f(x)\|$.*

PROPOSITION VII-8. *Soit $(F, \|\cdot\|)$ un espace vectoriel normé complet, et (X, μ) un espace mesuré. Pour tout $p \in [1, \infty]$, on note $L^p(X; F)$ l'espace des fonctions mesurables $f : X \rightarrow F$ telles que $\|f\| \in L^p(X, \mu)$, quotienté par la relation d'équivalence "coïncider presque partout". Alors $L^p(X; F)$ est un espace vectoriel normé complet*

quand on le munit de la norme

$$\begin{cases} \|f\|_{L^p} = \left\| \|f\| \right\|_{L^p(X)} = \left(\int_X \|f(x)\|^p d\mu(x) \right)^{1/p} & (1 \leq p < \infty), \\ \|f\|_{L^\infty} = \left\| \|f\| \right\|_{L^\infty(X)} = \operatorname{esssup}_{x \in X} \|f(x)\|. \end{cases}$$

REMARQUE VII-9. Cet énoncé contient la Proposition VII-5 comme cas particulier (prendre $X = \mathbb{N}$ et $\mu =$ mesure de comptage).

PROPOSITION VII-10. Soient $(E, \|\cdot\|)$ un espace vectoriel normé, et F un sous-espace vectoriel normé complet. Alors l'espace $L(E, F)$ des applications linéaires continues de E dans F est complet quand on le munit de la norme

$$\|T\|_{L(E, F)} = \|T\|_{E \rightarrow F} = \sup_{x \in E \setminus \{0\}} \frac{\|T(x)\|}{\|x\|} = \sup_{\|x\|=1} \|T(x)\| = \sup_{\|x\| \leq 1} \|T(x)\|.$$

COROLLAIRE VII-11 (espace dual). Soit E un espace vectoriel normé; l'espace $L(E, \mathbb{R})$ des formes linéaires continues de E dans \mathbb{R} est un espace de Banach, appelé **dual** de E et noté E^* ; il est muni de la norme

$$\|\Lambda\| := \sup_{\|x\| \leq 1} |\Lambda(x)| = \sup_{\|x\|=1} |\Lambda(x)|.$$

La dualité est un des ingrédients essentiels de la théorie des espaces de Banach. Si E est un espace vectoriel normé, pour tous $x \in E$ et $\Lambda \in E^*$ on note

$$\Lambda(x) = \langle \Lambda, x \rangle = \langle \Lambda, x \rangle_{E \times E'}.$$

Les crochets utilisés dans cette notation sont appelés **crochets de dualité** entre E et E^* ; ne pas confondre avec un produit scalaire. (Les deux notations sont cependant compatibles au sens où l'application "produit scalaire par un vecteur x " est une forme linéaire.)

Il peut être plus ou moins difficile d'"identifier" le dual d'un espace de Banach. Quand on parvient à trouver une façon simple de décrire les éléments d'un espace dual, on dit qu'on a un **théorème de représentation**. Citons dès maintenant les théorèmes de représentation les plus célèbres :

- (i) le dual d'un espace de Hilbert H "est" l'espace H lui-même ;
- (ii) le dual de L^p "est" $L^{p'}$ pour tout $p \in]1, +\infty[$;
- (iii) sur un espace mesuré σ -fini, le dual de L^1 "est" L^∞ ;
- (iv) sur un espace localement compact X , le dual de $C_0(X)$ "est" l'espace $M(X)$ des mesures signées de variation totale finie

Les énoncés (i) à (iii) seront prouvés dans la suite de ce chapitre ; quant à l'énoncé (iv), c'est exactement le contenu du Théorème VI-61).

Reprenons maintenant notre liste de critères de complétude :

PROPOSITION VII-12. Soient $(E, \|\cdot\|)$ un espace vectoriel normé complet, et F un sous-espace vectoriel fermé de E . Alors l'espace quotient E/F est un espace vectoriel normé complet quand on le munit de la norme quotient $\|X\|_{E/F} = \inf\{\|x\|; x \in X\}$, ou, ce qui revient au même, en désignant par \bar{x} la classe d'équivalence de x ,

$$\|\bar{x}\|_{E/F} = \inf \{ \|x + y\|; y \in F \}.$$

PROPOSITION VII-13. Soit V un espace vectoriel normé et soient $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ deux espaces de Banach s'injectant continûment dans V . On définit $E+F$ comme l'ensemble de toutes les sommes $x+y$ où x varie dans E et y dans F . On munit $E+F$ de la norme

$$\|z\|_{E+F} = \inf \left\{ \|x\|_E + \|y\|_F; x \in E; y \in F; x+y=z \right\}.$$

Alors $E+F$ est un espace de Banach.

PROPOSITION VII-14. Soit (E, N) un espace vectoriel normé, non nécessairement complet. On définit $\mathcal{C}(E)$ comme l'espace des suites de Cauchy à valeurs dans E . On injecte E dans $\mathcal{C}(E)$ en identifiant $x \in E$ à la suite constante égale à x ; et on étend N en une seminorme $\mathcal{C}(E)$ en définissant $N((x_k)_{k \in \mathbb{N}}) = \lim_{k \rightarrow \infty} N(x_k)$. On définit \overline{E} comme le quotient de $\mathcal{C}(E)$ par le noyau de N ; et on le munit de la norme N . L'espace \overline{E} est alors un espace vectoriel normé complet qui admet E comme sous-espace dense. On appelle \overline{E} le complété (ou la complétion) de E pour la norme N .

REMARQUE VII-15. Si F est un sous-espace dense de E , alors le complété de F (pour la distance induite par E) est E , comme on s'y attendait. Par exemple, le complété de \mathbb{Q} (muni de la distance usuelle) est \mathbb{R} ; en fait c'est ainsi que l'on construit traditionnellement \mathbb{R} .

REMARQUE VII-16. De manière imagée, le processus de complétion consiste à "ajouter à E toutes les limites de suites de Cauchy". En pratique, les normes auxquelles on applique ce procédé sont souvent de la forme $N(x) = \|T(x)\|_F$, où T est une application linéaire injective, et $(F, \|\cdot\|_F)$ un espace vectoriel normé. (On vérifie facilement qu'une telle application N est bien une norme.)

EXEMPLE VII-17. Soit $p \in [1, \infty]$ et soit $E = C_c^1(\mathbb{R}^n; \mathbb{R})$ (l'espace des fonctions C^1 à support compact, de \mathbb{R}^n dans \mathbb{R}), muni de la norme $\|T(f)\|$, où

$$T(f) = \left(f, \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right).$$

L'application linéaire T est injective de $C^1(\mathbb{R}^n, \mathbb{R})$ dans $L^p(\mathbb{R}^n) \times \dots \times L^p(\mathbb{R}^n)$. On définit alors l'**espace de Sobolev** $W^{1,p}(\mathbb{R}^n)$ comme la complétion de E . En clair, $W^{1,p}(\mathbb{R}^n)$ est l'espace complet défini par la norme

$$\|f\|_{W^{1,p}(\mathbb{R}^n)} = \|f\|_{L^p(\mathbb{R}^n)} + \sum_{j=1}^n \left\| \frac{\partial f}{\partial x_j} \right\|_{L^p(\mathbb{R}^n)}.$$

Les espaces $W^{1,p}$ sont les plus simples représentants de la grande famille des espaces de Sobolev et de leurs généralisations.

DÉMONSTRATION DE LA PROPOSITION VII-4. Soit $(x_k)_{k \in \mathbb{N}}$ une suite de Cauchy dans F ; c'est alors une suite de Cauchy dans E , elle converge donc vers $x \in E$, qui appartient à F puisque F est fermé. \square

DÉMONSTRATION DE LA PROPOSITION VII-5. Cette proposition est un cas particulier de la Proposition VII-8, cependant sa démonstration est plus élémentaire. Notons tout d'abord que $\|\cdot\|_{\ell^p}$ est bien une norme, l'inégalité triangulaire étant une conséquence de l'inégalité de Minkowski (pour l'espace L^p associé à la mesure de comptage sur \mathbb{N}). Si x est une suite, on note $x = (x_k)_{k \in \mathbb{N}}$. Soit alors $(x^\ell)_{\ell \in \mathbb{N}}$

une suite de Cauchy pour cette norme. (C'est donc une suite de suites.) Comme $\|y\|_{\ell^p} \geq N_k(y_k)$, la suite $(x_k^\ell)_{\ell \in \mathbb{N}}$ est de Cauchy dans E_k , quel que soit ℓ . Elle converge donc au sens de la norme N_k vers un élément x_k de E_k . Pour tout $\varepsilon > 0$ il existe $N \in \mathbb{N}$ tel que pour tous $\ell, m \geq N$,

$$\sum_{k \in \mathbb{N}} N_k(x_k^\ell - x_k^m)^p \leq \varepsilon^p.$$

En particulier,

$$\sum_{k \leq K} N_k(x_k^\ell - x_k^m)^p \leq \varepsilon^p,$$

et on peut passer à la limite dans cette expression quand $m \rightarrow \infty$:

$$\sum_{k \leq K} N_k(x_k^\ell - x_k)^p \leq \varepsilon^p.$$

En faisant tendre K vers l'infini, on en déduit que $\|x^\ell - x\|_{\ell^p} \leq \varepsilon$. La suite x^ℓ converge donc vers x , ce qui conclut la preuve. \square

DÉMONSTRATION DE LA PROPOSITION VII-7. Soit $(f_k)_{k \in \mathbb{N}}$ une suite de Cauchy dans $C_b(X; F)$. Alors, pour tout x , la suite $(f_k(x))_{k \in \mathbb{N}}$ est de Cauchy dans X , et converge donc vers un élément de X que l'on note $f(x)$. Pour tout $\varepsilon > 0$ on peut trouver $N \geq 1$ tel que pour tous $k, \ell \geq N$, et pour tout $x \in X$, $\|f_k(x) - f_\ell(x)\| \leq \varepsilon$. En passant à la limite dans cette inégalité quand $\ell \rightarrow \infty$, on voit que f_k converge uniformément vers f ; il s'ensuit en particulier que f est continue. L'espace $C_b(X; F)$ est donc bien complet. On laisse en exercice l'adaptation de cette démonstration à l'espace $C_0(X; F)$. \square

DÉMONSTRATION DE LA PROPOSITION VII-8. Elle est en tout point semblable à celle du Théorème VI-16 (de Riesz-Fischer), modulo le remplacement de \mathbb{R} par l'espace complet F . \square

DÉMONSTRATION DE LA PROPOSITION VII-10. On vérifie aisément que la formule donnée dans l'énoncé définit bien une norme. La preuve de la complétude est la même que celle de la complétude de $C_b(X)$: soit $(T_n)_{n \in \mathbb{N}}$ une suite de Cauchy dans $L(E, F)$; alors, pour tout $x \in E$, la suite $(T_n(x))$ est de Cauchy dans F , et converge donc vers un vecteur que l'on peut noter $T(x)$. En passant à la limite dans la définition de la linéarité de T_n , on vérifie que T est linéaire. Par ailleurs, la suite (T_n) étant de Cauchy, elle est bornée en norme par une constante $M \geq 0$; pour tout x de norme 1 on a $|T_n(x)| \leq M$, et en passant à la limite on trouve $|T(x)| \leq M$, donc $\|T\| \leq M$, ce qui montre que T est continue. \square

DÉMONSTRATION DE LA PROPOSITION VII-12. Rappelons que E/F est l'espace E quotienté par la relation d'équivalence " x est en relation avec y si et seulement si $x - y \in F$ ". On vérifie facilement que cette définition permet d'étendre les opérations d'addition et de multiplication scalaire à E/F , et que l'on a défini ainsi un espace vectoriel. Un élément de E/F est de la forme $x + F$, où $x \in E$.

Supposons donc E complet et F fermé. Pour vérifier la complétude de E/F , on vérifie la convergence de toute série absolument convergente (Proposition VII-3). Soit donc $(X_k)_{k \in \mathbb{N}}$ une suite à valeurs dans E/F , telle que

$$\sum_{k \in \mathbb{N}} \|X_k\|_{E/F} < +\infty;$$

nous allons montrer que la suite des sommes partielles $\sum_{k \leq n} X_k$ converge dans E/F . Pour tout k , on peut choisir $x_k \in X_k$ tel que $\|x_k\| \leq 2\|X_k\|_{E/F}$ (si $X_k = F$ c'est évident, et sinon c'est une conséquence de la définition de la norme quotient). Alors la série $\sum \|x_k\|$ converge, et la suite des sommes partielles $s_k = x_1 + \dots + x_k$ converge vers un certain $s \in E$. L'inégalité $\|s_k - s\| \leq \varepsilon$ implique $\|S_k - S\| \leq \varepsilon$, où $S_k = X_1 + \dots + X_k$ et S est la classe de s . Il s'ensuit que S_k converge vers S .

Il semble que nous n'ayons pas utilisé l'hypothèse de fermeture de F ...! Elle intervient en fait pour vérifier que N est bien une norme sur E . En effet, si $N(\bar{x}) = 0$ cela veut dire qu'il existe une suite y_k d'éléments de F tels que $\|x + y_k\| \rightarrow 0$; on en déduit que $x = -\lim y_k$, et pour en déduire que $\bar{x} = 0$, c'est-à-dire $x \in F$, il nous faut supposer que F est fermé.

Les autres propriétés de la norme N (homogénéité, inégalité triangulaire) ne dépendent pas de ce que F est fermé :

$$\inf_{x \in \lambda X} \|x\| = \inf_{x' \in X} \|\lambda x'\| = |\lambda| \inf_{x' \in X} \|x'\|;$$

$$\inf_{z \in X+Y} \|z\| \leq \inf_{x \in X, y \in Y} \|x + y\| \leq \inf_{x \in X, y \in Y} (\|x\| + \|y\|) = \inf_{x \in X} \|x\| + \inf_{y \in Y} \|y\|.$$

□

DÉMONSTRATION DE LA PROPOSITION VII-13. L'espace $E + F$ s'identifie au quotient de l'espace complet $E \times F$ par le sous-espace fermé $\Delta = \{(x, -x); x \in E \cap F\}$. C'est donc un espace complet. □

DÉMONSTRATION DE LA PROPOSITION VII-14. La démonstration ne présente pas de difficulté mais elle est assez lourde à écrire et à lire; on pourra l'omettre en première lecture.

Étant donnée une suite de Cauchy (x_k) dans E , la suite $(N(x_k))_{k \in \mathbb{N}}$ est une suite de Cauchy dans \mathbb{R} , grâce à l'inégalité $|N(x) - N(y)| \leq N(x - y)$. Il s'ensuit que cette suite converge vers une limite que l'on peut noter $N((x_k))$. On vérifie sans problème que N , définie sur $\mathcal{C}(E)$, est homogène et vérifie l'inégalité triangulaire, mais ce n'est bien sûr pas une norme : son noyau est exactement l'espace vectoriel Z des suites qui convergent vers 0. On définit \bar{E} comme le quotient de E par Z . Si $(x_k)_{k \in \mathbb{N}}$ et $(x'_k)_{k \in \mathbb{N}}$ sont deux représentants d'une même classe d'équivalence $x \in \bar{E}$, on vérifie facilement que $N((x_k)) = N((x'_k))$, ce qui permet de définir $N(x)$ sans ambiguïté pour $x \in \bar{E}$. L'application N reste homogène et vérifie toujours l'inégalité triangulaire, mais maintenant par construction $N(x) = 0$ si et seulement si x est (la classe d'équivalence de) 0. L'espace (\bar{E}, N) est donc un espace vectoriel normé.

Vérifions maintenant que \bar{E} est complet. Soit $(x^\ell)_{\ell \in \mathbb{N}}$ une suite de Cauchy de suites de Cauchy (!). Pour tout $\varepsilon > 0$ il existe $N \geq 1$ tel que pour $\ell, \ell' \geq N$ on ait $\lim_{k \rightarrow \infty} N(x_k^\ell - x_k^{\ell'}) \leq \varepsilon$; en particulier il existe k_0 tel que $N(x_k^\ell - x_k^{\ell'}) \leq \varepsilon$ pour tout $k \geq k_0$. Quitte à augmenter k_0 , on peut supposer $k_0 \geq N$. Alors $N(x_k^\ell - x_k^k) \leq \varepsilon$ pour tout $k \geq k_0$. La suite x^ℓ converge donc vers la suite y définie par $y_k = x_k^k$.

Si $(x_k)_{k \in \mathbb{N}}$ est une suite de Cauchy, alors la suite $\lim_{k \rightarrow \infty} N(x_\ell - x_k)$ tend vers 0 quand $\ell \rightarrow \infty$, comme conséquence de la définition des suites de Cauchy. Il s'ensuit que $N(x_\ell - x) \rightarrow 0$, si l'on note x la suite $(x_k)_{k \in \mathbb{N}}$. Cette propriété subsiste quand on passe au quotient : la classe d'équivalence d'une suite de Cauchy est donc limite dans \bar{E} de cette même suite de Cauchy, vue comme suite d'éléments de E ; et l'espace E est donc effectivement dense dans \bar{E} . □

VII-1.3. Théorèmes de continuité automatique. Les théorèmes de continuité automatique constituent une manifestation spectaculaire des propriétés liées à la complétude. Ces théorèmes garantissent que sous certaines conditions, certaines applications linéaires entre espaces de Banach sont continues ; ce sont les théorèmes de **Banach–Steinhaus** et de **l’application ouverte**, ainsi que leurs conséquences (théorème de Banach, théorème du graphe fermé). Tous ces résultats reposent crucialement sur l’hypothèse de complétude. Avant de les exposer, je vais commencer par rappeler quelques propriétés élémentaires des applications linéaires.

PROPOSITION VII-18. *Soient E et F deux espaces vectoriels normés.*

(i) *Une application linéaire $T : E \rightarrow F$ est continue si et seulement si elle est bornée, au sens où elle est bornée sur la sphère unité, c’est-à-dire*

$$\|T\|_{E \rightarrow F} := \sup_{\|x\|=1} \|T(x)\| = \sup_{x \neq 0} \frac{\|T(x)\|}{\|x\|} < +\infty.$$

(ii) *Une application linéaire $T : E \rightarrow F$ est ouverte (au sens où elle envoie un ouvert sur un ouvert) si et seulement si l’image réciproque par T de la boule unité $B_1(0)$ dans E contient une boule $B_r(0)$ ($r > 0$) dans F .*

(iii) *Si T est une application linéaire bijective $E \rightarrow F$, sa réciproque T^{-1} est continue si et seulement si elle est minorée en norme sur la sphère unité de E , c’est-à-dire*

$$\inf_{\|x\|=1} \|T(x)\| = \inf_{\|x\| \neq 0} \frac{\|T(x)\|}{\|x\|} > 0.$$

DÉMONSTRATION. (i) Remarquons tout d’abord que T , étant linéaire, est bornée sur la sphère unité si et seulement si elle est bornée sur la boule unité, et qu’alors on a

$$\|T(x)\| \leq \left(\sup_{\|y\| \leq 1} \|T(y)\| \right) \|x\|.$$

Il est évident également que si $\|T\|_{E \rightarrow F}$ est finie, alors T est lipschitzienne, donc continue. C’est l’implication inverse qui est plus subtile : si T est continue, montrer qu’elle est bornée sur la boule unité. Comme $T(0) = 0$ et T est continue en 0, il existe $\delta > 0$ tel que

$$\|x\| \leq \delta \implies \|T(x)\| \leq 1.$$

On peut alors écrire $x = (\|x\|/\delta)y$, avec $\|y\| = \delta$, d’où $\|T(x)\| \leq \|x\|/\delta$.

(ii) Si T est ouverte, alors l’image de la boule unité $B_1(0)$ est un ouvert contenant 0, et donc une boule $B_r(0)$ de rayon non nul. Réciproquement, si l’image de $B_1(0)$ contient une boule $B_r(0)$, soit O un ouvert de E , et $x \in O$, il existe alors une boule $B_\delta(x) = x + \delta B_1(0) \subset O$, et $T(O)$ contient $T(B_\delta(x)) = T(x) + \delta T(B_1(0))$, qui contient donc $T(x) + \delta B_r(0)$, c’est-à-dire une boule centrée en $T(x)$. L’image de O par T est donc un ouvert.

(iii) Il suffit d’écrire $x = T(T^{-1}(x))$ et de lire le critère (i) à l’envers. \square

THÉORÈME VII-19 (Théorème de Banach–Steinhaus). *Soit $(T_\alpha)_{\alpha \in A}$ une famille d’opérateurs linéaires continus de E , espace vectoriel normé complet, dans F , espace vectoriel normé quelconque. On suppose que pour tout $x \in E$, l’ensemble des $\|T_\alpha(x)\|$ est borné. Alors la famille T_α est uniformément bornée : il existe une constante C telle que*

$$\forall \alpha \in A, \forall x \in E, \quad \|T_\alpha(x)\| \leq C\|x\|.$$

REMARQUE VII-20. Ce théorème peut se reformuler comme une interversion de quantificateurs : l'énoncé

$$\forall x \in E, \|x\| \leq 1, \quad \exists C > 0; \quad \sup_{\alpha \in A} \|T(x)\| \leq C$$

implique l'énoncé, à première vue plus fort,

$$\exists C > 0; \quad \forall x \in E, \|x\| \leq 1, \quad \sup_{\alpha \in A} \|T(x)\| \leq C.$$

On peut le résumer par la formule “ponctuellement borné implique uniformément borné”, ou encore (selon une terminologie qui deviendra claire quand nous parlerons plus loin de convergence faible) “faiblement borné implique fortement borné”.

COROLLAIRE VII-21 (Théorème de continuité de Banach–Steinhaus). *Soit $(T_n)_{n \in \mathbb{N}}$ une famille d'opérateurs linéaires continus de E , espace vectoriel normé complet, dans F , espace vectoriel normé quelconque. On suppose que pour tout $x \in E$, la suite $T_n(x)$ converge dans F . Alors l'opérateur linéaire $T := \lim T_n$ est continu.*

COROLLAIRE VII-22 (Continuité des applications bilinéaires). *Soient E et F deux espaces vectoriels normés, dont l'un au moins est complet, et soit B une application bilinéaire définie sur $E \times F$, continue en chaque variable séparément. Alors B est continue au sens des applications bilinéaires : il existe $C > 0$ tel que $\|B(x, y)\| \leq C \|x\| \|y\|$ pour tout $(x, y) \in E \times F$.*

DÉMONSTRATION DU THÉORÈME VII-19. Soit, pour tout $n \in \mathbb{N}$,

$$F_n := \left\{ x; \sup_{\alpha \in A} \|T_\alpha(x)\| \leq n \right\}.$$

Pour chaque α , la fonction $x \mapsto \|T_\alpha(x)\|$ est continue $E \rightarrow \mathbb{R}$; l'ensemble F_n est donc une intersection de fermés, donc un fermé. Par hypothèse, tout $x \in E$ appartient à l'un des F_n , c'est-à-dire que la réunion des F_n est E tout entier. Puisque E est complet, l'un au moins des F_n est d'intérieur non vide, par contraposée du théorème de Baire. Il existe donc $n_0 \in \mathbb{N}$, $x_0 \in E$, $r_0 > 0$ tels que pour tous $\alpha \in A$, et $y \in E$ de norme unité,

$$\|T_\alpha(x_0 + r_0 y)\| \leq n_0.$$

On en déduit $\|T_\alpha(y)\| \leq n_0 + r_0^{-1} \|T_\alpha(x_0)\| \leq n_0(1 + r_0^{-1})$. \square

REMARQUE VII-23. On trouvera dans [Lieb–Loss, pp.52-53] une preuve légèrement plus directe, sans utilisation du théorème de Baire (mais pas plus constructive pour autant).

DÉMONSTRATION DU COROLLAIRE VII-21. C'est une conséquence immédiate du Théorème VII-19, puisqu'une suite convergente est forcément bornée. \square

DÉMONSTRATION DU COROLLAIRE VII-22. Supposons par exemple que F est complet, et soit G l'espace d'arrivée de l'application B . Pour tout $y \in F$, on pose $B_y(x) = B(x, y)$. L'application B_y étant continue linéaire $E \rightarrow G$, il existe $C_y > 0$ tel que pour tout $x \in E$, de norme au plus 1, $\|B_y(x)\| \leq C_y$. L'inégalité $\|B(x, y)\| \leq C_y$ montre que la famille d'applications linéaires $y \rightarrow B(x, y)$ (paramétrée par x) vérifie les hypothèses du Théorème VII-19; il existe donc $C > 0$, indépendante de y , telle que $\|B(x, y)\| \leq C$ pour tous x, y de norme au plus 1. Ceci montre que B est continue au sens des applications bilinéaires. \square

Après le Théorème de Banach–Steinhaus, passons maintenant au deuxième résultat frappant de cette section :

THÉORÈME VII-24 (Théorème de l'application ouverte). *Soient E et F deux espaces de Banach, et $T : E \rightarrow F$ une application surjective continue. Alors T est ouverte.*

COROLLAIRE VII-25 (Théorème de Banach). *Soient E et F deux espaces de Banach, et $T : E \rightarrow F$ une application bijective continue. Alors son inverse T^{-1} est continu $F \rightarrow E$.*

Le corollaire qui suit, correspondant au choix $T = \text{Id}$, semble peut-être plus surprenant encore que le précédent :

COROLLAIRE VII-26. *Soit E un espace vectoriel et soient N, N' deux normes qui font de E un espace de Banach. S'il existe $C > 0$ tel que $N' \leq CN$, alors il existe $C' > 0$ tel que $N \leq C' N'$.*

Avant de donner les démonstrations, on va énoncer un lemme simple, dont la preuve est laissée en exercice.

LEMME VII-27. *Si A et B sont deux parties d'un espace vectoriel normé E et λ un nombre réel, on définit*

$$A + B = \{x + y; a \in A, b \in B\}, \quad \lambda A = \{\lambda a, a \in A\}.$$

Alors $\overline{A} + \overline{B} \subset \overline{A + B}$, $\overline{\lambda A} = \lambda \overline{A}$, et, si $\lambda \neq 0$, $\text{Int}(\lambda A) = \lambda \text{Int}(A)$. En particulier, si A et B sont fermés, alors $A + B$ est fermé. En outre, si A est ouvert (et B est quelconque), alors $A + B$ est ouvert.

DÉMONSTRATION DU THÉORÈME VII-24. D'après la Proposition VII-18, pour montrer que T est ouverte, il suffit de montrer que $T(B_r(0))$ contient un voisinage de 0 pour un (ou, de manière équivalente, pour tout) $r > 0$. On va noter pour abrégé $B_r = B_r(0)$.

1. Montrons que $\overline{T(B_1)}$ contient un voisinage de 0. Cette étape utilise la complétude de F et la surjectivité de T . Par hypothèse, $F = T(E) = \cup T(B_n) \subset \cup \overline{T(B_n)}$. L'espace complet F est union des fermés $\overline{T(B_n)}$; par le théorème de Baire (sous forme contraposée) il existe n_0 tel que $\overline{T(B_{n_0})}$ soit d'intérieur non vide. Alors l'intérieur de $\overline{T(B_{1/2})}$, qui est égal à $(1/(2n_0))\overline{T(B_{n_0})}$, est également d'intérieur non vide, et contient un ouvert O . Cet ouvert n'est pas nécessairement voisinage de 0, mais la différence $O - O = \{x - y; x \in O, y \in O\}$ est un ouvert contenant 0. Il est clair que $O - O \subset \overline{T(B_{1/2})} - \overline{T(B_{1/2})}$, et par le Lemme VII-27, $O - O \subset \overline{T(B_{1/2}) + T(B_{1/2})} = \overline{T(B_{1/2} + B_{1/2})} \subset \overline{T(B_1)}$.

2. Comme conséquence de l'étape 1, pour tout $r > 0$, $\overline{T(B_r)}$ contient un voisinage de 0. On va maintenant montrer que $\overline{T(B_{1/2})} \subset T(B_2)$, ce qui conclura l'argument. Cette étape utilise la complétude de E et la continuité de T . Soit donc $y \in \overline{T(B_{1/2})}$. Puisque $\overline{T(B_{1/4})}$ est un voisinage de 0, $y - \overline{T(B_{1/4})}$ est un voisinage de y , et doit donc intersecter $T(B_{1/2})$, puisque y appartient à son adhérence. Il existe donc $x_1 \in B_{1/2}$ tel que $T(x_1) \in y - \overline{T(B_{1/4})}$, c'est-à-dire

$$y - T(x_1) \in \overline{T(B_{1/4})}.$$

Par le même argument, on peut trouver $x_2 \in B_{1/4}$ tel que

$$(y - T(x_1)) - T(x_2) \in \overline{T(B_{1/8})}.$$

Par itération, on construit ainsi une suite $(x_k)_{k \in \mathbb{N}}$ telle que $x_k \in B_{1/2^k}$ et

$$(48) \quad y - (T(x_1) + \dots + T(x_k)) \in \overline{T(B_{1/2^{k+1}})}.$$

Par continuité de T , il existe $C > 0$ telle que $\overline{T(B_1)}$ soit inclus dans la boule fermée $B[0, C]$, donc dans la boule ouverte B_{2C} . Par conséquent, $\overline{T(B_{1/2^{k+1}})} \subset \overline{T(B_{(2C)/2^{k+1}})}$, qui converge vers $\{0\}$ quand $k \rightarrow \infty$. En reportant dans (48), on en déduit que la série $\sum T(x_k)$ converge et a pour somme y . D'autre part, $\|x_k\| \leq 2^{-k}$, donc la série $\sum x_k$ est absolument convergente, et par complétude de E , converge vers un certain $x \in E$ tel que $\|x\| \leq \sum \|x_k\| \leq 1$. On peut alors passer à la limite grâce à la continuité de T : $T(x) = T(\sum x_k) = \sum T(x_k)$. En conclusion, il existe $x \in B_2$ tel que $y = T(x)$. Ceci conclut la preuve du Théorème VII-24.

3. Supposons maintenant que T est une bijection linéaire continue entre espaces de Banach. D'après le Théorème VII-24, T est ouverte ; on en déduit que l'image réciproque par T^{-1} de tout ouvert est ouverte, et donc T^{-1} est continue. \square

DÉMONSTRATION DU COROLLAIRE VII-25. Il suffit d'appliquer le Théorème VII-24 après avoir noté que la réciproque d'une bijection ouverte est continue (pour tout ouvert O , $(T^{-1})^{-1}(O) = T(O)$ est ouvert). \square

Nous allons maintenant examiner un dernier résultat célèbre de continuité automatique :

THÉORÈME VII-28 (Théorème du graphe fermé). *Soient E et F des espaces de Banach, et T une application linéaire de E dans F . Alors T est continu si et seulement si son graphe $G(T) = \{(x, T(x)); x \in E\}$ est fermé dans $E \times F$ (muni de la norme $\|(x, y)\| = \|x\| + \|y\|$).*

DÉMONSTRATION. Si T est continu, alors $\varphi : (x, y) \mapsto y - T(x)$ est continue de E dans $E \times F$, et le graphe $G(T) = \varphi^{-1}(0)$ est fermé.

Réciproquement, supposons $G(T)$ fermé. On note que $G(T)$ est un espace vectoriel par linéarité de T ; comme sous-espace vectoriel fermé de l'espace de Banach $E \times F$, il est lui-même un espace de Banach. Considérons les applications de projection π_E et π_F définies par $\pi_E(x, y) = x$, $\pi_F(x, y) = y$. L'application π_E est une bijection continue de E dans $G(T)$; par le Théorème VII-25 (de Banach) son inverse π_E^{-1} est continu de $G(T)$ dans E . L'application T est alors continue de E dans F puisque $T = \pi_F \circ \pi_E^{-1}$. \square

REMARQUE VII-29. Ici on a démontré le Théorème du graphe fermé comme conséquence du Théorème de Banach ; mais on peut aussi prouver le Théorème de Banach à l'aide du Théorème du graphe fermé. En effet, si $T : E \rightarrow F$ est une application linéaire bijective entre espaces de Banach, alors

$$\Gamma(T) = \{(x, T(x)); x \in E\}, \quad \Gamma(T^{-1}) = \{(T(x), x); x \in E\},$$

et il est clair que $\Gamma(T)$ est fermé si et seulement si $\Gamma(T^{-1})$ l'est.

Le Théorème du graphe fermé s'utilise habituellement par l'un des deux critères qui suivent.

CRITÈRE PRATIQUE VII-30. Soit $T : E \rightarrow F$ une application linéaire entre espaces de Banach. On suppose que pour toute suite $(x_n)_{n \in \mathbb{N}}$ telle que $x_n \rightarrow x$ dans E et $T(x_n) \rightarrow y$ dans F , on a $y = T(x)$. Alors T est continue.

CRITÈRE PRATIQUE VII-31. Soient $E, \tilde{E}, F, \tilde{F}$ des espaces de Banach tels que $E \subset \tilde{E}$, $F \subset \tilde{F}$ avec injections continues. On se donne T une application linéaire continue de \tilde{E} dans \tilde{F} , telle que $T(E) \subset F$. Alors T définit par restriction une application linéaire continue de E dans F .

DÉMONSTRATION. Le Critère VII-30 découle du Théorème VII-28 et de la caractérisation séquentielle de la propriété de fermeture du graphe. Nous allons voir qu'il implique le Critère VII-31. Supposons en effet que $x_n \rightarrow x$ dans E et $T(x_n) \rightarrow y$ dans F ; alors $x_n \rightarrow x$ dans \tilde{E} et $T(x_n) \rightarrow y$ dans \tilde{F} , puisque $E \subset \tilde{E}$ et $F \subset \tilde{F}$ avec injections continues. La continuité de $T : \tilde{E} \rightarrow \tilde{F}$ implique $y = T(x)$, d'où la continuité de T vue comme une application $E \rightarrow F$. \square

Voici quelques commentaires pour conclure. Les théorèmes de continuité automatique sont puissants et tirent pleinement parti de la complétude. Il convient cependant d'y prendre garde : leur grande généralité est fatalement compensée par leur caractère non-constructif. En pratique, il est souvent important d'obtenir des *majorations explicites* sur les normes des opérateurs linéaires que l'on rencontre ; les théorèmes de Banach–Steinhaus ou de l'application ouverte ne peuvent en aucun cas mener à de telles estimations, ils se contenteront de prouver la finitude de ces normes. Si l'objet d'étude est une application linéaire particulière, il est donc recommandé d'éviter autant que possible le recours à ces théorèmes de continuité automatique, et de leur préférer une estimation “à la main”, en apparence moins élégante mais souvent beaucoup plus informative.

En fait, on peut envisager les théorèmes de continuité automatique de la même manière que les théorèmes de mesurabilité dans \mathbb{R}^n . Les applications que l'on peut construire “explicitement” sur \mathbb{R}^n sont Lebesgue-mesurables ; de même, les applications linéaires que l'on peut définir explicitement entre espaces de Banach sont toutes continues : si on les définit par des limites de suites, c'est une conséquence du théorème de Banach–Steinhaus ; si on les définit comme des bijections réciproques c'est une conséquence du théorème de Banach ; si on les définit à partir d'un graphe c'est une conséquence du théorème du graphe fermé.

Le parallèle peut être poussé plus loin : de même que l'on peut construire une axiomatique pour laquelle toutes les parties de \mathbb{R} sont mesurables, il est également possible de construire une axiomatique pour laquelle toutes les applications linéaires entre espaces de Banach sont continues !

VII-2. Régularité des espaces de Banach

Nous allons maintenant étudier les principales propriétés de “régularité” (définies soit en termes analytiques, soit en termes géométriques) utilisées pour décrire les espaces de Banach. Je me concentrerai sur trois propriétés particulièrement importantes : la **séparabilité**, la **différentiabilité** (de la norme), et la **convexité uniforme**. Les espaces de Hilbert séparables, qui sont en quelque sorte les plus réguliers de tous les espaces de Banach, possèdent toutes ces qualités et encore bien d'autres.

VII-2.1. Séparabilité.

DÉFINITION VII-32 (séparabilité). *On dit qu'un espace métrique X est séparable s'il existe une suite $(x_n)_{n \in \mathbb{N}}$ dense dans X .*

Autrement dit, un espace métrique est dit séparable s'il existe une suite $(x_n)_{n \in \mathbb{N}}$ telle que tout $x \in X$ est limite d'une suite extraite $(x_{n(k)})_{k \in \mathbb{N}}$.

La séparabilité s'hérite par restriction :

PROPOSITION VII-33. *Soient (X, d) un espace métrique séparable, et Y une partie quelconque de X ; alors Y , muni de la restriction de d , est un espace métrique séparable.*

DÉMONSTRATION. Pour tous $m, n \in \mathbb{N}$, si $B_{1/m}(x_n) \cap Y$ est non vide, on choisit un élément x_{mn} dans cet ensemble. La suite double $(x_{m,n})_{m,n \in \mathbb{N}}$ ainsi définie est alors dense dans Y . En effet, si $y \in Y$, pour tout m assez grand il existe $n \in \mathbb{N}$ tel que $d(x_n, y) \leq 1/m$, et alors $d(y, x_{m,n}) \leq 2/m$. \square

Dans le cas où X est un espace vectoriel, il existe une autre formulation équivalente de la propriété de séparabilité :

PROPOSITION VII-34 (critère de séparabilité). *Soit E un espace vectoriel normé. Alors E est séparable si et seulement si il existe dans E un sous-espace vectoriel dense de dimension au plus dénombrable.*

Rappelons qu'un espace F est de dimension dénombrable s'il existe une base algébrique dénombrable, i.e. une famille $(e_k)_{k \in \mathbb{N}}$ telle que tout $x \in F$ puisse s'écrire de manière unique comme combinaison linéaire *finie* des e_k .

EXEMPLE VII-35. L'espace des polynômes à coefficients réels est un espace vectoriel de dimension dénombrable. (On ne parle pas de norme ici.)

DÉMONSTRATION DE LA PROPOSITION VII-34. Supposons qu'il existe un sous-espace F dense dans E et admettant une base dénombrable $(e_k)_{k \in \mathbb{N}}$. La famille $D = \{\sum_{k=1}^N \alpha_k e_k; \alpha_k \in \mathbb{Q}; N \in \mathbb{N}\}$ est dénombrable car D est en bijection avec l'union des \mathbb{Q}^m ($m \in \mathbb{N}$), qui est dénombrable. D'autre part D est dense : si $x = \sum_{k=1}^N \alpha_k e_k$, on peut approcher chaque α_k par un rationnel q_k avec une erreur au plus ε , de sorte que $\|x - \sum q_k e_k\| \leq (N \sup \|e_k\|)\varepsilon$. L'espace E admet donc un sous-ensemble dénombrable dense.

Réciproquement, supposons E séparable et soit $(x_k)_{k \in \mathbb{N}}$ une suite dense dans E . Par récurrence, on choisit parmi les x_k une famille libre $(x_k)_{k \in I}$ qui engendre le même espace vectoriel F que les $(x_k)_{k \in \mathbb{N}}$ (si x_k n'appartient pas à l'espace vectoriel engendré par les éléments déjà sélectionnés parmi x_1, \dots, x_{k-1} , on le retient ; sinon on l'élimine). On vérifie facilement que cette famille constitue une base de F , qui est donc de dimension au plus dénombrable. Comme cet espace contient une suite dense, il est lui-même dense. \square

Le théorème suivant examine la séparabilité des espaces fonctionnels déjà rencontrés.

THÉORÈME VII-36 (séparabilité des espaces usuels). *Soit X un espace métrique séparable. Alors*

(i) *Si μ est une mesure de Borel régulière et σ -finie sur X , alors $L^p(X, \mu)$ est séparable pour $1 \leq p < \infty$;*

(ii) Si X est localement compact, alors $C_c(X)$ est dense dans $C_0(X)$; si en outre X est σ -compact, alors les espaces $C_c(X)$ et $C_0(X)$ sont séparables ;

(iii) Les espaces $L^\infty(X, \mu)$, $M(X)$, $C_b(X)$ ne sont en général pas séparables. Ainsi,

- $C_b(X)$ est non séparable dès que X est localement compact, σ -compact mais non compact ;

- $M(X)$ est non séparable dès que X est non dénombrable ;

- $L^\infty(X, \mu)$ est non séparable dès que X contient une infinité d'ensembles disjoints de mesure positive.

EXEMPLES VII-37. Si Ω est un ouvert de \mathbb{R}^n muni de la mesure de Lebesgue, les espaces $L^p(\Omega)$, $1 \leq p < \infty$, sont séparables, de même que $\ell^p = L^p(\mathbb{N})$. En revanche, les espaces $C_b(\mathbb{R}^n)$, $\ell^\infty = L^\infty(\mathbb{N})$, $L^\infty(\Omega)$ ne sont pas séparables.

REMARQUE VII-38. Ces exemples d'espaces non séparables ne sont pas si inquiétants que l'on pourrait le craindre : comme on le verra par la suite, sous certaines hypothèses de régularité sur X ces espaces peuvent être vus comme *duals* d'espaces séparables, ce qui leur confère de bonnes propriétés.

DÉMONSTRATION DU THÉORÈME VII-36. 1. L'espace vectoriel des fonctions simples (combinaisons linéaires finies de fonctions indicatrices) est dense dans $L^p(X)$ au vu des Théorèmes II-29 et VI-15. Pour démontrer l'énoncé (i), il suffit donc d'exhiber un sous-espace dénombrable dense dans l'espace des fonctions simples, et pour cela il suffit bien sûr de trouver une famille dense dans l'espace des fonctions indicatrices d'ensembles mesurables de mesure finie. L'identité

$$\|1_A - 1_B\|_{L^p(\mu)} = \left(\mu[A \setminus B] + \mu[B \setminus A] \right)^{1/p}$$

montre qu'il suffit d'exhiber une famille dénombrable \mathcal{B} qui soit "dense dans la tribu borélienne", au sens où pour tout borélien A on puisse trouver une famille $(B_k)_{k \in \mathbb{N}}$ d'éléments de \mathcal{B} tels que $\mu[B_k \setminus A] + \mu[A \setminus B_k] \rightarrow 0$.

Une telle famille \mathcal{B} est formée par l'ensemble des **unions finies de boules** $B_r(y)$, où y varie dans une partie dénombrable dense D de X , et r dans l'ensemble des nombres rationnels positifs. En effet, soit A un ensemble mesurable, et soit $\varepsilon > 0$. Par régularité, on peut trouver un ensemble ouvert O contenant A tel que $\mu[O \setminus A] \leq \varepsilon$. Par le Théorème I-32 (i), on peut écrire $O = \cup_{k \in \mathbb{N}} B_{r_k}(y_k)$, où les r_k sont rationnels et les y_k appartiennent à D . Par σ -additivité, il existe N tel que

$$\mu \left[O \setminus \left(\cup_{1 \leq k \leq N} B_{r_k}(y_k) \right) \right] \leq \varepsilon.$$

En particulier, il existe $B \in \mathcal{B}$ tel que $\mu[O \setminus B] \leq \varepsilon$. On a alors

$$\mu[A \setminus B] + \mu[B \setminus A] \leq \mu[O \setminus B] + \mu[O \setminus A] \leq 2\varepsilon,$$

ce qui conclut l'argument.

2. Passons maintenant à l'énoncé (ii). Soit $f \in C_0(X)$; pour tout $\varepsilon > 0$ on peut trouver un compact K de X telle que

$$\|f - f1_K\|_{L^\infty(X)} \leq \varepsilon.$$

Par le lemme d'Urysohn, il existe une fonction φ , à support compact et à valeurs dans $[0, 1]$, identiquement égale à 1 sur K . En particulier, la fonction $f\varphi$, continue

et à support compact, vérifie

$$\|f - f\varphi\|_{L^\infty(X)} \leq \varepsilon.$$

L'espace $C_c(X)$ est donc bien dense dans $C_0(X)$.

3. Supposons maintenant que X est en outre σ -compact. Par le Théorème I-36, il existe une suite croissante $(K_n)_{n \in \mathbb{N}}$ de compacts, dont la réunion est égale à X , telle que chaque K_n soit inclus dans l'intérieur de K_{n+1} , et telle que tout compact K de X soit inclus dans l'un des K_n . L'espace $C_c(X)$ est donc la réunion croissante des espaces $C_c(K_n)$, et pour démontrer le résultat il suffit de traiter le cas où X est compact.

4. Supposons donc X compact. Soit \mathcal{B} définie comme dans la démonstration de (i). Pour tout $B \in \mathcal{B}$ et tout entier $\ell \geq 1$ on définit $B^{(\ell)}$ comme l'ensemble des $x \in X$ tels qu'il existe $y \in B$ avec $d(x, y) < \ell^{-1}$. L'ensemble $B^{(\ell)}$ est ouvert, puisque c'est l'union de toutes les boules ouvertes de rayon $1/\ell$ centrées en un élément de B . L'ensemble \overline{B} est fermé dans le compact X , c'est donc un compact. Par ailleurs, on a bien sûr $\overline{B} \subset B^{(\ell)}$; par le lemme d'Urysohn on peut donc introduire une fonction $\chi = \chi_{B, \ell}$ telle que

$$1_{\overline{B}} \leq \chi \leq 1_{B^{(\ell)}}.$$

En clair, χ est une approximation de la fonction indicatrice de B , qui s'annule sur une distance de l'ordre de $1/\ell$. L'ensemble de toutes les fonctions $\chi_{B, \ell}$ ($B \in \mathcal{B}$, $\ell \in \mathbb{N}$) est dénombrable puisque \mathcal{B} l'est.

Soit maintenant $f \in C(X)$; on veut montrer qu'il existe un élément de l'espace vectoriel engendré par les $\chi_{B, \ell}$ qui soit très proche de f au sens de la norme du supremum. En décomposant f en partie positive et partie négative, on se ramène au cas où f est à valeurs positives.

Soit alors m un entier positif. Étant continue sur un compact, f est uniformément continue, et il existe $\ell \in \mathbb{N}$ tel que

$$(49) \quad d(x, y) \leq \frac{3}{\ell} \implies |f(x) - f(y)| \leq 2^{-m}.$$

Pour tout $k \in \{0, \dots, M\}$, soit

$$E_k := \{x \in X; k 2^{-m} \leq f(x)\}.$$

L'ensemble E_k est compact, on peut donc le recouvrir par un nombre fini de boules de rayon $1/\ell$ centrées en des éléments de E_k ; soit B_k l'union de ces boules, et $\chi_k := \chi_{B_k, \ell}$. Tout élément de B_k est situé à une distance au plus $1/\ell$ de E_k , et donc la fonction χ_k s'annule en tout point x tel que $d(x, E_k) > 2/\ell$.

Soit $x \in E_k$, alors $x \in B_k$, et donc $\chi_k(x) = 1$. Par ailleurs, soit x tel que $f(x) < (k-1) 2^{-m}$: alors, pour tout $y \in E_k$ on a, par (49), $d(x, y) > 3/\ell$. Il s'ensuit que $d(x, E_k) \geq 3/\ell$, et donc $\chi_k(x) = 0$. Résumons: χ_k est identiquement égale à 1 là où $f \geq k 2^{-m}$ et identiquement nulle là où $f < (k-1) 2^{-m}$.

Soient maintenant $x \in X$, et k_x un entier tel que $k_x 2^{-m} \leq f(x) < (k_x + 1) 2^{-m}$. Pour tout $k \leq k_x$, on a $\chi_k(x) = 1$, et pour tout $k \geq k_x + 2$ on a $\chi_k(x) = 0$. Il s'ensuit que $\sum_{k=0}^M \chi_k(x)$ est comprise entre k_x et $k_x + 2$, d'où

$$\left| f(x) - \sum_{k=0}^M (2^{-m} \chi_k(x)) \right| \leq 2 \times 2^{-m}.$$

Cette majoration est indépendante de x . La conclusion en découle, puisque $\sum_{k=0}^M 2^{-m} \chi_k$ appartient effectivement à l'espace vectoriel engendré par les $\chi_{B,\ell}$.

5. Pour montrer qu'un espace E n'est pas séparable, il suffit d'exhiber une famille non dénombrable d'éléments de E qui sont tous à distance au moins 1 les uns des autres. Soit X un espace localement compact, σ -compact, non compact. En utilisant le Théorème I-36, on montre facilement que X est union croissante dénombrable de compacts K_n tels que K_{n+1} soit un voisinage de K_n , distinct de K_n (en effet, la suite de compacts fournie par ce théorème ne peut être stationnaire que si X est compact). Pour tout $n \in \mathbb{N}$, on peut construire grâce au lemme d'Urysohn, des fonctions f_n non nulles, à valeurs dans $[0, 1]$, telles que f_n soit à support compact dans $O_n = \text{Int}(K_n)$, et identiquement égale à 1 sur un compact non vide de O_n ; en particulier $\sup |f_n| = 1$. Pour tout $x \in X$, il existe au plus un $n \in \mathbb{N}$ tel que $f_n(x) \neq 0$; on peut donc définir, pour toute suite $(\varepsilon_n)_{n \in \mathbb{N}}$ à valeurs dans $\{0, 1\}$, la somme

$$f_\varepsilon := \sum_{n=0}^{\infty} \varepsilon_n f_n.$$

Si ε et ε' sont deux suites distinctes à valeurs dans $\{0, 1\}$, alors $\|f_\varepsilon - f_{\varepsilon'}\|_\infty = 1$. On conclut à la non-séparabilité de $C_b(X)$ en remarquant que l'ensemble des applications $\varepsilon : \mathbb{N} \rightarrow \{0, 1\}$ est non dénombrable.

6. Les autres cas évoqués dans l'énoncé (iii) se traitent de même. Si (X, μ) est un espace mesuré et $(B_k)_{k \in \mathbb{N}}$ une famille de boules disjointes de mesure positive, on peut répéter le même raisonnement que ci-dessus en remplaçant la fonction f_k par la fonction indicatrice de B_k . Enfin, si X est non dénombrable, la famille des $(\delta_x)_{x \in X}$ est une famille non dénombrable de $M(X)$; or $\|\delta_x - \delta_y\|_{VT} = 2$ dès que $x \neq y$. \square

REMARQUE VII-39. La démonstration du Théorème VII-36 fournit des sous-espaces denses assez explicites. Ainsi, elle montre que si X est un espace métrique séparable muni d'une mesure μ , régulière et σ -finie, alors l'espace vectoriel engendré par les fonctions indicatrices de boules ouvertes est dense dans $L^p(X, \mu)$; et l'on peut se limiter aux boules ouvertes de rayon rationnel, dont les centres appartiennent à un ensemble dense fixé a priori. A titre d'application, nous allons utiliser ce résultat pour démontrer le Théorème VI-20.

DÉMONSTRATION DU THÉORÈME VI-20. Il s'agit de montrer que l'espace vectoriel engendré par les produits tensoriels de fonctions $L^p(X_1, \mu_1)$ et $L^p(X_2, \mu_2)$ est dense dans $L^p(X_1 \times X_2, \mu_1 \otimes \mu_2)$. On munit l'espace produit $X = X_1 \times X_2$ de la distance $d((x_1, x_2), (y_1, y_2)) = \max(d_1(x_1, y_1), d_2(x_2, y_2))$, qui induit bien la topologie produit. Alors une boule ouverte B dans X s'écrit $B = B_1 \times B_2$, où B_1 est une boule ouverte dans X_1 et B_2 une boule ouverte dans X_2 . Il s'ensuit que $1_B = 1_{B_1} \otimes 1_{B_2}$ est un produit tensoriel. Comme l'espace vectoriel engendré par les fonctions indicatrices de boules ouvertes est dense dans $L^p(X)$, il en est de même de l'espace vectoriel engendré par les produits tensoriels. \square

REMARQUE VII-40. Les familles denses construites pour les besoins de la démonstration du Théorème VII-36 sont assez naturelles dans le cadre abstrait de ce théorème. Mais souvent, d'autres choix naturels pourront se présenter dans des cas particuliers. Par exemple, si Ω est un ouvert borné de \mathbb{R}^n , la famille des polynômes en n variables, à coefficients rationnels, restreints à $\overline{\Omega}$, est une famille dénombrable dense dans $C(\overline{\Omega})$ (théorème de Weierstrass). Cette même famille est également dense dans

$L^p(\Omega)$ puisque $C(\overline{\Omega})$ est dense dans $L^p(\Omega)$. De même, la famille des polynômes trigonométriques à coefficients rationnels est dense dans $C(\mathbb{T}^n)$ au vu du théorème de densité de Féjer (Cf. Chapitre ??).

Terminons cette sous-section avec un énoncé sur la séparabilité des espaces fonctionnels à valeurs vectorielles.

THÉORÈME VII-41 (Séparabilité des espaces usuels à valeurs vectorielles). *Soient X un espace métrique séparable et E un espace vectoriel normé séparable. Alors*

(i) *Si μ est une mesure de Borel régulière et σ -finie sur X , alors $L^p(X, \mu; E)$ est séparable pour $1 \leq p < \infty$;*

(ii) *Si X est localement compact et σ -compact, alors les espaces $C_c(X; E)$ et $C_0(X; E)$ sont séparables.*

DÉMONSTRATION. Ici je vais seulement démontrer l'énoncé (ii) qui est plus facile, et remettre la preuve de l'énoncé (i) à plus tard (**À PRÉCISER** – utiliser le théorème de désintégration?). On se ramène comme dans la preuve du Théorème VII-36 au cas où X est compact. On note d la distance sur X , $\|\cdot\|$ la norme de E , et on introduit une famille dénombrable dense $(e_k)_{k \in \mathbb{N}}$ dans E . Soit $f \in C(X; E)$, soit $\varepsilon > 0$ et soit $\delta > 0$ tel que

$$d(x, y) < \delta \implies \|f(x) - f(y)\| \leq \varepsilon.$$

Soit $(x_\ell)_{1 \leq \ell \leq L}$ une famille finie telle que tout $y \in X$ se trouve à une distance au plus δ de l'un au moins des x_ℓ . Les boules ouvertes $B(x_\ell, \delta)$ forment donc un recouvrement ouvert fini de X , et on peut introduire $(\chi_\ell)_{1 \leq \ell \leq L}$ une partition continue de l'unité subordonnée à ce recouvrement. Pour tout ℓ on choisit $e_{k(\ell)}$ tel que $\|f(x_\ell) - e_{k(\ell)}\| \leq \varepsilon$. On pose alors

$$g = \sum_{1 \leq \ell \leq L} \chi_\ell(x) e_{k(\ell)}.$$

Comme

$$f - g = \sum \chi_\ell e_{k(\ell)} - \left(\sum \chi_\ell \right) f,$$

on a

$$\sup_{y \in X} \|f(y) - g(y)\| \leq \sum_{1 \leq \ell \leq L} \chi_\ell \sup_{y \in B_\delta(x_\ell)} \|f(y) - e_{k(\ell)}\|.$$

Si $y \in B_\delta(x_\ell)$ alors $\|f(y) - e_{k(\ell)}\| \leq \|f(y) - f(x_\ell)\| + \|f(x_\ell) - e_{k(\ell)}\| \leq 2\varepsilon$. Il s'ensuit que

$$\sup_{y \in X} \|f(y) - g(y)\| \leq (2\varepsilon) \sum_{1 \leq \ell \leq L} \chi_\ell = 2\varepsilon.$$

En conclusion, la famille des sommes finies $\sum f_\ell e_k$, où $f_\ell \in C(X; \mathbb{R})$ et e_k appartient à une famille dense de fixée a priori dans E , est elle-même dense dans $C(X; E)$. En combinant ce résultat avec la séparabilité de $C(X; \mathbb{R})$ on conclut facilement à la séparabilité de $C(X; E)$. \square

REMARQUE VII-42. Ce que l'on a montré par l'argument précédent, c'est que le produit tensoriel $C(X) \otimes E$ (défini comme l'adhérence des sommes finies de produits tensoriels d'un élément de $C(X)$ par un élément de E) coïncide avec $C(X; E)$. On montrera plus tard que $L^p(X) \otimes E = L^p(X; E)$, ce qui permettra de conclure la preuve de l'énoncé (i).

VII-2.2. Différentiabilité de la norme. Nous allons maintenant nous intéresser à la forme de la **boule unité** $B_1(0)$. Cette boule peut être lisse ou pointue ; elle peut être “bien convexe” ou présenter des parties plates. Ces attributs peuvent être quantifiés par les propriétés de la norme. L’exemple à garder en tête est celui de la boule unité de $(\mathbb{R}^n, d_{\ell^p})$ (l’espace \mathbb{R}^n muni de la distance produit ℓ^p) : pour $p = 2$ on a une boule “ronde”, à la fois parfaitement régulière et parfaitement convexe ; pour $p = 1$ et $p = \infty$ la sphère présente des parties plates (défaut de stricte convexité) et des “coins” (défaut de différentiabilité) ; quand p est compris strictement entre 1 et ∞ , mais différent de 2, la norme est différentiable et strictement convexe, tout en étant “moins différentiable” et “moins convexe” que dans le cas $p = 2$.

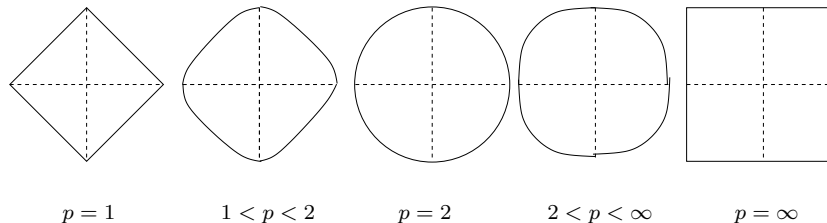


FIG. 1. Forme approximative de la boule ℓ^p en dimension 2, pour différentes valeurs de p

Dans cette sous-section, on s’intéresse à la différentiabilité de la norme. En dimension infinie, il convient de préciser le concept de différentiabilité.

DÉFINITION VII-43. Soit φ une application définie sur un ouvert O d’un espace vectoriel normé E , à valeurs dans un autre espace vectoriel normé F . On dit que φ est

- *Fréchet-différentiable (ou différentiable)* en $x \in O$, s’il existe une application linéaire continue de E dans F , notée $D\varphi(x)$ (ou $D_x\varphi$ ou $d\varphi(x)$ ou $d_x\varphi$), telle que

$$\lim_{h \rightarrow 0 \ (h \in E \setminus \{0\})} \frac{\|\varphi(x+h) - \varphi(x) - D\varphi(x) \cdot h\|_F}{\|h\|_E} = 0;$$

- *Gâteaux-différentiable* en $x \in O$, s’il existe une application linéaire continue de E dans F , notée $D\varphi(x)$ (ou $D_x\varphi$ ou $d\varphi(x)$ ou $d_x\varphi$), telle que

$$\forall e \in E, \quad \lim_{t \rightarrow 0 \ (t \in \mathbb{R} \setminus \{0\})} \frac{\|\varphi(x+te) - \varphi(x) - t D\varphi(x) \cdot e\|_F}{|t|} = 0.$$

L’application $D\varphi(x)$ est alors appelée *Fréchet-différentielle* ou *Gâteaux-différentielle* (ou tout simplement *différentielle*) de φ en x .

REMARQUE VII-44. La Gâteaux-différentiabilité est la différentiabilité “selon n’importe quelle direction”. Pour apprécier la différence entre les deux notions, on pourra réécrire la propriété de Fréchet-différentiabilité comme suit :

$$\lim_{t \rightarrow 0} \sup_{\|e\|_E=1} \frac{\|\varphi(x+te) - \varphi(x) - t D\varphi(x) \cdot e\|_F}{|t|} = 0;$$

et la Gâteaux-différentiabilité :

$$\|e\|_E = 1 \implies \lim_{t \rightarrow 0} \frac{\|\varphi(x+te) - \varphi(x) - t D\varphi(x) \cdot e\|_F}{|t|} = 0.$$

EXEMPLE VII-45. Même en dimension finie, les deux concepts diffèrent, comme le montre l'exemple suivant : soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ telle que $f(x, y) = y$ si $x = y$, et $f(x, y) = 0$ si $x \neq y$. (Peut-on construire un exemple discontinu ?)

Voici quelques propriétés à garder en tête, que je ne démontrerai pas.

PROPOSITION VII-46. Avec les notations de la Définition VII-43,

- La Fréchet-différentiabilité en x implique la Gâteaux-différentiabilité en x ;
- La Gâteaux-différentielle $D\varphi(x)$, si elle existe, est unique ; en particulier, si φ est Fréchet-différentiable en x , la Fréchet-différentielle et la Gâteaux-différentielle coïncident ;
- Si E est de dimension finie et φ est Gâteaux-différentiable sur un ouvert $O \subset E$, de Gâteaux-différentielle continue, alors φ est également Fréchet-différentiable.
- Soient $\varphi : O \rightarrow O'$ et $O' \rightarrow G$, où O (resp. O') est un ouvert d'un espace vectoriel normé E (resp. F). On suppose que φ est Fréchet-différentiable (resp. Gâteaux-différentiable) en x , et ψ est Fréchet-différentiable en $\varphi(x)$. Alors $\varphi \circ \psi$ est Fréchet-différentiable (resp. Gâteaux-différentiable) en x . En abrégé :

Fréchet \circ Fréchet est Fréchet ; Fréchet \circ Gâteaux est Gâteaux.

En revanche, la composition de deux applications Gâteaux-différentiables n'est pas a priori Gâteaux-différentiable.

Appliquons maintenant ces notions aux espaces de Lebesgue :

THÉORÈME VII-47. Soit (X, μ) un espace mesuré ; on note N_p la norme dans $L^p(X, \mu)$. Alors,

(i) Pour $1 < p < \infty$, la norme N_p est Fréchet-différentiable en tout $f \in L^p(X, \mu) \setminus \{0\}$, et sa différentielle est définie par la formule

$$\forall g \in L^p(X, \mu), \quad DN_p(f) \cdot g = \frac{\int_X |f|^{p-2} f g \, d\mu}{\|f\|_{L^p}^{p-1}};$$

(ii) Ni la norme N_1 , ni la norme N_∞ ne sont en général Gâteaux-différentiables hors de l'origine.

REMARQUE VII-48. Avant de lire la preuve de la Fréchet-différentiabilité des normes L^p ($1 < p < \infty$), le lecteur peut essayer de prouver la Gâteaux-différentiabilité, qui est plus simple (noter que l'on peut tirer parti de la convexité de $t \rightarrow |t|^p$).

Le lemme suivant sera utile dans la preuve du Théorème VII-47 :

LEMME VII-49. Soit $p \in [1, +\infty[$, alors il existe une constante C_p telle que pour tous nombres réels x et y ,

$$(50) \quad 0 \leq |x + y|^p - |x|^p - p|x|^{p-2}xy \leq C_p(|x| + |y|)^{\max(p-2, 0)}|y|^{\min(p, 2)}.$$

DÉMONSTRATION DU LEMME VII-49. Si $x = 0$ ou $y = 0$, l'inégalité est triviale ; on supposera donc $x \neq 0$, $y \neq 0$. La fonction $f_p : x \mapsto |x|^p$ est convexe sur \mathbb{R} et différentiable en $x \neq 0$, de dérivée $p|x|^{p-2}x$. L'inégalité de convexité

$$f_p(x + y) \geq f_p(x) + yf'_p(x)$$

est donc précisément l'inégalité de gauche dans (50).

On se concentre maintenant sur l'inégalité de droite. Quitte à remplacer x par $-x$ et y par $-y$, on peut supposer que x est positif. On a alors $|x + y|^p \leq (x + |y|)^p$, il suffit donc de démontrer l'inégalité dans le cas où y aussi est positif. En divisant les deux membres de l'inégalité par y^p , et en posant $t := y/x$, on voit que l'inégalité est équivalente à

$$(51) \quad \forall t > 0, \quad (1+t)^p - 1 - pt \leq C_p(1+t)^{\max(p-2,0)} t^{\min(p,2)}.$$

(C'est un exemple simple d'**argument d'homogénéité**.)

Le cas $p = 1$ est trivial, on supposera donc $p > 1$.

La formule de Taylor avec reste intégral, appliquée à la fonction $\varphi(t) = (1+t)^p$ à l'ordre 2, s'écrit

$$(1+t)^p = 1 + pt + p(p-2)t^2 \int_0^1 (1-s)(1+st)^{p-2} ds.$$

Il suffit donc de prouver l'existence d'une constante $C'_p > 0$ telle que

$$\int_0^1 (1-s)(1+st)^{p-2} ds \leq C(1+t)^{\max(p-2,0)} t^{\min(p-2,0)}.$$

Si $p \geq 2$, c'est évident : u^{p-2} est une fonction croissante de u , donc

$$\int_0^1 (1-s)(1+st)^{p-2} ds \leq (1+t)^{p-2} \int_0^1 (1-s) ds = \frac{(1+t)^{p-2}}{2}.$$

Si $p < 2$, au contraire u^{p-2} est une fonction décroissante de u , d'où

$$\int_0^1 (1-s)(1+st)^{p-2} ds \leq \int_0^1 (1-s)(st)^{p-2} ds = \left(\int_0^1 (1-s)s^{p-2} ds \right) t^{p-2},$$

et l'intégrale entre parenthèses est convergente puisque $p-2 > -1$. La conclusion en découle. \square

DÉMONSTRATION DU THÉORÈME VII-47. L'application $\psi_p : z \mapsto z^{1/p}$ est différentiable sur $]0, +\infty[$; pour prouver la Fréchet-différentiabilité de N_p en $f \neq 0$, il suffit donc de prouver la Fréchet-différentiabilité de N_p^p en f ; la formule finale pour DN_p s'obtiendra par composition des différentielles de N_p^p et de ψ_p .

Soit $g \in L^p(X, \mu)$, de norme 1 ; on doit établir

$$\int_X |f + tg|^p d\mu - \int_X |f|^p d\mu - tp \int_X |f|^{p-2} fg d\mu = o(t),$$

où le $o(t)$ est uniforme en g (sans cette uniformité on ne prouverait que la Gâteaux-différentiabilité).

Grâce au Lemme VII-49, on peut écrire

$$(52) \quad \int_X |f + tg|^p d\mu - \int_X |f|^p d\mu - tp \int_X |f|^{p-2} fg d\mu \\ = \int_X (|f + tg|^p - |f|^p - tp|f|^{p-2} fg) d\mu \leq C_p t^{\min(p,2)} \int_X (|f| + t|g|)^{\max(p-2,0)} |g|^{\min(p,2)} d\mu.$$

Quand $p \geq 2$, on majore l'intégrale du membre de droite par inégalité de Hölder :

$$\begin{aligned} \int_X (|f| + t|g|)^{p-2} |g|^2 d\mu &\leq \left(\int_X (|f| + t|g|)^p d\mu \right)^{\frac{p-2}{p}} \left(\int_X |g|^p d\mu \right)^{\frac{2}{p}} \\ &= \left\| |f| + t|g| \right\|_{L^p}^{p-2} \|g\|_{L^p}^2, \end{aligned}$$

que l'on majore par $(\|f\|_{L^p} + t\|g\|_{L^p})^{p-2} \|g\|_{L^p}^2 \leq (\|f\|_{L^p} + t)^{p-2}$. Le membre de droite de (52) est donc en $O(t^2)$ quand $t \rightarrow 0$, et a fortiori c'est un $o(t)$. Supposons maintenant que $1 < p < 2$, le membre de droite de (52) se réduit alors à $C_p t^p \|g\|_{L^p}^p = C_p t^p$, qui est également un $o(t)$. Ceci achève la preuve de (i).

Pour prouver l'énoncé (ii), il est facile de construire des contre-exemples : considérons l'espace à deux points $X = \{0, 1\}$, muni de la mesure de comptage. Toute fonction sur X s'identifie à un couple $x = (x_0, x_1)$ de nombres réels, et la norme N_p à la norme ℓ^p sur \mathbb{R}^2 ; en particulier, $\|x\|_1 = (|x_0| + |x_1|)$, et $\|x\|_\infty = \max(|x_0|, |x_1|)$. La première n'est pas différentiable en x dès que $x_0 = 0$ ou $x_1 = 0$, tandis que la seconde n'est pas différentiable en x dès que $x_0 = x_1$. \square

VII-2.3. Uniforme convexité. Si la différentiabilité de la norme quantifie l'idée que la boule unité est "lisse", l'uniforme convexité en revanche quantifiera l'idée qu'elle est "bien ronde".

DÉFINITION VII-50 (uniforme convexité). Soit $(E, \|\cdot\|)$ un espace vectoriel normé. On dit que E est uniformément convexe si pour tout $\varepsilon > 0$ on peut trouver $\kappa > 0$ tel que

$$(53) \quad \left[\|x\| \leq 1, \|y\| \leq 1, \|x - y\| \geq 2\varepsilon \right] \implies \left\| \frac{x + y}{2} \right\| \leq 1 - \kappa.$$

Cette définition mérite quelques explications. Tout d'abord, la boule unité d'un espace vectoriel normé est toujours convexe : comme conséquence de l'inégalité triangulaire, on a $\|(x + y)/2\| \leq 1$ dès que $\|x\| \leq 1$ et $\|y\| \leq 1$. Si la sphère unité ne contient pas de partie plate, ou plus précisément pas de segment, cela signifie que l'on ne peut pas avoir simultanément $\|x\| = 1$, $\|y\| = 1$, $x \neq y$ et $\|(x + y)/2\| = 1$; autrement dit,

$$\left[\|x\| \leq 1, \|y\| \leq 1, x \neq y \right] \implies \left\| \frac{x + y}{2} \right\| < 1.$$

L'inégalité (53) est une version *quantitative* de cet énoncé : si l'on a une borne inférieure sur l'écart entre x et y , alors on peut minorer la distance entre $(x + y)/2$ et la sphère unité.

EXEMPLE VII-51. Il est facile de se convaincre visuellement que $\ell^p(\{0, 1\}) \simeq (\mathbb{R}^2, d_{\ell^p})$ est uniformément convexe pour $1 < p < \infty$, mais pas pour $p = 1$ ou $p = \infty$.

REMARQUE VII-52. Soient E un espace vectoriel normé et $\Phi : E \rightarrow \mathbb{R} \cup \{+\infty\}$. On dit que Φ est uniformément convexe s'il existe $K > 0$ telle que pour tous $x, y \in E$ et $\lambda \in [0, 1]$,

$$\Phi(x) + \Phi(y) - \Phi(\lambda x + (1 - \lambda)y) \geq \frac{K\lambda(1 - \lambda)}{2} \|x - y\|^2.$$

Il est tentant de relier l'uniforme convexité d'un espace vectoriel normé à l'uniforme convexité de sa norme. Mais la terminologie est trompeuse : une norme n'est jamais

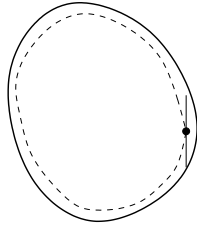


FIG. 2. Signification de l'uniforme convexité : le milieu d'un segment (ici un point épais) ne peut jamais s'approcher du bord de la boule.

uniformément convexe, ni même strictement convexe, comme on peut s'en convaincre en choisissant $y = tx$ ($t > 0$) dans l'inégalité ci-dessus.

Si $(E, \|\cdot\|)$ est un espace uniformément convexe, on peut définir son **module de convexité** κ par la formule

$$\kappa(\varepsilon) = \inf \left\{ 1 - \left\| \frac{x+y}{2} \right\| ; \|x\| \leq 1, \|y\| \leq 1, \|x-y\| \geq 2\varepsilon \right\}.$$

En d'autres termes, κ est la plus grande fonction croissante vérifiant

$$[\|x\| \leq 1, \|y\| \leq 1] \implies \left\| \frac{x+y}{2} \right\| \leq 1 - \kappa\left(\frac{\|x-y\|}{2}\right).$$

Le comportement de κ près de 0 indique "à quel point l'espace est uniformément convexe".

THÉORÈME VII-53 (uniforme convexité des espaces de Lebesgue). *Soit (X, μ) un espace mesuré. Alors*

(i) *Pour tout $p \in]1, \infty[$ l'espace $L^p(X, \mu)$ est uniformément convexe, avec un module de convexité*

$$\kappa_p(\varepsilon) \geq K_p \varepsilon^{\max(2,p)},$$

où $K_p > 0$ ne dépend que de p ;

(ii) *En général ni $L^1(X, \mu)$ ni $L^\infty(X, \mu)$ ne sont uniformément convexes.*

Ce théorème découlera des **inégalités de Hanner**, qui renforcent en quelque sorte l'inégalité de Minkowski :

THÉORÈME VII-54 (inégalités de Hanner). *Soit (X, μ) un espace mesuré, et soient f, g deux fonctions dans $L^p(X, \mu)$, $1 \leq p < \infty$. Alors*

$$\begin{aligned} p \leq 2 &\implies \left(\frac{\|f\|_{L^p} + \|g\|_{L^p}}{2} \right)^p + \left| \frac{\|f\|_{L^p} - \|g\|_{L^p}}{2} \right|^p \leq \left\| \frac{f+g}{2} \right\|_{L^p}^p + \left\| \frac{f-g}{2} \right\|_{L^p}^p ; \\ p \geq 2 &\implies \left(\frac{\|f\|_{L^p} + \|g\|_{L^p}}{2} \right)^p + \left| \frac{\|f\|_{L^p} - \|g\|_{L^p}}{2} \right|^p \geq \left\| \frac{f+g}{2} \right\|_{L^p}^p + \left\| \frac{f-g}{2} \right\|_{L^p}^p . \end{aligned}$$

La preuve des inégalités de Hanner s'appuiera elle-même sur le lemme suivant.

LEMME VII-55. *La fonction*

$$\Phi : (x, y) \longmapsto (x^{1/p} + y^{1/p})^p + |x^{1/p} - y^{1/p}|^p,$$

définie sur $\mathbb{R}_+ \times \mathbb{R}_+$, est homogène de degré 1, convexe pour $p \leq 2$, et concave pour $p \geq 2$.

DÉMONSTRATION DU LEMME VII-55. Soit, pour $x \geq 0$,

$$\varphi(x) := \Phi(x, 1) = (x^{1/p} + 1)^p + |x^{1/p} - 1|^p.$$

Par calcul direct, pour tout $x > 0$,

$$\varphi''(x) = - \left(\frac{p-1}{p} \right) x^{\frac{1}{p}-2} \left[(x^{1/p} + 1)^{p-2} - |x^{1/p} - 1|^{p-2} \right].$$

Comme x est positif, on a toujours $(x^{1/p} + 1) \geq |x^{1/p} - 1|$, donc $(x^{1/p} + 1)^{p-2}$ est toujours plus grand que $|x^{1/p} - 1|^{p-2}$ si $p \geq 2$, et toujours plus petit si $p \leq 2$. On en déduit que φ est convexe pour $p \geq 2$, et concave pour $p \leq 2$. Par la Remarque III-76, la fonction $\Phi(x, y) = y \varphi(x/y)$ est homogène et convexe (pour $p \leq 2$) ou concave (pour $p \geq 2$) sur $\mathbb{R}_+ \times \mathbb{R}_+$. \square

DÉMONSTRATION DU THÉORÈME VII-54. On se concentre sur le cas $p \leq 2$, la démonstration étant rigoureusement similaire dans le cas $p \geq 2$. Appliquons l'inégalité de Jensen, sous la forme du Théorème III-74, avec la mesure μ , la fonction convexe homogène Φ , et la fonction $F : X \rightarrow \mathbb{R}_+ \times \mathbb{R}_+$ définie par $F(x) = (|f(x)|^p, |g(x)|^p)$: on trouve

$$\Phi \left(\int_X |f|^p d\mu, \int_X |g|^p d\mu \right) \leq \int_X \Phi(|f|^p, |g|^p) d\mu.$$

Le membre de gauche vaut alors $(\|f\|_{L^p} + \|g\|_{L^p})^p + \|\|f\|_{L^p} - \|g\|_{L^p}\|^p$, ce qui est bien, à un facteur 2^p près, le membre de gauche de l'inégalité de Hanner. Pour calculer le membre de droite, on remarque que $\Phi(|f|^p, |g|^p) = (|f| + |g|)^p + \||f| - |g|\|^p$, et en considérant séparément le cas où f et g ont même signe et celui où ils ont des signes opposés, on constate que c'est également $|f + g|^p + |f - g|^p$; en intégrant par rapport à μ , on retrouve donc le membre de droite de l'inégalité de Hanner, toujours à un facteur 2^p près. \square

DÉMONSTRATION DU THÉORÈME VII-53. 1. Commençons par le cas où $1 < p \leq 2$. En substituant f et g dans l'inégalité de Hanner par $(f + g)/2$ et $|f - g|/2$, et en utilisant la notation $\|\cdot\|$ pour la norme L^p , on obtient

$$(54) \quad \frac{1}{2} \left(\left\| \frac{f+g}{2} \right\| + \left\| \frac{f-g}{2} \right\| \right)^p + \frac{1}{2} \left| \left\| \frac{f+g}{2} \right\| - \left\| \frac{f-g}{2} \right\| \right|^p \leq \frac{1}{2} (\|f\|^p + \|g\|^p),$$

Si maintenant on suppose $\|f\| \leq 1$, $\|g\| \leq 1$, et qu'on note $m := (f + g)/2$, on trouve

$$(55) \quad \frac{1}{2} \left[\left(\|m\| + \frac{\|f-g\|}{2} \right)^p + \left| \|m\| - \frac{\|f-g\|}{2} \right|^p \right] \leq 1.$$

Admettons provisoirement l'inégalité

$$(56) \quad a \geq b \geq 0 \implies \frac{(a+b)^p + (a-b)^p}{2} \geq a^p + p(p-1)a^{p-2}b^2;$$

en combinant (56) avec (55), on obtient

$$\|m\|^p + \frac{p(p-1)}{4} \|m\|^{p-2} \|f-g\|^2 \leq 1.$$

En multipliant par $\|m\|^{2-p}$ on trouve $\|m\|^2 + (p(p-1)/4) \|f-g\|^2 \leq \|m\|^{2-p} \leq 1$, d'où

$$\|m\| \leq \sqrt{1 - \frac{p(p-1)}{4} \|f-g\|^2} \leq 1 - \frac{p(p-1)}{8} \|f-g\|^2.$$

(La dernière inégalité provient de la concavité de la fonction $t \rightarrow \sqrt{1-t}$; on majore cette fonction par sa tangente en 0.)

Reste à prouver (56). En appliquant la formule de Taylor avec reste intégral à la fonction $\psi(t) = t^p$, on trouve

$$\begin{aligned}\varphi(a+b) + \varphi(a-b) &= 2\varphi(a) + b^2 \int_0^1 (1-t)[\varphi''(a+tb) + \varphi''(a-tb)] dt \\ &= 2a^p + p(p-1)b^2 \int_0^1 (1-t)[(a+tb)^{p-2} + (a-tb)^{p-2}] dt.\end{aligned}$$

Comme p est plus petit que 2, la fonction $z \mapsto z^{p-2}$ est convexe sur $]0, +\infty[$, et donc, dès que $a+b \geq a-b \geq 0$ on a

$$\frac{(a+tb)^{p-2} + (a-tb)^{p-2}}{2} \geq a^{p-2}.$$

L'inégalité (56) en découle.

2. Considérons maintenant le cas où $p \geq 2$. Soit φ la fonction introduite dans la preuve du Lemme VII-55. Comme $p \geq 2$, φ est concave, et donc majorée par sa tangente en $x = 1$, ce qui peut se réécrire

$$\left(\frac{x^{1/p} + 1}{2}\right)^p + \left|\frac{x^{1/p} - 1}{2}\right|^p \leq \frac{x+1}{2}.$$

Par homogénéité, pour tous nombres positifs a et b , on a

$$(57) \quad \left(\frac{a^{1/p} + b^{1/p}}{2}\right)^p + \left|\frac{a^{1/p} - b^{1/p}}{2}\right|^p \leq \frac{a+b}{2}.$$

L'inégalité de Hanner implique donc

$$(58) \quad \left\|\frac{f+g}{2}\right\|^p + \left\|\frac{f-g}{2}\right\|^p \leq \frac{1}{2}(\|f\|^p + \|g\|^p).$$

Supposons maintenant $\|f\| \leq 1$, $\|g\| \leq 1$, en notant toujours $m := (f+g)/2$ on a

$$\|m\|^p + \left\|\frac{f-g}{2}\right\|^p \leq 1.$$

La fonction $t \mapsto (1-t)^{1/p}$, étant concave, se majore par sa tangente en 1, la fonction $1-t/p$; on en déduit

$$\|m\| \leq \left(1 - \left\|\frac{f-g}{2}\right\|^p\right)^{1/p} \leq 1 - \frac{1}{p} \left(\frac{\|f-g\|}{2}\right)^p,$$

ce qui implique bien une minoration du module de convexité en ε^p/p .

3. Terminons en considérant les cas $p = 1$, $p = \infty$ mentionnés dans l'énoncé (ii). En général, on peut trouver deux parties mesurables disjointes A et B de mesure strictement positive. On considère alors les fonctions $f := 1_A/\mu[A]$, $g := 1_B/\mu[B]$: $\|f\|_{L^1} = \|g\|_{L^1} = 1$ et $\|(f+g)/2\|_{L^1} = 1$; il s'ensuit que L^1 n'est pas uniformément convexe. Si on considère les fonctions $f := 1_A$, $g := 1_A + 1_B$, alors $\|f\|_{L^\infty} = \|g\|_{L^\infty} = 1$, et $\|(f+g)/2\|_{L^\infty} = 1$, ce qui montre que L^∞ n'est pas uniformément convexe non plus. \square

REMARQUE VII-56. Si l'on ne peut pas trouver de parties disjointes A et B de mesure strictement positive, alors toute fonction mesurable est égale à une constante μ -presque partout (si f prenait deux valeurs distinctes avec des mesures strictement positives, on pourrait construire, en passant aux images réciproques, deux ensembles disjoints de mesure strictement positive). Si μ ne prend pas ses valeurs dans $\{0, +\infty\}$, l'espace $L^p(\mu)$ s'identifie alors à une droite, et la boule unité de $L^p(\mu)$ à un segment, qui ne saurait être uniformément convexe.

REMARQUE VII-57. L'inégalité (58) est l'**inégalité de Clarkson**; on l'a obtenue comme conséquence de l'inégalité de Hanner, mais on peut aussi la démontrer directement [Brézis, p. 60] à partir de l'inégalité élémentaire

$$a^p + b^p \leq (a^2 + b^2)^{p/2}.$$

Cette inégalité ne couvre que le cas où $p \geq 2$; pour $p \leq 2$ il existe une "deuxième inégalité de Clarkson", nettement plus délicate :

$$(59) \quad 1 < p \leq 2 \implies \left\| \frac{f+g}{2} \right\|_{L^p}^{p'} + \left\| \frac{f-g}{2} \right\|_{L^p}^{p'} \leq \frac{1}{2} \left(\|f\|_{L^p}^p + \|g\|_{L^p}^p \right)^{\frac{p'}{p}}.$$

Cette inégalité peut aussi se déduire de l'inégalité de Hanner, sous la forme (54) : pour s'en convaincre, il suffit de vérifier la validité de l'inégalité

$$(60) \quad (a^{p'} + b^{p'})^{\frac{p}{p'}} \leq \frac{(a+b)^p + |a-b|^p}{2},$$

pour tous $a, b \geq 0$. Si l'on introduit $x := a+b$, $y := a-b$, on peut réécrire (60) sous la forme

$$(61) \quad \left[\left(\frac{x+y}{2} \right)^{p'} + \left| \frac{x-y}{2} \right|^{p'} \right]^{1/p'} \leq \left(\frac{x^p + y^p}{2} \right)^{1/p}.$$

Pour vérifier (61), on va utiliser un argument élémentaire d'interpolation complexe : soit $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ l'opérateur linéaire

$$T : (x, y) \longmapsto \left(\frac{x+y}{2}, \frac{x-y}{2} \right).$$

On identifie \mathbb{R}^2 à l'espace des fonctions mesurables sur l'espace à deux points $Z = \{0, 1\}$. L'inégalité (61) exprime le fait que T a pour norme au plus 1 de $L^p(Z, \mu)$ dans $L^{p'}(Z, \nu)$, avec $\mu = (\delta_0 + \delta_1)/2$ et $\nu = \delta_0 + \delta_1$. Par le Théorème VI-33 (de Riesz–Thorin), il suffit de vérifier que la norme de T est majorée par 1 aussi bien $L^2(Z, \mu) \rightarrow L^2(Z, \nu)$ que $L^1(Z, \mu) \rightarrow L^\infty(Z, \nu)$; autrement dit, que

$$\sqrt{\left(\frac{x+y}{2} \right)^2 + \left(\frac{x-y}{2} \right)^2} \leq \sqrt{\frac{x^2 + y^2}{2}};$$

et

$$\max \left(\frac{|x+y|}{2}, \frac{|x-y|}{2} \right) \leq \frac{|x| + |y|}{2}.$$

Ces deux inégalités sont évidemment vraies (ce sont des égalités!), ce qui achève la preuve de (59).

REMARQUE VII-58. Les inégalités de Hanner (comme celles de Clarkson) se réduisent, quand $p = 2$, à l'**identité du parallélogramme** :

$$\|f + g\|_{L^2}^2 + \|f - g\|_{L^2}^2 = 2(\|f\|_{L^2}^2 + \|g\|_{L^2}^2);$$

qui est caractéristique des espaces de Hilbert. Par ailleurs, dans le cas $p = 1$, l'inégalité de Hanner se ramène à l'inégalité triangulaire.

REMARQUE VII-59. Dans le cas où $p \leq 2$, on peut obtenir une estimation plus précise du module de convexité en combinant l'inégalité de Hanner avec une inégalité subtile due à Gross, que nous ne démontrerons pas (voir les références dans [Ball-Carlen-Lieb]) : pour tous nombres réels a et b ,

$$\left(\frac{|a + b|^p + |a - b|^p}{2} \right)^{1/p} \geq (a^2 + (p - 1)b^2)^{1/2}.$$

On trouve alors que le module de convexité $\kappa_p(\varepsilon)$ pour $1 < p \leq 2$, est au moins $(p - 1)\varepsilon^2/2$, ce qui est optimal. En fait, sauf dans des cas “pathologiques”, on a exactement

$$\kappa_p(\varepsilon) = \begin{cases} \left(\frac{p-1}{2} \right) \varepsilon^2 & \text{si } 1 \leq p \leq 2 \\ \frac{\varepsilon^p}{p} & \text{si } p \geq 2. \end{cases}$$

On note que la constante se dégrade quand $p \rightarrow 1$, ce qui est normal puisque L^1 n'est pas uniformément convexe.

REMARQUE VII-60. Il existe une notion “duale” du module de convexité, appelée “module de régularité (uniforme)”, qui quantifie à quel degré la boule unité “ne présente pas de pointes”. Les inégalités de Hanner permettent encore de calculer ces modules de régularité ; pour plus d'informations on renvoie à [Ball-Carlen-Lieb].

VII-2.4. Liens avec la topologie. Il existe de nombreux théorèmes de nature topologique faisant intervenir la séparabilité, la différentiabilité de la norme et l'uniforme convexité. Voici quelques-uns des plus notables (les trois premiers seront démontrés dans la suite du chapitre ; les autres sont mentionnés seulement pour mémoire) :

- Si E est uniformément convexe, alors $E^{**} = E$; c'est le **théorème de Milman–Pettis**.

- Si E est uniformément convexe, alors pour tout convexe fermé $C \subset E$ on peut définir une projection continue de E dans C .

- Si E est uniformément convexe et que sa norme est Gâteaux-différentiable en dehors de l'origine, alors $E^* = \{\lambda DN(x) ; \lambda \in \mathbb{R} ; x \neq 0\}$.

- La même conclusion est vraie si la norme de E est Fréchet-différentiable en-dehors de l'origine, et si $E^{**} = E$.

- Si E^* est séparable, alors il existe une norme N , équivalente à la norme $\|\cdot\|$ de E (c'est-à-dire qu'il existe $C > 0$ telle que $C^{-1}\|\cdot\| \leq N \leq C\|\cdot\|$), telle que N soit Fréchet-différentiable en-dehors de l'origine.

- Si E est uniformément convexe, alors la norme de E^* est Fréchet-différentiable en-dehors de l'origine.

VII-2.5. Espaces de Hilbert séparables. Entre tous les espaces de Banach, on distingue une catégorie qui possède à un degré extrême toutes les qualités : ce sont les espaces de Hilbert séparables (voire l'espace de Hilbert séparable, car ils sont tous isométriques entre eux).

DÉFINITION VII-61. *Un espace vectoriel normé complet $(H, \|\cdot\|)$ est appelé espace de Hilbert s'il existe une forme bilinéaire symétrique $b : H \times H \rightarrow \mathbb{R}$ telle que*

$$\forall x \in H, \quad b(x, x) = \|x\|^2.$$

La forme bilinéaire b est alors uniquement déterminée par la formule

$$b(x, y) = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2);$$

on l'appelle produit scalaire dans H , et on note souvent $b(x, y) = \langle x, y \rangle_H$; ou, s'il n'y a pas de confusion possible, $\langle x, y \rangle$ ou $x \cdot y$.

La preuve de la proposition suivante est immédiate :

PROPOSITION VII-62 (identité du parallélogramme). *Soit H un espace de Hilbert. Alors, pour tous $x, y \in H$ on a*

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

REMARQUE VII-63. Il s'agit en fait d'une *caractérisation* des espaces de Hilbert.

Nous allons facilement en déduire que les espaces de Hilbert séparables possèdent toutes les qualités recensées jusqu'ici :

THÉORÈME VII-64 (régularité des espaces de Hilbert). *Soit H un espace de Hilbert muni d'un produit scalaire $\langle \cdot, \cdot \rangle$. Alors la norme N sur H est Fréchet-différentiable en tout $x \neq 0$, et*

$$DN(x) \cdot h = \left\langle \frac{x}{\|x\|}, h \right\rangle.$$

En outre, H est uniformément convexe, de module de convexité $\kappa(\varepsilon) = \varepsilon^2/2$.

DÉMONSTRATION DU THÉORÈME VII-64. De l'identité

$$N(x + h)^2 = N(x)^2 + 2 \langle x, h \rangle + \|h\|^2,$$

on déduit que N^2 est Fréchet-différentiable en tout $x \in H$, et que sa différentielle en x est l'application linéaire "produit scalaire par $2x$ ". L'application "racine carrée" est dérivable en $N(x)^2$ dès que $N(x) \neq 0$, i.e. $x \neq 0$, et sa dérivée vaut alors $1/(2\|x\|)$. Par composition des différentielles, l'application N est différentiable en tout $x \neq 0$, et sa différentielle est l'application linéaire "produit scalaire par $(2x)/(2\|x\|)$ ", comme indiqué dans l'énoncé.

D'autre part, la formule du parallélogramme et la concavité de la fonction racine carrée entraînent que pour tous x, y de norme au plus 1,

$$\left\| \frac{x + y}{2} \right\| \leq \left(\frac{\|x\|^2 + \|y\|^2}{2} - \left\| \frac{x - y}{2} \right\|^2 \right)^{1/2} \leq 1 - \frac{1}{2} \left\| \frac{x - y}{2} \right\|^2.$$

En choisissant x et y très proches et de normes égales, on vérifie facilement que cette estimation du module de convexité ne peut être améliorée. \square

Le Théorème VII-64 garantit que tous les énoncés abstraits que nous rencontrons par la suite s'appliquent aux espaces de Hilbert séparables. En fait, la plupart de ces résultats pourront se démontrer de manière très directe dans les espaces de Hilbert, grâce à l'outil puissant que constituent les **bases Hilbertiennes**. Nous reprendrons cette discussion dans la section VII-8, après avoir développé la théorie générale des espaces de Banach.

VII-3. Projection

La projection d'un point x sur un ensemble X est une meilleure approximation de ce point par un élément de X . Plus précisément :

DÉFINITION VII-65 (projection). *Soient $(E, \|\cdot\|)$ un espace vectoriel normé, X une partie de E , et $x \in E$. On dit que $y \in X$ est une projection de x sur E si*

$$\|x - y\| = \inf_{z \in X} \|x - z\|.$$

Il est clair que si $x \in X$ alors x est l'unique projection de x . En revanche si $x \notin X$ et que X est un ensemble quelconque, on ne peut garantir ni l'existence ni l'unicité d'une projection. Nous allons voir cependant que la projection existe de manière unique si X est *convexe fermé* et que l'espace ambiant E est *uniformément convexe*.

VII-3.1. Théorème de projection dans les espaces uniformément convexes.

THÉORÈME VII-66 (Théorème de projection). *Soient E un espace de Banach uniformément convexe et C un convexe fermé de E . Alors, pour tout x il existe une unique projection de x sur C , notée $\Pi_C(x)$. Cette relation définit une application Π_C continue de E dans C , appelée "projection sur C ".*

REMARQUE VII-67. C'est l'espace ambiant E qui doit être uniformément convexe ; le convexe C quant à lui peut être "plat" ; on l'autorise même à être un espace vectoriel.

REMARQUE VII-68. La convexité et la fermeture de C sont essentielles, même en dimension finie.

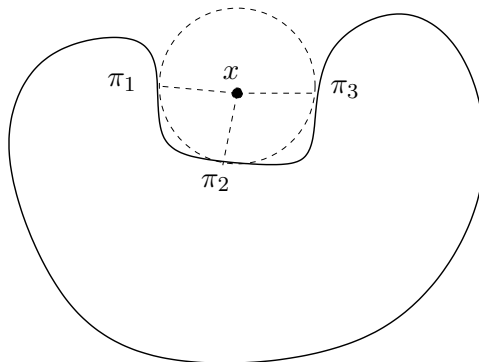


FIG. 3. La projection sur un fermé non convexe n'est en général pas unique.

DÉMONSTRATION DU THÉORÈME VII-66. Si A est une partie de E et $x \in E$, on note $d(x, A) = \inf\{\|x - y\|; y \in A\}$ la distance de x à A . Posons $D = d(x, C)$.

Pour tout $n \in \mathbb{N}$ on peut trouver $z_n \in C$ tel que

$$(62) \quad \|x - z_n\| \leq D + \frac{1}{n}.$$

Vérifions que la suite $(z_n)_{n \in \mathbb{N}}$ est de Cauchy. On se donne $m \geq n$; par convexité $(z_m + z_n)/2 \in C$, et donc

$$\|x - z_n\| \leq D + \frac{1}{n}, \quad \|x - z_m\| \leq D + \frac{1}{m}, \quad \left\|x - \frac{z_n + z_m}{2}\right\| \geq D.$$

En posant

$$X = \frac{x - z_n}{D + n^{-1}}, \quad Y = \frac{x - z_m}{D + m^{-1}},$$

on obtient

$$\|X\| \leq 1, \quad \|Y\| \leq 1, \quad \left\|\frac{X + Y}{2}\right\| \geq \frac{D}{D + n^{-1}} = 1 - \frac{n^{-1}}{D + n^{-1}} \geq 1 - \frac{1}{nD}.$$

Soit $\varepsilon > 0$ arbitraire, et $\kappa = \kappa(\varepsilon)$ comme dans la définition de l'uniforme convexité. Si $1/(nD) < \kappa$, la contraposée de cette définition montre que $\|X - Y\| < 2\varepsilon$, en particulier

$$\|z_n - z_m\| < 2\varepsilon(D + n^{-1}) \leq 2\varepsilon(D + 1).$$

La suite $(z_n)_{n \in \mathbb{N}}$ est donc bien de Cauchy.

Par complétude de E , cette suite admet une limite, notée z . En passant à la limite dans (62), on obtient $\|x - z\| \leq D$. Puisque C est fermé, on a $z \in C$, et en particulier $\|x - z\| \geq D$. On conclut que $\|x - z\| = D$, et z est donc une projection de x sur C .

Vérifions maintenant l'unicité de z . Si $D = 0$, c'est que $z = x$ et il n'y a rien à prouver. Si $D > 0$, soit z' tel que $\|x - z'\| = \|x - z\|$ et $z' \in C$; on pose $m := (z + z')/2$, par convexité m appartient à C . Alors les vecteurs $u := (z - x)/D$, $u' := (z' - x)/D$ sont de norme 1, et leur demi-somme est de norme $\|m - x\|/D \geq 1$. Par uniforme convexité, on a $u' = u$, soit $z' = z$.

Montrons maintenant la continuité de l'application Π_C . Le raisonnement est assez similaire à celui qui a déjà été utilisé pour l'existence. Notons $d(z, C) = \inf\{\|z - c\|; c \in C\}$.

Si x, y sont deux éléments quelconques de E , on a

$$\|x - \Pi_C(x)\| = d(x, C); \quad \|y - \Pi_C(y)\| = d(y, C).$$

Si x, y n'appartiennent pas tous deux à C , on peut définir

$$u = \frac{x - \Pi_C(x)}{\max(d(x, C), d(y, C))}, \quad v = \frac{y - \Pi_C(y)}{\max(d(x, C), d(y, C))}.$$

Alors

$$\frac{u + v}{2} = \frac{1}{\max(d(x, C), d(y, C))} \left[\frac{x + y}{2} - \left(\frac{\Pi_C(x) + \Pi_C(y)}{2} \right) \right],$$

de sorte que

$$\|u\| \leq 1; \quad \|v\| \leq 1; \quad \left\|\frac{u + v}{2}\right\| \geq \frac{d\left(\frac{x+y}{2}, C\right)}{\max(d(x, C), d(y, C))}.$$

Par convexité uniforme, il s'ensuit

$$\kappa \left(\frac{\|u - v\|}{2} \right) \leq 1 - \frac{d\left(\frac{x+y}{2}, C\right)}{\max(d(x, C), d(y, C))},$$

d'où

$$\|u - v\| \leq 2\kappa^{-1} \left(1 - \frac{d\left(\frac{x+y}{2}, C\right)}{\max(d(x, C), d(y, C))} \right),$$

et par conséquent

$$(63) \quad \|\Pi_C(x) - \Pi_C(y)\| \leq \|x - y\| + 2 \max(d(x, C), d(y, C)) \kappa^{-1} \left(1 - \frac{d\left(\frac{x+y}{2}, C\right)}{\max(d(x, C), d(y, C))} \right).$$

La continuité de Π_C en résulte. En effet, soit $x \in C$, alors

- si $x \in C$, alors $\Pi_C(x) = x$, donc

$$\|\Pi_C(y) - \Pi_C(x)\| = \|x - \Pi_C(y)\| \leq \|x - y\| + \|y - \Pi_C(y)\| \leq \|x - y\| + \|y - x\| = 2\|x - y\|,$$

et Π_C est donc continue en x ;

- si $x \notin C$, alors $d(x, C) > 0$ et $d((x+y)/2, C) / \max(d(x, C), d(y, C)) \rightarrow 1$ quand $y \rightarrow x$, car $d(\cdot, C)$ est continue (et même 1-lipschitzienne); donc le membre de droite de (63) tend vers 0, à une vitesse qui dépend du module de continuité κ et de la distance $d(x, C)$. \square

REMARQUE VII-69. Dans le cas particulier où E est un espace de Hilbert, on peut améliorer considérablement l'estimation (63). En effet, d'après l'inégalité (??) ci-dessous, on a

$$\begin{cases} \langle y - \Pi_C(y), \Pi_C(x) - \Pi_C(y) \rangle \leq 0; \\ \langle x - \Pi_C(x), \Pi_C(y) - \Pi_C(x) \rangle \leq 0; \end{cases}$$

d'où

$$\begin{aligned} \|\Pi_C(y) - \Pi_C(x)\|^2 &= \langle \Pi_C(y) - \Pi_C(x), \Pi_C(y) - \Pi_C(x) \rangle \\ &\leq \langle x - y, \Pi_C(y) - \Pi_C(x) \rangle \\ &\leq \|x - y\| \|\Pi_C(y) - \Pi_C(x)\|. \end{aligned}$$

On conclut que $\|\Pi_C(x) - \Pi_C(y)\| \leq \|x - y\|$, et Π_C est donc 1-lipschitzienne.

VII-3.2. Caractérisation différentielle. Si la norme de E est différentiable, les projections satisfont une équation (ou une inégalité) "différentielle", qui est le principal critère pratique permettant d'identifier les projections.

THÉORÈME VII-70. Soit E un espace de Banach, et soient C un convexe fermé de E , $x \in E \setminus C$, π une projection de x sur C . On suppose que la norme N de E est Gâteaux-différentiable en $x - \pi$. Alors

$$(64) \quad \forall z \in C, \quad DN(x - \pi) \cdot (z - \pi) \leq 0.$$

Réciproquement, si (64) est vérifiée pour tout $z \in C$, alors π est la projection de x sur C .

Si C est un espace vectoriel F , on peut remplacer (64) par

$$(65) \quad \forall y \in F, \quad DN(x - \pi) \cdot y = 0;$$

en d'autres termes, $F \subset \ker DN(x - \pi)$.

REMARQUE VII-71. Dans cet énoncé on peut remplacer DN par $D(N^2)$ (ou $D(N^2/2)$), ce qui a l'avantage de garder un sens même quand $x - \pi = 0$ (i.e. $z \in C$).

En combinant la remarque précédente avec le calcul de la différentielle de la norme dans un espace de Hilbert (Cf. le Théorème VII-64 et sa preuve), on obtient le

COROLLAIRE VII-72 (Projection dans les espaces de Hilbert). Soient H un espace de Hilbert, C un convexe fermé de H , et Π l'opérateur de projection de H sur C . Alors, pour tout $x \in H$ on a

$$(66) \quad \forall z \in C, \quad \langle x - \Pi(x), z - \Pi(x) \rangle \leq 0;$$

c'est-à-dire que l'angle formé par les directions $x - \Pi(x)$ et $z - \Pi(x)$ est obtus. En outre, cette inégalité caractérise $\Pi(x)$.

Si C est un espace affine F , l'inégalité (66) peut se remplacer par l'identité

$$(67) \quad \forall y \in F, \quad \langle x - \Pi(x), y \rangle = 0;$$

en d'autres termes, $x - \Pi(x)$ est orthogonal à F .

REMARQUE VII-73. Le Corollaire VII-72 justifie la terminologie de **projection orthogonale** pour désigner la projection Π_F sur un sous-espace vectoriel fermé F dans un espace de Hilbert. Comme l'équation (67) est linéaire et caractérise $\Pi_F(x)$, on voit que Π_F est une application linéaire. Ce n'est pas le cas a priori si H n'est pas un espace de Hilbert !

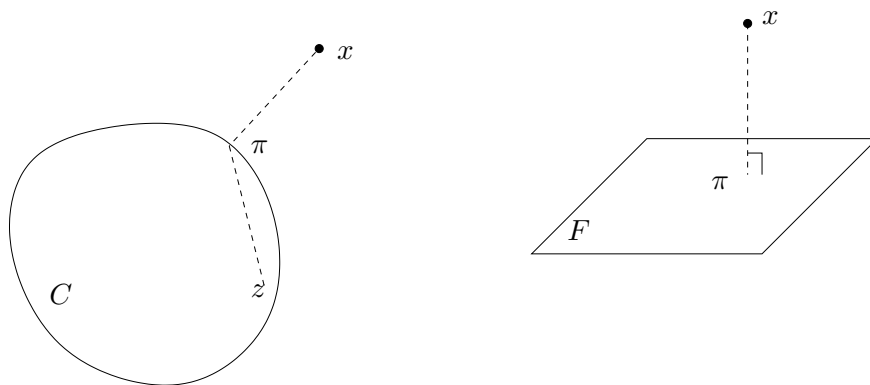


FIG. 4. Image géométrique de la projection dans un espace de Hilbert ; ici π est la projection de x sur le convexe C (resp. l'espace affine fermé F). L'angle formé par les vecteurs issus de π et aboutissant en x d'une part, en un élément du convexe d'autre part, est toujours obtus (et droit dans le cas d'un espace affine).

DÉMONSTRATION DU THÉORÈME VII-70. Soit $z \in C$ et soit $\varepsilon \in]0, 1]$. Par convexité, $(1 - \varepsilon)\pi + \varepsilon z \in C$, donc

$$N\left(x - [(1 - \varepsilon)\pi + \varepsilon z]\right) - N(x - \pi) \geq 0$$

Par différentiabilité de N en $x - \pi$, le membre de gauche vaut $-\varepsilon DN(x - \pi) \cdot (z - \pi) + o(\varepsilon)$. En divisant par ε et en faisant tendre ε vers 0, on obtient l'inégalité $DN(x - \pi) \cdot (z - \pi) \leq 0$, ce qui prouve (64).

Réciproquement, supposons (64) vérifié. On pose $\varphi(z) = N(z - x)$; c'est une fonction convexe, différentiable en $z = \pi$ (car $\pi \neq x$ par hypothèse). Pour tout $z \in C$ et pour tout $t \in [0, 1[$, on a

$$\varphi((1 - t)\pi + tz) \leq (1 - t)\varphi(\pi) + t\varphi(z),$$

d'où, après division par $1 - t$,

$$\frac{\varphi((1 - t)\pi + tz) - \varphi(\pi)}{t} \leq \varphi(z) - \varphi(\pi).$$

Quand on fait tendre t vers 0, le membre de gauche tend vers $D\varphi(\pi) \cdot (z - \pi)$, qui est positif par hypothèse. On conclut que $\varphi(z) \geq \varphi(\pi)$, soit $N(z - \pi) \geq N(x - \pi)$, ce qui montre bien que π est projection de x sur C .

Si maintenant C est un espace affine F , on peut établir une bijection entre $z \in C$ et $y = z - \pi$ appartenant à \tilde{F} , l'espace vectoriel qui dirige F . On a alors $DN(x - \pi) \cdot y \leq 0$; en changeant y en $-y$, on trouve $DN(x - \pi) \cdot y \geq 0$, d'où (65). \square

VII-4. Dualité

L'espace dual E^* d'un espace vectoriel normé a été introduit dans le Corollaire VII-11; c'est un espace de Banach. Dans cette section, on va passer en revue des outils liés à la notion de dualité, et certains théorèmes de représentation de l'espace dual.

VII-4.1. Théorème de Hahn–Banach. Le théorème d'extension de Hahn–Banach permet de prolonger (étendre) des formes linéaires définies sur de “petits” sous-espaces d'un espace vectoriel normé E , en des formes linéaires définies sur E tout entier. Conçu à l'origine pour démontrer des résultats surprenants ou paradoxaux (par exemple le Théorème IV-43, en liaison avec le paradoxe de Banach–Tarski), le théorème de Hahn–Banach a pris au cours du vingtième siècle une place fondamentale dans l'analyse fonctionnelle, en rapport avec l'axiome du choix. Dans ce cours, je trahirai l'esprit du théorème en me contentant de l'énoncer dans un espace **séparable**, ce qui évitera l'usage de l'axiome du choix. (Plus précisément, je commencerai par un énoncé purement algébrique sous une hypothèse de dimension au plus dénombrable, et j'en déduirai par densité le résultat pour des espaces séparables.)

THÉORÈME VII-74 (Théorème d'extension de Hahn–Banach). *Soient E un espace vectoriel de dimension au plus dénombrable, et $p : E \rightarrow \mathbb{R}_+$ une application homogène de degré 1, sous-additive :*

$$\forall x, y \in E, \quad \forall \alpha \geq 0, \quad p(\alpha x) = \alpha p(x), \quad p(x + y) \leq p(x) + p(y).$$

On se donne F un sous-espace vectoriel de E , et λ une forme linéaire sur F telle que

$$\forall x \in F, \quad \lambda(x) \leq p(x).$$

Alors il existe une forme linéaire Λ sur E , qui coïncide avec λ sur F , et vérifie

$$\forall x \in E, \quad \Lambda(x) \leq p(x).$$

COROLLAIRE VII-75. Soient E un espace vectoriel normé séparable, et λ une forme linéaire continue sur un sous-espace vectoriel F de E . Alors il existe une forme linéaire continue Λ sur E telle que

$$\Lambda|_{F=\lambda} \quad \text{et} \quad \|\Lambda\|_{E^*} = \|\lambda\|_{F^*}.$$

DÉMONSTRATION DU THÉORÈME VII-74. On montre par récurrence que F , sous-espace d'un espace vectoriel de dimension au plus dénombrable, est lui-même de dimension au plus dénombrable (c'est un exercice facile d'algèbre linéaire). Toujours par récurrence, on peut construire une base de F , au plus dénombrable, et la compléter en une base de E , de sorte que E sera engendré par F et par une famille finie ou dénombrable de vecteurs $(e_n)_{n \geq 1}$, tous indépendants entre eux, et indépendants de F . On va définir Λ sur E tout entier en imposant $\Lambda(e_n)$ pour tout n , par récurrence finie ou infinie.

Supposons Λ définie sur l'espace vectoriel F_k engendré par F et e_1, \dots, e_k , et vérifiant sur cet espace l'inégalité $|\Lambda| \leq p$. On cherche à déterminer $\Lambda(e_{k+1})$ pour que cette inégalité soit encore vraie sur l'espace engendré par F_k et e_{k+1} , autrement dit :

$$\forall x \in F_k, \quad \forall t \in \mathbb{R}, \quad \Lambda(x + te_{k+1}) \leq p(x + te_{k+1}).$$

On pose $\alpha := \Lambda(e_{k+1})$, l'inégalité souhaitée se réécrit

$$\Lambda(x) + t\alpha \leq p(x + te_{k+1}).$$

Pour $t = 0$ elle est évidemment vérifiée, tandis que pour $t \neq 0$ elle devient

$$\sup_{x \in F_k} \sup_{t < 0} \left(\frac{p(x + te_k) - \Lambda(x)}{t} \right) \leq \alpha \leq \inf_{x \in F_k} \inf_{t < 0} \left(\frac{p(x + te_k) - \Lambda(x)}{t} \right).$$

En changeant t en $-t$ dans le membre de gauche et en utilisant la propriété d'homogénéité de p , la double inégalité précédente se transforme en

$$\sup_{x \in F_k} [\Lambda(x) - p(x - e_{k+1})] \leq \alpha \leq \inf_{x \in F_k} [p(x + e_{k+1}) - \Lambda(x)].$$

Cette équation en α a une solution si et seulement si le supremum dans le membre de gauche est plus petit que l'infimum dans le membre de droite, i.e.

$$\forall x, y \in F_k, \quad \Lambda(x) - p(x - e_{k+1}) \leq p(y + e_{k+1}) - \Lambda(y).$$

Or pour tous x, y dans F_k on a également $x + y \in F_k$, donc

$$\Lambda(x) + \Lambda(y) = \Lambda(x + y) \leq p(x + y) \leq p(x + e_{k+1}) + p(y - e_{k+1}),$$

où l'on a utilisé successivement la linéarité de Λ , l'hypothèse de récurrence, et la sous-additivité de p . Le théorème est démontré. \square

DÉMONSTRATION DU COROLLAIRE VII-75. Soit \tilde{F} un sous-espace dense dans F , de dimension au plus dénombrable, et \tilde{E} un sous-espace dense dans E , de dimension au plus dénombrable également. Quitte à remplacer \tilde{E} par l'espace engendré par \tilde{E} et \tilde{F} , on peut supposer que \tilde{E} contient \tilde{F} . La forme linéaire λ est définie sur F , a fortiori sur \tilde{F} , et elle y est majorée par l'application sous-linéaire homogène p définie par $p(x) = \|\lambda\| \|x\|$. Par le Théorème VII-74, elle se prolonge en une forme linéaire $\tilde{\Lambda}$ définie sur \tilde{E} , vérifiant

$$\forall x \in \tilde{E}, \quad \Lambda(x) \leq \|\lambda\| \|x\|.$$

En changeant x en $-x$, on voit qu'en fait

$$(68) \quad \forall x \in \tilde{E}, \quad |\Lambda(x)| \leq \|\lambda\| \|x\|.$$

L'application Λ étant uniformément continue sur une partie dense de l'espace complet E , elle admet un prolongement continu Λ à E tout entier. En passant à la limite dans (68), on voit que cette inégalité est vérifiée pour tout $x \in E$, ce qui montre bien que $\|\Lambda\| \leq \|\lambda\|$. Par ailleurs, on a forcément $\|\Lambda\| \geq \|\lambda\|$, puisque E contient F . Le corollaire est donc démontré. \square

En pratique, dans la grande majorité des cas on applique le Théorème de Hahn-Banach avec une application p donnée

- soit par une norme, comme dans le Corollaire VII-75 ;
- soit par la **jauge d'un convexe**, définie ci-après (ce deuxième cas est plus général car la norme n'est autre que la jauge de la boule unité) :

DÉFINITION VII-76 (jauge). *Soit, dans un espace vectoriel normé E , un convexe C contenant un voisinage de 0 . Pour tout $x \in E$ on définit*

$$p(x) := \inf \left\{ \alpha > 0; \quad \frac{x}{\alpha} \in C \right\}.$$

L'application $p : E \rightarrow \mathbb{R}_+$, homogène de degré 1 et sous-additive, est appelée jauge du convexe C .

DÉMONSTRATION DES PROPRIÉTÉS DE LA JAUGE. Par hypothèse, C contient une boule centrée en 0 ; donc, pour tout x , on peut trouver $\alpha > 0$ assez grand pour que $x/\alpha \in C$, et l'application p est bien à valeurs dans \mathbb{R}_+ . Il est évident que $p(\lambda x) = \lambda p(x)$ pour tout $\lambda > 0$. Enfin, si x et y sont deux vecteurs de E , pour tout $\varepsilon > 0$ on a

$$\frac{x}{p(x) + \varepsilon} \in C, \quad \frac{y}{p(y) + \varepsilon} \in C,$$

et donc par convexité

$$\frac{x + y}{p(x) + p(y) + 2\varepsilon} = \left(\frac{p(x) + \varepsilon}{p(x) + p(y) + 2\varepsilon} \right) \frac{x}{p(x) + \varepsilon} + \left(\frac{p(y) + \varepsilon}{p(x) + p(y) + 2\varepsilon} \right)$$

appartient aussi à C . Il s'ensuit que $p(x + y) \leq p(x) + p(y) + 2\varepsilon$, et la sous-additivité vient en faisant tendre ε vers 0 . \square

VII-4.2. Théorèmes de séparation. Les formes linéaires sont souvent utilisées pour “séparer” des objets dans un espace de Banach. Les principaux théorèmes en la matière sont des conséquences du théorème de prolongement de Hahn–Banach, et sont eux-mêmes souvent appelés théorèmes de Hahn–Banach.

La définition suivante, peu classique, précise les deux notions-clé relatives à la séparation des points par des formes linéaires.

DÉFINITION VII-77 (séparabilité linéaire). *Soient A une partie d'un espace vectoriel normé E , et \mathcal{L} une partie de E^* . On dit que \mathcal{L} sépare A si, étant donnés deux points distincts x et y dans A , on peut trouver $\Lambda \in \mathcal{L}$ tel que $\Lambda(x) \neq \Lambda(y)$.*

On dit que E est linéairement séparable si E^ sépare E ; ou de manière équivalente si, étant donné $x \in E \setminus \{0\}$ on peut trouver $\Lambda \in E^*$ tel que $\Lambda(x) \neq 0$.*

Soit $x \in E \setminus \{0\}$; on dit que $\Lambda \in E^$ normalise x si $\|\Lambda\| = 1$ et $\Lambda(x) = \|x\|$. On dit que x est linéairement normalisable s'il existe $\Lambda \in E^*$ normalisant x ; on dit que E est linéairement normalisable si tout $x \in E \setminus \{0\}$ est linéairement normalisable. (Cette notion est donc plus forte que celle de séparabilité linéaire.)*

REMARQUE VII-78. L'axiome du choix implique que tout espace vectoriel normé est linéairement normalisable; les notions introduites dans la Définition VII-77 ne présentent donc d'intérêt que si l'on ne souhaite pas utiliser l'axiome du choix.

La proposition suivante donne plusieurs conditions suffisantes pour qu'un espace vectoriel normé soit linéairement séparable, ou linéairement normalisable.

PROPOSITION VII-79. *Soit E un espace vectoriel muni d'une norme N . Si E remplit l'une quelconque des trois conditions suivantes :*

- (a) E est séparable;*
- (b) E est le dual d'un espace vectoriel normé *E ;*
- (c) la norme de E est Gâteaux-différentiable hors de l'origine;*

alors E est linéairement séparable. En outre,

- sous l'hypothèse (a) ou (c), E est linéairement normalisable;*
- sous l'hypothèse (b), *E (vu comme sous-espace de E^*) sépare E ;*
- sous l'hypothèse (c), tout $x \neq 0$ est normalisé par la forme linéaire $DN(x)$;*
- sous l'hypothèse (a), il existe une suite $(\Lambda_k)_{k \in \mathbb{N}}$ de formes linéaires séparant E .*

Plus généralement, si S est un sous-espace séparable de E , on peut trouver une suite de formes linéaires sur E qui sépare S .

REMARQUE VII-80. On notera que la terminologie est cohérente, au sens où un espace séparable est automatiquement linéairement séparable.

EXEMPLE VII-81 (espaces L^p). Soit (X, \mathcal{A}, μ) un espace mesuré; on note $L^p = L^p(X, \mu)$. Pour tout $p \in (1, +\infty)$, L^p est linéairement normalisable puisqu'il satisfait l'hypothèse (c). (Cette propriété a d'ailleurs été utilisée au cours de la démonstration de ce que le dual de L^p est $L^{p'}$.) Si X est σ -fini, alors L^1 sépare son dual L^∞ grâce au critère (b); en revanche L^1 ne normalise pas L^∞ (exercice). Enfin L^1 est linéairement normalisable : pour $f \in L^1$, on choisit la forme linéaire représentée par $g = (\text{sign } f) \in L^\infty$.

EXEMPLE VII-82. Si X est un espace localement compact, alors $C_0(X)$ sépare son dual $M(X)$ grâce au critère (b). Plus subtil : si X est un espace polonais, alors $C_b(X)$ sépare l'espace $M(X)$ (qui a priori n'est le dual de personne). En effet, soit μ une mesure finie non nulle sur X ; par le Théorème I-54, ν est automatiquement

régulière. Soit A un ensemble mesurable A tel que $\mu_+[A] = \delta > 0$, $\mu_-[A] = 0$; par régularité il existe un ensemble compact $K \subset A$ tel que $\mu_+[K] > \delta/2$, et bien sûr $\mu_-[K] = 0$, donc $\mu[K] > \delta/2$. On pose $f_k(x) := \max(k^{-1} - d(x, K), 0)$. Pour tout $x \notin K$, $f_k(x) = 0$ pour k assez grand; par convergence dominée on a donc $\mu[K] = \lim \int f_k d\mu$. Il s'ensuit que $\int f_k d\nu \neq 0$ pour k assez grand.

EXERCICE VII-83. Montrer qu'en général $C(X)$ ne normalise pas $M(X)$, même si X est compact.

DÉMONSTRATION DE LA PROPOSITION VII-79. 1. Commençons par supposer E séparable. Soit z un vecteur non nul, et $F = \mathbb{R}z$ la droite vectorielle engendrée par z . Sur F on définit l'application

$$\lambda : tz \mapsto t\|z\|.$$

C'est bien sûr une forme linéaire, évidemment de norme 1. Par le Corollaire VII-75 (Théorème de Hahn–Banach), on peut étendre λ en une forme linéaire $\Lambda \in E^*$, de norme 1, telle que

$$\Lambda(z) = \lambda(z) = \|z\|.$$

Ceci montre que E est linéairement normalisable (et en particulier linéairement séparable).

2. Toujours sous l'hypothèse (a), soit ensuite $(z_k)_{k \in \mathbb{N}}$ une suite dense dans E . Pour tout $\ell \in \mathbb{N}$ on note E_ℓ le sous-espace vectoriel engendré par les $(z_k)_{k \leq \ell}$. C'est un espace de dimension finie; soit $(\Lambda_{j,\ell})_{j \leq \ell}$ une base de son dual $(E_\ell)^*$. Chacune de ces formes linéaires s'étend (encore par le Théorème de Hahn–Banach) en un élément de E^* . Soit \mathcal{L} l'espace vectoriel engendré par l'union de tous les $\Lambda_{j,\ell}$ quand j et ℓ varient : c'est un espace de dimension au plus dénombrable, et cet espace normalise n'importe quel vecteur non nul de l'espace vectoriel engendré (algébriquement) par les z_k .

Soit alors $y \in E \setminus \{0\}$, et soit $y_k \rightarrow y$ tel que $y_k \in E_k$. Pour tout k on peut trouver $\Lambda_k \in E_k$ tel que $\|\Lambda_k\| = 1$ et $\Lambda_k(y_k) = \|y_k\|$. Alors, quand $k \rightarrow \infty$, $\|\Lambda_k(y_k) - \Lambda_k(y)\| \leq \|\Lambda_k\| \|y_k - y\| = \|y_k - y\| \rightarrow 0$; donc $\lim_{k \rightarrow \infty} \Lambda_k(y) = \lim_{k \rightarrow \infty} \|y_k\| = \|y\| \neq 0$. En particulier, $\Lambda_k(y)$ est non nul pour k assez grand. Ceci montre bien que \mathcal{L} sépare E . Pour conclure il suffit de choisir une suite dense dans \mathcal{L} .

3. Si maintenant E est le dual d'un espace vectoriel normé *E , soient $\Lambda \neq \Lambda'$ deux formes linéaires distinctes. Par définition, cela veut dire qu'il existe $x_0 \in {}^*E$ tel que $\Lambda(x_0) \neq \Lambda'(x_0)$; donc *E , vu comme sous-espace de E^* , sépare bien E .

4. Considérons finalement l'hypothèse (c). Soit $z \neq 0$; alors pour tout $t > 0$,

$$(1+t)\|z\| = N((1+t)z) = N(z) + DN(z) \cdot (tz) + o(t) = \|z\| + t DN(z) \cdot z + o(t).$$

On simplifie par $\|z\|$ et on divise par $t > 0$, puis on fait tendre t vers 0, il vient

$$DN(z) \cdot z = \|z\|.$$

Pour conclure que $DN(z)$ normalise z , il suffit de vérifier que $\|DN(z)\| \leq 1$. Pour cela on utilise le même argument : on se donne $y \in E$ et on écrit, pour $\varepsilon > 0$,

$$N(z) + \varepsilon N(y) = N(z + \varepsilon y) = N(z) + \varepsilon DN(z) \cdot y + o(\varepsilon).$$

En faisant encore tendre ε vers 0, on obtient $DN(z) \cdot y \leq N(y)$. En appliquant cette inégalité à $-y$, on conclut que $|DN(z) \cdot y| \leq \|y\|$, ce qui achève la preuve. \square

Voici maintenant le principal théorème de cette sous-section ; il permet de séparer des convexes disjoints par des formes linéaires.

THÉORÈME VII-84. *Dans un espace vectoriel normé séparable E , soient C et C' deux convexes disjoints.*

(i) *Si C est ouvert, alors on peut trouver une forme linéaire $\Lambda \in E^*$ non nulle, telle que*

$$\sup_C \Lambda \leq \inf_{C'} \Lambda;$$

on dit que Λ sépare C et C' .

(ii) *Si C est fermé et C' compact, alors on peut trouver une forme linéaire $\Lambda \in E^*$, telle que*

$$\sup_C \Lambda < \inf_{C'} \Lambda;$$

on dit que Λ sépare strictement C et C' .

DÉMONSTRATION. (i) On note $C - C'$ la réunion de tous les $C - y$ quand y décrit C' : il est facile de vérifier que c'est un convexe ouvert (car réunion d'ensembles ouverts), ne contenant pas 0 (car $C \cap C' = \emptyset$). Soit x_0 tel que $-x_0 \in C - C'$, on pose $\tilde{C} := C - C' + x_0$. Alors \tilde{C} est un convexe ouvert contenant 0 mais ne contenant pas x_0 . Soit p la jauge de \tilde{C} , selon la Définition VII-76. Puisque $x_0 \notin \tilde{C}$, on a $p(x_0) \geq 1$, et par homogénéité $p(tx_0) \geq t$ pour tout $t > 0$ (et donc également pour $t \geq 0$, puisque $p \geq 0$). Sur $F := \mathbb{R}x_0$, on définit une forme linéaire λ par $\lambda(tx_0) = t$; il est clair que λ est majorée par p . Grâce au Corollaire VII-75 (Théorème de Hahn-Banach) on peut donc l'étendre en une forme linéaire Λ sur E , majorée par p . Alors $\Lambda(x_0) = 1$, mais $\Lambda(\tilde{x}) \leq p(\tilde{x}) \leq 1$ pour tout $x \in \tilde{C}$. Soient maintenant $x \in C$ et $x' \in C'$, on peut poser $\tilde{x} := x - x' + x_0$ et on a $\Lambda(x) - \Lambda(x') + \Lambda(x_0) \leq \Lambda(x_0)$, soit $\Lambda(x) \leq \Lambda(x')$.

(ii) Puisque C' est compact, la fonction continue $x' \mapsto d(x', C)$ atteint son minimum en un certain $x'_0 \in C'$, et ce minimum est non nul puisque $x'_0 \notin C$ et que C est fermé. Il existe donc $\varepsilon > 0$ tel que $d(x, x') \geq 2\varepsilon$ pour tous $x \in C$, $x' \in C'$. On pose alors

$$C_\varepsilon := \{x \in E; d(x, C) < \varepsilon\} = C + B_\varepsilon(0); \quad C'_\varepsilon := C' + B_\varepsilon(0).$$

Ces deux ensembles sont convexes et ouverts ; par (i) il existe une forme linéaire $\Lambda \neq 0$ et $\alpha \in \mathbb{R}$ tels que $\Lambda(x) \leq \alpha \leq \Lambda(x')$ pour tous $x \in C$, $x' \in C'$. On a donc, pour tous z, z' de norme 1,

$$\Lambda(x) + \varepsilon\Lambda(z) = \Lambda(x + \varepsilon z) \leq \alpha \leq \Lambda(x' + \varepsilon z') = \Lambda(x') + \varepsilon\Lambda(z').$$

En passant au supremum en z et à l'infimum en z' , on obtient

$$\Lambda(x) + \varepsilon\|\Lambda\| \leq \alpha \leq \Lambda(x') - \varepsilon\|\Lambda\|,$$

ce qui entraîne bien l'inégalité stricte puisque $\|\Lambda\| \neq 0$. □

Pour conclure cette sous-section, je mentionnerai quelques **critères de densité** qui découlent des théorèmes de séparation rencontrés précédemment. Pour ajouter encore à la confusion, ces critères sont aussi souvent appelés Théorèmes de Hahn-Banach.

THÉORÈME VII-85. *Soient E un espace vectoriel normé séparable, C un convexe de E et $x \in E$. Alors $x \in \overline{C}$ si et seulement si*

$$\forall \Lambda \in E^*, \quad \Lambda(x) \leq \sup_{y \in C} \Lambda(y).$$

Dans le cas où C est un sous-espace vectoriel F de E , on peut remplacer ce dernier critère par

$$[\forall y \in F, \Lambda(y) = 0] \implies [\Lambda(x) = 0].$$

En particulier, F est dense dans E si et seulement si la forme linéaire nulle est la seule forme linéaire continue s'annulant sur F .

DÉMONSTRATION DU THÉORÈME VII-85. (i) Si $x \in \overline{C}$, alors par continuité on a $|\Lambda(x)| \leq \sup_{y \in C} |\Lambda(y)|$. Au contraire, si $x \notin \overline{C}$, alors par le Théorème VII-84(ii) on peut séparer le convexe fermé \overline{C} et le convexe compact $\{x\}$ par une forme linéaire non nulle Λ : on a alors

$$\sup_{y \in C} \Lambda(y) < \Lambda(x).$$

(ii) Si $x \in \overline{F}$ et que Λ s'annule identiquement sur F , alors par continuité $\Lambda(x) = 0$: le critère est donc vérifié.

À l'opposé, si $x \notin \overline{F}$, par le Théorème VII-84(ii) on peut séparer strictement le convexe fermé \overline{F} du convexe compact $\{x\}$; il existe donc $\Lambda \in E^*$ telle que

$$\sup_{\overline{F}} \Lambda < \Lambda(x).$$

En particulier $\sup_{\overline{F}} \Lambda < +\infty$. Mais s'il existait $y \in \overline{F}$ tel que $\Lambda(y) \neq 0$, alors on pourrait donc trouver $t \in \mathbb{R}$ tel que $\Lambda(ty)$ soit arbitrairement grand, et $\sup_{\overline{F}} \Lambda = +\infty$, ce qui est impossible. La forme linéaire Λ s'annule donc identiquement sur \overline{F} mais pas en x . \square

VII-4.3. Théorème de représentation par projection. Le théorème qui suit est l'un des outils les plus utiles pour identifier l'espace dual d'un espace de Banach.

THÉORÈME VII-86. *Soit E un espace de Banach muni de sa norme N . Si E est uniformément convexe et N est Gâteaux-différentiable en-dehors de l'origine, alors*

$$E^* = \{\lambda DN(x); \quad \lambda \in \mathbb{R}, x \neq 0\}.$$

DÉMONSTRATION. Tout d'abord, il est clair que pour tous $x \neq 0$ et $\lambda \in \mathbb{R}$, on a bien $\lambda DN(x) \in E^*$; c'est la réciproque qui est non triviale : montrer que tout $\Lambda \in E^*$ s'écrit sous la forme $\lambda DN(x)$.

Soit donc Λ une forme linéaire continue sur E ; on supposera Λ non nulle (sinon il suffit de prendre $\lambda = 0$). Son noyau $\ker \Lambda$ est un hyperplan fermé, et puisque E est uniformément convexe on peut définir (Théorème VII-66) l'application de projection $\Pi_{\ker \Lambda}$ de E dans $\ker \Lambda$. Soit $x_0 \in E$ tel que $\Lambda(x_0) \neq 0$, et $\pi_0 := \Pi_{\ker \Lambda}(x_0)$; en particulier $\Lambda(\pi_0) = 0$. Bien sûr $x_0 \neq \pi_0$, puisque $\Lambda(x_0) \neq 0 = \Lambda(\pi_0)$; N est donc Gâteaux-différentiable en $x_0 - \pi_0$. D'après le Théorème ??(iii) $DN(x_0 - \pi_0)$ normalise $x_0 - \pi_0$; en particulier $DN(x_0 - \pi_0)$ est une forme linéaire *non nulle*.

D'autre part, par (65) on a

$$\ker \Lambda \subset \ker DN(x_0 - \pi_0).$$

En résumé : le noyau de la forme linéaire Λ est inclus dans le noyau de la forme linéaire non nulle $DN(x_0 - \pi_0)$. Par un résultat élémentaire d'algèbre linéaire, il existe $\lambda \in \mathbb{R}$ tel que $\Lambda = \lambda DN(x_0 - \pi_0)$. \square

VII-4.4. Dualité des espaces L^p . Je vais maintenant utiliser les résultats de la sous-section précédente pour identifier les duals des espaces de Lebesgue. La formule à retenir est

$$(L^p)^* = L^{p'}, \quad \frac{1}{p} + \frac{1}{p'} = 1,$$

du moins dès que $p > 1$. Voici un énoncé plus précis :

THÉORÈME VII-87 (dual de L^p). *Soit (X, μ) un espace mesuré; on note L^p l'espace de Lebesgue $L^p(X, \mu)$, quotienté par la relation d'égalité presque partout. Alors,*

(i) *Pour tout $p \in]1, \infty[$ on peut identifier isométriquement le dual de $L^p(X, \mu)$ à l'espace $L^{p'}(X, \mu)$, où $p' = p/(p-1)$. Plus précisément, pour toute forme linéaire continue Λ sur $L^p(X, \mu)$ il existe un unique $g \in L^{p'}(X, \mu)$ tel que*

$$\forall f \in L^p(X, \mu), \quad \Lambda(f) = \int fg \, d\mu.$$

(ii) *Si X est σ -fini, on peut de même identifier isométriquement le dual de $L^1(X, \mu)$ à $L^\infty(X, \mu)$.*

REMARQUE VII-88. Quand on dit que $(L^p)^*$ s'identifie à $L^{p'}$, cela veut dire que l'application qui à $g \in L^{p'}$ associe la forme linéaire Λ_g définie par

$$\langle \Lambda_g, f \rangle = \int fg \, d\mu$$

est une *isométrie* de $L^{p'}$ sur $(L^p)^*$.

REMARQUE VII-89. Comme cas particulier du Théorème VII-87, on retrouve le théorème de Riesz selon lequel le dual de L^2 (espace de Hilbert) s'identifie à L^2 .

REMARQUE VII-90. Même dans des situations concrètes (par exemple si X est un ouvert de \mathbb{R}^n muni de la mesure de Lebesgue), l'identification du dual de $L^\infty(X)$ est un problème insoluble.

- Si l'on admet l'axiome du choix on peut montrer que $L^\infty(\mathbb{R}^n)^*$ est strictement plus grand que $L^1(\mathbb{R}^n)$. C'est une construction facile : on introduit la forme linéaire "limite à l'infini" (définie sur le sous-espace fermé de $L^\infty(\mathbb{R}^n)$ formé des applications qui convergent à l'infini vers une limite réelle) ; puis on utilise le théorème de Hahn-Banach en une forme linéaire définie sur $L^\infty(\mathbb{R}^n)$ tout entier. Cette forme linéaire attribue une limite à toute fonction bornée, et il est facile de vérifier qu'elle ne peut être représentée par aucune fonction intégrable. Si X est, disons, un ouvert borné, on peut effectuer une construction similaire avec l'application "limite en x_0 ", où x_0 est un point de X .

- Si l'on exclut l'axiome du choix, le problème de savoir si $(L^\infty)^*$ est plus grand que L^1 devient *indécidable*!

Conclusion : Mieux vaut ne pas se poser de questions sur le dual de L^∞ , et s'interdire d'utiliser cet espace...

REMARQUE VII-91. Si X n'est pas σ -fini, la nature du dual de $L^1(X)$ n'est pas claire non plus.

REMARQUE VII-92. Même si L^1 n'est (peut-être) pas le dual de L^∞ , la formule

$$\|f\|_{L^{p'}} = \sup \left\{ \int_X fg \, d\mu; \quad \|g\|_{L^p} \leq 1 \right\}$$

reste vraie pour toutes les valeurs de $p \in [1, \infty]$, comme on l'a vu dans l'énoncé du Théorème VI-24 (pour $p = 1$ il suffit de choisir $g = 1_{f>0} - 1_{f<0}$).

DÉMONSTRATION DU THÉORÈME VII-87. Commençons par le cas $1 < p < \infty$. Par le Théorème VII-53, L^p est uniformément convexe. Par le Théorème VII-47, sa norme N_p est Fréchet-différentiable (a fortiori Gâteaux-différentiable) hors de l'origine. Par le Théorème VII-86 $(L^p)^*$ coïncide avec le cône engendré par les $DN_p(f)$, $f \in L^p \setminus \{0\}$, c'est-à-dire par les fonctions $|f|^{p-2}f$. Si $f \in L^p$, alors $g = |f|^{p-2}f \in L^{p/(p-1)} = L^{p'}$; on a donc $(L^p)^* \subset L^{p'}$. L'inclusion réciproque est déjà connue par le Théorème VI-22 (inégalité de Hölder). Le Théorème VI-24 montre en outre que pour tout $g \in L^{p'}$ on a $\|g\|_{(L^p)^*} = \|g\|_{L^{p'}}$, ce qui achève la démonstration dans le cas $1 < p < \infty$.

Passons maintenant au cas $p = 1$. On sait que toute fonction essentiellement bornée définit une forme linéaire continue sur L^1 ; il reste à vérifier la réciproque. Soit donc Λ une forme linéaire continue sur L^1 . Pour tout ensemble mesurable $Y \subset X$, de mesure finie, on a $L^2(Y, \mu) \subset L^1(A, \mu) \subset L^1(X, \mu)$ (ici on aurait pu choisir n'importe quel indice $p > 1$ à la place de 2); donc Λ induit une forme linéaire continue sur $L^2(Y, \mu)$; et il existe $g \in L^2(A, \mu) = L^2(A, \mu)^*$ tel que pour tout $f \in L^2(Y, \mu)$,

$$\Lambda(f) = \int_Y fg \, d\mu.$$

Soit $\varepsilon > 0$, et

$$A_\varepsilon := \{x \in Y; \quad |g| \geq \|\Lambda\| + \varepsilon\}.$$

On pose $f := 1_{A_\varepsilon} \operatorname{sign}(g)$; alors $f \in L^1(X) \cap L^2(Y)$, et $\|f\|_{L^1} = \mu[A_\varepsilon]$; d'où

$$\int_Y fg \, d\mu \leq \|\Lambda\| \mu[A_\varepsilon].$$

Mais par définition de f ,

$$\int_Y fg \, d\mu = \int_{|g| \geq \|\Lambda\| + \varepsilon} |g| \geq (\|\Lambda\| + \varepsilon) \mu[A_\varepsilon].$$

On en déduit que $\mu[A_\varepsilon] = 0$. En considérant $\varepsilon = 1/k$ et en faisant tendre k vers l'infini, on trouve que

$$\mu[\{x \in Y; |g(x)| > \|\Lambda\|\}] = 0.$$

Donc $|g| \leq \|\Lambda\|$ presque partout sur Y . Soit maintenant $f \in L^1(Y, \mu)$, on peut trouver une famille de fonctions simples $(f_k)_{k \in \mathbb{N}}$, en particulier dans $L^1 \cap L^2(Y, \mu)$, convergeant vers f dans $L^1(Y, \mu)$, et on peut alors passer à la limite :

$$\Lambda f = \lim_{k \rightarrow \infty} \Lambda f_k = \lim_{k \rightarrow \infty} \int f_k g \, d\mu = \int fg \, d\mu.$$

Par hypothèse X est union dénombrable de parties Y_k de mesure finie, que l'on peut supposer disjointes. Par le résultat précédent, pour chaque Y_k on peut trouver g_k tel que pour tout $f_k \in L^1(Y_k, \mu)$, on ait $|g_k| \leq \|\Lambda\|$ presque partout sur Y_k , et $\Lambda f_k = \int f_k g_k \, d\mu$. Posons $g := \sum g_k$; on a bien sûr $|g| \leq \|\Lambda\|$ presque partout. Soit

$f \in L^1(\mu)$, et $f_k = f1_{Y_k}$. Par convergence dominée, $f = \lim f_m$ dans $L^1(\mu)$, où $f_m = \sum_{k \leq m} f_k$. Alors

$$\begin{aligned} \Lambda f &= \lim_{m \rightarrow \infty} \Lambda f_m = \lim_{m \rightarrow \infty} \sum_{k \leq m} \Lambda f_k \\ &= \lim_{m \rightarrow \infty} \sum_{k \leq m} \int f_k g_k d\mu = \lim_{m \rightarrow \infty} \sum_{k \leq m} \int f_k g d\mu = \lim_{m \rightarrow \infty} \int f_m g d\mu = \int f g d\mu. \end{aligned}$$

La forme linéaire Λ s'identifie donc bien à la fonction bornée g , et en particulier $\|\Lambda\|_{L^\infty} \leq \|g\|_{L^\infty}$. Or on sait déjà que $\|g\|_{L^\infty} \leq \|\Lambda\|$; on a donc bien égalité des normes. \square

Pour conclure cette sous-section je vais mentionner une généralisation simple au cas des espaces L^p à valeurs vectorielles.

THÉORÈME VII-93. *Soient (X, μ) un espace mesuré, $p \in (1, \infty)$ et E un espace vectoriel normé séparable. Alors $L^p((X, \mu); E)^* = L^{p'}((X, \mu); E^*)$. La même conclusion reste valable pour $p = 1$ si X est σ -fini.*

REMARQUE VII-94. Dans le cas particulier où $E = L^p(Y, \nu)$ (Y σ -fini si $p = 1$), cet énoncé est facile car $L^p((X, \mu); L^p(Y, \nu))$ s'identifie alors à $L^p(X \times Y; \mu \otimes \nu)$ (exercice).

DÉMONSTRATION DU THÉORÈME VII-93. PREUVE À REVOIR. J'admettrai ici la Remarque VII-42. Si ξ est une forme linéaire continue sur $L^p(X; E)$, elle définit en particulier une forme linéaire sur l'ensemble des produits tensoriels $f e$, $f \in L^p(X; \mathbb{R})$, $e \in E$. Fixons $e \in E$, on pose $\alpha_e(f) = \xi(f e)$; α_e est alors une forme linéaire continue sur $L^p(X; \mathbb{R})$ puisque $|\alpha_e(f)| \leq \|\xi\| \|f\|_{L^p} \|e\|_E$. Il existe donc $g = g_e \in L^{p'}(X; \mathbb{R})$ tel que $\alpha_e(f) = \int f g_e d\mu$.

Si e et e' sont deux vecteurs de E , et $\lambda, \lambda' \in \mathbb{R}$, alors $\int g_{\lambda e + \lambda' e'} f = \xi(f(\lambda e + \lambda' e')) = \lambda \xi(f e) + \lambda' \xi(f e') = \lambda \int (g_e) f + \lambda' \int (g_{e'}) f$. Cette identité étant valable pour tout $f \in L^p(X)$, on en déduit que $g_{\lambda e + \lambda' e'} = \lambda g_e + \lambda' g_{e'}$, μ -presque partout. Quitte à écarter un ensemble $X_0 \subset X$ de mesure nulle, on peut donc supposer que cette identité est vraie pour tous $\lambda, \lambda' \in \mathbb{Q}$ et pour tous e, e' variant dans un ensemble dense $D = \{e_k; k \in \mathbb{N}\}$ fixé a priori.

Pour tous $e, e' \in D$ on a $\|g_e - g_{e'}\|_{L^{p'}} \leq C \|e - e'\|$. On en déduit que la fonction $e \rightarrow g_e$ se prolonge uniquement en une fonction continue $E \rightarrow L^{p'}$, linéaire continue en E , bien définie à un ensemble de mesure nulle près. On pose alors $\Lambda(x)(e) = g_e(x)$.

On vérifie que $\Lambda \in L^{p'}(X; E')$ et que $\Lambda(f e) = \int f g_e = \xi(f e)$; Λ s'identifie donc bien à ξ . \square

VII-4.5. Bidualité et réflexivité.

DÉFINITION VII-95. *Soit E un espace vectoriel normé. On appelle bidual de E et on note E^{**} (ou parfois E'') l'espace (de Banach) des formes linéaires continues sur E^* .*

Il s'avère que le bidual E^{**} "contient" toujours automatiquement l'espace E , et ce de manière canonique :

THÉORÈME VII-96. *Soit E un espace vectoriel normé linéairement normalisable au sens de la Définition VII-77. Alors E est isométrique à un sous-espace de E^{**} , également noté E , via l'identification $x \longrightarrow \widehat{x}$, où*

$$\widehat{x}(\Lambda) = \Lambda(x).$$

DÉMONSTRATION. Il faut vérifier que pour tout $x \in E$, $\|\widehat{x}\|_{E^{**}} = \|x\|_E$. Il est clair que

$$\forall \Lambda \in E^*, \quad |\widehat{x}(\Lambda)| = |\Lambda(x)| \leq \|\Lambda\|_{E^*} \|x\|_E,$$

ce qui montre que $\|\widehat{x}\| \leq \|x\|$. Si $x = 0$, alors $\widehat{x} = 0$. Sinon, on peut normaliser x : il existe $\Lambda \in E^*$ tel que $\|\Lambda\| = 1$ et $|\Lambda(x)| = \|x\|$. En particulier, $|\widehat{x}(\Lambda)| = \|x\|$, d'où $\|\widehat{x}\| \geq \|x\|$, ce qui conclut l'argument. \square

Dans la suite, on notera en abrégé $\widehat{x}(\Lambda) = \langle x, \Lambda \rangle$, de sorte que $\langle x, \Lambda \rangle = \langle \Lambda, x \rangle$. Quand on envisagera E comme sous-espace de E^{**} , c'est *toujours* au moyen de l'application $x \longmapsto \widehat{x}$.

L'énoncé qui suit montre que tout élément de E^{**} peut être approché, en un certain sens, par des éléments de E .

THÉORÈME VII-97 (Lemme de Goldstine). *Soient $\xi \in E^{**}$, $N \in \mathbb{N}$, $\theta_1, \dots, \theta_N$ des formes linéaires continues sur E , et $\varepsilon_1, \dots, \varepsilon_N > 0$. Alors il existe $x \in E$ tel que*

$$\|x\| \leq \|\xi\|; \quad \forall j \in \{1, \dots, N\}, \quad |\langle \xi, \theta_j \rangle - \langle x, \theta_j \rangle| \leq \varepsilon_j.$$

DÉMONSTRATION. Sans perte de généralité on peut supposer $\|\xi\| = 1$. Soit $\Phi : E \rightarrow \mathbb{R}^N$ et $z \in \mathbb{R}^N$ définis par

$$\Phi(x) := (\theta_1(x), \dots, \theta_N(x)), \quad z := (\langle \xi, \theta_1 \rangle, \dots, \langle \xi, \theta_N \rangle).$$

La conclusion à prouver est : $z \in \overline{\Phi(B_1(0))}$. Comme $C := \Phi(B_1(0))$ est un convexe de \mathbb{R}^N , la conclusion pourra être obtenue grâce au Théorème VII-85.

Soit Λ une forme linéaire sur \mathbb{R}^N : elle s'écrit comme le produit scalaire par un vecteur $(\alpha_1, \dots, \alpha_N)$. D'où

$$\Lambda(z) = \sum \alpha_j \langle \xi, \theta_j \rangle = \langle \xi, \sum \alpha_j \theta_j \rangle \leq \left\| \sum \alpha_j \theta_j \right\|.$$

D'autre part,

$$\sup_{y \in C} \Lambda(y) = \sup_{\|x\| < 1} \sum \alpha_j \theta_j(x) = \sup_{\|x\| \leq 1} \left(\sum \alpha_j \theta_j \right)(x) = \left\| \sum \alpha_j \theta_j \right\|.$$

Donc $\Lambda(z) \leq \sup_C \Lambda$; comme Λ est arbitraire on déduit du Théorème VII-85 que $z \in \overline{C}$. \square

REMARQUE VII-98. Cette preuve ne nécessite ni la séparabilité, ni l'axiome du choix.

REMARQUE VII-99. On pourra plus tard interpréter le lemme de Goldstine en disant que E est dense dans E^{**} au sens de la topologie faible-* de E^{**} .

C'est le moment d'introduire la propriété de **réflexivité** :

DÉFINITION VII-100. *Un espace de Banach E est dit réflexif si $E^{**} = E$.*

EXEMPLE VII-101. Tout espace de dimension finie est réflexif. Le Théorème VII-87 montre que les espaces L^p sont réflexifs pour $1 < p < \infty$ (car $(p')' = p$), mais ne dit rien de tel sur les espaces L^1 et L^∞ . L'espace $c_0 = C_0(\mathbb{N})$ des suites qui tendent vers 0 à l'infini n'est pas réflexif : son dual est l'espace $\ell^1 = L^1(\mathbb{N})$ des suites sommables (c'est un cas particulier du théorème de Riesz, en effet une mesure finie sur \mathbb{N} s'identifie à une fonction intégrable); l'espace $(c_0)^{**} = (\ell^1)^*$ contient donc toutes les suites bornées, et pas seulement celles qui tendent vers 0.

Le critère le plus utilisé pour démontrer la réflexivité est le remarquable théorème de Milman–Pettis :

THÉORÈME VII-102 (Théorème de Milman–Pettis). *Tout espace de Banach uniformément convexe est réflexif.*

EXEMPLE VII-103. En appliquant ce théorème aux espaces de Lebesgue et en utilisant le Théorème VII-53, on retrouve la réflexivité des espaces de Lebesgue L^p pour $1 < p < \infty$, sans avoir eu besoin d'appliquer le Théorème de représentation (Théorème VII-87).

DÉMONSTRATION. Soient $\xi \in E^{**}$; il s'agit de montrer qu'il existe $x \in E$ tel que $\xi = \hat{x}$ (on notera simplement $\xi = x$). Sans perte de généralité, on peut supposer $\|\xi\| = 1$. L'espace E étant fermé, il suffit de montrer que $\xi \in \overline{E}$, c'est-à-dire que pour tout $\varepsilon > 0$ il existe $x \in E$ avec $\|\xi - x\| \leq \varepsilon$.

Soit donc $\varepsilon > 0$, et $\eta := \kappa(\varepsilon/4)$, où κ est le module de convexité de E , comme défini dans la sous-section VII-2.3. Puisque $\|\xi\| = 1$, on peut trouver $\Lambda \in E^*$ tel que $\|\Lambda\| = 1$ et

$$(69) \quad \langle \xi, \Lambda \rangle \geq 1 - \frac{\eta}{2}.$$

D'après le Lemme VII-97 (de Goldstine), il existe $x \in E$ tel que $\|x\| \leq \|\xi\| = 1$ et

$$(70) \quad |\langle x, \Lambda \rangle - \langle \xi, \Lambda \rangle| < \frac{\eta}{2}.$$

Si $\|x - \xi\| \leq \varepsilon$, la démonstration est terminée.

Sinon, c'est que $\|x - \xi\| > \varepsilon$; par définition de la norme dans E^{**} on peut donc trouver une forme linéaire Λ' de norme 1 telle que

$$(71) \quad |\langle x, \Lambda' \rangle - \langle \xi, \Lambda' \rangle| > \varepsilon.$$

En appliquant le Lemme de Goldstine de nouveau, on peut trouver $y \in E$ tel que $\|y\| \leq 1$ et

$$(72) \quad |\langle y, \Lambda \rangle - \langle \xi, \Lambda \rangle| < \frac{\eta}{2}, \quad |\langle y, \Lambda' \rangle - \langle \xi, \Lambda' \rangle| < \frac{\varepsilon}{2}.$$

En combinant cela avec (70) on en déduit

$$\left| \left\langle \frac{x+y}{2}, \Lambda \right\rangle - \langle \xi, \Lambda \rangle \right| \leq \frac{1}{2} \left(|\langle x, \Lambda \rangle - \langle \xi, \Lambda \rangle| + |\langle y, \Lambda \rangle - \langle \xi, \Lambda \rangle| \right) < \frac{\eta}{2}.$$

Par (69) il résulte

$$\left\| \frac{x+y}{2} \right\| \geq \left\langle \frac{x+y}{2}, \Lambda \right\rangle \geq \langle \xi, \Lambda \rangle - \frac{\eta}{2} > 1 - \eta = 1 - \kappa(\varepsilon/4).$$

Par définition du module d'uniforme convexité, on en déduit que $\|x - y\| \leq \varepsilon/2$. En particulier, grâce à (72),

$$|\langle \xi, \Lambda' \rangle - \langle x, \Lambda \rangle| \leq |\langle \xi, \Lambda' \rangle - \langle y, \Lambda' \rangle| + |\langle x - y, \Lambda \rangle| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

On obtient une contradiction avec (71), ce qui montre que l'hypothèse $\|x - \xi\| > \varepsilon$ était absurde. \square

VII-5. Convergence et compacité

VII-5.1. Convergences faibles et fortes : Définitions et propriétés fondamentales. Soit E un espace vectoriel normé ; la norme définit naturellement une structure métrique sur E , donc une topologie. Par convention, on l'appelle **topologie forte**, et on dit qu'une suite x_n à valeurs dans E converge **fortement** vers $x \in E$ si elle converge vers x au sens de la norme ($\|x_n - x\| \rightarrow 0$).

Soit alors $\Lambda \in E^*$ une forme linéaire continue sur E ; par continuité de Λ , il est clair que $\Lambda x_n \rightarrow \Lambda x$. C'est cette dernière propriété qui est utilisée pour définir les convergences faibles : on considère les propriétés de convergence "à travers certaines formes linéaires continues". Pour que cette notion de convergence soit acceptable, et en particulier qu'il y ait unicité de la limite éventuelle, il est important que l'on puisse "distinguer les points" au moyen de ces formes linéaires. La notion de séparabilité linéaire, introduite dans la Définition VII-77 est alors naturelle.

DÉFINITION VII-104. Soient E un espace vectoriel normé, et F un sous-espace de E^* qui sépare E . On dit qu'une suite (x_n) à valeurs dans E converge vers $x \in E$ au sens faible $\sigma(E, F)$ si

$$\forall \Lambda \in F, \quad \langle \Lambda, x_n \rangle \xrightarrow{n \rightarrow \infty} \langle \Lambda, x \rangle.$$

REMARQUE VII-105. Il est clair que la convergence forte implique la convergence faible.

REMARQUE VII-106. Soit \tilde{F} une partie de F qui engendre un espace vectoriel dense (dense dans F au sens de la norme duale sur E^*), et soit $(x_n)_{n \in \mathbb{N}}$ une suite à valeurs dans E , bornée. Alors il est équivalent de dire que x_n converge vers x au sens $\sigma(E, F)$ ou que

$$\forall \tilde{\Lambda} \in \tilde{F}, \quad \tilde{\Lambda} x_n \xrightarrow{n \rightarrow \infty} \tilde{\Lambda} x.$$

En effet, si $\Lambda \in F$ et $\varepsilon > 0$ est donné, on peut trouver $\tilde{\Lambda}$ dans l'espace vectoriel engendré par \tilde{F} avec $\|\tilde{\Lambda} - \Lambda\| \leq \varepsilon$, et alors, par linéarité,

$$|\Lambda x_n - \Lambda x| \leq |\tilde{\Lambda} x_n - \tilde{\Lambda} x| + \varepsilon \|x_n - x\|.$$

Le premier terme du membre de droite tend alors vers 0 car $\tilde{\Lambda}$ appartient à l'espace vectoriel engendré par les éléments de \tilde{F} , tandis que le deuxième est majoré par $2C\varepsilon$, où C est un majorant de $\|x_n\|$ et $\|x\|$. Il s'ensuit que Λx_n tend vers Λx .

Comme conséquence immédiate de la définition, une limite au sens faible $\sigma(E, F)$ est unique. Sous une hypothèse additionnelle de séparabilité sur F , il va s'avérer que les convergences faibles sont en fait métriques.

PROPOSITION VII-107. Soient E un espace vectoriel normé, et F un sous-espace séparable de E^* qui sépare les points. Soit $(\Lambda_k)_{k \geq 1}$ une suite dense dans la boule unité de F , et soit

$$d_{\sigma(E,F)}(x, y) := \sum_{k \geq 1} \frac{|\langle \Lambda_k, y - x \rangle|}{k^2}.$$

Alors $d_{\sigma(E,F)}$ est une distance, qui métrise la convergence faible $\sigma(E, F)$ sur toute boule $B_R(0) \subset E$.

DÉMONSTRATION. Notons $d := d_{\sigma(E,F)}$. Tout d'abord, puisque les Λ_k sont tous de norme au plus 1, on a $|\langle \Lambda_k, y - x \rangle| \leq \|x - y\|$, et donc la série définissant $d(x, y)$ est convergente. Il est clair que $d(x, y) = d(y, x)$ et que d vérifie l'inégalité triangulaire. Vérifions maintenant que si x_n est une suite à valeurs dans $B_R(0)$, alors

$$d(x_n, x) \xrightarrow{0} \text{ si et seulement si } \Lambda x_n \rightarrow \Lambda x \text{ pour tout } \Lambda \in F.$$

Il en résultera bien l'équivalence entre les deux notions de convergence, et cela montrera par la même occasion que $d(x, y) = 0$ implique $x = y$.

D'une part, si $k \in \mathbb{N}$ est fixé, on a $|\Lambda_k(x_n - x)| \leq k^2 d(x_n, x)$, qui tend vers 0 quand $n \rightarrow \infty$ si $d(x_n, x) \rightarrow 0$. Par densité, $\Lambda x_n \rightarrow \Lambda x$ pour tout Λ dans la boule unité de F . Si $\Lambda \in F$ est non nul, on peut appliquer ce résultat à $\Lambda/\|\Lambda\|$, et on en déduit que $\Lambda x_n \rightarrow \Lambda x$, comme on le souhaitait.

Réciproquement, supposons que $\Lambda_k(y - x) \rightarrow 0$ quand $n \rightarrow \infty$. Soit $\varepsilon > 0$, et soit K tel que $(2R) \sum_{k \geq K} k^{-2} \leq \varepsilon$; pour tout $k \leq K$ on peut trouver N_k tel que $\|\Lambda_k(x_n - x)\| \leq \varepsilon$. On choisit alors $N := \sup\{N_k; k \leq K\}$, et on a, pour $n \geq N$,

$$d(x_n, x) \leq \sum_{k \leq K} \frac{\varepsilon}{k^2} + \varepsilon.$$

La suite x_n converge donc vers x au sens de la distance d . □

Par convention, la convergence $\sigma(E, E^*)$ est appelée **convergence faible** sur E , et la convergence $\sigma(E^*, E)$ est appelée **convergence faible-étoile** (faible-*) sur E^* . On distingue donc trois notions classiques de convergence :

- la **convergence forte**, ou normique : $x_n \rightarrow x$ fortement si $\|x_n - x\| \rightarrow 0$; cette convergence est donc associée à la norme;

- la **convergence faible** sur E : $x_n \rightarrow x$ faiblement (ce que l'on note parfois $x_n \rightharpoonup x$) si pour tout $\Lambda \in E^*$, $\Lambda x_n \rightarrow \Lambda x$; dès que E^* est séparable, cette convergence est associée à une distance sur les parties bornées de E ;

- la **convergence faible-*** sur un dual E^* : $\Lambda_n \rightarrow \Lambda$ au sens faible-* (ce que l'on note parfois $x_n \rightharpoonup_* x$) si pour tout $x \in E$, $\Lambda_n x \rightarrow \Lambda x$; dès que E est séparable, cette convergence est associée à une distance sur les parties bornées de E^* .

Cette restriction à des parties bornées n'est pas gênante; il existe deux bonnes raisons pour ne pas considérer l'espace tout entier. La première est que, sous une hypothèse de complétude, les suites faiblement convergentes sont automatiquement bornées; la seconde est qu'il n'y a pas, en général, de distance associée à la convergence faible sur l'espace E tout entier. C'est le contenu de la proposition suivante.

PROPOSITION VII-108. Soit E un espace vectoriel normé.

(i) Si E est complet, toute suite de E^* convergeant au sens faible-* est automatiquement bornée;

(ii) Si E est séparable, toute suite de E convergeant au sens faible est automatiquement bornée ;

(iii) Dans ce cas, soit la convergence faible-* (resp. faible) équivaut à la convergence forte, soit elle n'est associée à aucune distance.

REMARQUE VII-109. La convergence faible est équivalente à la convergence forte en dimension finie (pour le voir, il suffit de considérer les formes linéaires coordonnées), mais également dans un espace tel que $\ell^1(\mathbb{N})$ (c'est une conséquence du Théorème de Schur ci-dessous). Mais dans la grande majorité des cas, c'est la deuxième possibilité qui prévaut dans l'énoncé (ii) : la convergence faible n'est associée à aucune distance.

DÉMONSTRATION. 1. Considérons d'abord une suite $(\Lambda_n)_{n \in \mathbb{N}}$ dans E^* , convergeant au sens faible-* vers Λ . On suppose E complet. Alors, par le Théorème de Banach–Steinhaus, la famille des Λ_n est bornée.

2. Soit maintenant (x_n) une suite convergeant au sens faible dans E . Pour tout $\Lambda \in E^*$, l'ensemble $\{\Lambda x_n\}_{n \in \mathbb{N}}$ est donc borné. En identifiant les x_n à des éléments de E^{**} , on peut appliquer le théorème de Banach–Steinhaus dans E^* (qui est toujours complet) pour conclure à l'existence d'une constante C telle que

$$\sup_{n \in \mathbb{N}} \sup_{\Lambda \in E^*} |\Lambda x_n| \leq C \|\Lambda\|.$$

Si x_n est non nul, on peut le normaliser par une forme linéaire Λ_n de norme 1 (c'est ici qu'intervient l'hypothèse de séparabilité, via le Théorème ??), d'où

$$\|x_n\| = \|\Lambda_n x_n\| \leq C \|\Lambda_n\| = C.$$

La suite (x_n) est donc bien bornée.

3. Dans la suite on considère par exemple la convergence faible ; le même raisonnement fonctionnerait avec la convergence faible-*. Supposons que la convergence faible n'équivaut pas à la convergence forte ; on peut donc trouver une suite x_n qui converge faiblement vers x , avec $\limsup \|x_n - x\| > 0$. Quitte à extraire une sous-suite, on peut donc supposer que $\|x_n - x\| \geq \delta > 0$; on pose alors $y_n := (x_n - x)/\|x_n - x\|$, de sorte que $\|y_n\| = 1$, et pour tout $\Lambda \in E^*$,

$$|\Lambda y_n| \leq \delta^{-1} |\Lambda x_n - \Lambda x| \xrightarrow{n \rightarrow \infty} 0.$$

La suite y_n est donc une suite normée qui converge faiblement vers 0.

Soit maintenant, pour m, n entiers non nuls, $u_{m,n} := x_m + m x_n$, et $A := \{u_{m,n} ; m, n \geq 1\}$. Bien sûr $\|u_{m,n}\| \geq m - 1$; si $u_{m,n}$ converge vers un élément v , on a donc forcément m borné, et quitte à extraire une sous-suite on peut supposer que m est constant. Alors, soit n est borné, et quitte à extraire une sous-suite on peut supposer que n est constant ; soit n est non borné, et quitte à extraire une sous-suite on peut supposer que $n \rightarrow \infty$. Dans le premier cas, $v = u_{m,n}$, dans le deuxième cas $v = x_m$. L'ensemble $\lim A$ de toutes les limites de suites d'éléments de A est donc exactement constitué de A et des x_m . Si la convergence faible était associée à une distance, l'ensemble $\lim A$ serait l'adhérence \overline{A} de A pour cette distance, et en particulier il serait fermé au sens faible ; mais ce n'est pas le cas puisque la suite x_m converge faiblement vers 0, qui n'appartient pas à $\lim A$. \square

Voici une conséquence simple des résultats précédents :

PROPOSITION VII-110. Soit E un espace complet, on suppose que $\Lambda_n \longrightarrow \Lambda$ au sens faible-* dans E^* et $x_n \longrightarrow x$ dans E ; alors

$$\Lambda_n x_n \longrightarrow \Lambda x.$$

La même conclusion reste vraie si E est séparable, $\Lambda_n \longrightarrow \Lambda$ au sens fort dans E^* et $x_n \longrightarrow x$ au sens faible dans E .

DÉMONSTRATION. Traitons le premier cas, le deuxième se démontrant de même. Comme E est complet, on sait que Λ_n est bornée : il existe C tel que $\sup \|\Lambda_n\| \leq C$. On a alors

$$|\Lambda_n x_n - \Lambda x| = |\Lambda_n(x_n - x) + (\Lambda_n - \Lambda)x| \leq |\Lambda_n(x_n - x)| + |(\Lambda_n - \Lambda)x| \leq C\|x_n - x\| + |(\Lambda_n x - \Lambda x)|.$$

On conclut en notant que, par hypothèse, $\|x_n - x\| \rightarrow 0$, et $\Lambda_n x \rightarrow \Lambda x$. \square

REMARQUE VII-111 (topologies). Les convergences faibles ne sont pas toujours associées à des distances, mais elles sont toujours associées à des topologies. Si E est un espace vectoriel normé, et F un sous-espace de E^* , on définit la topologie faible $\sigma(E, F)$ comme la topologie la plus grossière qui rende continus tous les éléments de F . Les ouverts de $\sigma(E, F)$ peuvent se décrire ainsi : ce sont les unions (quelconques) d'intersections finies d'ensembles de la forme $\Lambda^{-1}(] \alpha, +\infty[)$ ($\Lambda \in F$, $\alpha \in \mathbb{R}$). La topologie $\sigma(E, E^*)$ est aussi appelée topologie faible et parfois notée $w - E$ [le “w” signifiant “weak”] ; alors que la topologie $\sigma(E^*, E)$ est appelée topologie faible-*, et parfois notée $w * - E$.

Si E est de dimension finie, la topologie faible coïncide avec la topologie forte. En revanche, si E est de dimension infinie, elle est strictement plus grossière que la topologie forte : en effet, on peut montrer [Brézis, p.37] que l'adhérence de la sphère unité ($\|x\| = 1$) est la boule unité fermée entière ($\|x\| \leq 1$), et que la boule unité ouverte ($\|x\| < 1$) n'est jamais ouverte. Cela traduit le fait que les voisinages de la topologie faible sont extrêmement grossiers : ils ne tiennent compte que d'un nombre fini de formes linéaires (un nombre fini de coordonnées). Cela n'est pas contradictoire avec le fait que la convergence faible coïncide parfois avec la topologie forte : cela montre seulement que les topologies faibles et faible-* ne sont jamais métrisables en dimension infinie.

L'approche topologique est développée dans de nombreux ouvrages tels que [Brézis] ; cependant, c'est bien la notion de convergence faible qui sert en pratique. Dans le même ordre d'idées, c'est la notion de compacité faible séquentielle qui est utile, bien plus que la notion de compacité faible abstraite.

VII-5.2. Convergence faible et forte dans les espaces classiques. Après avoir défini des concepts généraux, nous allons les mettre en oeuvre sur les espaces de Lebesgue et les espaces de mesures.

Soit (X, μ) un espace mesuré σ -fini.

- Dire que f_n converge faiblement vers f dans $L^1(X, \mu)$, c'est dire que pour tout $g \in L^\infty(X, \mu)$, on a $\int f_n g d\mu \longrightarrow \int f g d\mu$. Cette notion de convergence n'est en général associée à aucune métrique, comme le montre l'Exercice VII-135 ci-dessous.

- Pour $1 < p < \infty$, dire que f_n converge f au sens faible, ou au sens faible-étoile, dans $L^p(X, \mu)$, c'est dire que pour tout $g \in L^{p'}(X, \mu)$ on a $\int f_n g d\mu \longrightarrow \int f g d\mu$. Si X est un espace métrique séparable et μ une mesure de Borel, cette notion de

convergence est associée à une famille de métriques définies sur les boules de L^p . (En effet, la séparabilité de $E = L^p$ garantit que les suites convergentes sont bornées ; et la séparabilité de $E' = L^{p'}$ garantit que la convergence faible est métrisable sur les parties bornées.)

- Dire que f_n converge au sens faible-étoile vers f dans $L^\infty(X, \mu)$, c'est dire que pour tout $g \in L^1(X, \mu)$ on a $\int f_n g d\mu \longrightarrow \int f g d\mu$. Si X est un espace métrique séparable et μ une mesure de Borel, cette notion de convergence est associée à une métrique. (En effet, la complétude de L^1 garantit que les suites convergeant au sens faible-* dans L^∞ sont bornées ; et la séparabilité de L^1 garantit que la convergence faible-* est métrisable sur les parties bornées.)

- Nous n'avons rencontré aucun espace dont L^1 soit le dual (en fait il est impossible de démontrer l'existence d'un tel espace, qui d'ailleurs serait nécessairement non séparable) ; il n'y a donc pas lieu de parler de convergence faible-étoile dans L^1 . Quant à la convergence faible dans L^∞ , en pratique elle n'est jamais utilisée.

- Supposons maintenant que X est localement compact ; alors, par le Théorème de Riesz on sait que l'espace $M(X)$ des mesures de Radon finies s'identifie au dual de $C_0(X)$. Alors, dire qu'une famille de mesures μ_n converge au sens faible-* vers μ , c'est dire que pour tout $f \in C_0(X)$ on a

$$\int f d\mu_n \longrightarrow \int f d\mu.$$

Si X est en outre σ -compact, on sait par le Théorème VII-36(ii) que $C_0(X)$ est séparable, et la convergence faible-* est une convergence métrique.

- En théorie des probabilités, on utilise couramment une autre notion de convergence faible, qui ne rentre pas dans la liste précédente : Soit (X, μ) un espace métrique, on dit que μ_n converge faiblement vers μ si pour tout $f \in C_b(X)$ on a $\int f d\mu_n \longrightarrow \int f d\mu$. Il s'agit donc de la convergence $\sigma(M(X), C_b(X))$, qu'en pratique on applique à des suites de mesures de probabilités. Malgré la similitude des notations, il ne faut pas s'y méprendre : ce n'est pas une convergence faible au sens où l'espace $C_b(X)$ n'est pas en général le dual de $M(X)$ (toute fonction bornée, non nécessairement continue, appartient à cet espace dual). Cette notion est acceptable au sens où l'espace $C_b(X)$ sépare $M(X)$ d'après l'Exemple VII-82. Comme l'espace $C_b(X)$ n'est en général pas séparable, il n'est pas clair a priori que cette notion de convergence soit métrique ; c'est pourtant le cas, et d'ailleurs il existe quantité de distances "naturelles" sur l'espace des mesures de probabilité, qui métrisent la convergence faible des mesures. Nous en reparlerons dans le paragraphe VII-5.7.

Il y a certaines implications entre ces différentes notions :

PROPOSITION VII-112. (i) Soit (X, μ) un espace de mesure finie. Alors la convergence faible dans L^p (faible-étoile pour $p = \infty$) implique la convergence faible dans L^q dès que $p \geq q$.

(ii) Soit (X, μ) un espace dénombrable tel que $\mu[\{x\}] \geq \varepsilon > 0$ pour tout x . Alors la convergence faible dans L^p implique la convergence faible dans L^q (faible-étoile pour $q = \infty$) dès que $p \leq q$.

(iii) Soit (X, μ) un espace mesuré. Si f_n converge faiblement dans L^1 vers f , alors $f_n \mu$ converge vers $f \mu$ faiblement, où cette convergence est à comprendre au sens faible-étoile si X est localement compact, ou au sens faible si X est métrique.

Si X est une espace métrique localement compact, la convergence faible au sens des mesures implique la convergence faible-étoile au sens des mesures.

DÉMONSTRATION. Les parties (i) et (ii) de ce théorème sont des conséquences directes des relations d'emboîtement entre espaces de Lebesgue, décrites dans les Théorèmes VI-25 et VI-26. La partie (iii) découle des inclusions $C_0(X) \subset C_b(X) \subset L^\infty(X)$. \square

EXEMPLES VII-113. On se place maintenant en dimension 1, avec la mesure de Lebesgue comme mesure de référence. Il est assez facile de se représenter des suites convergeant fortement dans L^p : par exemple, si $p \in [1, +\infty[$ et (f_n) est une suite de fonctions convergeant presque partout vers f , dominée par une fonction fixée dans L^p , alors f_n converge fortement vers f . Quels sont les exemples typiques de convergence faible non forte ?

- On considère les fonctions $f_n(x) = \sin(2\pi nx)$ sur $[0, 1]$. Cette suite de fonctions est bien sûr bornée uniformément, donc dans tous les espaces L^p . Pour tout intervalle $[a, b]$, il est facile de vérifier que $\int_a^b f_n(x) dx \rightarrow 0$ quand $n \rightarrow \infty$. Comme l'espace vectoriel engendré par les fonctions indicatrices est dense dans $L^{p'}$ pour tout $p' \in [1, +\infty[$, i.e. $p \in]1, +\infty]$, on en déduit en appliquant la Remarque VII-106 que f_n converge faiblement vers 0 dans L^p faible pour $1 < p < \infty$, et dans L^∞ faible-étoile. Comme l'espace $[0, 1]$ est de mesure finie, la convergence L^p faible ($p \geq 1$) implique qu'il y a également convergence dans L^1 faible. Si λ désigne la mesure de Lebesgue sur \mathbb{R} , on a donc également convergence de $f_n\lambda$ vers 0 aux sens faible et faible-étoile des mesures sur $[0, 1]$.

- On considère les fonctions $f_n(x) = 1_{[n, n+1]}$. Cette suite est encore bornée dans tous les espaces L^p . Un raisonnement similaire à l'argument précédent montre que f_n converge vers 0 dans L^p faible pour $1 \leq p < \infty$, et dans L^∞ faible-étoile. En revanche, elle ne converge pas dans L^1 faible, puisque $\int f_n = 1 \neq 0 = \int 0$. La famille $f_n\lambda$ converge vers 0 au sens faible-étoile des mesures, mais pas au sens faible (comme on le voit encore en intégrant la fonction constante 1, qui n'appartient pas à $C_0(\mathbb{R})$ mais appartient à $C_b(\mathbb{R})$).

- On considère la suite de fonctions $f_n(x) = 2n1_{[-n^{-1}, n^{-1}]}$ ($n \geq 1$). Alors la famille f_n est bornée dans $L^1([0, 1])$, mais dans aucun autre espace de Lebesgue. Son intégrale contre la fonction indicatrice $1_{[a, b]}$ tend vers 0 si $0 \notin [a, b]$. Une limite éventuelle dans L^1 faible devrait donc être nulle presque partout en-dehors de 0, donc nulle presque partout. Mais cela est impossible puisque $\int f_n = 1$. On en déduit que f_n ne converge pas dans L^1 faible. En revanche, $f_n\lambda$ converge aux sens faible et faible-étoile des mesures vers la masse de Dirac δ .

- On considère la suite de fonctions $f_n(x) = 2n^{-1}1_{[-n, n]}$ sur \mathbb{R} ($n \geq 1$). La famille f_n est bornée dans tous les espaces L^p pour $p \geq 1$, et son intégrale sur tout intervalle $[a, b]$ tend vers 0. On en déduit, comme plus haut, que f_n converge vers 0 dans L^p faible pour tout $p > 1$ (faible-étoile pour $p = \infty$). Il est également facile de vérifier que $f_n\lambda$ converge au sens faible-étoile, mais pas au sens faible dans l'espace des mesures.

- La famille de fonctions $f_n(x) = 1_{[-n, n]}$ converge dans L^∞ au sens faible-étoile vers la fonction constante 1.

- Soit, dans \mathbb{R} , une suite x_n convergeant vers x . Alors la famille δ_{x_n} converge vers δ_x au sens faible (et donc faible-étoile).

- La famille de mesures $\sum_{1 \leq k \leq n} \delta_{k/n}$ converge au sens faible vers la mesure de Lebesgue sur $[0, 1]$. En fait, n'importe quelle mesure de probabilité sur $\mathcal{B}(\mathbb{R})$ peut être approchée au sens faible par une combinaison convexe de masses de Dirac : on peut le voir comme une conséquence du Théorème de Krein–Milman, que nous rencontrerons plus tard.

- Tous ces exemples admettent des généralisations faciles aux dimensions supérieures. On peut d'ailleurs construire de nombreux exemples par tensorisation : en utilisant, pour $1 < p \leq \infty$, l'identité $L^{p'}(\mathbb{R}^{n+m}) = L^{p'}(\mathbb{R}^n) \otimes L^{p'}(\mathbb{R}^m)$ (Cf. Théorème VI-20) on vérifie que si les suites f_k et g_k convergent faiblement vers f et g respectivement dans $L^p(\mathbb{R}^n)$ et $L^p(\mathbb{R}^m)$, alors $f_k \otimes g_k$ converge faiblement dans $L^p(\mathbb{R}^{n+m})$ vers $f \otimes g$. Il s'agit là d'un cas particulier de théorèmes plus généraux que nous mentionnerons après avoir étudié quelques critères de compacité faible.

REMARQUE VII-114. L'espace \mathbb{R}^n est muni d'une structure différentiable, et on peut considérer d'autres classes de “fonctions-test” (c'est-à-dire les fonctions contre lesquelles on intègre pour définir la convergence faible) que les classes de fonctions L^p ou les fonctions continues : par exemples, les fonctions continûment différentiable, ou k fois continûment différentiables, etc. En **théorie des distributions**, on considère la classe test très restreinte des *fonctions infiniment dérivables à support compact* pour définir la convergence au sens des distributions. Toutes les convergences que nous avons rencontré jusqu'à présent sont plus fortes que la convergence au sens des distributions.

VII-5.3. Compacité faible et forte.

DÉFINITION VII-115. Soit E un espace vectoriel normé, et F une partie de E^* qui sépare E . On dit que $C \subset E$ est séquentiellement $\sigma(E, F)$ -compact si de toute suite à valeurs dans C on peut extraire une sous-suite qui converge dans C au sens faible $\sigma(E, F)$.

On dit que $C \subset E$ est relativement séquentiellement $\sigma(E, F)$ -compact si de toute suite à valeurs dans C on peut extraire une sous-suite qui converge dans E au sens faible $\sigma(E, F)$.

CONVENTION VII-116. Pour alléger les notations, on remplacera, en s'inspirant de la terminologie anglo-saxonne, “relativement séquentiellement compact” par **précompact** ; et on précisera **précompact faible** (ou w -précompact) si la topologie considérée est la topologie faible, et **précompact faible-étoile** (ou w^* -précompact) si c'est la topologie faible-étoile.

REMARQUE VII-117. Pour une topologie non métrique, les notions de compacité séquentielle et de compacité ne coïncident pas forcément. Cependant, en pratique c'est (quasiment toujours) la compacité séquentielle qui est utile, et c'est donc la seule notion qui sera développée ici.

Nous avons rencontré de nombreux exemples où une suite bornée x_n converge faiblement vers une limite x sans converger fortement ; et même sans qu'aucune suite extraite ne converge fortement. Être borné n'est donc pas en général une condition suffisante pour la compacité relative. Le **théorème de non-compacité de Riesz** assure que c'est la règle en dimension infinie.

THÉORÈME VII-118 (non-compacité en dimension infinie). Dans un espace vectoriel normé de dimension infinie, les boules fermées ne sont pas compactes.

DÉMONSTRATION. Soit E un espace vectoriel normé de dimension infinie (c'est-à-dire qui n'est pas de dimension finie). Nous allons construire par récurrence une suite $(x_n)_{n \in \mathbb{N}}$ de vecteurs de norme 1, tels que $\|x_n - x_m\| \geq 1/2$ pour tous $m \neq n$; il s'ensuivra que (x_n) n'admet aucune sous-suite extraite convergente.

Supposons donc x_1, \dots, x_n construits; on note F_n l'espace vectoriel engendré par ces vecteurs. Puisque E n'est pas de dimension finie, il existe $y \notin F_n$. Comme F_n est fermé, la distance de y à F_n est atteinte : il existe $\tilde{y} \in F_n$ tel que $\|y - \tilde{y}\| \leq \|y - z\|$ pour tout $z \in F_n$; \tilde{y} est bien sûr distinct de y . On pose alors $x_{n+1} := (y - \tilde{y})/\|y - \tilde{y}\|$. Alors, pour tout $x \in F_n$ on a $\tilde{y} + \|y - \tilde{y}\|x \in F_n$, donc

$$\|x_{n+1} - x\| = \frac{\|y - \tilde{y} - \|y - \tilde{y}\|x\|}{\|y - \tilde{y}\|} \geq 1.$$

□

En fait, en dimension infinie, il n'y a pas vraiment de critères généraux pour garantir la précompacité, que ce soit pour la topologie forte ou pour la topologie faible; à l'exception notable de la précompacité faible-étoile dans le dual d'un espace séparable. Dans les deux sous-sections suivantes, on recensera les quelques critères qui sont utilisés en pratique.

Un concept souvent utilisé par commodité est celui d'**injection compacte**. Rappelons qu'un espace vectoriel normé (\tilde{E}, \tilde{N}) s'injecte continûment dans un espace vectoriel normé (E, N) s'il existe une constante C telle que $N \leq C\tilde{N}$; les parties bornées de \tilde{E} sont alors bornées dans E . On dit que l'injection est *compacte* si les parties bornées de \tilde{E} sont non seulement bornées, mais en outre précompactes dans E . On notera bien qu'un espace peut s'injecter à la fois compactement et densément dans un autre.

Pour obtenir des résultats de précompacité, on utilise souvent le critère qui suit, dont la démonstration est laissée en exercice :

PROPOSITION VII-119. *Soit A un sous-ensemble d'un espace métrique (E, d) . On suppose que pour tout $x \in A$ il existe une famille $R_n(x)$ ($n \in \mathbb{N}$) à valeurs dans E , telle que*

$$(i) \sup_{x \in A} d(x, R_n(x)) \longrightarrow [n \rightarrow \infty] 0;$$

$$(ii) \text{ pour tout } n \in \mathbb{N}, A_n := \{R_n(x), x \in A\} \text{ est précompact dans } E.$$

Alors A est précompact.

VII-5.4. Critères de précompacité forte. Le plus célèbre des critères de précompacité forte est le **théorème d'Ascoli**, qui s'applique aux espaces de fonctions continues. Il sera naturel pour nous de l'énoncer pour l'espace $C_0(X)$, où X est localement compact.

THÉORÈME VII-120 (théorème de compacité d'Ascoli). *Soit (X, d) un espace métrique localement compact, σ -compact. Alors, une partie \mathcal{A} de $C_0(X)$ est précompacte si et seulement si*

$$(i) \text{ Pour tout } x \in X, \text{ l'ensemble } \{f(x); f \in \mathcal{A}\} \text{ est borné;}$$

$$(ii) \mathcal{A} \text{ tend vers } 0 \text{ uniformément à l'infini : pour tout } \varepsilon > 0 \text{ il existe un compact } K_\varepsilon \text{ tel que pour tous } f \in \mathcal{A}, x \in X,$$

$$x \notin K_\varepsilon \implies |f(x)| \leq \varepsilon.$$

(iii) \mathcal{A} est localement équicontinu : pour tout compact K et pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour tous $f \in \mathcal{A}$, $x, y \in K$,

$$d(x, y) \leq \delta \implies |f(x) - f(y)| \leq \varepsilon.$$

En outre, dans cet énoncé on peut remplacer les hypothèses (i) et (iii) par les hypothèses apparemment plus fortes

$$(i') \sup_{f \in \mathcal{A}} \|f\|_{\infty} < +\infty ;$$

(iii') \mathcal{A} est (globalement) équicontinu : pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que pour tous $f \in \mathcal{A}$, $x, y \in K$,

$$d(x, y) \leq \delta \implies |f(x) - f(y)| \leq \varepsilon.$$

REMARQUES VII-121. L'hypothèse (ii) est bien sûr vide si X est compact. Quant à l'hypothèse (iii'), on peut la reformuler en disant que \mathcal{A} admet un **module de continuité uniforme** : il existe une fonction $m : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, croissante, $m(\delta) \rightarrow 0$ quand $\delta \rightarrow 0$, telle que pour tout $f \in \mathcal{A}$,

$$d(x, y) \leq \delta \implies |f(x) - f(y)| \leq m(\delta).$$

Toute fonction continue admet un module de continuité, l'important est qu'il existe un module *commun* pour tous les éléments de \mathcal{A} . Notons enfin qu'il existe des versions plus générales où l'espace d'arrivée est par exemple un espace métrique localement compact.

DÉMONSTRATION DU THÉORÈME VII-120. 1. Il est clair que la condition (i') (et donc la condition (i) aussi) est nécessaire : un ensemble précompact dans $C(X)$ est d'adhérence compacte, donc borné.

3. Supposons \mathcal{A} précompact et montrons par l'absurde que (ii) est vérifié. Si ce n'est pas le cas, il existe $\varepsilon > 0$ fixé, tel que pour tout compact K on puisse trouver $f \in \mathcal{A}$ et $x \notin K$ avec $|f(x)| \geq \varepsilon$. Soit (K_n) une suite exhaustive de compacts ; par récurrence on peut construire une suite $(f_n)_{n \in \mathbb{N}}$ à valeur dans \mathcal{A} , et une suite $(x_n)_{n \in \mathbb{N}}$, à valeurs dans X , telle que $x_n \notin K_n$ et $|f_n(x_n)| \geq \varepsilon$. Supposons que (f_n) admette une suite extraite convergeant vers $f \in C_0(X)$. Il existe un compact K tel que $|f(x)| \leq \varepsilon/2$ pour tout x en-dehors de K . Comme K est compact et qu'il est recouvert par l'union des ensembles ouverts croissants $\text{Int}(K_n)$, il existe un rang fini N tel que $K \subset K_N$. Pour tout $n \geq N$, l'élément x_n n'appartient donc pas à K , d'où $|f(x_n)| \leq \varepsilon/2$, et par conséquent $|f_n(x_n) - f(x_n)| \geq \varepsilon/2$. Cela contredit la convergence uniforme d'une sous-suite de (f_n) vers f . La condition (ii) est donc nécessaire.

4. Supposons maintenant que la condition (iii') n'est pas vérifiée. Il existe donc $\varepsilon > 0$ fixé tel que pour tout $k \in \mathbb{N}$ on puisse trouver une fonction f_k et des éléments x_k, y_k avec $d(x_k, y_k) \leq 1/k$ et $|f_k(x_k) - f_k(y_k)| \geq \varepsilon$. Grâce à la condition (ii), il existe un compact K tel que $|f(x)| \leq \varepsilon/2$ pour tout $x \notin K$; en particulier, x_k et y_k ne peuvent appartenir tous deux à K , et on pourra sans perte de généralité supposer que $x_k \in K$. Quitte à extraire une sous-suite, on peut supposer que x_k converge vers un certain $x \in K$; alors y_k converge aussi vers x . Soit f une valeur d'adhérence de la suite f_k ; quitte à extraire encore, on peut supposer que f_k converge uniformément vers f ; en outre, f est continue, donc $f(x_k)$ et $f(y_k)$ convergent vers la même limite

$f(x)$. Pour k assez grand on a donc $|f_k(x_k) - f(x_k)| < \varepsilon/3$, $|f_k(y_k) - f(y_k)| < \varepsilon/3$ et $|f(x_k) - f(y_k)| < \varepsilon/3$; ce qui est en contradiction avec $|f_k(x_k) - f_k(y_k)| \geq \varepsilon$.

5. Montrons finalement que les conditions (i) à (iii) impliquent la précompacité. Soit $(f_n)_{n \in \mathbb{N}}$ une suite à valeurs dans \mathcal{A} . Puisque X est union dénombrable de compacts, il est séparable; soit donc $(x_k)_{k \in \mathbb{N}}$ une suite dense dans X . La suite $f_n(x_1)$ est bornée; donc, quitte à extraire une sous-suite de (f_n) , on peut supposer que $f_n(x_1)$ converge dans X . Quitte à extraire à nouveau, on peut supposer que $f_n(x_2)$ converge aussi, etc. Par un procédé d'extraction diagonale, on peut supposer que pour tout k , la suite $f_n(x_k)$ converge vers un certain y_k . On pose $f(x_k) = y_k$, ce qui définit une fonction f sur $Z = \{x_k\}_{k \in \mathbb{N}}$. En passant à la limite dans la condition (iii), on voit que f est localement uniformément continue sur Z ; comme Z est dense, f admet un unique prolongement par continuité à toute partie K de X ; en utilisant une suite exhaustive de compacts on peut donc prolonger f à X tout entier en une fonction continue, que l'on note encore f . En passant à la limite dans la condition (ii), on voit que f tend vers 0 à l'infini. Il reste à montrer que la convergence est uniforme. Pour tout $\varepsilon > 0$, il existe un compact K tel que $|f(x)| \leq \varepsilon/2$ et $|f_n(x)| \leq \varepsilon/2$ pour tout $x \notin K$ et pour tout n . Pour prouver que la convergence est uniforme, il suffit donc de prouver qu'elle est uniforme sur tout compact.

6. Soient donc K un compact, et $\varepsilon > 0$. Il existe $\delta > 0$ tel que $|f_n(x) - f_n(y)| \leq \varepsilon$ et $|f(x) - f(y)| \leq \varepsilon$ dès que $d(x, y) \leq \delta$. On recouvre K par un nombre fini de boules $B_\delta(x_k)$. Comme f_n converge simplement vers f et que les x_k sont en nombre fini, il existe N tel que pour tout $n \geq N$, $|f_n(x_k) - f(x_k)| \leq \varepsilon$. Pour tout $x \in K$, on peut trouver k tel que $d(x_k, x) \leq \delta$, et alors on a $|f_n(x_k) - f(x_k)| \leq \varepsilon$, $|f_n(x_k) - f_n(x)| \leq \varepsilon$ et $|f(x_k) - f(x)| \leq \varepsilon$, d'où $|f_n(x) - f(x)| \leq 3\varepsilon$, ce qui prouve la convergence uniforme. \square

EXEMPLES VII-122. (i) Soient K un espace métrique compact, $L > 0$ et $x_0 \in K$, et soit $\mathcal{A} := \text{Lip}_{L; x_0}(K)$ l'ensemble des fonctions L -Lipschitziennes f sur K , telles que $f(x_0) = 0$. Alors cet ensemble est compact dans $C(K)$. En effet, il est facile de vérifier qu'il est fermé; pour tous $x \in K$ et $f \in \mathcal{A}$ on a $|f(x)| \leq Ld(x, x_0)$, de sorte que la condition (i) du critère d'Ascoli est satisfaite; et le module de continuité $m(\delta) = L\delta$ convient à tous les éléments de \mathcal{A} . En fait, si l'on définit $\text{Lip}_{x_0}(K)$ comme l'espace des fonctions Lipschitziennes s'annulant en x_0 , on peut en faire un espace vectoriel normé en le munissant de la norme

$$\|f\|_{\text{Lip}} := \sup_{x \neq y} \frac{|f(y) - f(x)|}{d(x, y)},$$

et le résultat précédent s'interprète en disant que l'injection de $\text{Lip}_{x_0}(K)$ dans $C(K)$ est compacte. De même, si l'on munit l'espace des fonctions Lipschitziennes $\text{Lip}(K)$ de la norme

$$\|f\|_{BL} := \|f\|_\infty + \|f\|_{\text{Lip}},$$

alors l'injection de $\text{Lip}(K)$ dans $C(K)$ est compacte.

(ii) Pour tout $\alpha \in]0, 1[$, l'espace $C^\alpha(K)$ des fonctions α -Hölderiennes, muni de la norme

$$\|f\|_{C^\alpha(K)} := \|f\|_\infty + \sup_{x \neq y} \frac{|f(y) - f(x)|}{d(x, y)^\alpha}$$

s'injecte compactement dans $C(K)$. Plus généralement, si on se donne une fonction m continue de \mathbb{R}_+ dans \mathbb{R}_+ , strictement croissante près de 0, la norme

$$\|f\|_{C_m(K)} := \|f\|_\infty + \sup_{x \neq y} \frac{|f(y) - f(x)|}{m(d(x, y))}$$

définit un sous-espace vectoriel normé complet, qui s'injecte compactement dans $C(K)$.

(iii) Terminons avec un exemple non compact. Soient $\alpha \in]0, 1[$ et $\beta > 0$. Dans $C_0(\mathbb{R}^n)$, la norme

$$\sup_{x \in \mathbb{R}^n} |f(x)|(1 + |x|^\beta) + \sup_{x \neq y} \frac{|f(y) - f(x)|}{|y - x|^\alpha}$$

définit un sous-espace vectoriel normé complet, qui s'injecte compactement dans $C_0(\mathbb{R}^n)$.

Il n'existe pas d'analogue "naturel" du théorème d'Ascoli pour des espaces L^p . La condition suffisante qui suit couvre cependant de nombreux cas intéressants. Comme le critère d'Ascoli, elle contient trois conditions : une borne "globale", une condition de "décroissance à l'infini", et une condition de "régularité locale". En fait, sa démonstration se ramène au critère d'Ascoli.

THÉORÈME VII-123 (condition suffisante de précompacité dans L^p). *Soient (X, d) un espace métrique localement compact, σ -compact, et μ une mesure borélienne doublante, finie sur les compacts de X , et soit $p \in [1, +\infty[$. Soit \mathcal{F} un sous-ensemble de $L^p(X, \mu)$ tel que*

(i) \mathcal{F} est borné dans L^p ;

(ii) Pour tout $\varepsilon > 0$ il existe un compact K_ε de X telle que

$$\sup_{f \in \mathcal{F}} \int_{X \setminus K_\varepsilon} |f|^p d\mu \leq \varepsilon.$$

(iii) Pour tout compact $K \subset X$ il existe une fonction η positive croissante, telle que

$$(73) \quad \lim_{\delta \rightarrow 0} \left(\frac{\eta(\delta)}{\inf_{x \in K} \mu[B_\delta(x)]} \right) = 0,$$

et

$$\sup_{f \in \mathcal{F}} \int_K \int_K \frac{|f(x) - f(y)|^p}{\eta(d(x, y))} d\mu(x) d\mu(y) < +\infty.$$

Alors \mathcal{F} est précompact dans $L^p(X, \mu)$.

DÉMONSTRATION. 1. Supposons l'énoncé démontré dans le cas où X est compact. Soit alors (f_n) une suite à valeurs dans \mathcal{F} . On introduit une suite exhaustive $(K_j)_{j \in \mathbb{N}}$ de compacts dans X , et on définit $\mu_j = 1_{K_j} \mu$, la restriction de μ à K_j . Quitte à extraire une sous-suite, on peut supposer que f_n converge dans $L^p(\mu_j)$ vers une fonction $f^j \in L^p(\mu_j)$, et quitte à extraire encore on peut supposer que cette convergence a lieu presque partout. Par une extraction diagonale, on peut supposer que f_n converge presque partout vers une fonction mesurable f , qui appartient à dans tous les $L^p(\mu_j)$. Par le lemme de Fatou et l'hypothèse (i),

$$\int |f|^p d\mu \leq \liminf_{n \rightarrow \infty} \int |f_n|^p d\mu;$$

il s'ensuit que $f \in L^p(X, \mu)$. En particulier, si $\varepsilon > 0$ est donné, pour j assez grand on a

$$\int_{X \setminus K_j} |f|^p d\mu \leq \varepsilon.$$

L'ensemble K_ε de l'hypothèse (ii) est recouvert par l'union croissante des ouverts $O_j := \text{Int}(K_j)$; il est donc inclus dans K_j pour j assez grand. On a alors

$$\sup_{n \in \mathbb{N}} \int_{X \setminus K_j} |f_n|^p d\mu \leq \varepsilon.$$

Il s'ensuit que, pour j assez grand,

$$\sup_{n \in \mathbb{N}} \int_{X \setminus K_j} |f_n - f|^p d\mu \leq 2\varepsilon;$$

j étant maintenant fixé, on sait que pour n assez grand on a

$$\int_{K_j} |f_n - f|^p d\mu \leq \varepsilon.$$

On conclut que f_n converge bien vers f dans L^p . Il suffit donc de prouver l'énoncé dans le cas où X est compact.

2. On suppose maintenant X compact; μ est alors finie par hypothèse; en particulier $C(X)$ s'injecte continûment dans $L^p(X)$.

Nous allons **régulariser** les éléments de \mathcal{F} comme suit. Soit $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ définie par $\psi(r) = 1$ pour $r \leq 1/2$, $\psi(r) = 0$ pour $r \geq 1$, ψ affine entre $1/2$ et 1 (les détails de ψ importent peu, tout ce qui compte c'est qu'elle soit à support compact et identiquement égale à 1 près de 0). Pour $\delta \in]0, 1[$ (δ petit) et f dans \mathcal{F} , on définit

$$f^\delta(x) := \int_X f(y) d\mu_x^\delta(y),$$

où μ_x^δ est la probabilité définie par

$$\mu_x^\delta(dy) = \frac{\psi\left(\frac{d(x,y)}{\delta}\right) \mu(dy)}{\int_X \psi\left(\frac{d(x,y')}{\delta}\right) \mu(dy')}.$$

La valeur de $f^\delta(x)$ a donc été obtenue par une *moyenne* des valeurs $f(y)$ dans laquelle seuls comptent les y qui sont proches de x (au sens où la distance est au plus δ).

On pose alors $\mathcal{F}^\delta := \{f^\delta\}_{0 < \delta < 1}$. Nous allons voir que (a) $\|f - f^\delta\|_{L^p} \rightarrow 0$ quand $\delta \rightarrow 0$, uniformément en $f \in \mathcal{F}$; et (b) pour tout $\delta > 0$, \mathcal{F}^δ est précompact dans $C_0(X)$, et donc dans $L^p(X)$. La conclusion en découlera grâce à la Proposition VII-119.

3. Montrons maintenant la propriété (a). On fixe $\delta > 0$; on va montrer que l'ensemble \mathcal{F}^δ est borné et uniformément Lipschitz, et donc précompact dans $C(X)$ grâce au théorème d'Ascoli. On supposera bien sûr que $\mu[X] > 0$, sans quoi il n'y a rien à démontrer.

Puisque $\psi(r)$ est uniformément égale à 1 pour $r \leq 1/2$, on a

$$\zeta(x) := \int_X \psi\left(\frac{d(x,y)}{\delta}\right) d\mu(y) \geq \int_X 1_{d(x,y) \leq \delta/2} d\mu(y) = \mu[B_{\delta/2}(x)].$$

Étant compact, X est borné ; soit D son diamètre. Pour tout $x \in X$, on peut inclure X dans $B_D(x)$. Soit N tel que $2^N(\delta/2) \geq D$, et C tel que μ soit C -doublante. On a alors

$$0 < \mu[X] = \mu[B_D(x)] \leq C^N \mu[B_{\delta/2}(x)],$$

ce qui montre que $\mu[B_{\delta/2}(x)]$ est minoré par une constante positive indépendante de x , et donc il en est de même de la fonction ζ . On écrit alors, en utilisant l'inégalité de Hölder,

$$|f^\delta(x)| \leq \frac{\int |f(y)| d\mu(y)}{\zeta(x)} \leq \frac{\|f\|_{L^p}^{1/p} \mu[X]^{1-1/p}}{\inf \zeta},$$

ce qui est uniformément borné grâce à l'hypothèse (i). L'ensemble \mathcal{F}^δ est donc bien uniformément borné.

4. Montrons maintenant que la fonction ζ est Lipschitzienne. Pour cela on note que la fonction ψ elle-même est 2-Lipschitzienne, d'où

$$\begin{aligned} |\zeta(x) - \zeta(x')| &= \left| \int \left[\psi\left(\frac{d(x,y)}{\delta}\right) - \psi\left(\frac{d(x',y)}{\delta}\right) \right] d\mu(y) \right| \\ &\leq \int \left| \psi\left(\frac{d(x,y)}{\delta}\right) - \psi\left(\frac{d(x',y)}{\delta}\right) \right| d\mu(y) \\ &\leq \frac{2}{\delta} \int |d(x,y) - d(x',y)| d\mu(y) \\ &\leq \frac{2}{\delta} \int d(x,x') d\mu(y) = \left(\frac{2\mu[X]}{\delta} \right) d(x,x'). \end{aligned}$$

Par le même raisonnement, et l'inégalité de Hölder, on montre que

$$\left| \int_X \psi\left(\frac{d(x,y)}{\delta}\right) f(y) d\mu(y) - \int_X \psi\left(\frac{d(x',y)}{\delta}\right) f(y) d\mu(y) \right| \leq \left(\frac{2\|f\|_{L^p}^{1/p} \mu[X]^{1-1/p}}{\delta} \right) d(x,x').$$

Il s'ensuit que f^δ est le quotient de deux fonctions Lipschitziennes, le dénominateur étant minoré par une constante strictement positive ; elle est donc Lipschitzienne. Toutes ces bornes sont uniformes en $f \in \mathcal{F}$, l'ensemble \mathcal{F}^δ est donc uniformément Lipschitz. (Bien sûr, la constante de Lipschitz tend vers l'infini quand $\delta \rightarrow 0$, ce qui était prévisible car l'ensemble \mathcal{F} lui-même n'est pas a priori Lipschitz.

5. Établissons finalement la propriété (b). D'après (iii), il existe une constante C telle que pour tout $f \in \mathcal{F}$,

$$\int_X \int_X \frac{|f(x) - f(y)|^p}{\eta(d(x,y))} d\mu(x) d\mu(y) \leq C.$$

En particulier, puisque $0 \leq \psi \leq 1$,

$$\int_X \int_X \frac{|f(x) - f(y)|^p}{\eta(d(x,y))} d\mu(x) d\mu(y) \psi\left(\frac{d(x,y)}{\delta}\right) \leq C.$$

Puisque $\psi(r)$ est nul pour $r \geq 1$, seuls comptent dans l'intégrale précédente les (x,y) tels que $d(x,y) \leq \delta$, ce qui implique $\eta(d(x,y)) \leq \eta(\delta)$. On a donc

$$\frac{1}{\eta(\delta)} \int_X \int_X |f(x) - f(y)|^p \psi\left(\frac{d(x,y)}{\delta}\right) d\mu(y) d\mu(x) \leq C;$$

en divisant par $\int \psi(d(x, y')/\delta) d\mu(y')$ on en déduit

$$\frac{1}{\eta(\delta)} \int_X \int_X |f(x) - f(y)|^p d\mu_x(y) d\mu(x) \leq \sup_{x \in X} \frac{C}{\int_X \psi\left(\frac{d(x, y')}{\delta}\right) d\mu(y')}.$$

Comme $\psi(r)$ est identiquement égale à 1 pour $r \leq 1/2$, et que μ est doublante, il existe $K > 0$ tel que

$$\int_X \psi\left(\frac{d(x, y')}{\delta}\right) d\mu(y') \geq \mu[B_{\delta/2}(x)] \geq K\mu[B_\delta(x)].$$

En combinant cela à l'inégalité de Jensen (appliquée à la fonction convexe $|\cdot|^p$ et à la mesure de probabilité μ_x , pour x fixé, on trouve

$$\begin{aligned} \int_X |f(x) - f^\delta(x)|^p d\mu(x) &\leq \int_X \int_X |f(x) - f(y)|^p d\mu_x(y) d\mu(x) \\ &\leq \sup_x \frac{C\eta(\delta)}{K\mu[B_\delta(x)]} \xrightarrow{\delta \rightarrow 0} 0. \end{aligned}$$

Ceci conclut la preuve du théorème. \square

EXEMPLES VII-124. (i) Soient O un ouvert Lipschitz de \mathbb{R}^n , $p \in [1, +\infty[$ et $s \in]0, 1[$; on définit l'**espace de Sobolev fractionnaire** $W^{s,p}(O)$ comme l'espace des fonctions $L^p(O)$ telles que

$$\|f\|_{W^{s,p}(O)} := \|f\|_{L^p(O)} + \left(\int_O \int_O \frac{|f(x) - f(y)|^p}{|x - y|^{n+sp}} dx dy \right)^{1/p} < +\infty.$$

La fonction $\|\cdot\|_{W^{s,p}(O)}$ est alors une norme sur $W^{s,p}(O)$, et l'espace ainsi défini s'injecte compactement dans $L^p(O)$ dès que O est borné. Pour le voir, on applique le Théorème VII-123 avec l'espace compact $X = \overline{O}$ et la fonction $\eta(\delta) = \delta^{n+sp}$. On note que l'hypothèse de régularité Lipschitzienne sur O implique l'existence d'une constante K telle que le volume de la boule de rayon δ dans \overline{O} soit minoré par $K\delta^n$; si cet ouvert comportait une "pointe très fine", les boules de rayon δ centrées en la pointe pourraient avoir un volume beaucoup trop petit.

Les espaces de Sobolev fractionnaires jouent un rôle important en analyse moderne; leur définition peut sembler compliquée, mais elle est en fait assez naturelle.

(ii) Si M est une variété Riemannienne compacte régulière, on définit l'espace de Sobolev fractionnaire $W^{s,p}(M)$ par la même formule, en remplaçant simplement $|x - y|$ par $d(x, y)$, où d est la distance géodésique. Cet espace s'injecte compactement dans $L^p(M)$.

(iii) Voici finalement un exemple dans le cas où X est non compact : le sous-espace de $L^1(\mathbb{R}^n)$ défini par la norme

$$\int_{\mathbb{R}^n} |f(x)|(1 + \log(1 + |x|)) dx + \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{|f(x) - f(y)|}{|x - y|^n \log\left(1 + \frac{1}{|x - y|}\right)} dx dy$$

s'injecte compactement dans $L^1(\mathbb{R}^n)$. En effet, si cette norme est bornée par C , l'inégalité de Chebyshev garantit que l'intégrale de $|f|$ sur le complémentaire de la boule de rayon R est majorée par $C/\log(1 + R)$, qui tend vers 0 quand $R \rightarrow \infty$.

On conclut ce paragraphe avec une variante du critère précédent : le critère de **Riesz-Fréchet-Kolmogorov**, qui est une condition nécessaire et suffisante dans le cas où X est un ouvert de \mathbb{R}^n , muni de la mesure de Lebesgue. Ce critère est basé sur l'invariance de la mesure de Lebesgue par translation ; en fait il reste valable dans un cadre plus général que nous étudierons au Chapitre ??, celui des **groupes topologiques localement compacts** munis d'une mesure finie sur les compacts et invariante par translation, i.e. une mesure de Haar. Par exemple, on peut remplacer dans l'énoncé qui suit l'espace \mathbb{R}^n par le tore \mathbb{T}^n (muni de la mesure de Lebesgue sur le tore).

THÉORÈME VII-125 (Théorème de Riesz-Fréchet-Kolmogorov). *Soient $\Omega \subset \mathbb{R}^n$ un ouvert, muni de la mesure de Lebesgue, et \mathcal{F} une partie de $L^p(\Omega)$, $1 \leq p < \infty$. Alors \mathcal{F} est précompact dans $L^p(\Omega)$ si et seulement si*

(i) \mathcal{F} est borné dans L^p ;

(ii) Pour tout $\varepsilon > 0$ il existe un compact $K_\varepsilon \subset \Omega$ tel que

$$\sup_{f \in \mathcal{F}} \|f\|_{L^p(\Omega \setminus K_\varepsilon)} \leq \varepsilon ;$$

(iii) Pour tout compact K de Ω et pour tout $\varepsilon > 0$ il existe $\delta > 0$ tel que

$$|h| \leq \delta \implies \sup_{f \in \mathcal{F}} \|f - f(\cdot + h)\|_{L^p(K)} \leq \varepsilon .$$

REMARQUE VII-126. L'hypothèse (iii) exprime le fait que les éléments de \mathcal{F} varient peu, en norme L^p , quand on les translate d'un vecteur h de petite norme. (Par "translater" on entend ici remplacer f par $\tau_h f$, où $\tau_h f(x) = f(x + h)$.) Notons que $f(\cdot + h)$ est bien défini dès que $|h| \leq d(K, \Omega^c)$.

DÉMONSTRATION DU THÉORÈME VII-125. Il est clair que la condition (i) est nécessaire ; pour montrer que la condition (ii) est nécessaire, on peut utiliser un raisonnement analogue à celui qui a été effectué dans la preuve du Théorème d'Ascoli. Montrons maintenant que la condition (iii) est nécessaire ; on va commencer par l'établir dans le cas où \mathcal{F} est réduit à un singleton. Soit donc $f \in L^p(\mathbb{R}^n)$; on sait qu'il existe $g \in C_c(\mathbb{R}^n)$ tel que $\|f - g\|_{L^p} \leq \varepsilon$. Par invariance de la mesure de Lebesgue par translation, on a aussi

$$\|f(\cdot + h) - g(\cdot + h)\|_{L^p} = \|f - g\|_{L^p} \leq \varepsilon .$$

Comme g est uniformément continue et à support compact, on a bien sûr

$$\lim_{h \rightarrow 0} \|g - g(\cdot + h)\|_{L^p} = 0 ,$$

et par suite

$$\limsup_{|h| \rightarrow 0} \|f - f(\cdot + h)\|_{L^p} \leq \limsup_{|h| \rightarrow 0} \|f - g\|_{L^p} + \limsup_{|h| \rightarrow 0} \|g(\cdot + h) - f(\cdot + h)\| \leq 2\varepsilon .$$

En faisant tendre ε vers 0, on conclut que $\|f - f(\cdot + h)\|_{L^p} \longrightarrow 0$.

Vérifions maintenant (iii) dans le cas général. Supposons que \mathcal{F} soit précompact et que (iii) ne soit pas vérifié ; il existe alors $\varepsilon > 0$ fixé, tel que l'on puisse construire des suites h_k (à valeurs dans \mathbb{R}^n) et f_k (à valeurs dans \mathcal{F}) avec $h_k \rightarrow 0$ et $\|f_k - f_k(\cdot + h_k)\|_{L^p} \geq \varepsilon > 0$. Par précompacité, on peut supposer (quitte à extraire une sous-suite) que f_k converge dans L^p vers $f \in L^p$. Alors, pour k assez grand on a $\|f_k - f_k(\cdot + h_k)\|_{L^p} = \|f_k - f\|_{L^p} < \varepsilon/3$ et $\|f - f(\cdot + h_k)\| < \varepsilon/3$, ce qui contredit

$\|f_k - f_k(\cdot + h_k)\|_{L^p} \geq \varepsilon$. On conclut que (iii) est également une condition nécessaire.

Montrons maintenant que les conditions (i) à (iii) sont suffisantes. La démonstration est exactement la même que celle du Théorème VII-119 : on se ramène à établir la précompacité sur un compact K de Ω , on effectue une **régularisation par convolution** en posant, pour $x \in K$, et $f \in \mathcal{F}$,

$$f^\delta(x) = \int_{\Omega} f(y) d\nu_x^\delta(y),$$

$$\nu_x^\delta(y) = \frac{\psi\left(\frac{|x-y|}{\delta}\right) dy}{\int_{\Omega} \psi\left(\frac{|x-y'|}{\delta}\right) dy'};$$

ce qui est bien défini pour δ suffisamment petit. On reprend alors le raisonnement du Théorème VII-119, avec une différence dans l'estimation de

$$\int_K \int_K |f(x) - f(y)|^p d\nu_x^\delta(y) dx.$$

En effet, la mesure $\nu_x^\delta(dy)$ s'écrit maintenant sous la forme $\phi^\delta(|x-y|) dy$, et on peut profiter de l'invariance de la mesure de Lebesgue sous l'effet des translations, pour effectuer un changement de variable $h = y - x$ (à x fixé); alors $dy = dh$, ce qui revient à dire que $\nu_x^\delta(dy) = \phi^\delta(|h|) dh \equiv \nu^\delta(dh)$, où la mesure ν^δ est fixée, et son support est contenu dans $(|h| \leq \delta)$. Alors

$$\begin{aligned} \int_K \int_K |f(x) - f(y)|^p d\nu_x^\delta(y) dx &= \int_K \left(\int_{K-h} |f(x) - f(x+h)|^p dx \right) \nu^\delta(dh) \\ &\leq \int_K \left(\int_O |f(x) - f(x+h)|^p dx \right) \nu^\delta(dh) \\ &\leq \sup_{|h| \leq \delta} \int_O |f(x) - f(x+h)|^p dx, \end{aligned}$$

et par hypothèse, cette quantité tend vers 0 quand $\delta \rightarrow 0$. \square

EXEMPLE VII-127. Soit $K \in L^1(\mathbb{R}^n)$ fixé; on sait en particulier que $\|\tau_h K - K\|_{L^1} \rightarrow 0$ quand $h \rightarrow 0$. Pour tout g dans $L^p(\mathbb{R}^n)$, on définit la convolution de K et g par la formule

$$(K * g)(x) = \int_{\mathbb{R}^n} g(y) K(x-y) dy.$$

Il découle des propriétés élémentaires de la convolution (que nous étudierons au chapitre suivant, mais que le lecteur peut établir en guise d'exercice) que $\|K * g\|_{L^p} \leq \|K\|_{L^1} \|g\|_{L^p}$

$$\|\tau_h(K * g) - K * g\|_{L^p} = \|(\tau_h K - K) * g\|_{L^p} \leq \|\tau_h K - K\|_{L^1} \|g\|_{L^p}.$$

Il s'ensuit que si g varie dans un ensemble borné de L^p et B est une partie bornée de \mathbb{R}^n , la famille des $(K * g)1_B$ satisfait au critère de Riesz-Fréchet-Kolmogorov, et est donc précompacte dans $L^p(B)$. On dit que l'application linéaire $g \mapsto K * g$, définie de $L^p(\mathbb{R}^n)$ dans $L^p(B)$, est un **opérateur compact** : il transforme les parties bornées en parties précompactes.

VII-5.5. Critères de compacité faible. La notion de convergence faible étant a priori beaucoup moins forte que celle de convergence en norme, on peut espérer que la précompacité au sens faible est beaucoup moins contraignante que la précompacité au sens fort. C'est effectivement le cas, au moins pour la topologie faible-étoile, comme le montre le **Théorème de Banach-Alaoglu**, principal résultat de compacité faible.

THÉORÈME VII-128 (Théorème de Banach-Alaoglu). *Soit E un espace de Banach séparable. Alors, une partie \mathcal{B} de E^* est précompacte faible-étoile si et seulement si elle est bornée pour la norme de E^* . En particulier, les parties compactes sont les parties bornées pour la norme, fermées pour la topologie faible-étoile.*

- REMARQUES VII-129.** (i) Si l'on accepte l'axiome du choix dans sa version forte, on peut utiliser le Théorème de Tychonov pour éliminer l'hypothèse de séparabilité dans le Théorème de Banach-Alaoglu.
(ii) L'hypothèse de complétude sur E sert uniquement à établir l'implication directe : i.e. la précompacité entraîne la bornitude.

En combinant le Théorème VII-128 avec le Théorème VII-36, on obtient :

COROLLAIRE VII-130. *Soit X un espace métrique séparable. Alors*

(i) *Si μ est une mesure σ -finie régulière, alors pour $1 < p \leq \infty$, les parties bornées de $L^p(X)$ sont précompactes pour la topologie faible $\sigma(L^p, L^{p'})$ (faible-étoile pour $p = \infty$) ;*

(ii) *Si X est en outre localement compact et σ -compact, alors les parties bornées de $M(X)$ sont précompactes pour la topologie faible-étoile $\sigma(M(X), C_c(X))$.*

DÉMONSTRATION DU THÉORÈME VII-128. Tout d'abord, si \mathcal{B} n'est pas bornée, on peut construire une suite (x_n) à valeurs dans \mathcal{B} , telle que $\|x_n\| \geq n$. Toute sous-suite extraite de x_n est alors non bornée, et donc non convergente d'après la Proposition VII-108(i), ce qui montre que \mathcal{B} n'est pas précompact.

Réciproquement, supposons que \mathcal{B} est borné en norme par une constante M . Pour tout $x \in E$, la famille des $\langle \Lambda, x \rangle_{\Lambda \in \mathcal{B}}$ est une famille de nombres réels bornée par $M\|x\|$, elle admet donc une sous-suite convergente. Soit maintenant $(x^k)_{k \in \mathbb{N}}$ une suite dense dans E . La famille $\langle \Lambda, x^1 \rangle_{\Lambda \in \mathcal{B}}$ est bornée, elle admet donc une sous-suite $\langle \Lambda_n, x^1 \rangle$ qui converge vers un nombre, que l'on note $\Lambda(x^1)$. La famille des $\langle \Lambda_n, x^2 \rangle$ est bornée, on peut donc extraire de la suite (Λ_n) une sous-suite, toujours notée Λ_n , telle que $\langle \Lambda_n, x^2 \rangle$ qui converge vers un nombre noté $\Lambda(x^2)$. Grâce à un procédé d'*extraction diagonale*, on construit par récurrence une sous-suite Λ_n telle que pour tout $k \in \mathbb{N}$, $\Lambda_n(x^k)$ converge vers $\Lambda(x^k)$. La fonction Λ est alors définie sur l'ensemble S des x^k , qui forme une partie dense dans E . Comme tous les Λ_n sont uniformément Lipschitziens, la fonction Λ est également Lipschitzienne sur S , et admet donc un unique prolongement Lipschitzien à E , toujours noté Λ ; alors, pour tout $x \in E$ on a $\Lambda_n(x) \rightarrow \Lambda(x)$. En passant à la limite dans les relations de linéarité de Λ_n , on vérifie que Λ est bien une forme linéaire continue, qui est donc la limite faible-étoile des Λ_n . \square

Le théorème de Banach-Alaoglu laisse en suspens les cas de deux espaces d'usage courant : l'espace $L^1(X)$, et l'espace $M(X)$ quand X n'est pas localement compact. Le premier est l'objet du **Théorème de Dunford-Pettis**, basé sur le concept d'**équii-intégrabilité** ; le second est abordé dans le **Théorème de Prokhorov**, qui

repose sur le concept de **tension**. de **tension** ; rappelons que nous avons étudié ces notions dans la section III-5.

THÉORÈME VII-131 (Théorème de compacité faible de Dunford-Pettis). *Soit (X, d) un espace métrique séparable, muni d'une mesure μ . Alors, une partie \mathcal{F} de $L^1(X, \mu)$ est précompacte pour la topologie $\sigma(L^1, L^\infty)$ si et seulement si*

(i) \mathcal{F} est bornée dans L^1 ;

(ii) Pour tout $\varepsilon > 0$ il existe un ensemble $Y_\varepsilon \subset X$, de mesure finie, tel que

$$\sup_{f \in \mathcal{F}} \int_{X \setminus Y_\varepsilon} |f| d\mu \leq \varepsilon;$$

(iii) \mathcal{F} est équi-intégrable sur toute partie $Y \subset X$ de mesure finie.

En outre, on peut remplacer l'énoncé (iii) par

(iii') \mathcal{F} est équi-intégrable (sur X tout entier) ;

REMARQUES VII-132. (i) Si (X, μ) est σ -fini, alors la topologie $\sigma(L^1, L^\infty)$ est la topologie faible dans L^1 . Notons que même si X n'est pas σ -fini, l'hypothèse (ii) implique qu'il existe une partie σ -finie $Y \subset X$ telle que tous les $f \in \mathcal{F}$ sont presque partout nuls en-dehors de Y (pour le voir, poser $Y = \cup Y_{1/n}$).

(ii) La séparabilité ne sert que dans la démonstration de l'implication réciproque : si un ensemble \mathcal{F} vérifie les conditions indiquées, alors il est précompact.

EXEMPLES VII-133. Soit (X, μ) un espace mesuré fini, alors tout ensemble borné en norme L^p ($p > 1$) est faiblement précompact dans L^1 . Dans \mathbb{R}^n , pour tout C l'ensemble des fonctions f vérifiant

$$\int |f|(1 + |\log(1 + |f|)|) d\mu + \int |f(x)|(1 + \log(1 + |x|))(x) dx \leq C$$

est faiblement précompact dans L^1 .

La démonstration de l'implication directe dans le Théorème VII-131 est élémentaire mais assez subtile ; on pourra l'omettre en première lecture, avec d'autant moins de remords que c'est l'implication réciproque qui sert en pratique. Cette preuve utilise le lemme élémentaire suivant :

LEMME VII-134. *Soit (X, μ) un espace mesurable et $f \in L^1(X, \mu)$. Alors il existe $A \subset X$, de mesure finie, tel que*

$$\left| \int_A f d\mu \right| \geq \frac{1}{4} \int_X |f| d\mu$$

(on peut remplacer le 1/4 par un 1/2 si X est de mesure finie).

DÉMONSTRATION. On a $\int |f| = \int_{X_+} f + \int_{X_-} f$, où $X_+ = \{x; f(x) > 0\}$ et $X_- = \{x; f(x) < 0\}$. En posant $B = X_+$ ou X_- , on a donc $|\int_B f d\mu| \geq (\int_X |f| d\mu)/2$. Sans perte de généralité, on peut supposer que $B = X_+$. Soit maintenant X_n l'ensemble des $x \in X$ tels que $0 \leq f(x) \leq n$. Par convergence dominée, $\int_{X_n} f$ converge vers $\int_B f$; et X_n est de mesure finie par inégalité de Chebyshev. On en déduit que l'inégalité est vraie si l'on choisit $A = B_n$ pour n assez grand. \square

DÉMONSTRATION DU THÉORÈME VII-131, IMPLICATION DIRECTE. Soit \mathcal{F} une partie précompacte de L^1 , nous allons vérifier les énoncés (i) à (iii).

1. Si \mathcal{F} n'est pas bornée dans L^1 , on peut construire une suite (f_k) à valeurs dans \mathcal{F} , avec $\|f_k\|_{L^1} \geq k$. Puisque L^∞ est complet, le Théorème de Banach–Steinhaus s'applique pour dire que toute suite de formes linéaires convergeant simplement est bornée en norme duale. On sait (Remarque VII-92) que la norme de f_k dans $(L^\infty)^*$ est $\|f_k\|_{L^1}$, qui n'est pas borné; donc aucune sous-suite extraite de (f_k) ne peut converger simplement, et \mathcal{F} ne peut être précompact.

2. Supposons maintenant que \mathcal{F} est précompact mais (ii) n'est pas vérifié. Il existe donc $\varepsilon > 0$ fixé, tel que pour tout ensemble Y de mesure finie, on puisse trouver $f \in \mathcal{F}$ avec $\int_{X \setminus Y} |f| d\mu \geq \varepsilon$, ce qui entraîne, par le Lemme VII-134, l'existence de A mesurable avec $A \cap Y = \emptyset$ et $|\int_A f d\mu| \geq \varepsilon/2$.

On construit par récurrence une suite à valeurs dans \mathcal{F} comme suit. On choisit f_1 dans \mathcal{F} et A_1 de mesure finie tel que

$$\left| \int_{A_1} f_1 d\mu \right| \geq \frac{\varepsilon}{4}.$$

On choisit alors f_2 et A_2 de mesure finie tel que $A_1 \cap A_2 = \emptyset$ et

$$\left| \int_{A_2} f_2 d\mu \right| \geq \frac{\varepsilon}{4}.$$

Par récurrence, on construit ainsi une suite croissante de fonctions f_k et d'ensembles A_k de mesure finie, disjoints, tels que pour tout k ,

$$\left| \int_{A_k} f_k d\mu \right| \geq \frac{\varepsilon}{4}.$$

On extrait de (f_n) une sous-suite convergeant au sens faible vers un certain $f \in L^1(X, \mu)$; on renumérote cette sous-suite par $1, 2, \dots$, et on pose $A := \cup A_k$. Par convergence dominée, il existe N tel que pour tout $k \geq N$,

$$\int_{A \setminus (A_1 \cup \dots \cup A_k)} |f| d\mu \leq \frac{\varepsilon}{8}.$$

On ne garde que les indices k qui sont supérieurs à N , et on pose alors $g_k := f_k - f$. La suite (g_k) ainsi construite vérifie les propriétés suivantes : g_k converge au sens faible vers 0 (de sorte que $\int_A g_k$ tend vers 0 pour toute partie mesurable A), et pour tout k on a

$$\left| \int_{A_k} g_k d\mu \right| \geq \frac{\varepsilon}{8} =: \alpha.$$

Sans perte de généralité, on continue à noter $k = 1, 2, \dots$

On va maintenant extraire à nouveau ! On pose $n_0 = 1$. Il existe n_1 tel que pour tout $n \geq n_1$,

$$\left| \int_{A_{n_0}} g_n d\mu \right| \leq \frac{\alpha}{4}.$$

On choisit alors $n_2 \geq n_1$ tel que pour tout $n \geq n_2$,

$$\left| \int_{A_{n_1}} g_n d\mu \right| \leq \frac{\alpha}{8}, \quad \int_{A \setminus (A_1 \cup \dots \cup A_{n_1})} |g_{n_1}| d\mu \leq \frac{\alpha}{8},$$

et ainsi de suite.

On a alors, pour tout $k \geq 1$,

$$\left| \int_{A_{n_k}} g_{n_k} d\mu \right| \geq \alpha,$$

$$\left| \int_{A_{n_j}} g_{n_k} d\mu \right| \leq \frac{\alpha}{4 \cdot 2^{k-j}} \quad \text{pour } j < k;$$

$$\left| \int_{A_{n_j}} g_{n_k} d\mu \right| \leq \int_{X \setminus Z_j} |g_{n_k}| d\mu \leq \frac{\alpha}{4 \cdot 2^{j-k}} \quad \text{pour } j > k.$$

Si l'on pose $B := \bigcup_{k \geq 1} A_{n_k}$, on a alors

$$\left| \int_B g_{n_k} d\mu \right| \geq \frac{\alpha}{2},$$

ce qui contredit la convergence faible de (g_n) vers 0.

3. Montrons maintenant la nécessité de la condition (iii'). Rappelons d'abord que d'après la Proposition III-87, elle est vérifiée quand \mathcal{F} est un singleton.

On applique encore le même schéma de démonstration : supposons par l'absurde que \mathcal{F} est précompact, et qu'il existe $\varepsilon > 0$ tel que pour tout δ on puisse trouver un ensemble mesurable A et une fonction $f \in \mathcal{F}$ tels que $\mu[A] \leq \delta$ et $\int_A |f| d\mu \geq \varepsilon$. En particulier, par le Lemme VII-134, on peut trouver A tels que $\mu[A] \leq \delta$ et $|\int_A f d\mu| \geq \varepsilon/2$.

On construit alors une suite de fonctions f_n et de parties A_n telles que

$$\mu[A_n] \leq \frac{1}{n}, \quad \left| \int_{A_n} f_n d\mu \right| \geq \frac{\varepsilon}{2}.$$

On extrait une sous-suite convergeant vers f au sens $\sigma(L^1, L^\infty)$. Il existe $\delta > 0$ tel que pour $\mu[B] \leq \delta$, $\int_B |f| d\mu \leq \varepsilon/4$. On pose $g_n = f_n - f$, et on ne garde que les indices n supérieurs à $1/\delta$. Après renumérotation, nous avons donc une suite de fonctions intégrables g_n et une suite de parties mesurables A_n , satisfaisant les propriétés suivantes : (a) g_n converge vers 0 au sens $\sigma(L^1, L^\infty)$, en particulier $\int_A g_n d\mu \rightarrow 0$ pour tout A fixé ; (b) $\mu[A_n] \rightarrow 0$, et donc en particulier pour tout k fixé on a $\int_{A_n} |g_k| d\mu \rightarrow 0$ pour $n \rightarrow \infty$; (c) $|\int_{A_n} g_n d\mu| \geq \varepsilon/4 =: \alpha$.

On extrait à nouveau ! On pose $n_1 = 1$, et on choisit n_2 tel que pour tout $n \geq n_2$,

$$\int_{A_n} |g_{n_1}| \leq \frac{\alpha}{8}, \quad \left| \int_{A_{n_1}} g_n \right| \leq \frac{\alpha}{8},$$

puis n_3 tel que pour tout $n \geq n_3$,

$$\int_{A_n} |g_{n_1}| \leq \frac{\alpha}{16}, \quad \int_{A_n} |g_{n_2}| \leq \frac{\alpha}{8},$$

$$\left| \int_{A_{n_1}} g_n \right| \leq \frac{\alpha}{16}, \quad \left| \int_{A_{n_2}} g_n \right| \leq \frac{\alpha}{8}, \quad \text{etc.}$$

On construit ainsi par récurrence une suite d'indices tels que

$$\begin{aligned} \left| \int_{A_{n_k}} g_{n_k} d\mu \right| &\geq \alpha, \\ \left| \int_{A_{n_j}} g_{n_k} d\mu \right| &\leq \frac{\alpha}{4 \cdot 2^{k-j}} \quad \text{pour } j < k; \\ \left| \int_{A_{n_j}} g_{n_k} d\mu \right| &\leq \int_{X \setminus Z_j} |g_{n_k}| d\mu \leq \frac{\alpha}{4 \cdot 2^{j-k}} \quad \text{pour } j > k. \end{aligned}$$

Si l'on pose $B := \bigcup_{k \geq 1} A_{n_k}$, on a alors

$$\left| \int_B g_{n_k} d\mu \right| \geq \frac{\alpha}{2},$$

ce qui contredit la convergence faible de (g_n) vers 0. \square

DÉMONSTRATION DU THÉORÈME VII-131, IMPLICATION RÉCIPROQUE. Montrons maintenant que les conditions (i)-(iii) suffisent à assurer la précompacité.

1. Quitte à écrire $f = f_+ - f_-$ et à traiter séparément les parties positives et négatives, on peut supposer que chaque $f \in \mathcal{F}$ est à valeurs positives.

2. On pose $A_k = Y_1 \cup Y_{1/2} \cup \dots \cup Y_{1/k}$. Pour $f \in \mathcal{F}$, $k \in \mathbb{N}$ et $M > 0$, soit

$$f^{k,M} = f 1_{A_k} 1_{f \leq M};$$

on note que $f^{k,M}$ est croissante en k et M . On a

$$\int_X |f - f^{k,M}| d\mu \leq \int_{X \setminus A_k} f d\mu + \int_{\{f \geq M\} \cap A_k} f d\mu,$$

et on majore séparément les deux termes du membre de droite. D'abord, par (ii), on a

$$\int_{X \setminus A_k} f d\mu \leq \frac{1}{k}.$$

Ensuite, par (i) et l'inégalité de Chebyshev,

$$\mu[\{f \geq M\}] \leq \frac{C}{M},$$

où $C = \sup\{\|f\|_{L^1}; f \in \mathcal{F}\}$; en appliquant (iii) on en déduit que

$$\sup_{f \in \mathcal{F}} \int_{\{f \geq M\} \cap A_k} f d\mu \xrightarrow{M \rightarrow \infty} 0.$$

On conclut que l'on peut trouver $M = M(k)$ tel que

$$\sup_{f \in \mathcal{F}} \int_X |f - f^{k,M}| d\mu \leq \frac{2}{k}.$$

Sans perte de généralité, $M(k)$ est une fonction croissante de k .

3. Si l'on fixe k et M , la famille $\{f^{k,M}; f \in \mathcal{F}\}$ est uniformément bornée sur A_k , donc précompacte pour la topologie $\sigma(L^\infty(A_k), L^1(A_k))$ par le Corollaire VII-130. Comme A_k est de mesure finie, on a $L^\infty(A_k) \subset L^1(A_k)$, donc la précompacité a aussi lieu au sens $\sigma(L^1(A_k), L^\infty(A_k))$.

4. À ce stade on ne peut conclure par la Proposition VII-119, car la topologie $\sigma(L^1, L^\infty)$ n'est pas métrique ! (voir l'Exercice VII-135) Nous allons cependant reprendre le raisonnement qui sous-tend cette proposition, en remplaçant l'argument de convergence métrique par un argument de monotonie.

Soit donc $(f_n)_{n \in \mathbb{N}}$ une suite d'éléments de \mathcal{F} , on pose $f_n^k = f_n^{k, M(k)}$. Pour tout k , on peut trouver une suite extraite $(f_{n'}^k)$ qui converge vers une fonction g^k dans la topologie faible de $L^1(A_k)$. Quitte à effectuer une extraction diagonale, on peut supposer que pour tout k , f_n^k converge vers g^k dans $L^1(A_k)$ faible.

On étend g_k en-dehors de A_k par la fonction nulle ; on pose en outre $f_n^0 = 0$, $g^0 = 0$. Comme f_n^k est croissante en k , on a, pour tout $\varphi \geq 0$ mesurable, et pour tous indices $k \geq \ell$, $\int (f_n^k - f_n^\ell) \varphi \geq 0$, d'où par passage à la limite $\int (g^k - g^\ell) \varphi \geq 0$. Le choix $\varphi = 1_{g^k \leq g^\ell}$ montre que $g^k \geq g^\ell$, μ -presque partout. La suite $(g^k)_{k \in \mathbb{N}}$ est donc positive et croissante, à un ensemble μ -négligeable près ; en outre $\int g^k$ est majorée par le supremum des intégrales des f_n . Posons $g = \sup_k g^k$; on a alors par convergence monotone

$$\int_X |g^k - g| d\mu \xrightarrow{k \rightarrow \infty} 0.$$

5. On conclut aisément : si $\chi \in L^\infty(X)$, alors

$$\begin{aligned} \left| \int f_n \chi d\mu - \int g \chi d\mu \right| &\leq \int_X |f_n - f_n^k| |\chi| d\mu + \left| \int_X f_n^k \chi d\mu - \int_X g^k \chi d\mu \right| \\ &\quad + \int_X |g^k - g| |\chi| d\mu \\ &\leq \|\chi\|_{L^\infty} \left(\int |f_n - f_n^k| d\mu + \int |g^k - g| d\mu \right) + \left| \int_{A_k} f_n^k \chi d\mu - \int_{A_k} g^k \chi d\mu \right|. \end{aligned}$$

Le premier terme du membre de droite converge vers 0 quand $k \rightarrow \infty$, uniformément en n ; et le second converge vers 0 quand $n \rightarrow \infty$, pour tout k fixé ; on conclut que $\int f_n \chi d\mu$ converge vers $\int g \chi d\mu$, ce qui montre que f_n converge faiblement vers g . \square

EXERCICE VII-135. On considère, dans $L^1([-\pi, \pi])$, la famille des $f_{m,n}$ ($m \geq n \geq 1$) définies par

$$f_{m,n}(x) = \sin(nx) + n1_{[-n^{-1}, n_{-1}]}$$

Montrer que $f_{m,n}$ est bornée dans L^1 . En utilisant le Théorème de Dunford-Pettis, identifier toutes les limites de suites extraites convergeant faiblement dans L^1 . Montrer que la fonction identiquement nulle n'est limite d'aucune suite extraite, mais que c'est une limite de limites de suites extraites. En déduire que la convergence dans $\sigma(L^1, L^\infty)$ n'est pas métrisable, même si on se restreint aux parties bornées de L^1 .

Passons maintenant, pour terminer ce paragraphe, au Théorème de Prokhorov. Sa démonstration a des points communs avec celle du Théorème de Dunford-Pettis, mais aussi d'importantes différences. Alors que le Théorème de Dunford-Pettis relevait purement de la théorie de la mesure, le Théorème de Prokhorov mélange théorie de la mesure et topologie.

THÉORÈME VII-136 (Théorème de Prokhorov). *Soient (X, d) un espace Polonais et \mathcal{M} un ensemble de mesures positives. Alors \mathcal{M} est précompact pour la topologie faible $\sigma(M(X), C_b(X))$ si et seulement si*

(i) \mathcal{M} est borné en variation totale ;

(ii) \mathcal{M} est tendu, i.e.

$$\forall \varepsilon > 0 \quad \exists K_\varepsilon \text{ compact}; \quad \sup_{\mu \in \mathcal{M}} \mu[X \setminus K_\varepsilon] \leq \varepsilon.$$

EXEMPLES VII-137. L'espace des mesures de probabilité sur un espace compact est lui-même compact (ce que nous savions déjà grâce au Théorème de Banach-Alaoglu). L'espace des mesures de probabilité sur \mathbb{R}^n vérifiant

$$\int |x|^\varepsilon d\mu(x) \leq C$$

est compact pour tous $C < +\infty$ et $\alpha > 0$.

REMARQUE VII-138. Une subtilité importante est à noter : le Théorème de Prokhorov s'applique à des mesures **positives**. Si l'on dispose d'un ensemble de mesures signées μ , telles que les mesures $|\mu|$ vérifient les conditions (i) et (ii), alors on peut obtenir la précompacité en appliquant le Théorème de Prokhorov aux parties positives et aux parties négatives séparément. En revanche, pour prouver que la précompacité implique la condition (ii), on va utiliser la positivité de manière importante. Je ne sais pas si le résultat reste vrai pour des ensembles de mesures signées !

DÉMONSTRATION DU THÉORÈME DE PROKHOROV. Montrons d'abord que la précompacité entraîne les conditions (i) et (ii). On rappelle que d'après le Théorème I-54, une mesure finie sur un espace Polonais est automatiquement régulière.

1. Si \mathcal{F} n'est pas bornée dans M , on peut construire une suite (μ_k) à valeurs dans \mathcal{M} , avec $\|\mu_k\|_{VT} \geq k$. Puisque $C_b(X)$ est complet, le Théorème de Banach–Steinhaus s'applique pour dire que toute suite de formes linéaires convergeant simplement est bornée en norme duale. On sait (Théorème VI-64) que la norme de μ_k dans $(C_b)^*$ est $\|\mu_k\|_{VT}$, qui n'est pas borné ; donc aucune sous-suite extraite de (μ_k) ne peut converger simplement, et \mathcal{M} ne peut être précompact. Cet argument reste d'ailleurs vrai pour des mesures signées.

2. Supposons maintenant que \mathcal{M} est précompact, et montrons qu'elle est tendue. Soit $(x_j)_{j \in \mathbb{N}}$ une suite dense dans X , on considère la famille des boules $B_j = B_\delta(x_j)$ où $\delta > 0$ est arbitrairement petit. Montrons que pour tout $\varepsilon > 0$ il existe N_ε tel que pour tout $\mu \in \mathcal{M}$, $\mu[X \setminus (B_1 \cup \dots \cup B_{N_\varepsilon})] \leq \varepsilon$. En effet, si ce n'était pas le cas, on pourrait construire par récurrence une suite μ_n à valeurs dans \mathcal{M} , telle que pour tout n , $\mu_n[X \setminus (B_1 \cup \dots \cup B_n)] > \varepsilon$. En particulier, pour tout $n \geq k$ on a $\mu_n[F_k] > \varepsilon$, où $F_k = X \setminus (B_1 \cup \dots \cup B_k)$ est un fermé.

On va en déduire que $\mu[F_k] \geq \varepsilon$; il s'agit d'un argument général qui repose uniquement sur le caractère fermé de F_k . Soit F_k^t l'ensemble des $x \in X$ qui sont à distance au moins t de F_k . Comme F_k et F_k^t sont des fermés disjoints dans un espace métrique, on peut construire une fonction φ^t à valeurs dans $[0, 1]$, identiquement égale à 1 sur F_k et à 0 sur F_k^t (se rappeler le paragraphe I-2.3). On a alors

$$\mu[X \setminus F_k^t] \geq \int \varphi^t d\mu = \lim_{n \rightarrow \infty} \int \varphi^t d\mu_n \geq \limsup_{n \rightarrow \infty} \mu_n[F_k^t] \geq \varepsilon.$$

En passant à la limite quand $t = 1/m \rightarrow 0$ par σ -additivité, on obtient

$$\mu[F_k] \geq \varepsilon.$$

Autrement dit, $\mu[X \setminus (B_1 \cup \dots \cup B_k)] \geq \varepsilon$, et ce pour tout k . On obtient une contradiction puisque l'union des B_k est l'espace X tout entier.

3. Pour tout $\delta > 0$ et pour tout $\varepsilon > 0$ nous avons démontré l'existence de boules ouvertes $B_1, \dots, B_{N_\varepsilon}$ de rayon δ , telles que pour tout $\mu \in \mathcal{M}$, $\mu[X \setminus (B_1 \cup \dots \cup B_{N_\varepsilon})] \leq \varepsilon$. Pour déduire que \mathcal{M} est tendue, on répète mot à mot le raisonnement de la fin de la démonstration du Théorème I-54 (le Théorème d'Ulam, qui est exactement le résultat que nous voulons montrer dans le cas où la famille \mathcal{M} est un singleton).

4. Montrons maintenant que si la famille \mathcal{M} est tendue, alors elle est précompact. Il suffit de prouver que toute suite tendue admet une sous-suite convergente.

Pour tout $m \geq 1$ on peut trouver un compact K_m tel que $\mu[X \setminus K_m] \leq 1/m$ pour tout $\mu \in \mathcal{M}$, et on peut supposer que les K_m forment une suite croissante. Pour chaque m , la suite (μ_n) définit une suite bornée dans $M(K_m)$, donc dans le dual de $C(K_m)$. Par le Théorème de Banach-Alaoglu, on peut extraire de μ_n une sous-suite qui converge au sens faible vers un certain $\mu^m \in M(K_m)$. Par une extraction diagonale, on peut extraire de μ_n une sous-suite telle que pour tout m , et pour tout $\varphi \in C(K_m)$,

$$\int_{K_m} \varphi d\mu_n \longrightarrow \int_{K_m} \varphi d\mu^m.$$

La mesure μ^m est bien sûr positive, car limite de mesures positives. Elle définit une mesure sur X tout entier, par restriction : $\mu^m[A] = \mu^m[A \cap K_m]$.

Soit $\varphi \in C(K_m)$, positive. Pour $m' \leq m$ on peut écrire

$$\int_{K_m} \varphi d\mu_n \geq \int_{K_{m'}} \varphi d\mu_n.$$

(La fonction φ induit bien sûr une fonction continue positive sur $K_{m'}$.) En passant à la limite quand $n \rightarrow \infty$, on obtient

$$\int_{K_m} \varphi d\mu^m \geq \int_{K_{m'}} \varphi d\mu^{m'}.$$

Si A est une partie fermée (donc compacte) de K_m , on peut introduire grâce au lemme d'Urysohn une suite $\varphi_k \in C(K_m)$, identiquement égale à 1 sur A , convergeant en décroissant vers la fonction indicatrice de A . En passant à la limite quand $k \rightarrow \infty$, on obtient

$$\mu^m[A] \geq \mu^{m'}[K_{m'} \cap A].$$

On en conclut que la suite de mesures $\mu^m[K_m \cap \cdot]$ est croissante en m .

Soit A une partie mesurable de X , on définit

$$\mu[A] := \lim_{m \rightarrow \infty} \mu^m[K_m \cap A].$$

Par la Remarque III-3, μ est une mesure. C'est une mesure finie puisque la suite $\|\mu^m\|_{VT}$ est bornée, et elle est bien sûr concentrée sur l'union croissante des K_m ; donc $\mu[X \setminus K_m] \rightarrow 0$ quand $m \rightarrow \infty$.

Montrons finalement que μ_n converge vers μ au sens faible. On se donne $\varphi \in C_b(X)$. Alors

$$\int_X \varphi d\mu_n = \int_{K_m} \varphi d\mu_n + \int_{X \setminus K_m} \varphi d\mu_n,$$

et le deuxième terme est majoré en valeur absolue par $\|\varphi\|_\infty \mu_n[X \setminus K_m] = O(1/m)$, uniformément en n . De même,

$$\int_X \varphi d\mu = \int_{K_m} \varphi d\mu + \int_{X \setminus K_m} \varphi d\mu,$$

et le deuxième terme tend vers 0 quand $m \rightarrow \infty$. Puisque d'autre part $\int_{K_m} \varphi d\mu_n$ converge vers $\int_{K_m} \varphi d\mu$, on conclut que

$$\int_X \varphi d\mu_n \xrightarrow{n \rightarrow \infty} \int_X \varphi d\mu.$$

□

Si la convergence dans L^1 faible n'est pas métrisable, la convergence au sens faible des mesures est en revanche métrisable (sur les parties bornées de l'espace des mesures positives). Soit d une distance bornée sur X (on peut toujours construire une telle distance, par exemple en remplaçant la distance d_0 initialement donnée par $d = d_0/(1 + d_0)$); alors la distance qui suit métrise la convergence faible :

$$d_{BL}(\mu, \nu) = \sup \left\{ \int f d\mu - \int f d\nu; \quad |f| \leq 1, \quad \|f\|_{\text{Lip}} \leq 1 \right\},$$

où l'on a noté

$$\|f\|_{\text{Lip}} = \sup_{x \neq y} \frac{|f(y) - f(x)|}{d(x, y)}.$$

On pourra penser aux lettres B et L comme aux initiales de “borné” et “Lipschitz”.

THÉORÈME VII-139. *Soit X un espace métrique Polonais, et soient $(\mu_n)_{n \in \mathbb{N}}$ et μ des mesures positives. Alors, les deux énoncés suivants sont équivalents :*

- (i) $d_{BL}(\mu_n, \mu) \rightarrow 0$;
- (ii) μ_n converge vers μ au sens faible $\sigma(M(X), C_b(X))$.

REMARQUES VII-140. (i) Le même énoncé resterait vrai si l'on remplaçait la condition de 1-Lipschitzianité par une condition de module de continuité uniforme.

(ii) Ici encore, la positivité joue un rôle important. Par exemple, dans $M(\mathbb{R})$, la suite de mesures signées $\mu_n := \delta_n - \delta_{n+1/n}$ vérifie $d_{BL}(\mu_n, \mu) \rightarrow 0$, mais ne converge pas faiblement vers 0 (exercice).

DÉMONSTRATION DU THÉORÈME VII-139. Supposons que $d(\mu_n, \mu) \rightarrow 0$. Alors, pour toute fonction Lipschitz ψ on a $\int \psi d\mu_n \rightarrow \int \psi d\mu$.

Soit maintenant φ une fonction continue, on pose

$$f_k(x) = \inf_{y \in X} \{\varphi(y) + kd(x, y)\}, \quad g_k(x) = \sup_{y \in X} \{\varphi(y) - kd(x, y)\}.$$

Il est clair que f_k et g_k sont Lipschitz, et que

$$g_k \leq \varphi \leq f_k,$$

en particulier

$$\begin{aligned} \int g_k d\mu &= \lim_{n \rightarrow \infty} \int g_k d\mu_n \leq \liminf_{n \rightarrow \infty} \int \varphi d\mu_n \\ &\leq \limsup_{n \rightarrow \infty} \int \varphi d\mu_n \leq \lim_{n \rightarrow \infty} \int f_k d\mu_n = \int f_k d\mu. \end{aligned}$$

D'autre part, on vérifie facilement que f_k converge en croissant vers φ quand $k \rightarrow \infty$, alors que la suite g_k converge en décroissant vers φ . Il est facile d'en déduire que

$$\lim_{k \rightarrow \infty} \int f_k d\mu = \int \varphi d\mu = \lim_{k \rightarrow \infty} \int g_k d\mu.$$

On conclut donc que

$$\lim_{n \rightarrow \infty} \int \varphi d\mu_n = \int \varphi d\mu.$$

Réciproquement, supposons que μ_n converge vers μ au sens faible. On introduit une suite (K_m) de compacts, choisis de telle sorte que $\mu_n[X \setminus K_m] \leq 1/m$ pour tout n , $\mu[X \setminus K_m] \leq 1/m$.

Soit maintenant φ_k une suite de fonctions bornées par 1, 1-Lipschitziennes, telles que

$$d_{BL}(\mu_k, \mu) \leq \int \varphi_k d(\mu_k - \mu) + \frac{1}{k}.$$

Pour tout compact K_m , la suite (φ_k) définit par restriction une suite bornée et 1-Lipschitzienne dans $C(K_m)$; grâce au Théorème d'Ascoli on peut en extraire une sous-suite qui converge uniformément dans $C(K_m)$. Par un procédé d'extraction diagonale, on peut extraire de la suite (φ_k) une sous-suite qui converge uniformément sur chaque compact K_m , vers une fonction φ , 1-Lipschitz, définie sur l'union des K_m . On prolonge cette fonction en une fonction 1-Lipschitz bornée définie sur X tout entier (Cf. les rappels du paragraphe I-2.3), toujours notée φ . En particulier, $\int \varphi d\mu_k \rightarrow \int \varphi d\mu$.

Pour montrer que $d_{BL}(\mu_k, \mu)$ tend vers 0, il suffit de montrer que $\int \varphi_k d\mu_k$ est très proche de $\int \varphi_k d\mu$. Pour cela on écrit

$$\begin{aligned} \left| \int_X \varphi_k d\mu_k - \int_X \varphi_k d\mu \right| &\leq \left| \int_{K_m} (\varphi_k - \varphi) d(\mu_k - \mu) \right| + \left| \int_{X \setminus K_m} (\varphi_k - \varphi) d(\mu_k - \mu) \right| \\ &\quad + \int_X \varphi d(\mu_k - \mu). \end{aligned}$$

Pour chaque m , le premier terme tend vers 0 puisque φ_k converge uniformément vers φ sur K_m ; le second terme tend vers 0 quand $m \rightarrow \infty$, uniformément en k , grâce à la tension des μ_k et de μ ; enfin le dernier terme tend vers 0 quand $k \rightarrow \infty$ grâce à l'hypothèse de convergence faible. \square

VII-5.6. De la convergence faible à la convergence forte. La convergence faible est (comme son nom l'indique!) plus faible que la convergence forte, néanmoins dans certaines situations on peut combiner l'information de convergence faible avec une autre information, pour obtenir la convergence forte. Voici le plus simple de ces résultats :

THÉORÈME VII-141. *Soit $(E, \|\cdot\|)$ un espace de Banach séparable, uniformément convexe, tel que x_n converge au sens faible $\sigma(E, E^*)$ vers x , et $\|x_n\| \rightarrow \|x\|$. Alors $\|x_n - x\| \rightarrow 0$.*

DÉMONSTRATION. Si $\|x_n\|$ tend vers 0, alors il n'y a rien à démontrer; dans le cas contraire, on pose $y_n = x_n / \|x_n\|$, de sorte que y_n converge au sens faible

vers $y = x/\|x\|$; il suffit de montrer que $\|y_n - y\| \longrightarrow 0$. Soit Λ une forme linéaire normalisant y . On a alors

$$\left\| \frac{y_n + y}{2} \right\| \geq \left\langle \Lambda, \frac{y_n + y}{2} \right\rangle \longrightarrow \langle \Lambda, y \rangle = \|y\| = 1.$$

Donc $\limsup \|(y_n + y)/2\| \geq 1$, ce qui montre que $\|(y_n + y)/2\|$ tend vers 1. L'uniforme convexité de E implique alors $\|y_n - y\| \longrightarrow 0$. \square

Comme nous le verrons dans la section suivante, ce critère admet dans les espaces de Lebesgue une généralisation en termes de fonctionnelles strictement convexes. L'autre critère utile de passage de la convergence faible à la convergence forte est le Théorème de Schur :

THÉORÈME VII-142 (Théorème de Schur). *Dans (X, μ) un espace mesuré, soit (f_n) une suite de fonctions convergeant presque partout vers f , précompacte dans $\sigma(L^1, L^\infty)$. Alors f_n converge vers f dans $L^1(X, \mu)$.*

DÉMONSTRATION. Quitte à remplacer f_n par $f_n - f$, on peut supposer que $f = 0$.

Soit $\varepsilon > 0$. La suite (f_n) étant précompacte, elle est uniformément équi-intégrable à l'infini par le Théorème de Dunford-Pettis ; il existe donc un ensemble de mesure finie, Y_ε , tel que pour tout n on ait

$$\int_{X \setminus Y_\varepsilon} |f_n| d\mu \leq \varepsilon.$$

De même, il existe $\delta > 0$ tel que pour tout A de mesure au plus δ on ait $\int_A |f_n| d\mu \leq \varepsilon$ pour tout n .

Sur l'ensemble Y_ε , la suite f_n converge presque partout vers 0, donc par le Théorème d'Egorov il existe un ensemble A_δ , de mesure au plus δ , en-dehors duquel f_n converge uniformément vers 0, et donc dans $L^1(Y_\varepsilon \setminus A_\delta)$ puisque Y_ε est de mesure finie.

Pour conclure, on écrit

$$\int_X |f_n| d\mu = \int_{X \setminus Y_\varepsilon} |f_n| d\mu + \int_{A_\delta} |f_n| d\mu + \int_{Y_\varepsilon \setminus A_\delta} |f_n| d\mu,$$

ou les deux premiers morceaux sont majorés par ε et le dernier tend vers 0 quand $n \rightarrow \infty$. Pour n assez grand, cette quantité est donc majorée par 3ε . On conclut que $\int |f_n|$ tend vers 0. \square

Le Théorème de Schur peut se reformuler ainsi : la convergence dans L et la compacité dans L^1 faible entraîne la convergence dans L^1 . On comparera avec le Théorème VI-42. Voici un corollaire simple et surprenante du Théorème de Schur :

COROLLAIRE VII-143 (équivalence des convergences faible et forte). *Soit (X, μ) un espace mesuré dans lequel tous les points sont de mesure strictement positive. Alors, dans $L^1(X, \mu)$, il est équivalent de dire qu'une suite (f_n) converge au sens de la topologie faible $\sigma(L^1, L^\infty)$, ou qu'elle converge au sens de la norme L^1 .*

EXEMPLE VII-144. La convergence faible dans $\ell^1(\mathbb{N})$ coïncide avec la convergence forte (cela n'empêche que la topologie faible dans $\ell^1(\mathbb{N})$ soit distincte de la topologie forte!).

VII-5.7. Compléments sur la convergence faible des mesures de probabilité. Dans ce paragraphe on revient sur la convergence de mesures de probabilité. On commence par un critère de convergence.

PROPOSITION VII-145. *Soit X un espace Polonais, et (μ_n) une famille de mesures de probabilité. Alors, les quatre énoncés suivants sont équivalents :*

- (i) μ_n converge vers μ au sens $\sigma(M(X), C_b(X))$;
- (ii) Pour tout fermé $F \subset X$, $\mu[F] \geq \limsup \mu_n[F]$;
- (iii) Pour tout ouvert $O \subset X$, $\mu[O] \leq \limsup \mu_n[O]$;
- (iv) Pour tout $A \subset X$, $\mu[\partial A] = 0 \implies \mu[A] = \lim \mu_n[A]$.

DÉMONSTRATION. Preuve à compléter. Voir [Billingsley 2, Theorem 2.1] \square

Un cas particulier de ce théorème concerne la convergence de mesures de probabilité sur \mathbb{R} . Une telle mesure de probabilité μ est entièrement définie par sa **fonction de répartition** : $F(x) = \mu[] - \infty, x]$, qui est croissante et continue à droite, tendant vers 0 en $-\infty$ et 1 en $+\infty$. La Proposition précédente implique :

PROPOSITION VII-146. *Soit $(\mu_n)_{n \in \mathbb{N}}$ et μ des mesures de probabilité sur \mathbb{R} , de fonctions de répartitions respectives $(F_n)_{n \in \mathbb{N}}$ et F . Alors μ_n converge faiblement vers μ si et seulement si $F_n(x) \rightarrow F(x)$ en tout point x où F est continue.*

DÉMONSTRATION. Preuve à compléter. Voir [Billingsley 2, Theorem 2.3] \square

Un autre point sur lequel on pourrait dire plus est la métrisabilité de l'espace $P(X)$ des mesures de probabilité sur un espace métrique (typiquement Polonais) X . Nous avons déjà rencontré une distance intéressante, la distance d_{BL} ; des dizaines d'autres sont utilisées, soit sur l'espace $P(X)$ tout entier, soit sur des sous-espaces tels que

$$P_p(X) := \left\{ \mu \in P(X); \quad \exists x_0 \in X; \quad \int_X d(x_0, x)^p d\mu(x) < +\infty. \right\}$$

(cet ensemble dépend bien sûr du choix de la distance d). Mentionnons par exemple

- la distance de Kantorovich-Rubinstein sur $P_1(X)$:

$$d_{KR}(\mu, \nu) = \sup_{\|\varphi\|_{\text{Lip}} \leq 1} \int_X \varphi d(\mu - \nu);$$

- les distances de Wasserstein W_p sur $P_p(X)$:

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu, \nu)} \int d(x, y)^p d\pi(x, y) \right)^{\min(1, 1/p)},$$

où $\Pi(\mu, \nu)$ désigne l'ensemble des mesures de probabilité sur $X \times Y$ dont les marginales sont μ et ν ;

- la distance de Lévy-Prokhorov sur $P(X)$:

$$\beta(\mu, \nu) = \inf \left\{ \varepsilon > 0; \quad \mu[A] \leq \mu[A_\varepsilon] + \varepsilon \quad \text{pour tout ensemble Borélien } A \subset X \right\},$$

où A_ε désigne l'ensemble $\{y \in X; d(y, A) < \varepsilon\}$.

Il s'avère que la distance de Kantorovich-Rubinstein coïncide (si X est un espace Polonais) avec la distance W_1 ; on en saura plus en consultant l'ouvrage de l'auteur, *Topics in Optimal Transportation* (American Mathematical Society, Providence, 2003). Pour la distance de Lévy-Prokhorov on pourra consulter par exemple

[Dudley, Chapitre 11]. Comme nous l'avons déjà mentionné, il en existe des dizaines d'autres....

Pour conclure, évoquons brièvement le **principe de compacité-concentration**, qui est d'une grande importance dans certains domaines du calcul des variations. Soit $(\mu_n)_{n \in \mathbb{N}}$ une suite de mesures de probabilité sur un espace métrique Polonais X . Alors il y a trois comportements possibles :

- l'*évanescence* : la masse de μ_n ne reste pas contenue dans une boule de rayon R :

$$\forall R > 0 \quad \lim_{n \rightarrow \infty} \sup_{x \in X} \mu_n[B_R(x)] = 0;$$

- la *concentration* : la masse de μ_n reste "concentrée" sur un ensemble de diamètre borné, autour d'un point x_n qui peut bouger avec n : il existe une suite $(x_n)_{n \in \mathbb{N}}$ telle que

$$\forall \varepsilon > 0 \quad \exists R > 0 \quad \forall n \in \mathbb{N} \quad \mu_n[B_R(x_n)] \geq 1 - \varepsilon;$$

- la *dichotomie* : la masse de μ_n se sépare en au moins deux morceaux non négligeables :

$$\exists \alpha \in]0, 1[\quad \forall \varepsilon > 0 \quad \exists R > 0, R_n \rightarrow \infty, \quad \exists x_n \in X;$$

$$|\mu_n[B_R(x_n)] - \alpha| \leq \varepsilon, \quad \left| \mu_n[B_{R_n}(x_n) \setminus B_R^f(x_n)] \right| \leq \varepsilon.$$

(Faire des dessins!)

Pour une démonstration de ce résultat, et de nombreuses applications à des problèmes variationnels faisant intervenir des énergies variées, on pourra consulter une série d'articles de référence de **P.L. Lions** : *The concentration-compactness principle in the Calculus of Variations* (Ann. Inst. H. Poincaré Anal. Non Linéaire **1** (1984), no.2, 109–145, no.4, 223–283, et Rev. Mat. Iberoamericana **1** (1985), no.1, 145–201 et no.2, 45–121).

VII-6. Intégration au sens faible

Dans le Chapitre II nous avons appris à intégrer des fonctions mesurables f à valeurs réelles, sous la condition $\int |f| < +\infty$. Si f est à valeurs dans un \mathbb{R} -espace vectoriel de dimension finie (tel que \mathbb{C} ou \mathbb{R}^n), on peut facilement définir l'intégrale de f composante par composante. En revanche, si f est à valeurs dans un \mathbb{R} -espace vectoriel de dimension infinie, l'intégration de f peut s'avérer beaucoup plus délicate. La notion d'**intégrale faible** permet de surmonter cette difficulté de manière très simple, par dualité, sans avoir à recourir à la théorie générale d'intégration à valeurs vectorielles.

THÉORÈME VII-147 (intégrale faible-*). Soient (X, \mathcal{A}, μ) un espace mesuré, E un espace vectoriel normé, et E^* le dual de E . Soit $f : X \rightarrow E^*$, mesurable quand E^* est muni de la topologie faible-*, telle que $\int \|f(x)\| d\mu(x) < +\infty$. Alors il existe un unique élément ξ de E^* tel que

$$\forall z \in E, \quad \langle z, \xi \rangle_{E \times E^*} = \int \langle z, f(x) \rangle_{E \times E^*} d\mu(x).$$

On dit que ξ est l'intégrale faible-* de f et on note

$$\xi = \int_X f(x) d\mu(x).$$

DÉMONSTRATION. On note d'abord que la fonction $\|f\|$ est mesurable puisque la norme est semi-continue inférieurement pour la topologie faible.

Pour tout $z \in E$ on a $|\langle z, f(x) \rangle| \leq \|z\|_E \|f(x)\|_{E^*} \in L^1(\mu)$; donc $\int \langle z, f(x) \rangle d\mu(x)$ est bien défini. On peut alors poser, pour tout $z \in E$,

$$\xi(z) = \int \langle z, f(x) \rangle d\mu(x).$$

Par bilinéarité du crochet de dualité, et linéarité de l'intégrale, la fonction $\xi : E \rightarrow \mathbb{R}$ est en fait une forme linéaire continue; ceci conclut la preuve du théorème. \square

REMARQUE VII-148. Comme on pourra le vérifier facilement en utilisant la caractérisation donnée dans la définition, les règles de calcul habituelles s'appliquent à cette notion : avec des notations évidentes,

$$\int_A f d\mu = \int 1_A f d\mu; \quad \int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu;$$

$$\left\| \int f d\mu \right\|_{E^*} \leq \int \|f\|_{E^*} d\mu.$$

En particulier, le théorème de convergence dominée et le théorème d'Egorov s'appliquent (on se ramène à \mathbb{R}_+ en utilisant la norme). Le théorème de Fubini s'applique également (on se ramène à \mathbb{R} en utilisant la dualité).

REMARQUE VII-149. Il existe une notion similaire d'intégrale faible (au lieu de faible-*) pour $f : X \rightarrow E$, définie par les relations $\langle \int f d\mu, \xi \rangle = \int \langle f, \xi \rangle d\mu$. Cependant, en général l'élément ainsi construit appartient à E^{**} . Dans le cas où E est réflexif, il est bien sûr équivalent d'utiliser cette notion ou d'utiliser l'intégrale faible-*.

REMARQUE VII-150. Dans les cas où le Théorème VII-147 ne s'applique pas, il est parfois possible de définir simplement l'intégrale "à la main".

EXEMPLE VII-151. Soient (X, \mathcal{A}, μ) et $(Y, \mathcal{B}, \lambda)$ deux espaces mesurés, $p \in [1, \infty]$, et $f \in L^1(X; L^p(Y))$ (c'est-à-dire que $\int \|f(x)\|_{L^p} d\mu(x) < +\infty$). On cherche à définir $\int f d\mu$ dans $L^p(Y)$. Si $1 < p < \infty$, on peut appliquer le Théorème VII-147 en utilisant le fait que L^p est le dual de $L^{p'}$ ($p' = p/(p-1)$). Si $p = \infty$ et Y est σ -fini, il en est de même.

Si $p = 1$, on ne peut appliquer le Théorème VII-147, cependant, en supposant X et Y σ -finis, on peut utiliser tout simplement le théorème de Fubini pour définir $\int f d\mu$. En effet, si l'on note $f(x, y) = f(x)(y)$, et $\xi(y) = \int f(x, y) d\mu(x)$, le théorème de Fubini garantit que $\xi \in L^1(\lambda)$, et que pour tout $g \in L^\infty(\lambda)$,

$$\int \xi(y) g(y) d\lambda(y) = \int \left(\int f(x, y) g(y) d\lambda(y) \right) d\mu(x).$$

EXEMPLE VII-152. Soient (X, \mathcal{A}, μ) un espace mesuré, (Y, \mathcal{B}) un espace mesurable, et $f \in L^1(X; M(Y))$. On cherche à donner un sens à $\int f d\mu$. Si X est localement compact, alors $M(Y) = C_0(Y)^*$ et on peut appliquer le Théorème VII-147. Dans le cas général où X n'est pas forcément localement compact, on ne peut plus appliquer ce théorème; cependant, on peut toujours définir $\int f(x, dy) \mu(dx)$ par le

même procédé que dans la sous-section III-2.4 : avec la notation $f(x, dy) = f(x)(dy)$, on peut poser, pour tout ensemble mesurable A ,

$$\left(\int f(x, dy) \mu(dx) \right) [A] = \int f(x, A) \mu(dx).$$

VII-7. Convexité

DÉFINITION VII-153. Soit E un \mathbb{R} -espace vectoriel. Un ensemble $C \subset E$ est dit convexe si

$$\forall x, y \in C, \quad \forall \lambda \in [0, 1], \quad (1 - \lambda)x + \lambda y \in C.$$

Une fonction φ définie sur un ensemble convexe C de E , à valeurs dans $\mathbb{R} \cup \{+\infty\}$, est dite convexe si

$$\forall x, y \in C, \quad \forall \lambda \in [0, 1], \quad \varphi((1 - \lambda)x + \lambda y) \leq (1 - \lambda)\varphi(x) + \lambda\varphi(y).$$

REMARQUE VII-154. Quitte à étendre φ par $+\infty$ en-dehors de C , on peut toujours considérer que la fonction φ est définie sur E tout entier.

EXEMPLES VII-155. Si C est convexe, la fonction indicatrice convexe de C est la fonction convexe \mathbf{I}_C définie par : $\mathbf{I}_C(x) = 0$ si $x \in C$, $+\infty$ sinon. Toute norme est convexe.

DÉFINITION VII-156. Si $\varphi : E \rightarrow \mathbb{R} \cup \{+\infty\}$ est une fonction convexe, on définit le domaine de φ comme étant $\varphi^{-1}(\mathbb{R})$; c'est un ensemble convexe.

La notion de convexité a déjà joué un rôle important à plusieurs reprises dans ce cours, en liaison avec la régularité des espaces de Banach ou le théorème de projection. Dans cette section je vais passer en revue quelques résultats plus spécifiquement associés aux ensembles et fonctions convexes ; on parle d'**analyse convexe**. Comme dans le reste du cours, j'utiliserai la séparabilité pour éviter le recours à la version la plus générale de l'axiome du choix.

On rappelle à toutes fins utiles une propriété utile de régularité automatique des fonctions convexes en dimension finie :

PROPOSITION VII-157. Soit $\varphi : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe. Alors φ est continue et localement lipschitzienne dans l'intérieur de son domaine.

Démonstration à rédiger.

VII-7.1. Convexité et topologie faible. Il est clair qu'un ensemble fermé pour la topologie faible est toujours fermé pour la topologie forte, la réciproque étant fautive en général. (Attention à la terminologie, la propriété de fermeture faible est donc plus forte que celle de fermeture forte!) Cependant, pour les ensembles convexes, il n'y a pas de différence :

THÉORÈME VII-158. Soient E un espace vectoriel normé séparable, et $C \subset E$ un convexe fortement fermé. Alors C est également fermé pour la topologie faible.

EXEMPLE VII-159. Soit H un espace de Hilbert séparable de dimension infinie. La sphère unité $S = \{x \in H; \|x\| = 1\}$, non convexe, est fermée au sens fort mais pas au sens faible. Son adhérence faible est la boule unité $B = \{x \in H; \|x\| \leq 1\}$, fermée à la fois au sens fort et au sens faible.

PREUVE DU THÉORÈME VII-158. Soient C un convexe fermé et $x \notin C$. Par le Théorème VII-84(ii) (de Hahn–Banach), il existe une forme linéaire continue Λ séparant C et $\{x\}$ au sens fort :

$$\Lambda(x) > \sup_{z \in C} \Lambda(z).$$

Puisque Λ est continue pour la topologie faible, cette inégalité stricte empêche x d'appartenir à l'adhérence de C . On conclut que C est fermé pour la topologie faible. \square

Le corollaire suivant est très utile :

COROLLAIRE VII-160. *Soit E un espace vectoriel normé séparable, et $\Phi : E \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, semicontinue inférieurement pour la topologie forte. Alors Φ est également semicontinue inférieurement pour la topologie faible.*

DÉMONSTRATION. Pour tout $\alpha \in \mathbb{R}$, $\Phi^{-1}([\alpha, +\infty])$ est convexe et fermé pour la topologie forte, donc également pour la topologie faible par le Théorème VII-158. Ceci veut dire que Φ est semicontinue inférieurement pour la topologie faible. \square

EXEMPLE VII-161. Soient (X, μ) un espace mesuré, $\psi : \mathbb{R} \rightarrow [0, +\infty]$ une fonction convexe, et soit $\Psi : L^p(X) \rightarrow [0, +\infty]$ définie par

$$\Psi(f) = \int \psi(f(x)) d\mu(x).$$

Il est clair que Ψ est convexe. Si $f_k \rightarrow f$ dans L^p , on peut extraire une sous-suite $(f_{k(\ell)})_{\ell \in \mathbb{N}}$ telle que

$$\Psi(f_{k(\ell)}) \xrightarrow{\ell \rightarrow \infty} \liminf_{k \rightarrow \infty} \Psi(f_k).$$

Quitte à extraire une nouvelle sous-suite, on peut supposer que $f_k \rightarrow f$ μ -presque partout. On a alors, par le Théorème III-9 (Lemme de Fatou),

$$\Psi(f) = \int \lim_{k \rightarrow \infty} \psi(f_k) d\mu \leq \lim_{k \rightarrow \infty} \int \psi(f_k) d\mu = \liminf_{k \rightarrow \infty} \Psi(f_k).$$

On conclut que Ψ est semicontinue inférieurement pour la topologie forte, et le Corollaire VII-160 garantit que Ψ est aussi semicontinue inférieurement pour la topologie faible.

REMARQUE VII-162. Ce résultat peut aussi se démontrer au moyen de la dualité de Legendre, que nous étudierons brièvement plus tard.

REMARQUE VII-163. Le cas particulier $\psi(f) = |f|^p$ redonne la semicontinuité inférieure de la norme L^p pour la topologie faible — mais l'Exemple VII-161 prouve également la semicontinuité inférieure de la norme L^q pour la topologie faible L^p .

En conclusion, si Φ est une fonction convexe semicontinue sur un espace vectoriel normé et $x_k \rightharpoonup x$, alors $\Phi(x) \leq \liminf \Phi(x_k)$. Pour $\Phi(x) = \|x\|$, dans un espace uniformément convexe, on sait bien que les cas d'égalité caractérisent la convergence forte : si $x_k \rightharpoonup x$ et $\|x_k\| \rightarrow \|x\|$, alors $x_k \rightarrow x$ (Théorème VII-141). On peut naturellement se demander si ce résultat se généralise ; et en effet on a par exemple l'énoncé suivant :

THÉORÈME VII-164. *Soit E un espace vectoriel normé séparable. Vérifier que c'est utile et $\Phi : E \rightarrow \mathbb{R}$ une fonction continue vraie continuité nécessaire ? et strictement convexe, au sens suivant : pour tous $x, y \in E$ il existe $\varepsilon = \varepsilon(x, y) > 0$ tel que*

$$\Phi((1 - \lambda)x + \lambda y) - (1 - \lambda)\Phi(x) - \lambda\Phi(y) \leq -\lambda(1 - \lambda)\varepsilon.$$

Si $x_k \rightarrow x$ et $\Phi(x_k) \rightarrow \Phi(x)$, alors $x_k \rightarrow x$.

Ce théorème sera démontré plus tard, comme application de la dualité de Legendre (voir p. ??).

VII-7.2. Points extrémaux.

DÉFINITION VII-165 (points extrémaux). *Soient E un espace vectoriel normé et C un convexe de E . On dit que $x \in C$ est extrémal (dans C) si x ne peut s'écrire comme combinaison convexe stricte de deux autres points de C . Plus explicitement :*

$$\exists(x_0, x_1, \lambda) \in C \times C \times [0, 1]; \quad x_0 \neq x_1, \quad \lambda \notin \{0, 1\}, \quad x = (1 - \lambda)x_0 + \lambda x_1.$$

De manière équivalente, x est un point extrémal s'il n'est milieu d'aucun segment non trivial contenu dans C .

NOTATION VII-166. Si C est un convexe on notera $\mathcal{E}(C)$ l'ensemble des points extrémaux de C .

EXEMPLE VII-167. La figure VII-7.2 représente l'ensemble des points extrémaux de quelques convexes fermés dans \mathbb{R}^2 ou \mathbb{R}^3 .

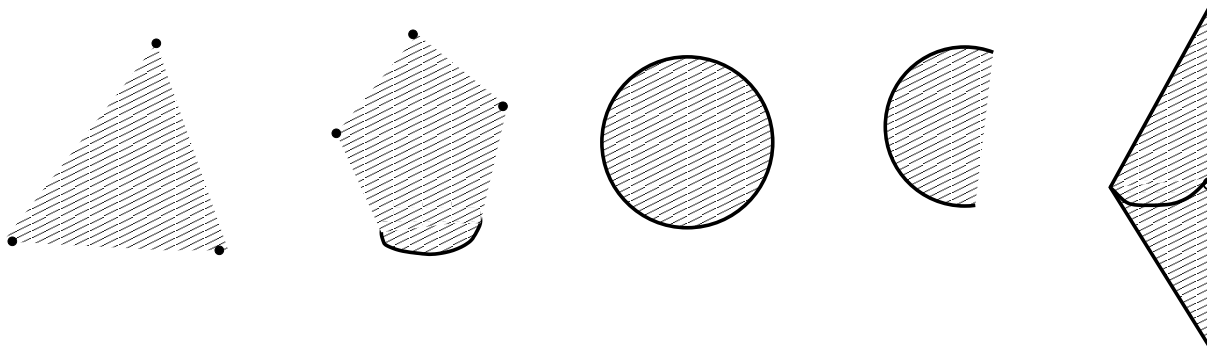


FIG. 5. En traits épais, les ensembles de points extrémaux de quelques convexes simples.

Le dernier exemple montre que $\mathcal{E}(C)$ n'est en général pas fermé, même en dimension finie. Pour autant, cela reste un ensemble raisonnable, comme le montre la proposition suivante :

PROPOSITION VII-168. *Soit C un convexe compact dans un espace vectoriel normé E . Alors $\mathcal{E}(C)$ est un borélien non vide, et plus précisément une intersection dénombrable d'ouverts.*

DÉMONSTRATION DE LA PROPOSITION VII-168. Soit $m : (x, y) \mapsto (x + y)/2$ (évidemment continue); et soit $\Delta = \{(x, x); x \in C\}$ la diagonale de C . L'ensemble $C \times C \setminus \Delta$ est l'union dénombrable des compacts

$$K_\ell = \{(x, y) \in C \times C; d(x, y) \geq 1/\ell\}$$

pour $\ell \in \mathbb{N}$. L'ensemble $C \setminus \mathcal{E}(C)$ coïncide avec $m(C \times C \setminus \Delta) = \cup_{\ell \in \mathbb{N}} m(K_\ell)$, qui est une union de compacts. L'ensemble $\mathcal{E}(C)$ est donc une intersection dénombrable d'ouverts.

Il reste à montrer que $\mathcal{E}(C)$ est non vide. Quitte à travailler dans l'espace de Banach engendré par C , on peut supposer que E est séparable. (En effet C est compact donc séparable.) Par la Proposition VII-79 il existe une suite $(\Lambda_k)_{k \in \mathbb{N}}$ de formes linéaires continues séparant C . On définit une suite $(C_k)_{k \in \mathbb{N}}$ de compacts par

$$C_0 = C;$$

$$C_1 = \left\{ x \in C; \Lambda_1(x) = \max_{y \in C_0} \Lambda_1(y) \right\} = \text{Argmax}_{C_0} \Lambda_1;$$

$$\dots C_k = \left\{ x \in C_{k-1}; \Lambda_k(x) = \max_{y \in C_{k-1}} \Lambda_k(y) \right\} = \text{Argmax}_{C_{k-1}} \Lambda_k.$$

La suite $(C_k)_{k \in \mathbb{N}}$ est une famille décroissante de compacts non vides, son intersection S est donc non vide. Soit $z \in S$; on va vérifier que z est extrémal. Supposons que $z = (x + y)/2$, le but est de montrer que $x = y$.

Pour commencer on a $\max_{C_0} \Lambda_1 = \Lambda_1(z) = (\Lambda_1(x) + \Lambda_1(y))/2 \leq \max_{C_0} \Lambda_1$; d'où $\Lambda_1(x) = \Lambda_1(y) = \Lambda_1(z)$, en particulier $x, y \in C_1$. En continuant ce raisonnement, on montre par récurrence que pour tout k on a $x, y \in C_k$ et $\Lambda_k(x) = \Lambda_k(y) = \Lambda_k(z)$. Puisque les Λ_k séparent C , on a forcément $x = y = z$, ce qui conclut la preuve. \square

EXEMPLE VII-169. Soit X un espace topologique séparé, et $P(X)$ l'espace des mesures de probabilités (de Borel) sur X . Alors les points extrémaux de $P(X)$ sont les masses de Dirac (exercice).

EXEMPLE VII-170. Soit \mathcal{B}_n l'espace des matrices $n \times n$ bistochastiques, c'est-à-dire les matrices $M = (a_{ij})$ telles que tous les coefficients a_{ij} sont positifs, et $\sum_i a_{ij} = 1$, $\sum_j a_{ij} = 1$. Un théorème célèbre dû à Birkhoff énonce que les points extrémaux de \mathcal{B}_n sont les matrices de permutation, c'est-à-dire telles que $a_{ij} \in \{0, 1\}$.

L'importance cruciale des points extrémaux vient de ce que, sous des hypothèses adéquates, ils suffisent à "engendrer" le convexe tout entier, via des combinaisons linéaires finies ou infinies. Le théorème suivant peut être considéré comme le plus simple de ces résultats; on notera $\overline{\text{co}}(A)$ l'enveloppe convexe fermée de A , c'est-à-dire le plus petit convexe fermé contenant A , ou encore l'adhérence de l'enveloppe convexe de A .

THÉORÈME VII-171 (Théorème de Krein–Milman). *Soit C un convexe compact d'un espace vectoriel normé E . Alors C est l'enveloppe convexe fermée de ses points extrémaux :*

$$\overline{\text{co}}(\mathcal{E}(C)) = C.$$

DÉMONSTRATION. Quitte à travailler dans le sous-espace vectoriel engendré par C , on peut supposer que E est séparable.

Soit $C' = \overline{\text{co}}(\mathcal{E}(C))$. Bien sûr $C' \subset C$, et C' est également un convexe compact. Si $C' \neq C$, soit $x \in C \setminus C'$; par Hahn–Banach (Théorème VII-84(ii)) il existe $\Lambda \in E^*$ tel que $\Lambda(x) > \sup_{C'} \Lambda$.

Soit alors $\alpha = \max_C \Lambda$ et $C_\alpha = C \cap \{\Lambda = \alpha\}$: C_α est un convexe non vide, n'intersectant pas C' . Soit z un point extrémal de C_α , il est facile de voir que c'est aussi un point extrémal de C . Il doit donc appartenir à C' , ce qui est une contradiction. \square

En dimension finie, le Théorème VII-171 peut être précisé : il suffit de considérer l'enveloppe convexe, on n'a pas besoin de passer à l'adhérence :

THÉORÈME VII-172 (Théorème de Krein–Milman en dimension finie). *Soit C un convexe compact de \mathbb{R}^n . Alors C est l'enveloppe convexe de ses points extrémaux :*

$$\text{co}(\mathcal{E}(C)) = C.$$

DÉMONSTRATION. On va raisonner par récurrence sur la dimension n . En dimension 1, le résultat est évident : un intervalle est l'enveloppe convexe de ses extrémités. Supposons donc le résultat démontré jusqu'à la dimension $n - 1$, et établissons-le en dimension n .

Soit C un convexe compact de \mathbb{R}^n , et soit k le plus grand entier tel qu'il existe $k + 1$ points affinement libres dans $\mathcal{E}(C)$. Si $k < n$, cela veut dire que $\mathcal{E}(C)$ est inclus dans un espace affine V de dimension k , et il en est donc de même de $C = \overline{\text{co}}(\mathcal{E}(C))$. Quitte à translater, le problème se repose donc dans \mathbb{R}^n , et l'hypothèse de récurrence assure le résultat. On peut donc supposer que $k = n$.

Soient z_0, \dots, z_n des points extrémaux affinement libres ; $C' = \text{co}(\mathcal{E}(C))$ contient l'ensemble convexe engendré par z_0, \dots, z_n ; en particulier, C' est un convexe d'intérieur non vide. Il s'ensuit facilement que $\text{Int}(C') = \overline{C'}$, qui n'est autre que C .

Le but est de montrer que $C' = C$. Supposons par l'absurde que ce soit faux, et soit $\bar{x} \in C \setminus C'$. On ne sait pas a priori que C' est fermé, mais on peut toujours séparer (au sens large) \bar{x} de $\text{Int}(C')$ (Théorème VII-84(i)) : il existe $\lambda \in \mathbb{R}^n$ tel que

$$\langle \lambda, \bar{x} \rangle \geq \sup_{x \in \text{Int}(C')} \langle \lambda, x \rangle.$$

En passant à l'adhérence, on obtient

$$\langle \lambda, \bar{x} \rangle \geq \sup_{x \in C} \langle \lambda, x \rangle.$$

Soit $\widehat{C} = C \cap \{x; \langle \lambda, x \rangle = \langle \lambda, \bar{x} \rangle\} \ni \bar{x}$. Si $\widehat{x} \in \widehat{C}$ est extrémal dans C , alors il est extrémal aussi dans C (en effet s'il était combinaison stricte de deux points dont un n'appartient pas à \widehat{C} , alors on aurait $\langle \lambda, x \rangle < \max_C \lambda$, d'où contradiction).

Or, par définition, \bar{x} n'est pas combinaison convexe de points extrémaux de C ; il n'est donc pas combinaison convexe de points extrémaux de \widehat{C} . Mais ceci contredit l'hypothèse de récurrence car \widehat{C} est de dimension au plus $n - 1$. \square

VII-7.3. Représentation barycentrique. Si on se donne des vecteurs x_1, \dots, x_k dans un espace vectoriel E , et des coefficients $\lambda_1, \dots, \lambda_k$ positifs, de somme égale à 1, on définit le barycentre des x_i , avec poids λ_i , par

$$\beta = \sum \lambda_i x_i;$$

ce vecteur est aussi appelé **combinaison convexe** des x_i , avec coefficients λ_i . La proposition suivante constitue un exercice facile :

PROPOSITION VII-173. Soient E un espace vectoriel et A une partie quelconque de E ; alors

$$\text{co}(A) = \left\{ \sum_{i=1}^N \lambda_i a_i; a_i \in A; \lambda_i \geq 0; \sum_{i=1}^N \lambda_i = 1; N \in \mathbb{N} \right\} :$$

en d'autres termes, l'enveloppe convexe de A coïncide avec l'ensemble de toutes les combinaisons convexes d'un nombre fini d'éléments de A .

Dans un contexte analytique, il est naturel de chercher à généraliser cette notion à un continuum de vecteurs x_i . Pour ce faire, on peut noter que, si E est un espace de dimension finie,

$$\sum \lambda_i x_i = \int x d\mu(x),$$

où $\mu = \sum \lambda_i \delta_{x_i}$, et on donne un sens à l'intégrale vectorielle en la considérant composante par composante. Il suffit alors de remplacer la mesure atomique μ par une mesure continue.

La définition suivante généralise cette construction en dimension infinie :

DÉFINITION VII-174 (barycentre). Soient E un espace vectoriel normé, F un sous-espace vectoriel de E^* , et μ une mesure de probabilité sur E (muni de la tribu borélienne). On dit que $\beta \in E$ est barycentre de μ relativement à la famille F si

$$(74) \quad \forall \Lambda \in F, \quad \langle \Lambda, \beta \rangle = \int \langle \Lambda, x \rangle d\mu(x).$$

PROPOSITION VII-175. Soient E un espace vectoriel normé, et F un sous-espace vectoriel séparable de E^* , qui sépare les points de E au sens de la Définition VII-77. Soit μ une mesure de probabilité sur E , supportée par un convexe compact C . Alors il existe un unique barycentre β de μ relativement à la famille F , et il appartient à C . On note

$$\beta = \beta_F(\mu) = \int x d\mu(x).$$

L'application β est alors continue de $P(C)$, muni de la topologie faible, dans E , muni de la topologie faible induite par F .

REMARQUE VII-176. L'existence de β est garantie par la Proposition VII-79, par exemple dès que E est séparable, ou le dual d'un espace séparable.

REMARQUE VII-177. Si E^* est séparable et sépare les points de E , alors l'argument utilisé pour prouver l'unicité dans la Proposition VII-175 montre que $\beta_F(\mu) = \beta_{E^*}(\mu)$; en particulier $\beta_F(\mu)$ devient indépendant du choix de F . On peut alors noter $\beta(\mu) = \beta_F(\mu)$ sans ambiguïté.

REMARQUE VII-178. Si l'on admet l'axiome du choix, les hypothèses de séparabilité dans la Proposition VII-175 deviennent superflues, et le même raisonnement que dans la Remarque VII-177 montre que le barycentre est indépendant du choix de l'espace vectoriel F séparant les points de E . C'est pourquoi la plupart des ouvrages abordant le sujet des barycentres ne font aucune mention d'un choix de sous-espace vectoriel F dans E^* .

PREUVE DE LA PROPOSITION VII-175. 1. On montre d'abord l'unicité : si β et β' sont deux barycentres relativement à F , on a $\langle \Lambda, \beta \rangle = \langle \Lambda, \beta' \rangle$ pour tout $\Lambda \in F$. Comme F sépare E , on en déduit $\beta = \beta'$.

2. Passons à l'existence. Soit $(\Lambda_k)_{k \in \mathbb{N}}$ engendrant un espace vectoriel dense dans F . Pour tout $\ell \in \mathbb{N}$ on pose

$$a_\ell = \left(\int \langle \Lambda_1, x \rangle d\mu(x), \dots, \int \langle \Lambda_\ell, x \rangle d\mu(x) \right) \in \mathbb{R}^\ell,$$

et on définit $\Phi_\ell : E \rightarrow \mathbb{R}^\ell$ par $\Phi_\ell = (\Lambda_1, \dots, \Lambda_\ell)$. Bien sûr $\Phi_\ell(C)$ est un convexe compact de \mathbb{R}^ℓ .

Si $a_\ell \notin \Phi_\ell(C)$ on peut séparer strictement a_ℓ et $\Phi_\ell(C)$ (Théorème VII-84(ii)) :

$$\exists \lambda \in \mathbb{R}^\ell; \quad \sum_{k=1}^{\ell} \lambda_k a_k > \sup_{z \in C} \sum_{k=1}^{\ell} \lambda_k \langle \Lambda_k, z \rangle.$$

En utilisant la définition de a_k et la linéarité de l'intégrale, on obtient

$$\int \sum_{k=1}^{\ell} \lambda_k \langle \Lambda_k, x \rangle d\mu(x) > \sup_{z \in C} \sum_{k=1}^{\ell} \lambda_k \langle \Lambda_k, z \rangle,$$

ce qui est bien sûr impossible. Il s'ensuit que $C_\ell = \Phi_\ell^{-1}(a_\ell)$ est non vide ; est c'est un convexe fermé de C , en particulier un compact. L'intersection des C_ℓ est donc non vide, et tout z appartenant à cette intersection vérifie

$$\forall k \in \mathbb{N}, \quad \langle \Lambda_k, z \rangle = \int \langle \Lambda_k, x \rangle d\mu(x).$$

Soit maintenant $\Lambda \in F$, on peut trouver une sous-suite $(\Lambda_{k(n)})_{n \in \mathbb{N}}$ convergeant vers Λ dans E^* . Alors

$$\langle \Lambda, z \rangle = \lim_{n \rightarrow \infty} \langle \Lambda_{k(n)}, z \rangle = \lim_{n \rightarrow \infty} \int \langle \Lambda_{k(n)}, x \rangle d\mu(x) = \int \langle \Lambda, x \rangle d\mu(x),$$

où l'inégalité de droite découle par exemple du théorème de convergence dominée (noter que $\Lambda_{k(n)}$ et x apparaissant sous l'intégrale sont bornés en norme). On conclut que z est un barycentre de μ .

3. Si $(\mu_k)_{k \in \mathbb{N}}$ est une suite de mesures de probabilité sur C convergeant faiblement vers μ , alors pour tout $\Lambda \in F$ on a

$$\int \langle \Lambda, x \rangle d\mu_k(x) \xrightarrow{k \rightarrow \infty} \int \langle \Lambda, x \rangle d\mu(x),$$

d'où $\langle \Lambda, \beta_F(\mu_k) \rangle \rightarrow \langle \Lambda, \beta_F(\mu) \rangle$, et ce pour tout $\Lambda \in F$. Ceci montre bien la continuité de β . \square

Les énoncés démontrés dans la sous-section précédente peuvent se traduire en termes de barycentres :

PROPOSITION VII-179 (Reformulation du théorème de Krein–Milman). *Soit C un convexe compact d'un espace vectoriel normé E . Alors*

(i) *Si C est de dimension finie, alors C est l'ensemble des combinaisons convexes finies d'éléments de $\mathcal{E}(C)$.*

(ii) Dans tous les cas,

$$C = \left\{ \beta_F(\mu); \quad \mu \in P(\overline{\mathcal{E}(C)}), \right.$$

où F est un sous-espace vectoriel séparable de E^* , séparant les points de E , choisi arbitrairement.

DÉMONSTRATION. L'énoncé (i) est une conséquence immédiate du Théorème VII-172 et de la Proposition VII-173.

Pour démontrer l'énoncé (ii), on note que d'après le Théorème VII-171 tout $x \in C$ est limite d'une suite d'éléments de $\text{co}(\mathcal{E}(C))$:

$$x = \lim_{k \rightarrow \infty} \sum_{j=1}^{N_k} \lambda_{j,k} x_{j,k}, \quad \lambda_{j,k} \geq 0, \quad \sum_j \lambda_{j,k} = 1, \quad x_{j,k} \in \mathcal{E}(C).$$

L'ensemble $\overline{\mathcal{E}(C)}$ est compact, donc $P(\overline{\mathcal{E}(C)})$ est compact pour la topologie faible ; quitte à extraire une sous-suite, on peut donc supposer que $\mu_k = \sum_j \lambda_{j,k} \delta_{x_{j,k}}$ converge faiblement vers une mesure de probabilité μ sur $\overline{\mathcal{E}(C)}$. En utilisant la continuité faible de β , on obtient

$$x = \lim_{k \rightarrow \infty} \beta_F(\mu_k) = \beta_F(\mu),$$

ce qui conclut la preuve. \square

Dans la suite de cette sous-section, nous allons démontrer les remarquables théorèmes de Carathéodory et de Choquet, qui renforcent les conclusions de la Proposition VII-179 ; le premier en fournissant une borne explicite et optimale sur le nombre de points impliqués dans la combinaison convexe en dimension finie ; le second en remplaçant l'intégrale sur $\overline{\mathcal{E}(C)}$ par une intégrale sur $\mathcal{E}(C)$ (ce qui est parfois un gain considérable).

THÉORÈME VII-180 (Théorème de combinaison convexe de Carathéodory). Soit $A \subset \mathbb{R}^n$, alors

$$\text{co}(A) = \left\{ \sum_{i=1}^{n+1} \lambda_i a_i; \quad a_i \in A \right\}.$$

COROLLAIRE VII-181. Si C est un convexe compact de \mathbb{R}^n , alors tout $x \in C$ est combinaison convexe de $n+1$ points extrémaux de C .

COROLLAIRE VII-182. Si K est un compact de \mathbb{R}^n , alors

$$\begin{cases} \overline{\text{co}}(K) = \text{co}(K); \\ \mathcal{E}(\text{co}(K)) \subset K. \end{cases}$$

PREUVE DU THÉORÈME VII-180. Soit $x \in \text{co}(A)$; par la Proposition VII-173 il existe $k \in \mathbb{N}$, des coefficients positifs λ_i ($1 \leq i \leq k$) de somme égale à 1, et des vecteurs $a_i \in A$, tels que

$$x = \sum_{i=1}^k \lambda_i a_i.$$

On va montrer que si $k \geq n+2$, alors x peut se réécrire $\sum_{i=1}^{k-1} \lambda'_i a'_i$, où les λ'_i sont des coefficients positifs de somme égale à 1, et les a'_i appartiennent à A . Le résultat en découlera immédiatement par récurrence.

Soit donc $x = \sum_{i=1}^k \lambda_i a_i$ avec $k \geq n+2$. Si l'un des λ_i est nul, le résultat est évident ; on peut donc supposer $\lambda_i > 0$ pour tout i . Les k vecteurs $(1, a_i)$ sont liés dans $\mathbb{R}^{n+1} = \mathbb{R} \times \mathbb{R}^n$; on peut donc trouver des coefficients μ_i non tous nuls tels que

$$\sum \mu_i = 0, \quad \sum \mu_i a_i = 0.$$

Pour $t > 0$ assez petit on a alors

$$\lambda_i - t\mu_i > 0; \quad \sum_{i=1}^k (\lambda_i - t\mu_i) = 0; \quad \sum_{i=1}^k (\lambda_i - t\mu_i) a_i = x.$$

En revanche quand $t \rightarrow \infty$, l'un au moins des $\lambda_i - t\mu_i$ devient négatif.

Quand on fait varier t de 0 à $+\infty$, il existe donc un premier temps t_0 tel que tous les $\lambda_i - t_0\mu_i$ sont positifs, et l'un au moins de ces nombres est nul ; disons $\lambda_j - t_0\mu_j = 0$. Alors x est combinaison convexe des a_i pour $i \neq j$, avec coefficients $\lambda_i - t_0\mu_i$. \square

PREUVE DU COROLLAIRE VII-181. L'énoncé découle instantanément des Théorèmes VII-172 et VII-180. \square

PREUVE DU COROLLAIRE VII-182. Soit K un convexe compact de \mathbb{R}^n . Par le Théorème VII-180 (de Carathéodory),

$$\text{co}(K) = \left\{ \sum_{i=1}^{n+1} \lambda_i x_i; \quad x_i \in K \right\},$$

qui est compact comme image continue d'un compact ; d'où $\overline{\text{co}}(K) = \text{co}(K)$.

Par suite, tout $x \in \overline{\text{co}}(K)$ s'écrit comme combinaison convexe d'éléments de K . Si cette combinaison n'est pas triviale (i.e. réduite à un élément de K), x n'est bien sûr pas extrémal. Donc $\mathcal{E}(\text{co}(K)) \subset K$. \square

On termine avec le point culminant de cette sous-section : le théorème de représentation barycentrique de Choquet.

THÉORÈME VII-183 (Théorème de Choquet). *Soit C un convexe compact d'un espace vectoriel normé E , et soit F un sous-espace vectoriel séparable de E^* , séparant C . Alors pour tout $x \in C$ il existe $\mu^x \in P(\mathcal{E}(C))$ tel que $x = \beta_F(\mu^x)$; soit*

$$x = \int_{\mathcal{E}(C)} \mu^x(z) dz.$$

REMARQUE VII-184. La Remarque VII-177 s'applique ici aussi : si E^* est séparable et sépare C , ou si l'on admet l'axiome du choix, alors $\beta_F(\mu)$ ne dépend plus du choix de F et l'énoncé devient : *Soit C un convexe compact d'un espace vectoriel normé E , alors pour tout $x \in C$ il existe $\mu^x \in P(\mathcal{E}(C))$ tel que $x = \beta(\mu^x)$.*

Pour démontrer le Théorème VII-183, on utilisera le lemme suivant :

LEMME VII-185. *Sous les hypothèses du Théorème VII-183, il existe une fonction continue $\Phi : C \rightarrow \mathbb{R}_+$, strictement convexe.*

PREUVE DU LEMME VII-185. Soit $(\Lambda_k)_{k \in \mathbb{N}}$ une suite de formes linéaires sur E de norme 1, séparant C . On pose

$$\Phi(x) = \sum_{k \in \mathbb{N}} 2^{-k} \langle \Lambda_k, x \rangle^2.$$

Cette fonction est bien sûr convexe. Si maintenant $\Phi((1-\lambda)x + \lambda y) = (1-\lambda)\Phi(x) + \lambda\Phi(y)$ pour un triplet $(x, y, \lambda) \in C \times C \times (0, 1)$, alors on a, pour tout $k \in \mathbb{N}$, $((1-\lambda)\langle \Lambda_k, x \rangle + \lambda\langle \Lambda_k, y \rangle)^2 = \langle \Lambda_k((1-\lambda)x + \lambda y) \rangle^2 = (1-\lambda)\langle \Lambda_k, x \rangle^2 + \lambda\langle \Lambda_k, y \rangle^2$, ce qui impose $\langle \Lambda_k, x \rangle = \langle \Lambda_k, y \rangle$, d'où $x = y$. La convexité de Φ est donc stricte. \square

PREUVE DU THÉORÈME VII-183. Tous les barycentres seront implicitement pris par rapport à la famille F , qui ne sera pas explicitement rappelée dans la suite. Pour tout $x \in C$ on note

$$P^x(C) = \beta^{-1}(x) = \{\mu \in P(C); \int z \mu(dz) = x\};$$

et pour tout $\mu \in P(C)$ on note

$$F(\mu) = \int_C \Phi(x) \mu(dx).$$

Puisque C est un compact métrique, $P(C)$ est compact pour la topologie faible (Théorème VII-136); et il est clair que $P^x(C)$ est compact pour cette même topologie, pour tout $x \in C$. De même F est continue pour la topologie faible.

Fixons $x \in C$; le problème de maximisation

$$\max\{F(\mu); \mu \in P^x(C)\}$$

admet une solution $\bar{\mu}$. On sait que $x = \beta(\bar{\mu})$, nous allons maintenant vérifier que $\bar{\mu}$ est concentrée sur $\mathcal{E}(C)$, ce qui achèvera la preuve.

Comme dans la démonstration de la Proposition VII-168, on sait que $C \setminus \mathcal{E}(C) = \bigcup m(K_\ell)$, où m est l'application milieu $C \times C \rightarrow C$, et $(K_\ell)_{\ell \in \mathbb{N}}$ une famille croissante de compacts inclus dans $(C \times C) \setminus \Delta$, Δ étant la diagonale de C . Par Théorème de sélection mesurable **À INCLURE DANS LES NOTES QUELQUES PART (CHAPITRE 1 ??)** on peut définir une application mesurable $g_\ell : m(K_\ell) \rightarrow K_\ell$, telle que $m \circ g_\ell = \text{Id}_{K_\ell}$. On définit g sur $C \setminus \mathcal{E}(C)$ en posant $g(z) = g_\ell(z)$ si $z \in K_\ell \setminus K_{\ell-1}$ (où $K_0 = \emptyset$ par convention). Enfin, pour tout $z \in \mathcal{E}(C)$ on pose $g(z) = z$. On a ainsi construit une application mesurable $g : z \mapsto (z_0, z_1)$ telle que $x = (z_0 + z_1)/2$, avec $z_0 \neq z_1$ si (et seulement si) $z \notin \mathcal{E}(C)$.

Pour tout $z \in C$ on pose alors $\mu^{(z)} = (1/2)(\delta_{z_0} + \delta_{z_1})$, où $(z_0, z_1) = g(z)$. La stricte convexité de Φ implique

$$(75) \quad F(\delta_z) = \Phi(z) \leq \Phi((z_0 + z_1)/2) = F(\mu^{(z)}),$$

avec inégalité stricte dès que $z \notin \mathcal{E}(C)$.

On définit alors $\mu(ds) = \int_C \mu^{(z)}(ds) \bar{\mu}(dz)$. On note que $\mu \in P^x(C)$ puisque (par Fubini)

$$\int_C s \mu(ds) = \int_C \left(\int_C s \mu^{(z)}(ds) \right) \bar{\mu}(dz) = \int_C z \bar{\mu}(dz) = x.$$

En appliquant encore Fubini et l'inégalité (75), on obtient

$$\begin{aligned} F(\mu) &= \int_C \Phi(s) \int_C \mu^{(z)}(ds) \bar{\mu}(dz) = \int_C \left(\int_C \Phi(s) \mu^{(z)}(ds) \right) \bar{\mu}(dz) = \int_C F(\mu^{(z)}) \bar{\mu}(dz) \\ &\geq \int_C \Phi(z) \bar{\mu}(dz) = F(\bar{\mu}). \end{aligned}$$

Puisque μ appartient à $P^x(C)$ et que $\bar{\mu}$ maximise F sur cet ensemble, l'inégalité ci-dessus est nécessairement une égalité. Ceci, et la caractérisation des cas d'égalité dans (75), impose que $z \in \mathcal{E}(C)$, $\bar{\mu}(dz)$ -presque partout. Ceci conclut la preuve. \square

VII-7.4. Dualité des fonctions convexes. Un des concepts les plus importants de l'analyse convexe est celui de dualité.

DÉFINITION VII-186. Soit E un espace vectoriel et $\varphi : E \rightarrow \mathbb{R} \cup \{+\infty\}$, non identiquement $+\infty$. On appelle transformée de Legendre–Fenchel de φ la fonction convexe $\varphi^* : E^* \rightarrow \mathbb{R} \cup \{+\infty\}$ définie par

$$\varphi^*(\xi) = \sup_{x \in E} (\langle \xi, x \rangle - \varphi(x)).$$

Comme conséquence directe de la définition, on a l'importante inégalité de Young : pour toute fonction convexe $E \rightarrow \mathbb{R} \cup \{+\infty\}$, non identiquement $+\infty$,

$$(76) \quad \forall x \in E, \quad \forall \xi \in E^*, \quad \langle \xi, x \rangle \leq \varphi(x) + \varphi^*(\xi).$$

Les cas d'égalité ont une importance considérable en analyse convexe :

DÉFINITION VII-187 (sous-différentiabilité). Soit E un espace vectoriel, $\varphi : E \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe, et $x \in E$. Le sous-différentiel de φ en x , noté $\partial\varphi(x)$, est l'ensemble des formes linéaires continues $\xi \in E^*$ vérifiant les deux conditions équivalentes suivantes :

$$\begin{aligned} \forall z \in E, \quad \varphi(z) &\geq \varphi(x) + \langle \xi, z - x \rangle; \\ \varphi(x) + \varphi^*(\xi) &= \langle x, \xi \rangle. \end{aligned}$$

Le théorème suivant est à la base des principaux résultats de dualité des fonctions convexes :

THÉORÈME VII-188. Soient E un espace vectoriel normé séparable, et Θ, Ξ deux fonctions convexes $E \rightarrow \mathbb{R} \cup \{+\infty\}$. Si

- (i) $\exists x_0 \in E; \quad \Theta(x_0) + \Xi(x_0) < +\infty;$
- (ii) les ensembles convexes disjoints

$$C = \{(x, \lambda) \in E \times \mathbb{R}; \quad \lambda > \Theta(x)\}, \quad C' = \{(y, \mu) \in E \times \mathbb{R}; \quad \mu \leq m - \Xi(y)\}$$

sont séparables (au sens large) par un hyperplan affine ;

alors

$$(77) \quad \inf_{x \in E} (\Theta + \Xi) = \max_{\Lambda \in E^*} [-\Theta^*(-\Lambda) - \Xi^*(\Lambda)].$$

COROLLAIRE VII-189 (dualité de Fenchel–Rockafellar). Si Θ, Ξ sont convexes $E \rightarrow \mathbb{R} \cup \{+\infty\}$ et s'il existe $x_0 \in E$ tel que $\Theta(x_0) + \Xi(x_0) < +\infty$ et Θ soit continue en x_0 , alors l'identité (77) a lieu.

COROLLAIRE VII-190 (sous-différentiabilité). Soit E un espace vectoriel normé séparable, et $\varphi : E \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe semicontinue inférieurement. Si φ est continue en x alors $\partial\varphi(x) \neq \emptyset$.

COROLLAIRE VII-191 (dualité de Legendre–Fenchel–Moreau). Soit E un espace vectoriel normé séparable, et $\varphi : E \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe semicontinue inférieurement. Alors

$$\varphi^{**} = \varphi.$$

REMARQUE VII-192. L'hypothèse de semicontinuité inférieure est cruciale dans le Corollaire VII-191. Dans \mathbb{R}^n , une fonction convexe est automatiquement continue dans l'intérieur de son domaine (Propriété VII-157), mais rien ne garantit la semicontinuité au bord du domaine (trouver des exemples).

PREUVE DU THÉORÈME VII-188. Soit $m = \inf(\Theta + \Xi)$; par hypothèse $m < +\infty$. On cherche à prouver que sous les hypothèses de l'énoncé,

$$\sup_{\xi \in E^*} \inf_{x, y \in E} (\Theta(x) + \Xi(y) + \langle \xi, x - y \rangle) = \inf_{x \in E} [\Theta(x) + \Xi(x)] = m.$$

En choisissant $x = y$, on voit que le membre de gauche est toujours plus petit que celui de droite il suffit donc de montrer que

$$\exists \Lambda \in E^*, \quad \forall x, y \in E, \quad \Theta(x) + \Xi(y) + \langle \Lambda, x - y \rangle \geq m.$$

Si les convexes C et C' peuvent se séparer, il existe $\xi \in E^*$ et $\alpha \in \mathbb{R}$, $(\xi, \alpha) \neq (0, 0)$, tels que

$$\langle \xi, x \rangle + \alpha \lambda \geq \langle \xi, y \rangle + \alpha \mu$$

dès que $\lambda > \Theta(x)$ et $\mu \leq m - \Xi(y)$. Ceci implique facilement $\alpha > 0$. On pose alors $\Lambda = \xi/\alpha$, de sorte que

$$\langle \Lambda, x \rangle + \lambda \geq \langle \Lambda, y \rangle + \mu.$$

En passant à la limite on trouve $\langle \Lambda, x \rangle + \Theta(x) \geq \langle \Lambda, y \rangle + m - \Xi(y)$, et le théorème est démontré. \square

Les preuves des Corollaires sont à rédiger, ainsi que la preuve du Théorème VII-164.

Terminons cette section avec une généralisation en dimension infinie du Théorème III-69 (inégalité de Jensen) : **Rédiger, bien penser aux cas d'égalité.**

THÉORÈME VII-193.

VII-8. Retour sur les espaces de Hilbert

ca vaut le coup de recapituler les prop. des hilberts ;

et de montrer ceux des thm precedents qui sont triviaux via base

on commence par rappeler un concept fondamental dont tout va decouler : l'orthogonalite. Def : deux vecteurs sont orthogonaux si ... Un sous-ensemble A de H etant donne, on definit son orthogonal par...

pour les projections, mentionner que c'est toujours Lipschitz et le montrer. En outre la proj sur un sous-espace est lineaire.

egalement on peut identifier H/F et F^\perp

VII-8.1. Baess hilbertiennes. Attention, ce ne sont pas des bases algébriques (quand on parle de base alg., il s'agit de combinaisons linéaires finies).

Def : une base hilb est un systeme orthonormé ($\langle \alpha, \beta \rangle = \delta_{\alpha, \beta}$) engendrant un sous-espace vectoriel **dense**.

Le thm fondamental est le suivant, mq proprietes familiares des espaces Euclidiens se generalisent bien :

THÉORÈME VII-194. Soit (e_k) une base hilb, $x \in H$, $x_j := \langle e_j, x \rangle$, alors la série $\sum x_j e_j$ est absolument convergente dans H , et sa somme vaut x :

$$x = \sum_{j=1}^{\infty} x_j e_j.$$

En outre,

$$\|x\|^2 = \sum x_j^2; \quad \langle x, y \rangle = \sum x_j y_j.$$

REMARQUE VII-195. Variante quand les coordonnées sont complexes. Plus joli. mettre des $\sum \overline{x_j} y_j$.

DÉMONSTRATION. Grâce à la définition, pour tout $\ell \leq j$,

$$\langle x - \sum_{j=0}^k x_j e_j, e_\ell \rangle = 0.$$

On en déduit

$$x - \sum_{j=0}^k x_j e_j \in E_k^\perp,$$

avec $E_k = \text{vect}(e_j)_{j \leq k}$. Donc

$$\sum_{j=1}^k x_j e_j \perp x - \sum_{j=0}^k x_j e_j.$$

Par la formule de Pythagore,

$$\left\| \sum_{j=1}^k x_j e_j \right\|^2 + \left\| x - \sum_{j=1}^k x_j e_j \right\|^2 = \|x\|^2.$$

Le terme de gauche est $\sum |x_j|^2$. En passant à la limite pour $k \rightarrow \infty$, on obtient l'**inégalité de Bessel** (vraie pour toute famille orthonormée, même si elle n'engendre pas un sous-espace dense) :

$$\sum_{j=1}^k |x_j|^2 \leq \|x\|^2.$$

... la série converge dans H . Notons $y := \sum x_j e_j$. On a

$$\forall \ell \quad x - \sum_{j=1}^k x_j e_j \perp e_\ell$$

d'où, en passant à la limite, $x - y \perp e_\ell$ pour tout ℓ . Le vecteur $x - y$, étant orthogonal à tous les éléments de la base, est aussi orthogonal au sous-espace vectoriel engendré par cette base, et par conséquent à l'adhérence de ce sous-espace vectoriel, i.e. H . Il est donc orthogonal à lui-même, et donc nul. On conclut que $x = y$. Pour conclure la preuve du théorème, il suffit de passer à la limite dans et on obtient l'**inégalité de Parseval** :

....

□

Le théorème suivant est fondamental, mais presque évident après le théorème précédent :

THÉORÈME VII-196. *Soit H un Hilbert séparable, alors*

(i) *A partir de n'importe quelle famille dénombrable libre $(x_k)_{k \in \mathbb{N}}$ engendrant un sous-espace vectoriel dense, on peut construire une base hilbertienne $(e_k)_{k \in \mathbb{N}}$ telle que $\text{vect}(x_k)_{k \leq \ell} = \text{vect}(e_k)_{k \leq \ell}$, par orthonormalisation de Gram-Schmidt. En particulier, il existe une base hilbertienne (et en fait, il en existe bien sûr une infinité).*

(ii) *Un espace de Hilbert admet une base hilbertienne si et seulement si il est séparable.*

(iii) *Toute base hilbertienne fixée fournit un isomorphisme entre H et $\ell^2(\mathbb{N})$, i.e. une application linéaire bijective préservant la norme et le produit scalaire.*

REMARQUE VII-197. Dans l'énoncé (i) : si on a une famille génératrice non libre, on peut toujours se ramener par récurrence à une famille libre. Pour (ii) : une implication est (i), l'autre est obtenue en prenant des combinaisons linéaires à coefficients rationnels. Les espaces de Hilbert non-séparables sont d'habitudes fuis avec horreur. L'énoncé (iii) est souvent paraphrasé par le slogan "il n'existe qu'un espace de Hilbert séparable".

Terminons cette section par des exemples de bases hilbertiennes :

- $H = \ell^2$, $e_j = (\delta_{jk})_{k \in \mathbb{N}}$;
- $H = L^2(\mathbb{T})$, $e_k = e^{ikx}$ (variables complexes) ou $e_k = \cos(kx), \sin(kx), \dots$
- $H = L^2(\mathbb{T}^n)$, $e_k = e^{ik \cdot x}$
- $H = L^2(\mathbb{R})$, système de Haar : $1_{[0,1/2[} - 1_{]1/2,1]}$ et toutes ses translatées (par \mathbb{Z}) et dilatées **faire un dessin !**
- $H = L^2(d\mu)$, polynômes orthogonaux, obtenus par Gram-Schmidt à partir de la base canonique des polynômes. Parfois c'est total : p. ex. Legendre pour dx sur $[-1, 1]$; Hermite pour $e^{-x^2/2} dx$ sur \mathbb{R} , Laguerre pour $e^{-x/2} dx$ sur \mathbb{R}_+ ; parfois ça ne l'est pas : p.ex. $e^{-\sqrt{|x|}}$. Cette alternative est liée à des propriétés de décroissance et de régularité de la mesure de référence.

Nous reparlerons plus tard de certains de ces résultats : par exemple, la totalité (définir ce mot) des séries de Fourier (définir). Un résultat d'analyse derrière : toute fonction continue est limite uniforme de séries de Fourier.

Expliquons juste ici que le système de Haar est total : la seule chose à comprendre c'est pourquoi la condition $\int f = 0$ ne gêne pas dans $L^2(\mathbb{R})$. En effet, soit f donnée dans L^2 , on peut trouver g d'intégrale nulle, tel que $\|f - g\|_{L^2} \leq \varepsilon$. Pour cela on choisit

$$g = f - \left(\int f \right) \left(\frac{1_{[0,B]}}{B} \right)$$

et on fait tendre B vers l'infini (exercice).

REMARQUE VII-198. (a mettre plus haut ?) Les polynômes orthogonaux sont très étudiés dans de nombreux contextes ; en particulier, ils satisfont des relations de récurrence que l'on peut obtenir par certains procédés systématiques, et s'expriment souvent en termes de valeurs propres d'opérateurs différentiels.

Exemples :

Legendre

$$P_n(x) = \frac{\sqrt{n + \frac{1}{2}}}{2^n n!} \frac{d^n}{dx^n} [(1 - x^2)^n]$$

Hermite

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} (e^{-x^2})$$

Laguerre

$$L_n(x) = \frac{e^x}{n!} \frac{d^n}{dx^n} (e^{-x} x^n)$$

Reference : Nikiforov-Ouvarov (rajouter cette référence dans la biblio)

Remarque : vérifier qu'on a défini la notation \perp .

VII-8.2. Propriétés géométriques dans les espaces de Hilbert. Beaucoup des prop. vues avant sont vraies, et peuvent se montrer très facilement.

- la norme est Fréchet-diff. $DN(x) = x/\|x\|$.
- l'espace est uniformément convexe, et en fait on a l'identité du parallélogramme
- le théorème de projection sur un convexe fermé est vrai : on peut donc définir

π_C

- le théorème de Riesz : $H^* = H$ et c'est canonique
- caractérisation différentielle de la projection : $x - \pi_F(x) \in F^\perp$ pour tout s.ev.

F

- l'op de projection est Lipschitzien (d'ailleurs cette prop. n'est vraie que dans un Hilbert !)

- $\text{Id} = \pi_F + \pi_{F^\perp}$. En particulier, F admet un supplémentaire (topologique!). Un cas particulier est la décomposition obtenue en choisissant $F = E_j$, i.e. on ne garde que les j premières composantes, on a vu ça dans la preuve plus haut...

- HB est maintenant trivial : le thm d'extension : on étend à \overline{F} par continuité uniforme, puis à E en utilisant la décomposition $F + F^\perp$.

L'existence d'une flc normalisante : on choisit le produit scalaire par $x/|x|$.

Les flc nulles sur F s'identifient à F^\perp , donc $\overline{\text{vect}(F)} = H \Leftrightarrow F^\perp = 0$.

- BS se voit facilement sur une base, et BA aussi.

- thm de classification : un seul Hilbert separable

CHAPITRE VIII

Analyse sur les groupes abéliens - À RÉDIGER

CHAPITRE IX

Fonction maximale de Hardy–Littlewood - À RÉDIGER

CHAPITRE X

Désintégration

Soit f une application mesurable positive sur un espace mesuré (X, \mathcal{A}, ν) , alors la formule

$$\mu[A] = \int_A f d\nu$$

définit une mesure $\mu = f\nu$ sur X . Cette construction s'étend aux fonctions mesurables de signe arbitraire si l'on définit μ comme une mesure signée : $\mu = (f_+\nu) - (f_-\nu)$ (différence formelle dans le cas général, vraie différence si f_+ ou f_- est ν -intégrable).

Réciproquement, étant donnée une mesure (signée ou pas) μ , peut-on trouver une fonction f telle que $\mu = f\nu$? Cette question est le point de départ de la théorie de la **désintégration** (pour passer de f à μ on intègre ; il est naturel de parler de désintégration pour le problème inverse).

Dans ce chapitre de nombreux résultats nécessiteront de mettre de côté des ensembles négligeables ; ces ensembles seront toujours implicitement choisis mesurables.

X-1. Théorème de Lebesgue–Radon–Nikodym

Commençons par introduire deux notions naturelles, et en quelque sorte opposées.

DÉFINITION X-1. Soit (X, \mathcal{A}, ν) un espace mesuré, et μ une mesure signée sur (X, \mathcal{A}) . On dit que μ est mesurable par ν (ou ν -mesurable) s'il existe une fonction mesurable $f : X \rightarrow \mathbb{R}$ telle que $\mu = f\nu$. On dit que μ est étrangère à ν si μ et ν sont concentrées sur des ensembles mesurables disjoints ; dans ce cas on note $\mu \perp \nu$.

EXEMPLE X-2. Soient $(X, \mathcal{A}, \lambda)$ un espace mesuré, et f, g deux fonctions mesurables positives sur X , telles que $g > 0$ presque partout. Si l'on pose $\mu = f\lambda$, $\nu = g\lambda$, alors $\mu = h\nu$, où $h = f/g$.

Le théorème suivant est l'un des résultats les plus importants de la théorie de la mesure :

THÉORÈME X-3 (décomposition de Lebesgue). Soient (X, \mathcal{A}, ν) un espace mesuré σ -fini, et μ une mesure signée σ -finie sur (X, \mathcal{A}) . Alors il existe une fonction mesurable $f : X \rightarrow \mathbb{R}$ et une mesure signée μ_s , telle que

$$\mu = f\nu + \mu_s, \quad \mu_s \perp \nu.$$

La mesure μ_s est unique, et la fonction f est unique à modification près sur un ensemble de ν -mesure nulle ; en particulier, la mesure $f\nu$ est unique.

NOTATION X-4. La mesure μ_s apparaissant dans le Théorème X-3 est dite partie singulière de μ par rapport à ν . La fonction f apparaissant dans ce même énoncé est appelée densité de Radon–Nikodym de μ par rapport à ν , et notée

$$f = \frac{d\mu}{d\nu};$$

cette fonction n'est bien définie qu'à un ensemble ν -négligeable près.

Intuitivement, $f(x)$ est le quotient de deux éléments infinitésimaux de mesure : en termes informels,

$$f(x) = \frac{\mu(dx)}{\nu(dx)}.$$

La notation est justifiée en un certain sens par l'observation selon laquelle si λ , μ et ν sont trois mesures telles que λ est μ -mesurable et μ est ν -mesurable, alors

$$\frac{d\lambda}{d\nu} = \frac{d\lambda}{d\mu} \frac{d\mu}{d\nu} \quad \nu\text{-presque partout.}$$

(C'est une variation sur l'Exemple X-2.)

Avant de prouver le Théorème X-3 nous allons discuter quelques remarques et conséquences.

COROLLAIRE X-5 (Théorème de Radon Nikodym). *Soient (X, \mathcal{A}, ν) un espace mesuré σ -fini, et μ une mesure signée σ -finie. Alors les deux énoncés suivants sont équivalents :*

(i) μ ne charge aucun ensemble ν -négligeable ; c'est-à-dire

$$(78) \quad \forall N \in \mathcal{A}, \quad \left[\nu[N] = 0 \implies \mu[N] = 0 \right].$$

(ii) μ est ν -mesurable ; c'est-à-dire

$$(79) \quad \exists f \in L(X, \nu); \quad \mu = f \nu.$$

EXEMPLE X-6. Soient $\nu = \mathcal{L}_{[0,1]} + \delta_1$ (mesure de Lebesgue sur \mathbb{R} , restreinte à l'intervalle $[0, 1]$, à laquelle on ajoute une masse de Dirac en 1) ; et $\mu = \mathcal{L}_{[0,2]} + \delta_0 + 2\delta_1$. On a alors

$$\frac{d\mu}{d\nu} = 1_{[0,1[} + 2\delta_{\{1\}}; \quad \mu_s = \mathcal{L}_{[1,2]} + \delta_0.$$

PREUVE DU COROLLAIRE X-5. Par le Théorème VI-50 (de décomposition de Hahn) il existe des parties mesurables disjointes S_+ et S_- telles que μ_+ et μ_- sont concentrées sur S_+ et S_- respectivement. Si N est un ensemble ν -négligeable, il en est de même de $S_+ \cap N$ et $S_- \cap N$, et l'hypothèse implique donc $\mu[S_+ \cap N] = \mu[S_- \cap N] = 0$; il suffit donc de démontrer le corollaire dans le cas où μ est une mesure non signée, ce que nous allons supposer.

Soit maintenant N un ensemble ν -négligeable sur lequel μ_s est concentrée. Pour tout $A \subset N$ on a $\nu[A] = 0$, d'où $\mu[A] = 0$ par hypothèse, et bien sûr $(f\nu)[A] = 0$; on en déduit $\mu_s[A] = 0$; μ_s est donc la mesure nulle, et la conclusion suit. \square

REMARQUE X-7. Une terminologie quelque peu malheureuse mais très bien implantée consiste à dire que μ vérifiant (78) est *absolument continue* par rapport à ν ,

ce que l'on note (de manière non moins malheureuse) $\mu \ll \nu$. La terminologie "absolue continuité" s'explique comme suit : si μ est finie, la propriété (78) est équivalente à la propriété en apparence plus forte

$$\forall \varepsilon > 0 \exists \delta > 0; \quad \nu[A] \leq \delta \implies |\mu[A]| \leq \varepsilon,$$

qui est analogue à la notion d'absolue continuité d'une fonction (voir la section X-4). Cet énoncé peut se démontrer de manière élémentaire ; on peut aussi le voir comme conséquence immédiate du Corollaire X-5 et d'un cas particulier du Théorème VII-131 (théorème de compacité faible de Dunford–Pettis, que l'on applique ici dans le sens le plus facile, et avec l'ensemble $\{f\}$ qui est trivialement compact).

REMARQUE X-8. Le Théorème de Radon–Nikodym se généralise pour des mesures à valeurs dans un espace vectoriel de dimension finie, mais pas pour des mesures à valeurs dans un espace de Banach quelconque. On dit précisément qu'un espace E vérifie la "PRN" (Propriété de Radon–Nikodym !) si le Théorème X-5 s'applique pour des mesures à valeurs dans E .

CONTRE-EXEMPLE X-9. Soient ν la mesure de comptage sur \mathbb{R} et μ la mesure de Lebesgue. Alors μ ne charge aucun ensemble ν -négligeable (seul l'ensemble vide vérifie cette condition !), pour autant μ n'est pas ν -mesurable. Ceci ne contredit pas le Théorème de Radon–Nikodym puisque ν n'est pas σ -finie.

PREUVE DU THÉORÈME X-3. 1. Si la conclusion du théorème est vraie, alors, puisque ν et μ_s sont étrangères on a

$$\mu_+ = (f\nu + \mu_s)_+ = f_+\nu + (\mu_s)_+; \quad \mu_- = (f\nu + \mu_s)_- = f_-\nu + (\mu_s)_-.$$

Il suffit donc de démontrer le résultat en supposant que μ et μ_s sont des mesures non signées, et que f est une fonction positive ; cet argument fonctionne aussi bien pour l'existence que pour l'unicité. Dans la suite on fera cette hypothèse.

2. Supposons que la mesure μ se décompose de deux manières différentes :

$$\mu = f\nu + \mu_s = f'\nu + \mu'_s;$$

alors

$$\mu|_E = (f1_E)\nu + \mu_s|_E = (f'1_E)\nu + \mu'_s|_E.$$

L'unicité de la décomposition de $\mu|_E$ entraînera donc $f1_E = f'1_E$, ν -presque partout, et $\mu_s|_E = \mu'_s|_E$. En introduisant une partition dénombrable $(E_k)_{k \in \mathbb{N}}$ de parties mesurables de X , telles que $\mu[E_k]$ et $\nu[E_k]$ soient finis, on en déduira que $f = f'$ ν -presque partout, et $\mu_s = \mu'_s$.

De même, si l'on a l'existence de la décomposition de $\mu|_E$ pour tout E de mesure finie, sous la forme $\mu_E = f_E + \mu_{s,E}$, alors on posera $f = \sum f_{E_k}$, $\mu_s = \sum \mu_{s,E_k}$, et ceci montrera l'existence de la décomposition. Il suffit donc de traiter le cas où X est de μ -mesure finie, et de ν -mesure finie, ce que l'on supposera par la suite.

Récapitulons : il reste à démontrer le Théorème X-3 pour des mesures finies μ et ν . Pour cela plusieurs voies sont possibles. L'importance du résultat justifie que l'on présente deux arguments différents.

Première preuve : Cette démonstration, due à Von Neumann, utilise des arguments d'analyse fonctionnelle. Soit Λ la forme linéaire définie sur $L^2(\mu + \nu)$ par

$$\langle \Lambda, \varphi \rangle = \int \varphi d\mu.$$

L'estimation

$$|\langle \Lambda, \varphi \rangle| \leq \sqrt{\int d\mu(x)} \sqrt{\int \varphi^2 d\mu(x)} \leq \sqrt{\int d\mu(x)} \sqrt{\int \varphi^2 d(\mu + \nu)(x)} = \mu[X]^{1/2} \|\varphi\|_{L^2(\mu + \nu)}$$

montre que Λ est bien définie (et à valeurs dans \mathbb{R}), et qu'en outre elle est continue. Par le Théorème de Riesz (c'est-à-dire le Théorème VII-87 dans le cas facile $p = 2$), il existe $h \in L^2(\mu + \nu)$ tel que

$$\forall \varphi \in L^2(\mu + \nu), \quad \langle \Lambda, \varphi \rangle = \int h \varphi d(\mu + \nu).$$

Le choix $\varphi = 1_A$, où A est un ensemble mesurable arbitraire, montre que

$$(80) \quad h(\mu + \nu) = \mu.$$

On déduit de (80) (par unicité de la décomposition de Hahn, ou en appliquant cette identité à l'ensemble $\{h < 0\}$) que $h \geq 0$, $(\mu + \nu)$ -presque partout.

Soit maintenant $\varepsilon > 0$, et $B_\varepsilon = \{h \geq 1 + \varepsilon\}$; en appliquant encore (80) on obtient

$$\mu[B_\varepsilon] \geq (1 + \varepsilon)(\mu + \nu)[B_\varepsilon] \geq \mu[B_\varepsilon] + \varepsilon(\mu + \nu)[B_\varepsilon];$$

d'où $(\mu + \nu)[B_\varepsilon] = 0$. Soit $B = \{h > 1\}$; comme $B = \cup_{k \in \mathbb{N}} B_{1/k}$ on a $(\mu + \nu)[B] = 0$.

Récapitulons : en-dehors d'un ensemble $(\mu + \nu)$ -négligeable, on a

$$0 \leq h \leq 1.$$

En outre, l'identité $\mu[B] = (\mu + \nu)[B]$ implique $\nu[B] = 0$.

Posons $\mu_m = 1_{X \setminus B} \mu$, $\mu_s = 1_B \mu$; il est évident que $\mu = \mu_m + \mu_s$, et μ_s est étrangère à ν . Il reste à démontrer que $\mu_m = 1_{\{h < 1\}} \mu$ est ν -mesurable. Pour cela on écrit

$$\mu_m = 1_{X \setminus B} h(\mu + \nu) = 1_{X \setminus B} h \mu + h \nu = h \mu_m + h \nu$$

(la première égalité découle de (80), la deuxième de ce que B est ν -négligeable, et la dernière de la définition de μ_m). On a donc $(1 - h) \mu_m = h \nu$, ou de manière équivalente (puisque μ_m est concentrée sur $h < 1$)

$$(1 - h) 1_{h < 1} \mu_m = h \nu.$$

En divisant les deux membres par la fonction strictement positive $(1 - h) 1_{h < 1}$, on trouve

$$\mu_m = \left(\frac{h}{1 - h} \right) 1_{h < 1} \nu;$$

ce qui donne le résultat avec $f = (h/(1 - h)) 1_{h < 1}$.

Deuxième preuve : Cette démonstration plus élémentaire (mais un peu plus délicate) utilise des arguments de monotonie; elle a été perfectionnée par divers auteurs, la version finale étant due à Loomis (voir [Folland] pour plus de détails).

Soit

$$\mathcal{F} = \left\{ f \text{ mesurable } X \rightarrow [0, +\infty]; \quad f \nu \leq \mu \right\}.$$

Bien sûr $\mathcal{F} \neq \emptyset$ ($0 \in \mathcal{F}$). En outre, si $f, g \in \mathcal{F}$, soit $A = \{f > g\}$ et $h = \max(f, g)$, alors

$$h \nu = f 1_A \nu + g 1_{X \setminus A} \nu = 1_A (f \nu) + 1_{X \setminus A} (g \nu) \leq 1_A \mu + 1_{X \setminus A} \mu = \mu;$$

donc $h \in \mathcal{F}$. Plus généralement, le maximum d'un nombre fini d'éléments de \mathcal{F} est lui-même élément de \mathcal{F} .

Pour tout $f \in \mathcal{F}$ on a $\int f d\nu \leq \mu[X]$; d'où

$$a := \sup \left\{ \int f d\nu; \quad f \in \mathcal{F} \right\} \leq \mu[X] < +\infty.$$

Soit maintenant $(f_n)_{n \in \mathbb{N}}$ une suite à valeurs dans \mathcal{F} telle que $\int f_n d\nu \rightarrow a$ quand $n \rightarrow \infty$. En remplaçant f_n par $\max(f_1, \dots, f_n) \in \mathcal{F}$ on peut supposer que la suite (f_n) est croissante; on pose alors $f = \sup f_n$. Par convergence monotone, pour tout A mesurable on a

$$\int_A f d\nu = \lim_{n \rightarrow \infty} \int_A f_n d\nu \leq \mu[A],$$

donc $f \in \mathcal{F}$. Une nouvelle application du théorème de convergence monotone montre que $\int f d\nu = a$.

Pour conclure la démonstration, il suffit de montrer que $\tilde{\mu} := \mu - f\nu$ est étrangère à ν ; alors on aura la décomposition recherchée sous la forme $\mu = f\nu + (\mu - f\nu)$.

Pour tout $k \in \mathbb{N}$, le Théorème VI-50 (de décomposition de Hahn) fournit un ensemble mesurable A_k tel que $(\tilde{\mu} - k^{-1}\nu)_+$ est concentrée sur A_k et $(\tilde{\mu} - k^{-1}\nu)_-$ sur $X \setminus A_k$. On pose alors $A = \cup_{k \in \mathbb{N}} A_k$ et $B = X \setminus A$. Pour tout $k \in \mathbb{N}$, la mesure $(\tilde{\mu} - k^{-1}\nu)_+$ est concentrée sur A , donc $(\tilde{\mu} - k^{-1}\nu)[B] \leq 0$, i.e. $\tilde{\mu}[B] \leq k^{-1}\nu[B]$. En faisant tendre k vers l'infini on conclut que $\tilde{\mu}[B] = 0$.

Si $\nu[A] > 0$, alors il existe $k \in \mathbb{N}$ tel que $\nu[A_k] > 0$. En outre, $\tilde{\mu}|_{A_k} \geq k^{-1}\nu|_{A_k}$; en utilisant la définition de $\tilde{\mu}$ on en déduit

$$\mu|_{A_k} \geq (f + k^{-1})\nu|_{A_k},$$

d'où

$$\mu \geq (f + k^{-1}1_{A_k})\nu.$$

En particulier $g := f + k^{-1}1_{A_k}$ appartient à \mathcal{F} ; mais ceci contredit

$$\int g d\nu = \int f d\nu + k^{-1}\nu[A_k] > \int f d\nu.$$

On conclut que $\nu[A] = 0$, donc $\tilde{\mu} \perp \nu$, ce qui achève la démonstration. \square

X-2. Désintégration

Le Théorème de Radon–Nikodym est le plus simple représentant d'une famille d'énoncés traitant du problème de la désintégration. Il existe plusieurs notions de désintégration; dans cette section on en considérera deux: d'abord la plus naturelle, la désintégration fibre à fibre, qui généralise le théorème de Radon–Nikodym; puis une notion plus faible qui nécessite moins d'hypothèses.

X-2.1. Désintégration fibre à fibre.

DÉFINITION X-10 (désintégration fibre à fibre). Soient (X, \mathcal{A}, μ) et (Y, \mathcal{B}, ν) deux espaces mesurés, et $p : X \rightarrow Y$ une application mesurable. On appelle ν -désintégration fibre à fibre de μ par l'application p une famille de mesures $\mu(\cdot | y)$ sur (X, \mathcal{A}) , indexées par $y \in Y$, telles que

- (i) pour tout $y \in Y$, $\mu(\cdot | y)$ est concentrée sur $p^{-1}(y)$;
- (ii) pour tout $A \in \mathcal{A}$, $\mu[A|y]$ est une fonction \mathcal{B} -mesurable de y , et

$$(81) \quad \mu[A] = \int_Y \mu[A|y] \nu(dy).$$

La mesure $\mu(dx|y)$ est appelée *mesures conditionnelles* (portée par la fibre $p^{-1}(y)$); on notera

$$\mu(dx) = \int_Y \mu(dx|y) \nu(dy).$$

REMARQUE X-11. Quand on parle de désintégration par p , l'application p ne doit pas être considérée comme une application entre deux ensembles, mais bien comme une application entre deux espaces mesurables. Prenons par exemple le cas où $Y = X$, \mathcal{B} est une sous-tribu de \mathcal{A} , et p est l'application $\text{Id} : X \rightarrow X$; la notion de désintégration dépendra alors crucialement du choix de \mathcal{B} (via l'hypothèse de \mathcal{B} -mesurabilité).

REMARQUE X-12. Dans la Définition X-10 les ensembles $p^{-1}(y)$ ne sont pas a priori mesurables (sauf bien sûr si les singletons sont \mathcal{B} -mesurables); la propriété (i) veut donc dire que pour tout ensemble mesurable A contenant $p^{-1}(y)$, $\mu[X \setminus A|y] = 0$ (Cf. Définition I-42). Cette généralité n'est pas fondamentale, mais permettra de faire le lien entre la notion de désintégration et le théorème de Radon–Nikodym (Cas Particulier X-18).

REMARQUE X-13. Les conditions (i) et (ii) de la Définition X-10 sont un moyen de traduire rigoureusement la définition informelle

$$\mu(dx|y) = \frac{\mu(dx) \delta_{p(x)=y}}{\nu(dy)}.$$

En langage imprécis, il s'agit de “conditionner” μ par l'événement $p(x) = y$.

REMARQUE X-14. On peut généraliser cette définition pour inclure le cas où μ est une mesure signée, de sorte que les mesures $\mu(dx|y)$ sont aussi signées; on peut également imposer que $\mu(dx|y)$ ait le même signe que $d(p_{\#}\mu)/d\nu$ en y . **À vérifier.**

REMARQUE X-15. La Définition X-10 est très utilisée en probabilités, dans le cas où μ est une mesure de probabilités, $\nu = p_{\#}\mu$; on a alors coutume d'imposer que les mesures $\mu(dx|y)$ soient des mesures de probabilités, que l'on appelle **probabilités conditionnelles régulières**. Intuitivement, si $\mu(dx)$ est la répartition attendue des valeurs d'une variable aléatoire X , $\mu(dx|y)$ est la répartition attendue des valeurs de X sachant que $p(X)$ vaut y .

CAS PARTICULIER X-16. Soient (Y, \mathcal{B}) et (Z, \mathcal{C}) deux σ -algèbres; soit $X = Y \times Z$, muni de la tribu produit $\mathcal{B} \otimes \mathcal{C}$, et soit $p : X \rightarrow Y$ la projection définie par $p(y, z) = y$. Soit μ une mesure de probabilités sur X , et $\nu = p_{\#}\mu$ sa marginale sur Y ; ν -désintégrer μ par p revient à écrire μ comme une superposition de mesures (de probabilités) $\mu_y = \mu(\cdot | y)$, de sorte que chaque μ_y est concentrée sur la fibre $\{y\} \times Z$ (et s'identifie donc à une probabilité sur Z). On peut noter

$$(82) \quad \mu(dy dz) = \mu_y(dz) \nu(dy) = \mu(dz|y) \nu(dy).$$

La désintégration peut ainsi se voir comme une opération inverse de celle décrite à la sous-section III-2.4.

EXEMPLE X-17. Avec les notations précédentes, soient λ et σ des mesures σ -finies sur Y et Z respectivement, soit $f(y, z)$ une fonction mesurable positive de (y, z) , et

$$\mu(dy dz) = f(y, z) \lambda(dy) \lambda(dz).$$

Alors on peut désintégrer μ sous la forme (82) en posant

$$\mu(dz|y) = \frac{f(y, z) \sigma(dz)}{\int_Z f(y, z') \sigma(dz')}, \quad \nu(dy) = \left(\int_Z f(y, z) \sigma(dz) \right) \lambda(dy).$$

CAS PARTICULIER X-18. Soient μ et ν deux mesures σ -finies sur un espace mesurable (X, \mathcal{A}) , telles que μ ne charge aucun ensemble ν -mesurable; et soit $f = d\mu/d\nu$ la densité de Radon–Nikodym de μ par rapport à ν . On obtient une ν -désintégration de μ par l'application $p = \text{Id}$ en posant $\mu(dx|y) = f(y)\delta_y(dx)$. En effet, cette mesure est évidemment concentrée sur y , et

$$\int \mu[A|y] \nu(dy) = \int f(y) \delta_y[A] \nu(dy) = \int f(y) 1_{y \in A} \nu(dy) = \int_A f(y) \nu(dy) = \mu[A].$$

Ceci est un cas particulier de l'Exemple X-16, dans lequel Z est réduit à un point.

L'énoncé suivant présente quelques propriétés très utiles des désintégrations :

PROPOSITION X-19. Soient (X, \mathcal{A}, μ) et (Y, \mathcal{B}, ν) deux espaces mesurés, $p : X \rightarrow Y$ une application mesurable, et $\mu(\cdot|y)$ une ν -désintégration de μ par p . Alors

$$(i) \forall A \in \mathcal{A}, \forall B \in \mathcal{B}, \quad \mu[A \cap p^{-1}(B)] = \int_B \mu[A|y] \nu(dy).$$

(ii) $p_{\#}\mu$ est ν -mesurable, et $y \mapsto \mu[X|y]$ est la densité de Radon–Nikodym de $p_{\#}\mu$ par rapport à ν :

$$\mu[X|y] = \frac{(p_{\#}\mu)(dy)}{\nu(dy)},$$

l'identité ayant lieu pour ν -presque tout y . En particulier, si $p_{\#}\mu = \nu$, alors $\mu(\cdot|y)$ est une mesure de probabilités, pour ν -presque tout y .

(iii) Pour tout $F : X \rightarrow [0, +\infty]$, mesurable pour la tribu produit $\mathcal{A} \otimes \mathcal{B}$,

$$\int_X F(x, p(x)) \mu(dx) = \int_Y \int_X F(x, y) \mu(dx|y) \nu(dy).$$

Cette formule reste vraie si F est mesurable $X \rightarrow \mathbb{R}$ et $x \mapsto F(x, p(x)) \in L^1(\mu)$ (alors $\int_X F(x, y) \mu(dx|y)$ est bien défini pour ν -presque tout y).

(iv) Pour toute fonction mesurable $f : Y \rightarrow [0, +\infty]$, la formule

$$(83) \quad f_Y(y) = \int_Y f(x) \mu(dx|y)$$

définit une fonction mesurable de y ; en outre, pour toute fonction mesurable $g : X \rightarrow [0, +\infty]$ on a

$$(84) \quad \int_X f(x) g(p(x)) \mu(dx) = \int_Y f_Y(y) g(y) \nu(dy).$$

Si $f \in L^1(\mu)$, l'intégrale (83) est bien définie pour ν -presque tout y , et (quitte à redéfinir f_Y sur un ensemble négligeable) c'est une fonction mesurable de y . La formule (84) reste en outre valable dès que $f \in L^q(w\mu)$ et $g \in L^{q'}(\nu)$, où $q \in [1, \infty]$, $q' = q/(q-1)$ $w(x) = h(p(x))^{q-1}$, $h = d(p_{\#}\mu)/d\nu$.

REMARQUE X-20. Si (X, \mathcal{A}, μ) est un espace de probabilité, \mathcal{B} est une sous-tribu de \mathcal{A} , p est l'identité $(X, \mathcal{A}) \rightarrow (X, \mathcal{B})$ et $\nu = p_{\#}\mu$, la fonction f_Y définie par (83) est appelée **espérance conditionnelle** de f par rapport à la sous-tribu \mathcal{B} . La formule (84) caractérise alors f_Y , et on l'utilise souvent comme définition de l'espérance conditionnelle (typiquement dans le cas où $f \in L^2(\mu)$, $g \in L^2(\nu)$) sans faire référence à une quelconque désintégration. En fait l'espérance conditionnelle, ainsi définie, est une notion plus générale que la probabilité conditionnelle, et se construit par un argument fonctionnel simple qui ressemble à celui que l'on a utilisé dans la première démonstration du Théorème de Radon–Nikodym (projeter $f \in L^2(X, \mathcal{A}, \mu)$ sur l'espace vectoriel $L^2(X, \mathcal{B}, \mu)$).

REMARQUE X-21. La formule apparaissant dans la Proposition X-19(i) est souvent prise comme définition de la désintégration d'une mesure, à la place des conditions (i) et (ii) de la Définition X-10. On obtient ainsi une notion plus générale de mesures conditionnelles, qui vérifie encore toutes les conclusions de la Proposition X-19. Nous étudierons cette notion dans la section suivante.

PREUVE DE LA PROPOSITION X-19. Soient $B \in \mathcal{B}$ et $y \in Y$. Si $y \in B$, alors $\mu(\cdot | y)$ est concentrée sur l'ensemble mesurable $p^{-1}(B)$, donc $\mu[A \cap p^{-1}(B) | y] = \mu[A | y]$. En revanche, si $y \notin B$, $\mu(\cdot | y)$ est concentrée sur $Y \setminus p^{-1}(B)$, donc $\mu[A \cap p^{-1}(B)] = 0$; on conclut que

$$\mu[A \cap p^{-1}(B) | y] = 1_{y \in B} \mu[A | y],$$

pour tout $y \in Y$. L'identité (81) implique alors

$$\mu[A \cap p^{-1}(B)] = \int_Y 1_{y \in B} \mu[A | y] \nu(dy) = \int_B \mu[A | y] \nu(dy),$$

ce qui prouve (i).

En particulier, pour tout $B \in \mathcal{B}$, on a

$$p_{\#}\mu[B] = \mu[p^{-1}(B)] = \int_B \mu[X | y] \nu(dy),$$

ce qui implique (ii).

Notons $\lambda(dx dy) = \mu(dx | y) \nu(dy)$ (soit $\nu \otimes \mu_y$ dans les notations de la sous-section III-2.4). La formule (i) montre que

$$(85) \quad \int_X F(x, p(x)) \mu(dx) = \int_X \times_Y F(x, y) \lambda(dx dy)$$

dans le cas particulier où $F(x, y) = 1_{x \in A} 1_{y \in B}$, $A \in \mathcal{A}$, $B \in \mathcal{B}$. Les deux membres de (85) sont stables par limite croissante et différence, et la classe monotone engendrée par les pavés $A \times B$ est la tribu produit entière (Théorème I-68); on en déduit que (85) est vrai dès que F est l'indicatrice d'un ensemble mesurable pour la tribu produit. Par un nouvel argument de limite monotone, cela est vrai pour toute fonction F mesurable positive. La généralisation à $F \in L^1$ se fait selon les mêmes lignes que le Théorème de Fubini (Théorème III-50).

Enfin, pour obtenir (iv), il suffit de poser $F(x, y) = f(x)g(y)$ dans (iii), et de noter que $\int F(x, y) \lambda(dx dy) = \int f_Y(y)g(y) \nu(dy)$. Dans le cas où f et g sont positives, cela ne demande aucune justification particulière; dans le cas général, il

convient de vérifier que les intégrales sont bien définies sous les conditions énoncées. Pour cela on remarque d'abord que

$$\int_Y \int_X |f(x)|^q \mu(dx|y) h(y)^{q-1} \nu(dy) = \int_X |f(x)|^q h(p(x))^{q-1} \mu(dx) < +\infty;$$

l'intégrale $\int_X |f(x)|^q \mu(dx|y)$ est donc finie pour ν -presque tout y (a priori seulement en-dehors de l'ensemble Z où h s'annule; mais cet ensemble est $(p_{\#}\mu)$ -négligeable, donc $\mu(\cdot|y) = 0$ pour ν -presque tout $y \in Z$). Comme $\mu[X|y]$ est fini pour ν -presque tout y , on a bien $\int_X |f(x)| \mu(dx|y) < +\infty$ presque partout. Ensuite on utilise l'inégalité de Hölder et la définition de la mesure image pour obtenir l'estimation

$$\begin{aligned} \int |f(x)| |g(p(x))| \mu(dx) &\leq \left(\int |f(x)|^q w(x)^{1/q} d\mu(x) \right)^{\frac{1}{q}} \left(\int |g(p(x))|^{q'} w(x)^{-1/q} d\mu(x) \right)^{\frac{1}{q'}} \\ &= \left(\int |f(x)|^q w(x) d\mu(x) \right)^{\frac{1}{q}} \left(\int |g(p(x))|^{q'} h(p(x))^{-1} d\mu(x) \right)^{\frac{1}{q'}} \\ &= \left(\int |f(x)|^q w(x) d\mu(x) \right)^{\frac{1}{q}} \left(\int |g(y)|^{q'} h(y)^{-1} d(p_{\#}\mu)(y) \right)^{\frac{1}{q'}} \\ &= \left(\int |f(x)|^q w(x) d\mu(x) \right)^{\frac{1}{q}} \left(\int |g(y)|^{q'} d\nu(y) \right)^{\frac{1}{q'}} < +\infty. \end{aligned}$$

Ceci conclut la preuve de la Proposition X-19. \square

Pour l'instant nous n'avons pas répondu à la question de savoir si l'on peut effectivement désintégrer des mesures sous des hypothèses assez générales. Ce problème est délicat, en contraste avec le Théorème de Radon–Nikodym pour lequel nous avons une solution complète sous l'hypothèse de σ -additivité. Il est d'ailleurs bien connu en probabilités que l'existence des probabilités conditionnelles est un problème beaucoup plus difficile que celui de l'existence des espérances conditionnelles.

Le théorème suivant, dont on peut trouver diverses variantes dans la littérature, apporte une solution suffisamment générale pour couvrir les besoins courants. On rappelle qu'un espace polonais est un espace métrique séparable complet. On dira qu'un espace mesurable (Y, \mathcal{B}) est *dénombrablement engendré* s'il existe une sous-famille \mathcal{B}' dénombrable incluse dans \mathcal{B} telle que $\sigma(\mathcal{B}') = \mathcal{B}$.

THÉORÈME X-22 (Théorème de désintégration). *Soit X un espace polonais, muni de sa tribu borélienne \mathcal{A} , et d'une mesure de Borel σ -finie μ . Soit (Y, \mathcal{B}) un espace mesurable dénombrablement engendré, dans lequel les singletons sont mesurables, et soit ν une mesure σ -finie sur Y . Soit p une application mesurable $X \rightarrow Y$; on suppose que $p_{\#}\mu$ ne charge pas les ensembles ν -négligeables. Alors il existe une ν -désintégration de μ par p :*

$$\mu(dx) = \mu(dx|y) \nu(dy),$$

satisfaisant les conditions de la Définition X-10.

En outre, cette désintégration est unique à modification sur un ensemble ν -négligeable près; de manière explicite, si $\mu(dx|y)$ et $\mu'(dx|y)$ sont deux ν -désintégrations de μ par p , alors il existe un ensemble négligeable N tel que pour tout $y \in Y \setminus N$ on ait $\mu(\cdot|y) = \mu'(\cdot|y)$.

EXEMPLE X-23. L'espace (Y, \mathcal{B}) vérifie bien sûr les hypothèses requises si c'est la tribu borélienne d'un espace métrique séparable, puisque cette tribu est engendrée par une famille dénombrable (voir le Théorème I-32).

REMARQUE X-24. Autant l'hypothèse selon laquelle \mathcal{B} contient les singletons est d'ordinaire facile à vérifier ou infirmer en pratique, autant l'hypothèse selon laquelle \mathcal{B} est dénombrablement engendrée peut s'avérer traîtresse. Il est important de noter à cet égard que cette propriété ne s'hérîte pas : ce n'est pas parce qu'une tribu \mathcal{A} est dénombrablement engendrée que ses sous-tribus le sont aussi.

EXEMPLE X-25. Soit \mathcal{A} la tribu borélienne de \mathbb{R} , et \mathcal{B} la tribu engendrée par les points : \mathcal{B} est constituée des parties qui sont dénombrables ou de complémentaire dénombrable. On pourra montrer en exercice que \mathcal{B} n'est pas dénombrablement engendrée.

REMARQUE X-26. Si \mathcal{B} ne vérifie pas les hypothèses demandées dans l'énoncé du Théorème X-22, on peut toujours prouver l'existence de mesures $\mu(\cdot|y)$ vérifiant (86) ; ce que de nombreux auteurs appellent mesures conditionnelles. C'est l'objet de la section suivante.

L'idée de la preuve du Théorème X-22 est naturelle : d'après la Proposition X-19(i) les quantités $\mu[A|y]$ recherchées doivent vérifier

$$(86) \quad \forall A \in \mathcal{A}, \quad \forall B \in \mathcal{B}, \quad \int_B \mu[A|y] \nu(dy) = \mu[A \cap p^{-1}(B)] = (p_{\#} \mu_A)[B],$$

où μ_A est la mesure définie par $\mu_A[C] = \mu[A \cap C]$. Il s'ensuit que $y \mapsto \mu[A|y]$ est la densité de Radon–Nikodym de μ_A par rapport à ν (bien définie grâce au Corollaire X-5). Le problème est que cette densité est seulement définie à un ensemble négligeable près, *qui dépend de A*. Comme \mathcal{A} n'est pas a priori dénombrable, on ne peut se permettre d'exclure l'union de tous ces ensembles négligeables ; à la place on va utiliser les hypothèses topologiques pour se ramener à une famille dénombrable d'ensembles mesurables A .

PREUVE DU THÉORÈME X-22. 1. On commence par se ramener au cas où μ est finie, de la même façon que dans la preuve du Théorème X-3. Pour cela on peut noter que X est union croissante de compacts $(K_\ell)_{\ell \in \mathbb{N}}$ de μ -mesure finie (Théorème I-54) et que chaque K_ℓ est un espace polonais ; de même, Y est union croissante d'ensembles mesurables Y_k de mesure finie, et chaque Y_k peut être vu comme un espace mesurable engendré par une partie dénombrable.

2. Soit D un ensemble dense dénombrable dans X , et soit \mathcal{C} la collection de toutes les unions finies de boules $B(x, r)$, où $x \in D$ et r est un nombre rationnel positif. La famille \mathcal{C} est (au plus) dénombrable, et on l'ordonne de manière arbitraire : $\mathcal{C} = \{C_k; k \in \mathbb{N}\}$. Par le Théorème I-32, \mathcal{C} engendre la tribu borélienne entière. L'algèbre \mathcal{A}' engendrée par \mathcal{C} est dénombrable, comme union croissante des algèbres finies engendrées par $\{C_1, \dots, C_k\}$ ($k \in \mathbb{N}$). On ordonne $\mathcal{A}' = \{A_\ell; \ell \in \mathbb{N}\}$.

3. À ce stade on peut déjà montrer l'unicité de la désintégration. Si $\mu(dx|y)$ est une désintégration admissible, alors, comme on l'a déjà remarqué, la Proposition X-19(i) implique

$$(87) \quad \int_B \mu[A|y] \nu(dy) = \mu[A \cap p^{-1}(B)] = (p_{\#} \mu_A)[B]$$

pour tous $A \in \mathcal{A}$, $C \in \mathcal{B}$, où $\mu_A[C] = \mu[A \cap C]$; et donc $y \mapsto \mu[A|y]$ est la densité de Radon–Nikodym de $p_{\#}\mu_A$. Si $\mu(dx|y)$ et $\mu'(dx|y)$ sont deux désintégrations admissibles, pour tout $A \in \mathcal{A}$ les fonctions $\mu[A|y]$ et $\mu'[A|y]$ coïncident pour ν -presque tout y , c'est-à-dire pour tout $y \in Y \setminus N(A)$, où $N(A)$ est un ensemble ν -négligeable. Soit N l'union des $N(A)$ pour $A \in \mathcal{A}'$: N est ν -négligeable et

$$\forall y \in Y \setminus N, \forall A \in \mathcal{A}', \quad \mu[A|y] = \mu'[A|y].$$

Si l'on fixe $y \in Y \setminus N$, les mesures $\mu(\cdot|y)$ et $\mu'(\cdot|y)$ coïncident sur \mathcal{A}' , donc sur la classe monotone engendrée par \mathcal{A}' , qui d'après le Théorème I-68 n'est autre que la tribu \mathcal{A} tout entière.

4. L'existence nous donnera plus de fil à retordre, et tout particulièrement la propriété de σ -additivité, pour laquelle nous allons utiliser la régularité via la Proposition I-52.

Par le Théorème I-54, μ est régulière, donc pour tout $\ell \in \mathbb{N}$ on peut trouver une suite croissante de compacts $(K_{\ell m})_{m \in \mathbb{N}}$ telle que $K_{\ell m} \subset A_{\ell}$ et

$$(88) \quad \mu[K_{\ell m}] \xrightarrow{m \rightarrow \infty} \mu[A_{\ell}].$$

Soit \mathcal{A}'' l'algèbre engendrée par les A_{ℓ} et les $K_{\ell m}$ ($\ell, m \in \mathbb{N}$); cette algèbre est encore dénombrable. Pour tout $A \in \mathcal{A}''$ on définit une mesure μ_A sur \mathcal{B} par la formule $\mu_A = p_{\#}(\mu|_A)$, soit

$$\mu_A[B] = \mu[A \cap p^{-1}(B)].$$

Bien sûr $\mu_A \leq p_{\#}\mu$; en particulier μ_A ne charge aucun ensemble ν -négligeable, et par le Théorème de Radon–Nikodym (Corollaire X-5) μ_A admet une densité f_A par rapport à ν : $\mu_A = f_A \nu$, soit encore

$$\mu[A \cap p^{-1}(B)] = \int_B f_A(y) \nu(dy).$$

Pour chaque $A \in \mathcal{A}''$, f_A est positive en-dehors d'un ensemble ν -négligeable $N(A)$.

Soient A_1, \dots, A_k des éléments disjoints de \mathcal{A}'' ($k \in \mathbb{N}$, $k \geq 2$); et soit $A = \cup A_j$. Bien sûr $\mu_A = \mu_{A_1} + \dots + \mu_{A_k}$, donc $f_A \nu = (f_{A_1} + \dots + f_{A_k}) \nu$. Il existe donc un ensemble négligeable $N(A_1, \dots, A_k) \subset Y$ tel que $\nu[N(A_1, \dots, A_k)] = 0$ et

$$(89) \quad \forall y \in Y \setminus N(A_1, \dots, A_k), \quad f_A(y) = f_{A_1}(y) + \dots + f_{A_k}(y).$$

Pour tous $\ell, m \in \mathbb{N}$ l'inclusion $K_{\ell m} \subset A_{\ell}$ implique $\mu_{K_{\ell m}} \leq \mu_{A_{\ell}}$, d'où $f_{K_{\ell m}} \leq f_{A_{\ell}}$ en-dehors d'un ensemble ν -négligeable $N(\ell, m)$. La relation (88) se réécrit

$$\int f_{K_{\ell m}}(y) \nu(dy) \xrightarrow{m \rightarrow \infty} \int f_{A_{\ell}}(y) \nu(dy),$$

soit encore

$$\int [f_{A_{\ell}}(y) - f_{K_{\ell m}}(y)] \nu(dy) \xrightarrow{\ell \rightarrow \infty} 0.$$

Quitte à restreindre l'intégrale au complémentaire de l'union des $N(\ell, m)$ ($m \in \mathbb{N}$), on peut supposer que l'intégrande est partout positif, et cette limite implique la convergence presque partout de $f_{A_{\ell}} - f_{K_{\ell m}}$ vers 0. Il existe donc un ensemble ν -négligeable $N(\ell)$ tel que

$$(90) \quad \forall y \in Y \setminus N(\ell), \quad f_{K_{\ell m}}(y) \xrightarrow{m \rightarrow \infty} f_{A_{\ell}}(y).$$

Soit N l'union de tous les $N(A)$ ($A \in \mathcal{A}''$), de tous les $N(A_1, \dots, A_k)$ ($A_j \in \mathcal{A}'', k \geq 2$), et de tous les $N(\ell)$ ($\ell \in \mathbb{N}$). Étant une union dénombrable d'ensembles négligeables, N est lui-même négligeable. Pour tout $y \in Y \setminus N$, on a donc par construction

- (i) Pour tout $A \in \mathcal{A}'', f_A(y) \geq 0$;
- (ii) Pour tout $k \in \mathbb{N}$ et toutes parties disjointes $A_1, \dots, A_k \in \mathcal{A}'', f_{A_1 \cup \dots \cup A_k}(y) = \sum_{j=1}^k f_{A_j}(y)$.
- (iii) Pour tout $A \in \mathcal{A}''$ il existe une suite $(K_m)_{m \in \mathbb{N}}$ de compacts inclus dans \mathcal{A} tels que $f_{K_m}(y) \longrightarrow f_A(y)$.

Pour tout $y \in Y \setminus N$ on définit $\mu_y : \mathcal{A}'' \rightarrow \mathbb{R}_+$ par la formule

$$\mu_y[A] = f_A(y);$$

les relations (i)–(iii) montrent que μ_y est une fonction additive d'ensembles. En outre, pour tout $A \in \mathcal{A}'$ on a

$$\mu_y[A] = \sup \left\{ \mu_y[K]; \quad K \text{ compact, } K \subset A, K \in \mathcal{A}'' \right\}.$$

Par la Proposition I-52, on déduit que μ_y est σ -additive sur \mathcal{A}' . Le Théorème de prolongement de Carathéodory (Théorème I-69) permet d'étendre μ_y en une mesure définie sur \mathcal{A} tout entier.

On déduit du théorème de convergence monotone que

$$\left\{ A \in \mathcal{A}; \quad y \longmapsto \mu_y[A] \text{ est mesurable sur } Y \setminus N \right\}$$

est une classe monotone; comme cet ensemble contient \mathcal{A}' , il doit contenir \mathcal{A} tout entier.

On pose $\mu_y = 0$ pour $y \in Y \setminus N$; comme N est mesurable, l'application μ_y reste mesurable en y . En outre on a, pour tout $A \in \mathcal{A}'$ et pour tout $B \in \mathcal{B}$,

$$(91) \quad \mu[A \cap p^{-1}(B)] = \int_B \mu_y[A] \nu(dy).$$

Si l'on fixe B , cette relation reste vraie (par σ -additivité et convergence monotone) sur la classe monotone engendrée par \mathcal{A}' , qui d'après le Théorème I-68 est \mathcal{A} tout entier; on conclut que la relation (91) est vraie pour tout $A \in \mathcal{A}$ et pour tout $B \in \mathcal{B}$. (Rappelons que ceci est une définition souvent adoptée pour la désintégration.)

5. Le choix $B = Y$ dans (91) donne

$$\mu[A] = \int_Y \mu_y[A] \nu(dy),$$

de sorte que la condition (i) dans la Définition X-10 est satisfaite.

La condition (ii) va demander plus de travail. La preuve de la Proposition X-19(iii) s'applique pour montrer que

$$(92) \quad \int_X F(x, p(x)) \mu(dx) = \int_{X \times Y} F(x, y) \mu_y(dx) \nu(dy)$$

pour toute fonction F mesurable $X \times Y \rightarrow \mathbb{R}_+$.

Admettons provisoirement que la diagonale $\Delta = \{(y, y); y \in Y\}$ est mesurable dans $Y \times Y$; alors

$$F : (x, y) \longmapsto 1_{p(x) \neq y}$$

est mesurable $X \times Y \rightarrow \mathbb{R}_+$ comme composée des fonctions mesurables $(x, y) \mapsto (p(x), y)$ et $(z, y) \mapsto 1_{z \neq y}$. En appliquant la formule (92) à F on trouve

$$\begin{aligned} \int_Y \mu_y[X \setminus p^{-1}(y)] \nu(dy) &= \int 1_{p(x) \neq y} \mu_y(dx) \nu(dy) \\ &= \int 1_{p(x) \neq p(x)} \mu(dx) = 0. \end{aligned}$$

On en déduit que pour tout y en-dehors d'un ensemble négligeable N' on a $\mu_y[X \setminus p^{-1}(y)] = 0$. En posant

$$\mu(\cdot | y) = \begin{cases} \mu_y & \text{si } y \notin N' \\ 0 & \text{si } y \in N' \end{cases}$$

on obtient une famille $\mu(\cdot | y)$ vérifiant toutes les conditions souhaitées.

Il reste seulement à montrer que Δ est mesurable dans $\mathcal{B} \times \mathcal{B}$. Par hypothèse \mathcal{B} est engendrée par une famille dénombrable $\mathcal{B}' = (B_k)_{k \in \mathbb{N}}$. Pour tout $k \in \mathbb{N}$, l'algèbre finie $\mathcal{B}'(k)$ engendrée par $\{B_1, \dots, B_k\}$ est associée à une partition finie $(Y_1^k, \dots, Y_{\ell(k)}^k)$ de Y (Exercice I-4(i)). Soient

$$\Delta_k = \bigcup_{i \leq \ell(k)} Y_i^k \times Y_i^k; \quad \tilde{\Delta} = \bigcap_{k \in \mathbb{N}} \Delta_k.$$

Bien sûr $\Delta_k \supset \Delta$ pour tout k , donc $\tilde{\Delta} \supset \Delta$. S'il existe $y \neq x$ tel que $(x, y) \in \tilde{\Delta}$, cela veut dire que pour tout $k \in \mathbb{N}$, aucun élément de $\mathcal{B}'(k)$ ne sépare x de y ; plus explicitement, pour tout $B \in \mathcal{B}'(k)$ on a soit $\{x, y\} \subset B$, soit $\{x, y\} \cap B = \emptyset$. Mais l'ensemble des $B \in \mathcal{B}$ vérifiant cette propriété est une σ -algèbre, et comme elle contient tous les B_k , elle coïncide avec \mathcal{B} tout entière. Ceci est en contradiction avec l'hypothèse de mesurabilité de $\{x\}$. On conclut que $\Delta = \tilde{\Delta}$, ce qui achève l'argument. \square

X-2.2. Désintégration généralisée. Cette sous-section peut être omise en première lecture. Le Théorème X-22 a beau sembler très général, il n'est pas assez puissant pour traiter toutes les situations d'intérêt. On ressent en effet le besoin, dans de nombreux problèmes (non académiques!), de désintégrer par rapport à un espace mesurable "pathologique" dans lequel les points ne sont pas mesurables, ou la tribu n'est pas dénombrablement engendrée.

Pour traiter de telles situations, le remède consiste à affaiblir la définition de la désintégration en utilisant la Propriété (i) dans la Proposition X-19 comme définition généralisée.

DÉFINITION X-27 (désintégration généralisée). Soient (X, \mathcal{A}, μ) , (Y, \mathcal{B}, ν) deux espaces mesurables, et (Z, \mathcal{C}) un espace mesuré. Soient $p : X \rightarrow Z$ et $q : Y \rightarrow Z$ deux applications mesurables. On appelle ν -désintégration de μ par (p, q) une famille de mesures $\mu(dx|z)$ sur X , indexées par $z \in Z$, telles que pour tout $A \in \mathcal{A}$, $z \mapsto \mu[A|z]$ est mesurable, et

$$\forall A \in \mathcal{A}, \forall C \in \mathcal{C}, \quad \mu[A \cap p^{-1}(C)] = \int_{q^{-1}(C)} \mu[A|q(y)] \nu(dy).$$

REMARQUE X-28. L'application q est incluse par commodité et ne joue pas un rôle crucial : en munissant Z de la mesure $q_{\#}\nu$, on peut toujours se ramener au cas où $q = \text{Id}$.

Voici maintenant une variante du Théorème X-22 :

THÉORÈME X-29 (désintégration généralisée). *Soient (X, \mathcal{A}, μ) un espace polonais σ -fini, (Y, \mathcal{B}, ν) un espace mesuré σ -fini, et (Z, \mathcal{C}) un espace mesurable. Soient $p : X \rightarrow Z$ une application mesurable, et $q : Y \rightarrow Z$ une application mesurable surjective. On suppose que pour tout $N \in \mathcal{C}$, tel que $q(N)$ est mesurable,*

$$\nu[N] = 0 \implies \mu[p^{-1}(q(N))] = 0.$$

Alors il existe une ν -désintégration de μ par (p, q) , c'est-à-dire une famille de mesures $\mu(dx|z)$ dépendant mesurablement de z , telles que

$$(93) \quad \forall A \in \mathcal{A}, \forall C \in \mathcal{C}, \quad \mu[A \cap p^{-1}(C)] = \int_{q^{-1}(C)} \mu[A|q(y)] \nu(dy).$$

Toutes les propriétés énoncées dans la Proposition X-19 sont alors vraies.

En outre, si $\mu(dx|z)$ et $\mu'(dx|z)$ sont deux désintégrations admissibles, alors il existe un ensemble ν -négligeable N tel que

$$\forall y \in Y \setminus N, \quad \mu(\cdot|q(y)) = \mu'(\cdot|q(y)).$$

De plus, si $C \in \mathcal{C}$ est fixé, alors pour $\nu(dy)$ -presque tout y on a

$$\mu[p^{-1}(C)|q(y)] = 1_{q(y) \in C} h(q(y)),$$

où $h = d(p_{\#}\mu)/d(q_{\#}\nu)$.

Si en outre \mathcal{C} est engendré par une famille dénombrable, alors pour ν -presque tout $y \in Y$ la mesure $\mu(\cdot|q(y))$ est concentrée sur $p^{-1}(a(q(y)))$, où

$$a(z) = \bigcap_{c \in C \in \mathcal{C}} C$$

est l'atome engendré par z .

REMARQUE X-30. On ne peut en général espérer que les mesures $\mu_z = \mu(\cdot|z)$ soient portées par les fibres $p^{-1}(y)$; l'obstacle technique étant la non-mesurabilité a priori de la diagonale $\Delta = \{(z, z)\} \subset Z \times Z$ (pour la tribu produit).

REMARQUE X-31. Si (Z, \mathcal{C}) est un ensemble mesurable quelconque, l'atome engendré dans Z par un point z n'est pas en général un ensemble mesurable ; mais cela devient vrai si \mathcal{C} est engendré par une famille dénombrable.

REMARQUE X-32. Dans le cas où (X, \mathcal{A}, μ) est un espace de probabilités, $(Z, \mathcal{C}) = (Y, \mathcal{B})$, $q = \text{Id}$ et $\nu = p_{\#}\mu$, les mesures $\mu(dx|y)$ sont appelées **probabilités conditionnelles** de μ par rapport aux valeurs de p . Dans le cas où (presque) toute mesure $\mu(\cdot|y)$ est concentrée sur l'atome $a(y)$ on parle de probabilités conditionnelles régulières.

PREUVE DU THÉORÈME X-29. La preuve est en grande partie semblable à celle du Théorème X-22, et je n'indiquerai donc que les différences. On commence par se ramener au cas où μ et ν sont finies, et on introduit \mathcal{A}' une algèbre dénombrable engendrant \mathcal{A} . Si l'on dispose d'une désintégration admissible, alors pour tous $C \in \mathcal{C}$

et $A \in \mathcal{A}$,

$$\begin{aligned} \int_C \mu[A|z] (q_{\#}\nu)(dz) &= \int_Z 1_C(z) \mu[A|z] (q_{\#}\nu)(dz) \\ &= \int_Y 1_C(q(y)) \mu[A|q(y)] \nu(dy) \\ &= \int_{q^{-1}(C)} \mu[A|q(y)] \nu(dy) \\ &= \mu[A \cap p^{-1}(C)] = \mu_A[C], \end{aligned}$$

où $\mu_A = p_{\#}(\mu|_A)$. Par conséquent, pour tout $A \in \mathcal{A}$ fixé, $z \mapsto \mu[A|z]$ est la densité de Radon–Nikodym de μ_A par rapport à $q_{\#}\nu$.

Si $\mu(\cdot|z)$ et $\mu'(\cdot|z)$ sont deux décompositions admissibles, pour tout A on peut trouver un ensemble mesurable $N_A \in \mathcal{C}$ tel que pour $q_{\#}\nu[N_A] = 0$ et

$$\forall z \in Z \setminus N_A, \quad \mu[A|z] = \mu'[A|z].$$

Soit N la réunion de tous les N_A quand A varie dans la famille dénombrable $\mathcal{A}' : N$ est toujours négligeable, et

$$\forall z \in Z \setminus N, \quad \forall A \in \mathcal{A}', \quad \mu[A|z] = \mu'[A|z].$$

Comme \mathcal{A}' engendre \mathcal{A} , on conclut que $\mu(\cdot|z) = \mu'(\cdot|z)$, pour $(q_{\#}\nu)$ -presque tout z ; de manière équivalente, $\mu(\cdot|q(y)) = \mu'(\cdot|q(y))$ pour ν -presque tout y .

On note ensuite que pour tout $A \in \mathcal{A}$, $\mu_A = p_{\#}(\mu|_A)$ ne charge pas les ensembles $(q_{\#}\nu)$ -négligeables. En effet, si $q_{\#}\nu[N] = 0$, i.e. $\nu[q^{-1}(N)] = 0$, alors par hypothèse $p^{-1}(N) = p^{-1}(q(q^{-1}(N)))$ est μ -négligeable, donc $\mu_A[N] = \mu[A \cap p^{-1}(N)] = 0$. (Ici on a utilisé la surjectivité de q via l'identité $N = q(q^{-1}(N))$.)

On peut donc définir f_A comme la densité de Radon–Nikodym de μ_A par rapport à $q_{\#}\nu$, et on pose

$$\mu[A|z] = f_A(z).$$

On démontre exactement comme dans la preuve du Théorème X-22 que pour tout z en dehors d'un ensemble de $q_{\#}\nu$ -mesure nulle, $f_A(z)$ est σ -additive en A et mesurable en z . On arrive ainsi à l'analogue de la formule (91) :

$$\begin{aligned} \mu[A \cap p^{-1}(C)] &= \int_C \mu_y[A] (q_{\#}\nu)(dy) \\ &= \int_{q^{-1}(C)} \mu[A|q(y)] \nu(dy). \end{aligned}$$

Ensuite, pour tous C et C' dans \mathcal{C} , on peut écrire

$$\begin{aligned} \int_{C'} 1_C(z) (p_{\#}\mu)(dz) &= \int_Z 1_{C'}(z) 1_C(z) (p_{\#}\mu)(dz) \\ &= p_{\#}\mu[C \cap C'] \\ &= \mu[p^{-1}(C \cap C')] \\ &= \mu[p^{-1}(C) \cap p^{-1}(C')] \\ &= \int_{C'} \mu[p^{-1}(C)|z] (q_{\#}\nu)(dz). \end{aligned}$$

Il s'ensuit que $z \mapsto \mu[p^{-1}(C)|z]$ est la densité de Radon–Nikodym de $1_C p_{\#}\mu$ par rapport à $q_{\#}\nu$; il existe donc un ensemble $q_{\#}\nu$ -négligeable N_C tel que

$$\forall z \in Z \setminus N_C, \quad \mu[p^{-1}(C)|z] = 1_C(z) h(z),$$

où $h = d(p_{\#}\mu)/d(q_{\#}\nu)$.

Supposons maintenant que \mathcal{C} est dénombrablement engendrée; on veut montrer que $\mu_z = \mu(\cdot|z)$ est concentrée sur l'atome $a(z)$, pour $q_{\#}\nu$ -presque tout z .

Montrons d'abord que $a(z)$ est mesurable. Soit \mathcal{C}' une algèbre dénombrable engendrant \mathcal{C} , et soit $a'(z)$ l'intersection de tous les éléments de \mathcal{C}' contenant z . Bien sûr $a(z) \subset a'(z)$, et $a'(z)$ est mesurable. Le problème est de montrer que $a(z) = a'(z)$. Soit

$$\tilde{\mathcal{C}} = \left\{ C \in \mathcal{C}; z \in C \implies a'(z) \subset C \right\}.$$

Par construction $\mathcal{C}' \subset \tilde{\mathcal{C}}$. On vérifie facilement que \mathcal{C}' est une classe monotone, et le Lemme de Classe Monotone (Théorème I-68) entraîne l'identité $\tilde{\mathcal{C}} = \mathcal{C}$, ce qui prouve le résultat.

Soient ensuite \mathcal{C}' une algèbre dénombrable engendrant \mathcal{C} , et N la réunion de tous les N_C . On a alors

$$(94) \quad \forall z \in Z \setminus N, \forall C \in \mathcal{C}' \quad \mu[p^{-1}(C)|z] = 1_C(z) h(z).$$

À z fixé, les deux membres de cette identité (considérées comme fonctions de $C \in \mathcal{C}$) définissent des mesures finies, qui coïncident sur \mathcal{C}' . Une nouvelle application du Lemme de Classe Monotone montre que ces deux mesures coïncident sur \mathcal{C} tout entier. On conclut que

$$\forall z \in Z \setminus N, \forall C \in \mathcal{C} \quad p_{\#}\mu_z[C] = 1_C(z) h(z),$$

et le résultat final s'obtient en choisissant $C = Z \setminus a(z)$. \square

Le Théorème X-29 rend de grands services. Il en existe diverses variantes montrant que si μ, ν vérifient certaines hypothèses (invariances, symétries, ...) alors on peut imposer aux mesures conditionnelles de vérifier des hypothèses en rapport. Le schéma de ces démonstrations est en général le même que celui que l'on a utilisé pour montrer la concentration atomique :

- en appliquant la formule (93) à des ensembles bien choisis, on montre que $\mu[A|z]$ est densité de Radon–Nikodym d'une certaine mesure, pour tout A dans un ensemble bien choisi;

- par unicité de cette densité, on tire des conclusions sur $\mu[A|z]$ hors d'un ensemble négligeable de points z ; cet ensemble peut être choisi indépendant de A si A varie dans une famille dénombrable;

- À z fixé, on montre que si la propriété est vraie pour A dans une famille dénombrable fixée a priori, elle reste vraie pour tout A dans la classe souhaitée.

La mise en oeuvre de cette stratégie peut s'avérer subtile. À titre d'exemple, essayons de démontrer l'énoncé suivant, fort utile en théorie des systèmes dynamiques :

THÉORÈME X-33. *Soit X un espace polonais muni de sa tribu borélienne \mathcal{A} et d'une mesure σ -finie μ . Soit $T : X \rightarrow X$ une bijection bimesurable; pour toute partie A de X , on note $S(A)$ le saturé de A pour l'action de $T : S(A) = \cup_{k \in \mathbb{Z}} T^k(A)$. On*

suppose que μ est quasi-invariante, c'est-à-dire que pour tout ensemble μ -négligeable N , on a $\mu[T(N)] = 0$. Alors on peut désintégrer μ sous la forme

$$\mu(dx) = \int_X \mu_y(dx) \mu(dy),$$

où les mesures conditionnelles μ_y vérifient (a) $\mu_y = \mu_{y'}$ si $y\mathcal{R}y'$; (b) μ_y est ergodique, c'est-à-dire que pour tout ensemble mesurable saturé S on a $\mu_y[S] \in \{0, 1\}$.

TENTATIVE DE PREUVE. Pour faire entrer cet énoncé dans le cadre du Théorème X-29, posons $(Y, \mathcal{B}, \nu) = (X, \mathcal{A}, \nu)$; définissons Z comme l'espace quotient X/\mathcal{R} , $p = q$ comme l'application qui à un point associe sa classe d'équivalence, et munissons Z de la tribu image $\mathcal{C} = p_{\#}\mathcal{A}$.

Le Théorème X-29 nous fournit une désintégration en mesures conditionnelles $\mu_y = \mu(\cdot | q(y))$ indexées par $y \in X$, et la condition (a) est évidemment vérifiée.

Soit maintenant A un ensemble saturé; de manière équivalente $A = p^{-1}(C)$ pour un certain $C \in \mathcal{C}$. Comme dans la preuve du Théorème X-29 on montre que $z \mapsto \mu[p^{-1}(C)|z]$ est la densité de Radon-Nikodym de $1_C p_{\#}\mu$ par rapport à $p_{\#}\mu$; autrement dit, pour tout z en-dehors d'un ensemble négligeable N_A ,

$$\mu_z[A] = 1_C(z).$$

En particulier, $\mu_z[A] \in \{0, 1\}$. Si l'on savait montrer que cette propriété est vraie pour tout z en-dehors d'un ensemble négligeable *indépendant de* A , on aurait la propriété (b)... mais comme $p_{\#}\mathcal{A}$ n'est pas a priori dénombrablement engendrée (et cela est faux en général), le raisonnement s'écroule.

En fait la preuve est plus subtile. **Indiquer une référence ou terminer la preuve.** \square

REMARQUE X-34. Pour certains exemples simples, Z , muni de la mesure image $p_{\#}\mu$, est un espace mesuré fortement pathologique : typiquement, tout ensemble mesurable est soit de mesure nulle soit de mesure pleine! La situation rappelle d'ailleurs celle de l'Exemple X-25.

X-3. Désintégration constructive

(seulement dans des espaces loc cpt....)

À RÉDIGER

X-4. Dérivation des fonctions absolument continues

À RÉDIGER

X-5. Décomposition des mesures d'une infinité de variables

À RÉDIGER

CHAPITRE XI

Analyse de Fourier – À COMPLÉTER

L'analyse de Fourier est un ensemble d'outils ayant pour dénominateur commun la décomposition fréquentielle des fonctions. Sa naissance remonte à un traité célèbre de Fourier (*Théorie analytique de la chaleur*, 1822).

Initialement conçue pour l'analyse mathématique des phénomènes physiques, l'analyse de Fourier s'est révélée d'une utilité prodigieuse dans à peu près toutes les branches des mathématiques, voire de toutes les sciences.

Dans ce chapitre nous allons passer en revue quelques aspects marquants de l'analyse de Fourier. Le style sera assez différent de celui des autres chapitres : on y sacrifiera la généralité à la simplicité et à la variété des applications ; on admettra en outre de nombreuses digressions, et un certain recouvrement avec le chapitre précédent. (**Analyse sur les groupes abéliens**)

Le sujet peut être complété par la lecture de [Körner] ou de [Dym–McKean] pour les applications (la première source est plus simple, et remarquable de pédagogie) ; de [Dhombres–Robert] et [Kahane] pour les aspects historiques ; de [Duoandikoetxea] et [Folland] pour les aspects analytiques plus avancés.

NdCV : Folland serait peut-être mieux à sa place dans les cmtr du chapitre précédent ; également vérifier que la TF de la gaussienne est quelque part

XI-1. Théorie de Fourier

XI-1.1. Contexte et motivations. En dépit des railleries de Hugo et Stendhal à son encontre, Fourier (Joseph de son prénom) laissera son nom à la postérité à plus d'un titre : en tant qu'égyptologue, une spécialité qu'il bouleverse à la suite de sa participation à l'expédition napoléonienne ; en tant qu'homme politique puisqu'il est préfet d'Isère sous Napoléon Bonaparte et sous la Restauration ; en tant qu'administrateur, comme réorganisateur des statistiques françaises ; et enfin en tant que scientifique, la consécration de sa carrière étant son élection au poste de secrétaire perpétuel de l'Académie des Sciences. Pour tous les mathématiciens, il reste le fondateur de l'analyse de Fourier, qu'il crée au début du dix-neuvième siècle pour étudier mathématiquement la répartition de chaleur dans un corps conducteur.

Quelques repères chronologiques :

1807 : Fourier écrit l'équation de la chaleur

1808 : Le mémoire de Fourier est accepté par l'Académie des Sciences en dépit de sérieuses critiques concernant la rigueur des démonstrations (Fourier a en particulier une controverse avec Poisson)

1822 : Publication du *Traité analytique de la chaleur*, version remaniée et augmentée du mémoire de Fourier, qui s'imposera comme l'un des ouvrages scientifiques majeurs du dix-neuvième siècle.

La lecture du *Traité analytique* est toujours profitable ; la première phrase sonne comme une profession de foi : *Les causes primordiales ne nous sont point connues ; mais elles sont assujetties à des lois simples et constantes, que l'on peut découvrir par l'observation, et dont l'étude est l'objet de la philosophie naturelle.* Fourier confesse donc son ambition de décrire des phénomènes physiques jusque là jugés inaccessibles, par le moyen des mathématiques ; il le confirme dans le choix d'une devise en latin (attribuée par Fourier à Aristote, mais probablement inventée par lui-même) : *Et ignem regunt numeri* (Même le feu est régi par les nombres).

L'idée maîtresse de Fourier, celle que l'on lui attribue le plus naturellement, consiste à décomposer une fonction périodique arbitraire sous forme de *série trigonométrique* :

$$f(x) = \sum a_n \cos nx + b_n \sin nx.$$

Des décompositions de ce style ne sont pas à proprement parler nouvelles : elles faisaient partie du répertoire des analystes de la fin du dix-huitième siècle, à commencer par le plus impressionnant d'entre eux, le Suisse Euler. Mais Fourier est le premier à avoir l'intuition du caractère universel et particulièrement commode de cette décomposition trigonométrique. (Dans la suite, on utilisera souvent la notation complexe, et la formule $e^{i\alpha} = \cos \alpha + i \sin \alpha$.)

Fourier donne quelques exemples frappants pour montrer que “n'importe quelle” fonction est décomposable en série trigonométrique : ainsi, les deux fonctions dont le graphe est représenté ci-dessous : une fonction “en dents de scie” et une autre “en créneaux”.

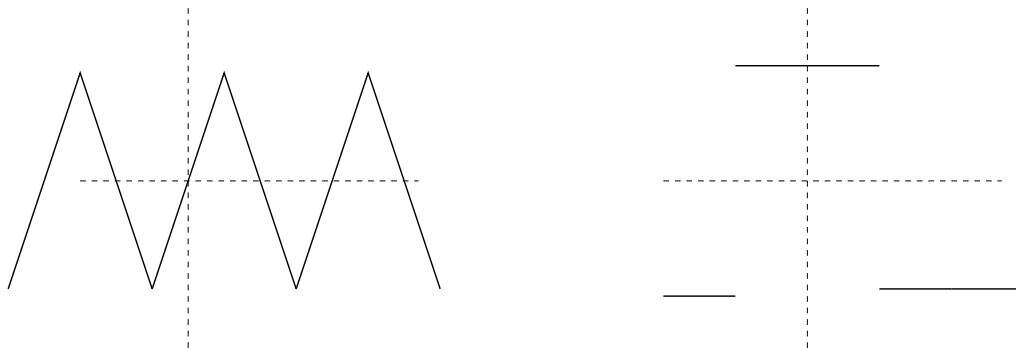


FIG. 1. Deux exemples de fonctions périodiques décomposables en séries trigonométriques

Pour la première, nous avons en effet la formule

$$\begin{aligned} \frac{\pi}{4} x &= \frac{\sin x}{1^2} - \frac{\sin 3x}{3^2} + \frac{\sin 5x}{5^2} - \frac{\sin 7x}{7^2} + \dots \\ &= \sum_{k \geq 0} (-1)^{2k+1} \frac{\sin(2k+1)x}{(2k+1)^2}, \end{aligned}$$

pour tout $x \in [-\pi, \pi]$. Pour la seconde fonction (qui n'est autre que la dérivée de la première),

$$\begin{aligned}\frac{\pi}{4} &= \frac{\cos x}{1} - \frac{\cos 3x}{3} + \frac{\cos 5x}{5} - \frac{\cos 7x}{7} + \dots \\ &= \sum_{k \geq 0} (-1)^{2k+1} \frac{\cos(2k+1)x}{2k+1},\end{aligned}$$

à nouveau pour tout $x \in [-\pi, \pi]$.

On voit ainsi que même des fonctions discontinues peuvent être exprimées en *séries de Fourier*, pourvu que l'on définisse

$$(95) \quad f(x) = \frac{1}{2}[f(x^+) + f(x^-)]$$

aux points de discontinuité (*condition de demi-somme*). Cela pouvait paraître surprenant à l'époque de Fourier, mais bien sûr ce n'est pas contradictoire avec les théorèmes connus : une série de fonctions continues n'est pas en général continue.

L'acte de foi suivant consiste à admettre que des décompositions similaires s'appliquent à des fonctions non périodiques (définies sur \mathbb{R} tout entier). Pour deviner la forme que doit prendre cette décomposition, considérons une fonction f un peu quelconque, et f_T sa périodisée de période $T > 0$ (on restreint f à $[0, T]$ et on étend le résultat par périodicité). Décomposant f_T en série de Fourier, on a

$$\forall x \in [0, T] \quad f(x) = \sum_{n \in \mathbb{Z}} c_{n,T} e^{2i\pi n \frac{x}{T}}.$$

Le coefficient $e^{2i\pi n x/T}$ dépend de n et T uniquement via le quotient n/T ; il n'est donc pas absurde de parier que les coefficients $c_{n,T}$ dépendront asymptotiquement de ce même quotient dans la limite $T \rightarrow \infty$. Au cours de ce processus, les coefficients n/T appartenant à un intervalle donné seront de plus en plus nombreux, et il est également naturel de remplacer la somme par sa version continue, l'intégrale. On arrive donc à deviner la forme suivante :

$$f(x) \simeq \sum_n c\left(\frac{n}{T}\right) e^{2i\pi x \frac{n}{T}} \simeq \int_{\xi} c(\xi) e^{2i\pi x \xi}.$$

C'est la forme des intégrales de Fourier, ou **transformées de Fourier**. Voici un exemple simple : soit h la fonction définie par

$$h(x) = \begin{cases} 0 & \text{si } |x| > 1 \\ -1-x & \text{si } -1 < x < 0 \\ 1-x & \text{si } 0 < x < 1 \end{cases} ;$$

alors

$$h(x) = \int (-2i) \left(\frac{1}{\xi} - \frac{\sin \xi}{\xi^2} \right) e^{i\xi x} dx.$$

Bien sûr, ces formules appellent instantanément quelques questions : comment trouver les coefficients ? (problème de l'**analyse**) En quel sens faut-il comprendre la convergence de ces séries ?

XI-1.2. Formules d'analyse. Soit f une fonction 1-périodique, considérée comme fonction sur le tore $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, et décomposée en série de Fourier : $f(x) = \sum c_n e^{2i\pi n x}$. Si l'on est autorisé à échanger somme et intégration, on a

$$\int_{\mathbb{T}} f(x) e^{2\pi i k x} dx = \sum_n c_n \left(\int e^{2i\pi(n+k)x} dx \right) = c_{-k}.$$

(En effet, le coefficient entre parenthèses vaut toujours 0, sauf si $n = -k$, auquel cas il vaut 1.)

On en déduit la formule d'inversion des séries de Fourier (déjà connue d'Euler) :

$$c_n = \int e^{-2i\pi n x} f(x) dx.$$

Par passage au continu, on devine également la formule d'inversion pour des fonctions définies sur \mathbb{R} entier (non périodiques) : si $f(x) = \int e^{2i\pi x \xi} g(\xi) d\xi$, alors

$$g(\xi) = \int e^{-2i\pi x \xi} f(x) dx.$$

Ces formules sont les formules de la **transformée de Fourier**.

Pour récapituler :

- La transformée de Fourier d'une fonction $f : \mathbb{T} \rightarrow \mathbb{C}$, Lebesgue-intégrable, est la suite de coefficients $(c_k(f))_{k \in \mathbb{Z}}$, définie par

$$c_k(f) = \int e^{-2i\pi k x} f(x) dx.$$

- La transformée de Fourier d'une fonction $f : \mathbb{R} \rightarrow \mathbb{C}$, Lebesgue-intégrable, est la fonction $\widehat{f} : \mathbb{R} \rightarrow \mathbb{C}$ définie par

$$\widehat{f}(\xi) = \int e^{-2i\pi x \xi} f(x) dx.$$

REMARQUE XI-1. Comme on l'a vu au Chapitre ??, ces formules se généralisent considérablement, le cadre naturel étant celui des groupes abéliens, (\mathbb{T} ou \mathbb{R} dans les exemples précédents). Ainsi on a une formule d'inversion de Fourier pour les fonctions définies sur le groupe (fini) des racines n -èmes de l'unité. **Mettre formules.**

REMARQUE XI-2. Les formules de la *synthèse* (ou inversion) sont très semblables à celles de l'analyse :

$$f(x) = \sum_{k \in \mathbb{Z}} c_k(f) e^{2i\pi k x}, \quad f(x) = \int e^{2i\pi x \xi} \widehat{f}(\xi) d\xi.$$

Cependant le signe dans l'exponentielle est différent !

REMARQUE XI-3. C'est un problème un peu subtil que de savoir quand les formules de synthèse sont valides. Mais on peut déjà noter que dans le processus d'analyse de Fourier, sous des hypothèses assez générales, *il n'y a pas de perte d'information*. Pour les fonctions périodiques c'est un théorème dû à Féjer : si f est périodique, continue par morceaux et vérifiant la condition de demi-somme (95), alors f est uniquement déterminée par les coefficients $c_k(f)$. De même, si $f : \mathbb{R} \rightarrow \mathbb{C}$ est Lebesgue-intégrable, continue par morceaux et satisfait la condition de demi-somme, alors f est uniquement déterminée par \widehat{f} .

REMARQUE XI-4. Il faut prendre bien garde à l'existence de différentes conventions, dont chacune a ses partisans acharnés. On peut mettre ou pas un coefficient 2π dans l'exponentielle complexe ; changer le signe dans cette même exponentielle ; ou encore mettre un facteur $1/(2\pi)$ ou $1/\sqrt{2\pi}$ devant l'exponentielle — soit douze possibilités, dont au moins la moitié sont effectivement utilisées. Quand les conventions changent, les formules changent aussi, et il ne faut pas se laisser piéger ! J'utiliserai moi-même dans la suite du cours plusieurs conventions différentes, selon les occasions.

XI-1.3. Théorie ponctuelle. La convergence des séries de Fourier est reconnue dès le début comme un problème important et délicat, sur lequel Fourier lui-même avait été attaqué. Le résultat qui fait référence en la matière est le **théorème de Dirichlet** (1829) :

THÉORÈME XI-5. *Soit $f : \mathbb{T} \rightarrow \mathbb{C}$, continue par morceaux, satisfaisant la condition de demi-somme, C^1 de part et d'autre de chaque point de discontinuité. Alors*

$$\forall x \in \mathbb{T} \quad \sum_{k \in \mathbb{Z}} c_k(f) e^{i2\pi kx} = f(x).$$

Faire la preuve. Une bonne référence est [Duoand.] p. 4-6

Jusqu'en 1876 les spécialistes pensaient que l'hypothèse de continuité C^1 de part et d'autre était superflue ; mais un contre-exemple de DuBois-Reymond prouva le contraire. Si l'on veut avoir la propriété de reconstruction ponctuelle, on doit supposer une certaine régularité ; tout au mieux on peut relaxer l'hypothèse C^1 . Ainsi, le critère de Jordan énonce que si f est à variation bornée au voisinage d'un point x , alors les sommes de Fourier partielles convergent vers $f(x)$.

Cependant, un résultat célèbre de Carleson (1966) énonce que la convergence des séries de Fourier d'une fonction continue $\mathbb{T} \rightarrow \mathbb{C}$ a lieu partout en-dehors d'un ensemble de mesure nulle. Cet énoncé est vrai aussi pour $f \in L^p(\mathbb{T})$ dès que $p > 1$; et si f est Riemann-intégrable. Et il est optimal au vu de résultats de Kahane et Katznelson, selon lesquels étant donné n'importe quel ensemble E de longueur nulle dans \mathbb{T} , on peut construire une fonction continue divergeant en tout point de E . Quant aux séries de Fourier des fonctions discontinues, elles peuvent très bien diverger *en tout point* de \mathbb{T} , comme le montre un autre contre-exemple dû à Kolmogorov (1926). On pourra consulter [Körner] ou [Kahane] pour plus d'information sur ces problèmes qui à une époque étaient au cœur des préoccupations des analystes.

Pour les intégrales de Fourier, on a un résultat simple et pratique :

THÉORÈME XI-6. *Soit $f \in C_b \cap L^1(\mathbb{R})$, alors*

$$\forall x \in \mathbb{R} \quad \int e^{2i\pi x\xi} \widehat{f}(\xi) d\xi = f(x).$$

Faire la preuve, ou renvoyer au bon passage dans le chapitre précédent ?

XI-1.4. Identités remarquables. La théorie de Fourier est basée sur trois identités remarquables :

Formule de convolution :

$$\widehat{f * g} = \widehat{f} \widehat{g}$$

Formule de conjugaison :

$$\int \widehat{f} g = \int f \widehat{\overline{g}}$$

Formule de dérivation :

$$\widehat{\nabla f} = -2i\pi \xi \widehat{f}$$

Ces formules nécessitent quelques précautions, et sont à réinterpréter éventuellement. Par exemple pour une fonction 1-périodique, la dernière formule deviendra $c_k(\partial f / \partial x_j) = -2i\pi k_j c_k(f)$.

Pour énoncer ces identités, le plus simple est, pour les deux premières, de se placer dans le cadre des *mesures*; et pour la troisième, des fonctions C^1 , de dérivée intégrable. Voici des énoncés précis :

..... **Que faire de l'échantillonnage ? fml de Poisson, fml de Shannon**

Voici un corollaire de la formule de conjugaison :

THÉORÈME XI-7. *Soit $(\mu_k)_{k \in \mathbb{N}}$ une famille de mesures signées sur \mathbb{R}^d , de variation totale bornée, et soit μ une autre mesure signée, telles que $\widehat{\mu}_k \rightarrow \widehat{\mu}$ simplement. Alors μ_k converge au sens faible-* vers μ .*

DÉMONSTRATION. Pour $\xi \in \mathbb{R}^d$, on pose $\varphi_n(\xi) = \int e^{i\xi \cdot x} \mu_k(dx)$, $\varphi(\xi) = \int e^{i\xi \cdot x} \mu(dx)$; l'hypothèse de variation totale sur les μ_k entraîne que φ_k est uniformément bornée. En outre, par hypothèse $\varphi_n(\xi)$ converge vers $\varphi(\xi)$ pour tout ξ quand $n \rightarrow \infty$.

Soit $\Psi \in C_c(\mathbb{R})$; on a alors la formule d'inversion $(\check{\Psi})^\wedge = \Psi$. (**Être plus clair sur les notations pour l'inversion ?**) De plus $\check{\Psi}$ est elle-même une fonction continue bornée et intégrable.

En appliquant la formule de conjugaison, on obtient

$$\int \Psi d\mu_k = \int \check{\Psi}(\xi) \varphi_k(\xi) d\xi.$$

Comme $\check{\Psi}$ est intégrable et φ_k bornée, on peut appliquer le théorème de convergence dominée :

$$\lim_{n \rightarrow \infty} \int \check{\Psi}(\xi) \varphi_k(\xi) d\xi = \int \check{\Psi}(\xi) \varphi(\xi) d\xi = \int \Psi d\mu,$$

la dernière égalité résultant d'une nouvelle application du théorème de conjugaison. Ceci conclut la preuve. \square

REMARQUE XI-8. Conditions suffisantes pour avoir la tension, exprimées en termes de la fonction caractéristique de probabilités.....

XI-1.5. Théorie L^2 . Même quand elle est vraie, la convergence ponctuelle des séries de Fourier peut être médiocre. Un exemple bien connu est le **phénomène de Gibbs**, qui pose des problèmes pratiques : la série de Fourier d'une fonction créneau présente des oscillations près des points de discontinuité, dont l'ampleur ne s'atténue pas quand le nombre de fréquences tend vers l'infini.

DESSIN

Au début du vingtième siècle, on aborde le problème sous un angle différent, plus abstrait et plus fonctionnel. Les théorèmes de Riesz–Fischer (1907) et de Plancherel (1910) montrent que l'analyse et la synthèse sont toujours possibles dans le cadre

hilbertien des espaces L^2 ; en outre ils aboutissent à la découverte fondamentale selon laquelle la transformée de Fourier est une **isométrie** $L^2 \rightarrow L^2$:

$$\|f\|_{L^2} = \|\widehat{f}\|_{L^2}$$

c'est l'**identité de Bessel–Parseval–Plancherel**, qui explique en partie la popularité extrême de l'analyse de Fourier dans un cadre hilbertien.

THÉORÈME XI-9 (Théorème de Riesz–Fischer). *Soit $f \in L^2(\mathbb{T})$, alors $f(x) = \sum c_k(f) e^{2i\pi kx}$ dans $L^2(\mathbb{T})$, au sens où*

$$\int_{\mathbb{T}} \left| f(x) - \sum_{|k| \leq N} c_k(f) e^{2i\pi kx} \right|^2 dx \xrightarrow{N \rightarrow \infty} 0.$$

En particulier, la série de Fourier de f converge presque partout vers f . En outre, pour toutes fonctions $f, g \in L^2(\mathbb{T})$ on a

$$\int_{\mathbb{T}} |f(x)|^2 dx = \sum_{k \in \mathbb{Z}} |c_k(f)|^2; \quad \int_{\mathbb{T}} f(x) \overline{g(x)} dx = \sum_{k \in \mathbb{Z}} c_k(f) \overline{c_k(g)}.$$

Ces énoncés se généralisent en plusieurs variables ($x \in \mathbb{T}^n$), avec d'évidentes modifications.

THÉORÈME XI-10 (Théorème de Plancherel). *Soit $f \in L^2(\mathbb{R})$, alors $f(x) = \int \widehat{f}(\xi) e^{2i\pi \xi x} d\xi$ dans $L^2(\mathbb{R})$, au sens où*

$$\int_{\mathbb{R}} \left| f(x) - \int_{-A}^A \widehat{f}(\xi) e^{2i\pi \xi x} d\xi \right|^2 dx \xrightarrow{A \rightarrow \infty} 0.$$

En particulier, l'intégrale de Fourier tronquée, $\int_{-A}^A \widehat{f}(\xi) e^{2i\pi \xi x} d\xi$, converge vers f presque partout. En outre, pour tous $f, g \in L^2(\mathbb{R}^n)$,

$$\int_{\mathbb{R}} |f(x)|^2 dx = \int_{\mathbb{R}} |\widehat{f}(\xi)|^2 d\xi; \quad \int_{\mathbb{R}} f(x) \overline{g(x)} dx = \int_{\mathbb{R}} \widehat{f}(\xi) \overline{\widehat{g}(\xi)} d\xi.$$

Ces énoncés se généralisent en plusieurs variables (pour des fonctions de $L^2(\mathbb{R}^n)$), avec d'évidentes modifications.

Faire les preuves

REMARQUE XI-11. Au cours de la preuve de l'identité de Parseval, on a obtenu un autre résultat fondamental : la somme partielle $S_N f$ est la meilleure approximation de f dans l'espace vectoriel engendré par les $(e^{2i\pi kx})_{|k| \leq N}$, au sens L^2 (approximation aux moindres carrés). De même, l'intégrale de Fourier de f tronquée aux fréquences $|\xi| \leq A$ est la meilleure approximation de f dans le sous-espace de $L^2(\mathbb{R})$ engendré par les fonctions $(e^{2i\pi \xi x})_{|\xi| \leq A}$ (**à expliquer car ces fonctions ne sont pas dans L^2**).

REMARQUE XI-12. On retrouve ici le concept, développé au Chapitre VII, des bases hilbertiennes.

REMARQUE XI-13. Parler de la convergence en norme L^p , $1 < p < \infty$, voir [Duoand. p.8] (Sur le tore seulement ?)

XI-1.6. Premières applications. Pour terminer cette section, nous allons considérer quelques applications simples et classiques des formules vues précédemment.

XI-1.6.1. *Solution formelle de l'équation de convolution.* Les opérations de convolution sont innombrables en ingénierie et traitement du signal :

$$g = K * f,$$

où f modélise le signal d'entrée, le noyau de convolution K représente l'effet d'une machine, et g est le signal de sortie. Le problème est de reconstituer le signal d'entrée, connaissant K . La solution pratique de ce problème peut souvent se faire au moyen de la transformée de Fourier :

$$\widehat{g} = \widehat{K} \widehat{f} \quad \implies \widehat{f} = \frac{\widehat{g}}{\widehat{K}},$$

donc f est donné par la transformée de Fourier inverse du quotient \widehat{g}/\widehat{K} .

XI-1.6.2. *L'inégalité isopérimétrique selon Hurwitz.* L'inégalité isopérimétrique traduit le fait que la boule minimise la surface, à volume fixé. Il en existe de nombreuses preuves, qui s'appliquent en plus ou moins grande généralité. En dimension 2, l'une des plus connues, due à Hurwitz (1901), est basée sur les séries de Fourier.

THÉORÈME XI-14 (Une version de l'inégalité isopérimétrique). *Soit Γ une courbe fermée du plan, de classe C^1 , sans auto-intersection, de longueur L , et soit A l'aire enserrée par Γ . Alors*

$$L^2 \geq 4\pi A,$$

avec égalité si et seulement si Γ est un cercle.

DÉMONSTRATION. On paramètre Γ par longueur d'arc ; cette courbe est donc représentée par une fonction $f : \mathbb{T} \rightarrow \mathbb{R}^2$, disons $f(t) = (x(t), y(t))$, de classe C^1 , avec

$$\sqrt{|\dot{x}|^2 + |\dot{y}|^2} = L$$

(le point désigne la dérivée par rapport au paramètre t) ; en particulier

$$\int (|\dot{x}(t)|^2 + |\dot{y}(t)|^2) dt = L^2.$$

Les coefficients de Fourier de \dot{x} sont $c_k(\dot{x}) = 2i\pi k c_k(x)$; en appliquant l'identité de Bessel–Parseval–Plancherel, et en notant $\widehat{x}(k) = c_k(x)$ pour simplifier, on obtient donc

$$4\pi^2 \sum_{k \in \mathbb{Z}} |k|^2 (|\widehat{x}(k)|^2 + |\widehat{y}(k)|^2) = L^2.$$

Par ailleurs, l'aire enserrée par la courbe Γ vaut, selon un avatar de la formule de Stokes,

$$\begin{aligned} A &= \frac{1}{2} \left| \int (\dot{x}y - x\dot{y}) dt \right| \\ &= \pi \sum_{k \in \mathbb{Z}} |k| (\widehat{x}(k) \overline{\widehat{y}(k)} + \overline{\widehat{x}(k)} \widehat{y}(k)) \\ &= 2\pi \sum_{k \in \mathbb{Z}} |k| |\Re(\widehat{x}(k) \widehat{y}(k))| \\ &\leq \pi \sum_{k \in \mathbb{Z}} |k| \left(\frac{|\widehat{x}(k)|^2 + |\widehat{y}(k)|^2}{2} \right), \end{aligned}$$

où la dernière inégalité résulte de Cauchy–Schwarz, sous la forme

$$\sum_k \alpha_k \beta_k \leq \left(\sum_k \alpha_k^2 \right)^{1/2} \left(\sum_k \beta_k^2 \right)^{1/2} \leq \frac{1}{2} \sum_k \alpha_k^2 + \frac{1}{2} \sum_k \beta_k^2;$$

ceci donne le résultat souhaité. On voit en outre qu'il y a égalité si et seulement si (a) $\widehat{x}(k) = \overline{\widehat{y}(k)}$ pour tout k , et (b) $|k| = k^2$ dans la somme, ce qui impose $k \in \{-1, 0, 1\}$, ce qui correspond à

$$x(t) = a \cos\left(\frac{L}{2\pi} t\right), \quad y(t) = a \sin\left(\frac{L}{2\pi} t\right),$$

soit l'équation d'un cercle. \square

EXERCICE XI-15. Soit Γ une courbe convexe lisse tracée dans le plan, de diamètre D et de longueur L , montrer que $L \geq 2D$, avec égalité pour le cercle... mais pas seulement ! (Cf. [Kahane, p.237]).

XI-1.6.3. *Le théorème d'équirépartition de Weyl.* Le théorème d'équirépartition de Weyl (1916) est un des premiers exemples de théorème ergodique. Plus que la théorie des séries de Fourier, il illustre l'intérêt d'approcher de l'approximation par des polynômes trigonométriques (second théorème d'approximation de Weierstrass). **(Remarque : faut-il les caser quelque part : théorème de Féjer ; thm d'approximation de Weierstrass ?)**

Exemple à rédiger

XI-2. Fourier calculateur : deux lois fondamentales des probabilités

Dans cette section on va voir la théorie de Fourier briller pour ses vertus calculatoires. On va l'utiliser en effet pour traiter par le calcul deux lois fondamentales des probabilités : le théorème de Pólya sur la récurrence de la marche aléatoire simple ; et le théorème central limite de Gauss. Un autre exemple intéressant, non abordé ici, est la multiplication rapide [Körner].

XI-2.1. La récurrence de la marche aléatoire. Un ivrogne se déplaçant "au hasard" va-t-il retrouver sa maison ? La version idéalisée de ce problème est un classique des probabilités. On considère donc une particule visitant les mailles d'un réseau régulier, disons \mathbb{Z}^d , où $d \in \mathbb{N}$.

Soient e_1, \dots, e_d les directions de la base (e_j est le vecteur dont toutes les composantes sont nulles, sauf la j -ème qui vaut 1) ; les directions autorisées sont les $(2d)$ vecteurs $\pm e_j$. On a donc des variables aléatoires $X_1, X_2, \dots, X_t, \dots$, toutes indépendantes et prenant la valeur $\pm e_j$ avec probabilité $1/(2d)$:

$$\forall t \in \mathbb{N}, \quad \mathbb{P}[X_t = \pm e_j] = \frac{1}{2d}.$$

La particule part de l'origine, sa position au temps t est donnée par

$$S_t = X_1 + \dots + X_t.$$

La question est de savoir si S_t va revenir à l'origine, ou bien s'éloigner pour ne jamais revenir. (Il n'est pas difficile de montrer que si S_t revient une infinité de fois près de l'origine, elle finira bien par retourner en l'origine exactement.)

C'est Pólya qui pose et résout ce problème, en 1921 : la solution dépend cruciallement de la dimension.

THÉORÈME XI-16 (Théorème de récurrence/transience de Polya). *Avec les notations précédentes,*

- Si $d \in \{1, 2\}$, alors avec probabilité 1, il existe une infinité de t tels que $S_t = 0$;
- Si $d \geq 3$, alors il y a une probabilité non nulle pour que S_t ne revienne jamais en 0 ; en outre, avec probabilité 1, $|S_t| \rightarrow \infty$ quand $t \rightarrow \infty$.

DÉMONSTRATION. Pour éviter toutes confusions dans les calculs, je rappelle que dans mes conventions, $\mathbb{N} = \{1, 2, 3, \dots\}$.

Soient p la probabilité de retour en 0, et T le temps de premier retour :

$$p = \mathbb{P}[\exists t > 0; S_t = 0]; \quad T = \inf\{t > 0; S_t = 0\}.$$

Clairement,

$$p = \mathbb{P}[T < \infty] = \sum_{k \in \mathbb{N}} \mathbb{P}[T = k].$$

Calculons la probabilité P_2 de revenir au moins deux fois en 0. Revenir au moins deux fois veut dire revenir au moins une fois, en un temps $T = k$, puis revenir en 0 en un temps $t > k$, sachant qu'on part de 0 au temps k . On a donc

$$\begin{aligned} P_2 &= \sum_{k \in \mathbb{N}} \mathbb{P}[T = k] \mathbb{P}[\exists t > k; S_t = 0 | T = k] \\ &= \sum_{k \in \mathbb{N}} \mathbb{P}[T = k] \mathbb{P}[\exists t > 0; X_{T+1} + \dots + X_t = 0 | T = k] \\ &= \sum_{k \in \mathbb{N}} \mathbb{P}[T = k] \mathbb{P}[\exists t > k; X_{k+1} + \dots + X_t = 0] \\ &= \sum_{k \in \mathbb{N}} \mathbb{P}[T = k] \mathbb{P}[\exists s > 0; X_1 + \dots + X_s = 0] \\ &= \sum_{k \in \mathbb{N}} \mathbb{P}[T = k] p = p^2. \end{aligned}$$

(On a utilisé le fait que l'événement $T = k$ est indépendant des variables X_{k+1}, X_{k+2}, \dots)
Un calcul similaire montrerait que la probabilité P_m de revenir au moins m fois en 0 est égale à p^m , pour n'importe quelle valeur de $m \in \mathbb{N}$.

Il devient alors facile de calculer la probabilité de revenir exactement k fois ($k \in \mathbb{N}$) en 0 :

$$p_k = p^k - p^{k+1} = p^k(1 - p).$$

En sommant ces probabilités, on obtient la probabilité de revenir un nombre fini de fois :

$$\mathbb{P}[\text{nombre fini de retours en 0}] = \sum_{k \in \mathbb{N}} p^k(1 - p) = \begin{cases} 0 & \text{si } p = 1 \\ 1 & \text{si } p < 1. \end{cases}$$

En particulier, dans le cas où $p = 1$, on voit que presque sûrement (avec probabilité 1), on revient une infinité de fois en 0 ; alors que si $p < 1$ cet événement est de probabilité nulle.

Dans ce dernier cas, on peut aller plus loin dans les conclusions : soit x un point fixé ; si on visite x un jour, la probabilité d'y revenir k fois est évidemment égale à p_k , et la probabilité d'y revenir une infinité de fois est également nulle. On en déduit que pour tout R fixé, la probabilité de revenir une infinité de fois dans la boule de

centre 0 et de centre R est nulle. Ceci revient à dire que S_t s'échappe à l'infini avec probabilité 1.

On peut également calculer l'espérance (la valeur moyenne) du nombre de retours N en 0 : si $p < 1$, c'est

$$\begin{aligned}\mathbb{E}N &= \sum_{k \in \mathbb{N}} k p^k (1-p) \\ &= \sum_{k \geq 1} k p^k - \sum_{k \geq 2} (k-1) p^k \\ &= p + \sum_{k \geq 2} p^k = \frac{p}{1-p};\end{aligned}$$

et cette formule reste valable pour $p = 1$ puisque dans ce cas, $\mathbb{E}N = \infty$.

Conclusion : tous les principaux paramètres du problème s'expriment en fonction de la probabilité p , et on a une dichotomie nette :

- si $p = 1$, alors presque sûrement on revient une infinité de fois au point de départ (comportement **récurrent**) ;
- si $p < 1$, alors presque sûrement on s'enfuit à l'infini (comportement **transient**).

Plutôt que de calculer p , on va calculer

$$\mathbb{E}N = \sum_{t \in \mathbb{N}} \mathbb{P}[X_1 + \dots + X_t = 0].$$

Pour tout $x \in \mathbb{Z}^d$ on pose

$$f(x) = \mathbb{P}[X_1 = x] = \begin{cases} \frac{1}{2d} & \text{si } x \in \{\pm e_j; 1 \leq j \leq d\} \\ 0 & \text{sinon.} \end{cases}$$

En d'autres termes, f est la distribution de probabilité (la **loi**) de X_1 (ou de n'importe quel X_t , $t \in \mathbb{N}$).

On définit alors

$$\begin{aligned}f_2(x) &= \mathbb{P}[X_1 + X_2 = x] = \sum_{y \in \mathbb{Z}^d} \mathbb{P}[X_1 = y \text{ et } X_2 = x - y] \\ &= \sum_{y \in \mathbb{Z}^d} \mathbb{P}[X_1 = y] \mathbb{P}[X_2 = y - x] = (f * f)(x).\end{aligned}$$

(Noter l'usage de l'indépendance de X_1 et X_2 !)

Un calcul similaire et un argument de récurrence montrent que

$$f_t(x) = \mathbb{P}[X_1 + X_2 + \dots + X_t = x] = (f_{t-1} * f)(x) = f^{*t}(x)$$

(produit de convolution appliqué t fois).

Ce que nous venons de redécouvrir est en fait un outil fondamental des probabilités : *la loi de la somme de variables aléatoires indépendantes est donnée par la convolution des lois*. Comme on sait que convolution et transformée de Fourier se marient bien, la suite est facile à deviner. Posons

$$\varphi(\xi) = \sum_{x \in \mathbb{Z}^d} f(x) e^{2i\pi x \cdot \xi}$$

(au signe près de i , c'est la transformée de Fourier de la loi f , que les probabilistes appellent souvent **fonction caractéristique** de la variable X_1). Ici $\xi \in \mathbb{T}^d$ et $\xi \cdot x$ est un produit scalaire dans \mathbb{R}^d . On peut calculer φ explicitement :

$$\varphi(\xi) = \frac{1}{2d} \sum_{1 \leq j \leq d} (e^{2\pi i \xi_j} + e^{-2\pi i \xi_j}) = \frac{1}{d} \sum_{1 \leq j \leq d} \cos(2\pi \xi_j),$$

où ξ_j est la j -ème composante du vecteur ξ . Noter que φ est à valeurs dans $[-1, 1]$, et $\varphi(0) = 1$.

Si l'on remplace f par f_t dans la définition de φ , on obtient alors

$$\begin{aligned} \varphi_t(\xi) &= \sum_{x \in \mathbb{Z}^d} f_t(x) e^{2i\pi x \cdot \xi} \\ &= \sum_{x \in \mathbb{Z}^d} \sum_{x_1 + \dots + x_t = x} f(x_1) \dots f(x_t) e^{2i\pi \xi_1 \cdot x} \dots e^{2i\pi \xi_t \cdot x} \\ &= \left(\sum_{x \in \mathbb{Z}^d} f(x) e^{2i\pi \xi \cdot x} \right)^t = \varphi(\xi)^t, \end{aligned}$$

où chaque x_j est une variable à valeurs dans \mathbb{R}^d . (On aurait pu s'y attendre : la transformée de Fourier transforme la convolution en produit, donc la convolution itérée en puissance.)

Pour obtenir $f_t(0)$, il suffit d'appliquer la formule d'inversion de Fourier :

$$f_t(0) = \int_{\mathbb{T}^d} \varphi_t(\xi) d\xi = \int_{\mathbb{T}^d} \varphi(\xi)^t d\xi.$$

On en déduit

$$\begin{aligned} \mathbb{E}N &= \sum_{t \in \mathbb{N}} f_t(0) = \sum_{t \in \mathbb{N}} \int_{\mathbb{T}^d} \varphi(\xi)^t d\xi \\ &= \lim_{\alpha \rightarrow 1^-} \sum_{t \in \mathbb{N}} \alpha^t \int_{\mathbb{T}^d} \varphi(\xi)^t d\xi \\ &= \lim_{\alpha \rightarrow 1^-} \sum_{t \in \mathbb{N}} \int_{\mathbb{T}^d} (\alpha \varphi(\xi))^t d\xi \\ &= \lim_{\alpha \rightarrow 1^-} \int_{\mathbb{T}^d} \left(\sum_{t \in \mathbb{N}} (\alpha \varphi(\xi))^t \right) d\xi, \end{aligned}$$

où l'interversion série-intégrale est justifié par le théorème de convergence dominée (domination par la série convergence $\sum \alpha^t$, $\alpha < 1$). D'où

$$\mathbb{E}N = \lim_{\alpha \rightarrow 1^-} \int_{\mathbb{T}^d} \frac{\alpha \varphi(\xi)}{1 - \alpha \varphi(\xi)} d\xi = \int_{\mathbb{T}^d} \frac{\varphi(\xi)}{1 - \varphi(\xi)} d\xi,$$

où cette fois l'échange limite-intégrale se justifie par convergence monotone, appliquée séparément sur les ensembles $\{\varphi < 0\}$ et $\{\varphi \geq 0\}$.

Pour savoir si on est dans le cas transient ou dans le cas récurrent, il suffit de vérifier si l'intégrale $\int \varphi/(1 - \varphi)$ converge ou diverge. Pour ce faire, on utilise le développement limité $\cos t = 1 - t^2/2 + O(t^4)$, ou encore l'inégalité élémentaire

$$\frac{1}{2} \left(\frac{2\pi^2}{d} |\xi|^2 \right) \leq 1 - \varphi(\xi) \leq \frac{2\pi^2}{d} |\xi|^2,$$

qui montre que la convergence ou la divergence de $\int \varphi/(1-\varphi)$ est équivalente à la convergence ou à la divergence de

$$\int_{[0,1]^d} \frac{d\xi}{|\xi|^2} \begin{cases} = +\infty & \text{si } d \in \{1, 2\} \\ < +\infty & \text{si } d \geq 3. \end{cases}$$

Ceci conclut la preuve du théorème. Dans le cas transient, on peut même calculer “explicitement” la probabilité p : sachant que

$$\frac{p}{1-p} = \mathbb{E}N = \int_{\mathbb{T}^d} \frac{\varphi}{1-\varphi},$$

et remplaçant φ par son expression, on obtient

$$p = 1 - \frac{1}{\int_{\mathbb{T}^d} \frac{1}{1-\varphi}} = 1 - \frac{1}{\int_{[0,1]^d} \left(\frac{1}{1 - \frac{1}{d} \sum_{j=1}^d \cos(2\pi\xi_j)} \right) d\xi}.$$

□

REMARQUE XI-17. En dimension 1, il est assez simple de calculer la probabilité de retour en 0 au temps $2n$:

$$\mathbb{P}[S_{2n} = 0] = \frac{1}{2^{2n}} C_{2n}^n = 4^{-n} \frac{(2n)!}{(n!)^2}$$

(noter qu’il s’agit de choisir n pas vers la gauche et n pas vers la droite). Un calcul un peu plus raffiné montre que la probabilité de revenir en 0 pour la première fois au temps $2n$ est exactement

$$\frac{1}{n} 4^{-n} C_{2n}^n.$$

Par la formule de Stirling, la probabilité de premier retour en $2n$ vaut environ

$$\frac{1}{\sqrt{\pi} n^{3/2}}.$$

En particulier, l’espérance du temps de premier retour diverge comme $\sum 1/\sqrt{\pi n}$. Conclusion : même en dimension 1 où la marche aléatoire est “très récurrente”, l’espérance du temps de premier retour est infinie !

REMARQUE XI-18. En dimension 2, on peut encore calculer simplement la probabilité de retour en 0 au temps $2n$: c’est

$$\mathbb{P}[S_{2n} = (0, 0)] = \frac{8^{-n} [(2n)!]^2}{(n!)^4};$$

en appliquant Stirling, on voit que cette somme est logarithmiquement divergente, ce qui est une façon de retrouver le résultat de Pólya par des moyens combinatoires. En dimension 3, le calcul exact ne semble plus possible, mais on peut encore majorer

$$\mathbb{P}[S_{2n} = (0, 0, 0)] \leq \frac{1}{2^{2n}} C_{2n}^n \left(\frac{1}{3^n} \frac{n!}{\left(\left[\frac{n}{3}\right]!\right)^3} \right),$$

et cette série converge comme $\sum n^{-3/2}$, ce qui redonne la transience. On pourra consulter [Doyle–Snell] pour plus d’informations. Cette référence contient également d’autres méthodes plus robustes, basées sur des analogies avec les circuits électriques !

On peut ainsi montrer que la marche aléatoire sur un sous-réseau de \mathbb{Z}^d ($d \geq 3$) est encore transiente (ce qui est intuitif mais non trivial).

XI-2.2. Le théorème central limite. Le théorème central limite est (avec peut-être la loi des grands nombres) le théorème le plus célèbre des probabilités : il énonce que *les fluctuations d'une moyenne de variables aléatoires indépendantes de variance finie sont gaussiennes*. L'effet magique de ce théorème magnifique a été quelque peu gâché par son succès phénoménal en statistiques et en sciences naturelles, où il est employé à tous propos, et parfois hors de son domaine d'application. Notons que la signification de “théorème central limite” est sujette à caution, certains auteurs préférant parler de “limite centrale” ou “limite centrée”.

On notera $\gamma = \mathcal{N}(0, 1)$ la loi normale (i.e. gaussienne) centrée (i.e. d'espérance nulle) réduite (i.e. de variance unité) ; en d'autres termes la densité de probabilité définie par

$$\gamma(t) = \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}.$$

THÉORÈME XI-19. *Soit $(X_k)_{k \in \mathbb{N}}$ une suite de variables aléatoires réelles, indépendantes et identiquement distribuées, telles que*

$$\mathbb{E}X_1^2 < +\infty;$$

on pose

$$\sigma^2 := \mathbb{E}[X_1^2 - (\mathbb{E}X_1)^2].$$

Alors

$$Z_N := \sqrt{\frac{N}{\sigma^2}} \left(\frac{1}{N} \sum_{j=1}^N X_j - \mathbb{E}X_j \right) \xrightarrow[N \rightarrow \infty]{Z} \simeq \gamma,$$

au sens où, pour tous $a, b \in \mathbb{R}$,

$$\mathbb{P}[a \leq Z_N \leq b] \xrightarrow[N \rightarrow \infty]{} \int_a^b \gamma(t) dt.$$

REMARQUE XI-20. En utilisant le fait que γ est une mesure de probabilité, on peut renforcer la convergence ci-dessus en convergence faible : pour toute fonction Ψ continue bornée on a

$$\mathbb{E}\Psi(Z_N) \xrightarrow[N \rightarrow \infty]{} \mathbb{E}\Psi(Z).$$

Démontrer cette propriété quelque part ??

REMARQUE XI-21. Il existe une version multidimensionnelle de ce théorème, dans laquelle on remplace le nombre réel σ^2 par la matrice de covariance

$$\Sigma_{ij} = \mathbb{E}[X_1^i X_1^j - (\mathbb{E}X_1)^i (\mathbb{E}X_1)^j],$$

X_1^j désignant la j -ème composante du vecteur $X_1 \in \mathbb{R}^d$; et la loi normale par

$$\gamma^{(d)}(t) = \frac{e^{-\frac{|t|^2}{2}}}{(2\pi)^{d/2}}.$$

La preuve est très similaire à celle du cas monodimensionnel.

REMARQUE XI-22. La preuve qui suit est la preuve qui s'est imposée comme la plus simple et la plus élégante. On peut cependant la trouver insatisfaisante au sens où la loi gaussienne y apparaît "comme par miracle", sans qu'on y comprenne rien. D'autres preuves sans doute plus profondes, mais nettement plus délicates, font intervenir par exemple l'information de Kullback [Cover-Thomas]. Un résultat de recherche assez récent, dû à Ball, Barthe et Naor, montre que l'entropie de Boltzmann de la variable aléatoire Z_N est croissante en N ; l'apparition de la gaussienne devient alors toute naturelle, comme la densité centrée réduite d'entropie maximale.

PREUVE DU THÉORÈME XI-19. Quitte à remplacer X_j par $(X_j - \mathbb{E}X_j)/\sigma$, on peut supposer que les X_j sont centrées réduites, i.e. $\mathbb{E}X_j = 0$ et $\sigma = 1$.

Soit φ la fonction caractéristique (transformée de Fourier) de la variable X (ou plutôt de sa loi) :

$$\varphi(\xi) = \mathbb{E}e^{i\xi X_1} = \int e^{i\xi t} \mu(dt),$$

où μ est la loi commune des X_j (c'est la mesure image par X_j de la mesure de probabilité \mathbb{P} ; cette dernière est une mesure sur l'espace de probabilités, alors que μ est une mesure sur \mathbb{R} , qui contient toute l'information sur la distribution des valeurs X_j).

De même, on définit φ_n comme la fonction caractéristique de Z_n :

$$\varphi_n(\xi) = \mathbb{E}e^{i\xi \left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right)} = \left(\mathbb{E}e^{i\xi \frac{X_1}{\sqrt{n}}}\right)^n = \varphi\left(\frac{\xi}{\sqrt{n}}\right)^n$$

(ici encore, l'intérêt de la transformée de Fourier est de changer "magiquement" les convolutions en produits).

En utilisant le théorème de convergence dominée, on vérifie facilement que $\varphi \in C^2(\mathbb{R})$, et

$$\varphi'(\xi) = i \int_{\mathbb{R}} t e^{i\xi t} \mu(dt); \quad \varphi''(\xi) = - \int_{\mathbb{R}} t^2 e^{i\xi t} \mu(dt);$$

en particulier

$$\varphi(0) = 1, \quad \varphi'(0) = 0, \quad \varphi''(0) = -1;$$

(vérifier les signes) d'où

$$\varphi(\xi) = 1 - \frac{\xi^2}{2} + o(\xi^2),$$

quand $\xi \rightarrow 0$. En particulier, $\log \varphi(\xi/\sqrt{n}) = -\frac{\xi^2}{2n} + o(\xi^2/n)$ quand $n \rightarrow \infty$; en multipliant par n et en prenant l'exponentielle, on obtient

$$\varphi\left(\frac{\xi}{\sqrt{n}}\right)^n = e^{-\frac{\xi^2}{2}} + o(1) \quad (n \rightarrow \infty).$$

Conclusion :

$$\forall \xi \in \mathbb{R}, \quad \varphi_n(\xi) \xrightarrow{n \rightarrow \infty} e^{-\frac{\xi^2}{2}}.$$

On rappelle que le membre de droite est la transformée de Fourier de γ :

$$\int e^{i\xi t} \gamma(t) dt = e^{-\frac{\xi^2}{2}};$$

Cf. Théorème..... **Faire un tel théorème, et insister sur les conventions.**

Le Théorème XI-7 montre que pour toute fonction $\varphi \in C_c(\mathbb{R})$,

$$\int \varphi d\mu_n \xrightarrow{n \rightarrow \infty} \int \varphi d\gamma,$$

où μ_n est la loi de Z_n et γ désigne la mesure gaussienne par abus de notation.

Soit $f = 1_{[a,b]}$; pour tout $\varepsilon > 0$ on peut trouver (exercice) des fonctions g_ε^+ , g_ε^- continues à support compact, telles que

$$g_\varepsilon^- \leq f \leq g_\varepsilon^+,$$

et g_ε^\pm coïncide avec f sauf sur un ensemble de longueur 2ε . On a alors

$$\begin{aligned} \int g_\varepsilon^- d\gamma &= \lim_{n \rightarrow \infty} \int g_\varepsilon^- d\mu_n \leq \liminf_{n \rightarrow \infty} \int f d\mu_n \\ &\leq \limsup_{n \rightarrow \infty} \int f d\mu_n \leq \limsup_{n \rightarrow \infty} \int g_\varepsilon^+ d\mu_n = \int g_\varepsilon^+ d\gamma. \end{aligned}$$

On conclut en faisant tendre ε vers 0, et en notant que, la probabilité γ n'ayant pas d'atome,

$$\lim_{\varepsilon \rightarrow 0} \int g_\varepsilon^\pm d\gamma = \int f d\gamma = \int_a^b \gamma(t) dt.$$

□

REMARQUE XI-23. Le théorème central limite conclut à la **convergence faible** des lois. La question de la convergence forte des densités de ces lois est beaucoup plus subtile; une réponse est apportée par le **théorème central limite local**, qui suppose Voir [Feller].

REMARQUE XI-24. Le théorème central limite est un théorème asymptotique, qui ne donne aucune information quantitative sur la convergence vers la loi gaussienne. La question d'obtenir de telles estimations est un problème célèbre en probabilités, auquel le **théorème de Berry–Esséen** apporte une réponse partielle : Voir [Feller].

Pour montrer la puissance heuristique du théorème central limite, nous allons maintenant voir comment l'utiliser pour deviner le Théorème XI-16 (transience/réurrence de la marche aléatoire). Reprenons les notations de la Sous-section XI-2.1; comme S_t est une somme de variables aléatoires indépendantes de variance finie, on s'attend à ce que

$$\mathbb{P}[S_t = 0] \simeq \frac{1}{t^{d/2}},$$

le membre de droite étant la valeur de la gaussienne en l'origine. (Exercice : Pourquoi n'est-ce pas rigoureux??) On s'attend donc à ce que l'espérance du nombre de retours en 0 soit

$$\mathbb{E} \sum_{t \in \mathbb{N}} 1_{S_t=0} = \sum_{t \in \mathbb{N}} \mathbb{E} 1_{S_t=0} = \sum_{t \in \mathbb{N}} \mathbb{P}[S_t = 0] \simeq \sum_{t \in \mathbb{N}} \frac{1}{t^{d/2}},$$

et cette somme diverge si $d \in \{1, 2\}$, converge si $d \geq 3$.

XI-3. Fourier physicien : l'Âge de la Terre

Le but premier de Fourier était de mettre au point des méthodes permettant d'étudier l'évolution de la chaleur dans les corps conducteurs, via certaines *équations aux dérivées partielles*. Dans cette section, nous allons présenter cette problématique à travers un exemple historiquement célèbre : l'estimation de l'âge de la Terre, une donnée vitale pour la géologie et la biologie.

Fourier n'est pas le premier à étudier l'âge de la Terre ; il avait au moins deux prédécesseurs célèbres. Tout d'abord, dans les années 1650, Usher, après des années à étudier la Bible, avait fixé à 4004 av J-C la date de la Création (*Annales de l'Ancien Testament, retracées depuis l'origine du Monde*). Ensuite, à la fin du dix-huitième siècle, Buffon (*Des époques de la Nature*, 1779) avait tenté de répondre à la même question en reprenant la conception de Descartes et Leibniz selon laquelle la Terre a un centre “en fusion” et est née “comme un Soleil” (ces croyances étaient basées en partie sur des principes philosophiques, en partie sur l'expérience qui montre que la température augmente quand on s'enfonce profondément dans le sol, en moyenne 3 degrés par 100 mètres). Buffon fait des expériences dans les forges de Montbard sur le temps de refroidissement de globes de fer portés à température de fusion ; après avoir appliqué des règles de proportionnalité, pris en compte l'échauffement parasite du Soleil, les variations de densité et bien d'autres paramètres, il aboutit à une estimation d'environ 75000 ans.

XI-3.1. La formule de Fourier. En 1820 Fourier modélise le problème au moyen de l'équation aux dérivées partielles qu'il a mise au point récemment, l'équation de la chaleur,

$$\frac{\partial \theta}{\partial t} = \kappa \Delta \theta,$$

où $\theta(t, x)$ est la température au temps t et au point x , κ est une constante déterminée par les paramètres physiques, et Δ est le Laplacien par rapport à la variable x , qui varie à l'intérieur d'une boule de rayon R (le rayon terrestre). Fourier suppose lui aussi que la Terre était initialement en fusion et de température uniforme :

$$\forall x \in B(0, R), \quad \theta(0, x) = \theta_0.$$

Il suppose en outre que la température à la surface est constante (fixée par les conditions extérieures) :

$$|x| = R \implies \theta(t, x) = \theta_e.$$

Des raisonnements basés sur les ordres de grandeur montrent que le refroidissement en profondeur est relativement faible, de sorte que l'essentiel du phénomène se produit près de la surface. On décide alors de négliger la courbure de la Terre, et de la modéliser non plus comme une boule, mais comme un demi-espace. Par symétrie, on peut se limiter à deux variables : $t \geq 0$ (temps), $y \geq 0$ (profondeur). Les équations deviennent

$$\begin{cases} \frac{\partial \theta}{\partial t} = \kappa \frac{\partial^2 \theta}{\partial y^2} \\ \theta(0, y) = \theta_0, \quad \forall y > 0 \\ \theta(t, 0) = \theta_e, \quad \forall t > 0. \end{cases}$$

Le programme est de résoudre cette équation, et d'en déduire le *gradient de θ en surface*, qui est accessible à l'expérience (on mesure l'augmentation de température avec la profondeur).

THÉORÈME XI-25. *Soient $\theta_0, \theta_e \geq 0$ et $\kappa > 0$; alors il existe une unique fonction $\theta = \theta(t, y)$ définie sur $\mathbb{R}_+ \times \mathbb{R}_+$ telle que*

(a) *sur $]0, +\infty[\times]0, +\infty[$, θ est continûment dérivable en t , deux fois continûment dérivable en x , et satisfait l'équation*

$$\frac{\partial \theta}{\partial t} = \kappa \frac{\partial^2 \theta}{\partial x^2};$$

(b) *θ est bornée, $\lim_{t \rightarrow 0} \theta(t, y) = \theta_0$ pour tout $y > 0$, $\lim_{y \rightarrow 0} \theta(t, y) = \theta_e$ pour tout $t > 0$.*

Pour $t > 0$, cette solution est donnée par la formule

$$\theta(t, y) = \frac{\theta_0}{\sqrt{4\pi\kappa t}} \int_0^{+\infty} e^{-\frac{|y-x|^2}{4\kappa t}} dx + \frac{2(\theta_0 - \theta_e)}{\sqrt{4\pi\kappa t}} \int_0^{+\infty} e^{-\frac{(y+x)^2}{4\kappa t}} dx;$$

et en particulier,

$$\lim_{y \rightarrow 0} \frac{\partial \theta}{\partial y}(t, y) = \frac{\theta_0 - \theta_e}{\sqrt{\pi\kappa t}}.$$

La preuve complète du Théorème XI-25 nécessite un minimum de familiarité avec les équations aux dérivées partielles, et dépasse le cadre de ce cours. Je me limiterai donc à une esquisse de preuve. **Faire une preuve complète ?**

ESQUISSE DE PREUVE DU THÉORÈME XI-25. On va utiliser un argument de symétrie pour ramener le problème posé dans le quadrant ($t > 0, y > 0$) à un problème posé dans le demi-espace entier ($t > 0, y \in \mathbb{R}$). Considérons en effet le problème auxiliaire

$$(96) \quad \begin{cases} \frac{\partial \theta}{\partial t} = \kappa \frac{\partial^2 \theta}{\partial y^2} \\ \theta(0, y) = \theta_0, & \forall y > 0 \\ \theta(0, y) = \theta_1, & \forall y < 0, \end{cases}$$

où θ_1 est choisi de telle sorte que

$$\frac{\theta_0 + \theta_1}{2} = \theta_e,$$

c'est-à-dire

$$\theta_1 = 2(\theta_0 - \theta_e).$$

La donnée initiale vérifie bien évidemment

$$\frac{\theta(0, y) + \theta(0, -y)}{2} = \theta_e$$

(pour tous y si on pose $\theta(0, 0) = \theta_e$), et par symétrie ceci sera préservé pour les temps positifs. (En effet, la fonction $(\theta(t, y) + \theta(t, -y))/2$ est solution de l'équation de la chaleur avec donnée initiale constante égale à θ_e , et elle reste donc constante pour tous temps; pour rendre cet argument rigoureux il faudrait un théorème adéquat d'unicité.) La figure ci-dessous précise l'allure qualitative de θ pour différentes valeurs de t .

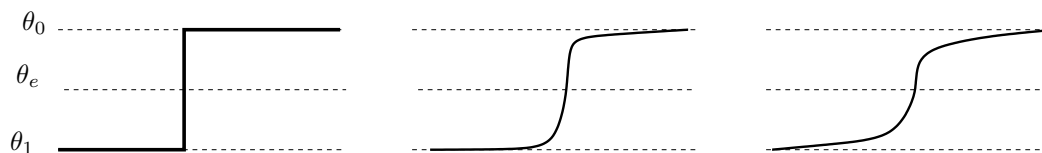


FIG. 2. Allure qualitative du graphe de $\theta(t, \cdot)$ pour trois valeurs successives de t ($t = 0$ à gauche, $t = t_1 > 0$ au milieu, $t = t_2 > t_1$ à droite)

La solution θ ainsi construite vérifiera donc, si on la restreint au quadrant ($t > 0, y > 0$), les solutions aux limites requises.

Pour résoudre (96), on applique la transformée de Fourier dans la variable y , que l'on notera avec un chapeau : cette opération commute avec la dérivation en temps, et transforme le laplacien en multiplication par $-4\pi^2\xi^2$. L'équation $\partial\theta/\partial t = \kappa\Delta\theta$ devient donc

$$\frac{\partial\hat{\theta}}{\partial t}(t, \xi) = -4\pi^2\kappa\xi^2\hat{\theta}(t, \xi) \quad (t > 0, \xi \in \mathbb{R}).$$

Il s'agit d'une équation différentielle ordinaire linéaire, *fréquence par fréquence*, que l'on peut intégrer exactement :

$$\hat{\theta}(t, \xi) = \hat{\theta}(0, \xi) e^{-4\pi^2\kappa t \xi^2}.$$

On résout cette équation en appliquant la transformée de Fourier inverse, qui change la multiplication en produit de convolution. Se souvenant que la transformée de Fourier inverse d'une gaussienne est une gaussienne, on obtient, tous calculs faits,

$$\theta(t, \cdot) = \theta(0, \cdot) * \frac{e^{-\frac{|\cdot|^2}{4\kappa t}}}{\sqrt{4\pi\kappa t}};$$

ou, plus explicitement,

$$\begin{aligned} \theta(t, y) &= \int_{-\infty}^{+\infty} \theta(0, x) \frac{e^{-\frac{(x-y)^2}{4\kappa t}}}{\sqrt{4\pi\kappa t}} dx \\ &= \int_{-\infty}^{+\infty} (\theta_0 1_{x>0} + \theta_1 1_{x<0}) \frac{e^{-\frac{(x-y)^2}{4\kappa t}}}{\sqrt{4\pi\kappa t}} dx \\ &= \frac{\theta_0}{\sqrt{4\pi\kappa t}} \int_0^{+\infty} e^{-\frac{(y-x)^2}{4\kappa t}} dx + \frac{\theta_1}{\sqrt{4\pi\kappa t}} \int_0^{+\infty} e^{-\frac{(y+x)^2}{4\kappa t}} dx, \end{aligned}$$

où la dernière identité résulte du changement de variable $x \rightarrow -x$ dans la seconde intégrale. Compte tenu de l'expression de θ_1 , ceci est le résultat annoncé dans le théorème.

Pour obtenir la dérivée partielle par rapport à y , on dérive l'expression sous le signe intégrale, ce qui est assez facile à justifier par convergence dominée : pour tout $t > 0$,

$$\frac{\partial\theta}{\partial y}(t, y) = -\frac{\theta_0}{\sqrt{4\pi\kappa t}} \int_0^{+\infty} \frac{(y-x)}{2\kappa t} e^{-\frac{(y-x)^2}{4\kappa t}} dx - \frac{\theta_1}{\sqrt{4\pi\kappa t}} \int_0^{+\infty} \frac{(y+x)}{2\kappa t} e^{-\frac{(y+x)^2}{4\kappa t}} dx;$$

en prenant la limite $y \rightarrow 0$ on obtient

$$\begin{aligned}\frac{\partial \theta}{\partial y}(t, 0) &= \frac{1}{\sqrt{4\pi\kappa t}} \left(\theta_0 \int_0^{+\infty} \frac{x}{2\kappa t} e^{-\frac{x^2}{4\kappa t}} dx - \theta_1 \int_0^{+\infty} \frac{x}{2\kappa t} e^{-\frac{x^2}{4\kappa t}} dx \right) \\ &= \frac{\theta_0 - \theta_1}{\sqrt{4\pi\kappa t}} \int_0^{+\infty} \frac{x}{2\kappa t} e^{-\frac{x^2}{4\kappa t}} dx \\ &= \frac{\theta_0 - \theta_1}{\sqrt{4\pi\kappa t}};\end{aligned}$$

ce qui est bien l'expression souhaitée. \square

XI-3.2. Controverse et épilogue. Résumons : Au moyen de son modèle, Fourier a obtenu la formule

$$\Gamma = \frac{\theta_0 - \theta_e}{\sqrt{\pi\kappa t}},$$

où Γ désigne le gradient de température en surface au temps t , t est le temps écoulé depuis les conditions initiales (que nous espérons être une bonne approximation de l'âge de la Terre), θ_0 est la température initiale (disons la température de fusion du fer, qui constitue une bonne part de la Terre), θ_e la température à la surface de la terre, et κ une constante physique (disons la conductivité du fer). On en déduit l'âge de la Terre, T , en fonction de $\theta_0, \theta_e, \kappa$ et Γ :

$$(97) \quad T = \frac{(\theta_0 - \theta_e)^2}{\pi\Gamma^2\kappa}.$$

L'élégance de cette formule, qui intègre harmonieusement tous les paramètres du problème, est peut-être l'une des raisons du grand crédit qui lui a été accordée.

C'est à peine croyable, mais Fourier n'effectue pas l'application numérique. Peut-être n'avait-il pas assez confiance dans les données fournies par la physique de son époque. Quarante ans après Fourier, Thomson, Lord Kelvin, considéré comme le meilleur physicien de son temps, reprend les calculs de Fourier, améliore l'argumentation et réalise l'application numérique. Le problème de l'âge de la Terre allait obséder Kelvin pendant plusieurs décennies, et le mettre au centre d'une controverse transcendant les disciplines — au point de faire oublier la contribution de Fourier, tout le monde ou presque étant maintenant persuadé que la formule (97) est due à Kelvin.

Utilisant les meilleurs valeurs connues à son époque, Kelvin obtient pour la Terre un âge de l'ordre de 100 millions d'années. Même en tenant compte des marges d'incertitude, il estime très improbable que la Terre soit âgée de plus de 400 millions d'années ; et il complète son calcul par d'autres arguments indépendants, mettant en jeu le bilan thermodynamique du Soleil, le ralentissement de la rotation terrestre dû aux marées, le temps d'agrégation des galaxies, etc. On pourra consulter [Thomson–Tait] pour certains de ces calculs ; noter le ton polémique de l'auteur.

Polémique car, et c'est là que le bât blesse, dans les années 1860 ces valeurs sont incompatibles avec les théories géologiques dites gradualistes (Lyell, Hutton) qui ont le vent en poupe ; et bien sûr avec la théorie de la sélection naturelle de Darwin et Wallace. Darwin estime que sa théorie ne peut tenir debout dans une échelle de temps aussi réduite ; entre la première édition de son *Origine des espèces* (1859) et la dernière édition (1872), il insérera des modifications prudentes, demandant une suspension de jugement jusqu'à ce que les connaissances en la matière soient plus sûres. Kelvin quant à lui semble avoir pensé être en possession d'une preuve de

l'existence de Dieu, ou peu s'en faut (si les échelles de temps sont trop courtes, c'est que la sélection naturelle ne peut pas être la seule responsable de l'évolution).

Les géologues essayèrent de revoir leurs chronologies pour les adapter au nouveau cadre que leur proposait Kelvin ; mais les choses empirèrent au cours des années, quand ce dernier révisa ses estimations à la baisse, jusqu'au chiffre incroyable de 24 millions d'année!! En 1893, l'ensemble du monde physicien accepte ces estimations, mais les géologues et évolutionnistes ne peuvent s'y résoudre, provoquant une crise interscientifique majeure.

Bien entendu, les calculs de Kelvin dépendaient de la validité du modèle ; en particulier de l'hypothèse de l'absence d'un mécanisme interne de production de la chaleur. En 1904, Rutherford annonce la découverte d'un tel mécanisme, la radioactivité ! En même temps, il remettait en question les conclusions de Kelvin. Le récit plaisant, fait par Rutherford lui-même, de la séance de l'Académie des Sciences à laquelle il a mené Kelvin à admettre son point de vue avec le sourire, est resté à la postérité... engendrant une nouvelle idée fausse particulièrement tenace ! Dans la culture scientifique collective, tout le monde ou presque "sait" maintenant que la radioactivité explique les estimations trop serrées de Kelvin ; et du reste, la datation par radioactivité donne des valeurs beaucoup plus fiables, et beaucoup plus grandes (de l'ordre de 4,6 milliards d'années, bien plus que les géologues et évolutionnistes n'en demandaient...). Pourtant, un examen soigneux montre que cette source de chaleur additionnelle ne modifie presque rien au calcul. La véritable erreur de Kelvin est d'avoir négligé la convection du manteau, qui du fait des échelles de temps colossales joue un rôle important malgré les viscosités énormes (penser à la viscosité de l'acier solide...) La convection entraîne une homogénéisation "rapide" de la température dans le manteau, augmentant le gradient de température à travers la croûte terrestre, gradient qui nous apparaît beaucoup plus grand qu'il ne devrait l'être. Ce phénomène est bien expliqué dans **mettre des références pour la conf de Le Moël et l'article de Pour la Science** ; il avait d'ailleurs été deviné par un contemporain et élève de Kelvin, Perry, qui pourtant n'osa pas crier ses objections trop fort.

Ceci conclut une histoire riche d'enseignements, dont le monde scientifique ne ressort guère à son avantage : de belles formules ruinées par un modèle trop grossier ; de grands scientifiques pêchant par excès d'orgueil ; des idées fausses ancrées pendant des décennies sans remise en question sérieuse.

XI-4. Fourier analyste : régularité

espaces de Sobolev

bebe injections de Sobolev

approximation par convolution